Statistical Approach to the Optimisation of the Technical Analysis Trading Tools: Trading Bands Strategies

Directeur de these: Alain Galli

Jury:

Delphine Lautier	Rapporteur
Yarema Okhrin	Rapporteur
Michel Schmitt	Examinateur
Frederic Philippe	Examinateur

Dedicated to my father Igor Ya. Oleksiv - the noblest person I know.

Acknowledgment

I want to thank my advisor Alain Galli for all his thoughtful guidance.

Many thanks to the Members of the Jury - Delphine Lautier, Yarema Okhrin, Michel Schmitt and Frederic Philippe for their time and valuable feedback.

I am grateful to Margaret Armstrong for her kind mentoring.

My special thanks go to:

- CERNA staff for giving me this opportunity, which made this research work possible
- professors Maisonneuve, Schmitt and Préteux of Ecole des Mines for their classes in probability theory and stochastic processes
- Matthieu Glachant, my fellow doctoral students and all participants of the research ateliers, for their constructive critiques
- Jerome Olivier and Frederic Philippe of Banque BRED for sharing their business perspectives and providing informational support for certain elements of this thesis that were completed during my project with them.
- Sesaria Ferreira for the great administrative and logistical help.

The five years I spent in Paris working on this thesis were more than just the time of my professional growth - it was a magnificent personal adventure. I thank my dear friends who shared this journey with me: Larysa Dzhankozova, Anne-Gaelle Geffroy, Yann Ménière, Sevdalina and Panayot Vasilev, and Biket and Doga Cagdas.

Many thanks to my parents-in-law, Elena and Yuri Razanau, for always having the words of encouragement in difficult moments.

I am thankful to my husband Aliaksei Razanau for his loving support, and to my son Mark who has been my greatest source of inspiration.

There are no words to fully express my indebtedness to my parents, Tamara and Igor Oleksiv, without whom none of my achievements would have been possible.

Abstract

In this thesis we have proposed several approaches to improve and optimize one of the most popular technical analysis techniques - trading bands strategies. Parts I and II concentrate on the optimization of the components of trading bands: the middle line (in the form of the moving average) and bandlines. Part III is dedicated to the improving of the decision-making process. In Part I we proposed the use of kriging method, a geostatistical approach, for the optimization of the moving average weights. The kriging method allows obtaining optimal estimates that depend on the statistical characteristics of the data rather than on the historical data itself as in the case of the simulation studies. Unlike other linear methods usually used in finance, this method can be applied to both equally spaced data (in our context, traditional time series) and data sampled at unequal intervals of time or other axis variables. Part II proposes a method based on the transformation of the data into a normal variable, which enables the definition of the extreme values and, therefore, the bands' values, without constraining assumptions about the distribution function of the residuals. Finally, Part III presents the application of disjunctive kriging method, another geostatistical approach, for more informative decision making about the timing and the value of a position. Disjunctive kriging allows estimating the probability of certain thresholds being reached in the future. The results of the analysis prove that the proposed techniques can be incorporated into successful trading strategies.

Cette these propose des approches pour ameliorer et optimiser un des instruments les plus populaire d'analyse techniques – bandes de trading. Les parties I et parties II se concentrent sur l'optimization des composantes des bandes de trading: ligne centrale (representée par la moyenne mobile) et lignes des bandes. La partie III est dédiée à l'amélioration du processus de prise de decision. Dans la partie I on proposes d'utilizer la méthode de krigeage, une approche geostatistique, pour l'optimization des poids des moyennes mobiles. La methode de krigeage permet d'obtenir l'estimateur optimal, qui incorpore les characteristiques statistiques des donnees. Contrarment aux methodes classiques, qui sont utilisées en finance, cette methode peux etre appliquée à deux types des données: echantillonées à distance régulière ou irrégulière. La partie II propose une methode, basée sur la transformation des données en une variable normale, qui permet de definir les valeurs extremes et en consequence les valeurs des bandes sans imposition des contraintes de la fonction de la distribution des residus. Enfin, la partie III presente l'application des methodes de krigeage disjonctif , une autre methode geostatistique, pour les decision plus informative sur le timing et type de position. Le krigeage disjonctif permet d'estimer les probabilités, que certain seuils seront atteints dans le futur. Les resultats d'analyse prouvent que les techniques proposées sont prometeuses et peuvent etre utilisées en pratique.

Resume

L'analyse technique consiste en l'ensemble des instruments, modèles graphiques et règles de trading, qui sont fondées sur la hypothese que les prix passés peuvent être utilisés pour anticiper les prix futurs. Les règles et modèles sont souvent developées par les traders-techniciens. L'analyse technique est largement ignorée par les traders-fondamentalistes, qui définissent leurs stratégies par les valeurs fondamentales (comme des macro- et micro-indicateurs). Ses idées sont aussi rejetées par la majorité de représentants academique, qui n'acceptent pas cette approche comme méthode pour la prevision des prix futurs.

Les chercheurs ont des difficultés pour accepter cette methode pour la raison suivante : L'analyse technique est fondée sur la hypothese que les prix passés peuvent être utilisés pour anticiper les prix futurs. Cette idée contredit l'hypothèse des marchés efficaces (« efficient market hypothesis ») sur laquelle la majorité des modèles financiers classique est basée. L'autre problème avec l'analyse technique est sa nature empirique: les règles de trading sont souvent dérivées d'observations empiriques plutôt que de modèles mathématiques. En plus, c'est plutôt la règle que l'exception pour les traders de déclarer que certains parametres de certaines stratégies sont optimaux sans aucune référence à des conditions, hypothèses et critères d'optimization. Finalement, les barrières linguistiques crées par le jargon et la terminologie technique utilisée par les traders et chercheurs compliquent encore le dialogue entre les deux parties. En conséquence, ce sujet est insuffisamment développé dans les recherches scientifiques, qui évoquent plutot le scepticisme que l'intérêt. Notre

motivation était donc de contribuer à la recherche pour essayer de combler le fossé entre praticiens et scientifiques.

Cette thèse propose des approches pour améliorer et optimiser un des instruments les plus populaires de l' analyse technique, les bandes de trading. La bande de trading est la ligne placée autour de l'estimateur de tendance centrale. Quatre composantes des bandes peuvent être définis : (1) la série de prix ; (2) la ligne centrale ; (3) la bande haute (supérieure) ; (4) la bande basse (inférieure). Deux types de stratégies peuvent être définis pour cet instrument : (1) « trend-following » ; et (2) « contrariant ». Pour la strategie « trend-following », les bandes servent de confirmations du signal de tendance établi. Au contraire, les bandes définissent les instruments qui sont trop chers ou moins chers pour la strategie « contrariant ». Certainment, les positions prise en contexte de ces types de strategie sont opposées.

Les traders utilisent différents types des bandes. Les classement des bandes peut être défini par les idées/hypothèses conceptuelles en ce qui concerne les prix pour lesquels les bandes sont définies. Par example, les bandes de trading, les plus simples obtenues par le déplacement parallèle de la ligne centrale en haut et en bas, supposent la volatilité constante des prix. Les bandes de Bollinger essaient d'incorporer la nature stochastique de la volatilité des prix. Les autres types de bandes prennent en compte la distribution statistique des residus calculés sur la base de prix.

L'avantage de bandes de trading consiste en la possibilité d'optimiser l'instrument par ses composantes. D'abord on peut optimiser la ligne centrale, comme estimateur optimal de la tendance (la partie I). Ensuite les bandes sont optimisées de façon à ce qu'elles contiennent *K%* des résidus (la partie II).

Enfin, la partie III est dédiée à l'amélioration du processus de prise de decision.

Dans la thèse on considère trois types des bandes qui correspondent aux groupes présentés plus haut. En première partie on estime la ligne centrale optimale sous la forme de la moyenne mobile krigée (KMA), la deuxième et la troisième parties utilisent respectivement la moyenne mobile simple (SMA) et moyenne mobile exponentielle (EMA) en tant que ligne centrale. La partie II a examiné les bandes définies par les caractéristiques statistiques des données (par exemple, variance). La partie III a analysé les stratégies de trading pour les bandes créées par le déplacement parallèle de la ligne centrale en haut et en bas. Le choix de différents types de la moyenne mobile (KMA et SMA) pour les deux premières parties est justifié par la nécessité d'éviter de mélanger des effets de l'amélioration des bandes de trading provoquées par les optimisations de ses composants. Quant au choix des bandes en partie III est expliqué par l'énorme popularité de ce type de bandes chez les traders.

Dans la partie I "Optimisation de l'indicateur de la moyenne mobile: méthode de krigeage" on propose d'utiliser le krigeage, une approche geostatistique, pour l'optimisation des poids des moyennes mobiles (MA). Cette méthode permet d'optimiser la structure des poids pour une longueur prédéfinie de la fenêtre sur lequelle la moyenne mobile est calculée. La méthode de krigeage permet d'obtenir l'estimateur optimal, qui incorpore les caractéristiques statistiques des données, telles que la covariance (autocovariance). Cela permet d'obtenir des estimateurs optimaux qui dépendent des caractéristiques statistiques des données plutôt que des valeurs des données historiques comme dans le cas des études de simulation. Contrairement aux méthodes classiques, qui sont utilisées en finance, cette méthode peux être appliquée à deux types de

données: échantillonées à distance régulière ou irrégulière. Cette approche propose de définir le meilleur estimateur de la moyenne comme une somme pondérée des observations dans un voisinage, qui coïncide avec la définition de la moyenne mobile. La méthode d'optimisation se base sur la minimisation de la variance d'estimation.

Nous avons vu que la meilleure moyenne mobile krigée (KMA), estimée sur les données régulières a une structure des poids spécifique pour certains modèles de covariance: les plus grand poids sont attachés à la première et la dernière observation, alors que tous les autres poids sont faibles. En conséquence, KMA oscille autour de la courbe de SMA. La volatilité et l'amplitude des oscillations est une fonction indirecte de la longueur du voisinage utilisé pour le KMA : le KMA sur un voisinage plus long est moins volatile et coïncide plus avec la courbe de SMA. Par conséquent, des stratégies de trend-following, qui sont basées sur les KMA et SMA prendront des positions différentes pour des voisinages courts et les même positions pour des voisinages grands. La structure des poids ne dépend pas de la longueur de la fenêtre mais du modèle de covariance. Ce dernier a un impact sur les valeurs des coefficients de pondération proches des bordures de la fenêtre : le moins régulier est le modèle de variogramme à l'origine le plus les poids de KMA sont proches des coefficients de pondération du SMA. Par example, le modèle effet de pépite amène aux poids optimaux qui correspondent aux poids de la moyenne mobile simple.

La structure des poids des échantillons à maille irrégulière est plus variable, elle dépend de l'écart entre les échantillons de la variable utilisée pour subordonner

les prix ou la distance entre les observations de prix temporaires: plus l'écart est grand, plus la structure des poids est volatile.

L'analyse du KMA en contexte de stratégies de trading montre que le KMA permet d'obtenir des résultats positifs et intéressants. Les résultats de l'application de stratégies « trend-following » définie par les croisements de moyenne mobile et de courbe des prix montrent que pour la majorité des instruments considérés KMA génère des résultats plus élevés que les moyennes mobiles simples ou exponentielles. En plus, le profit maximal etait obtenu pour des KMA sur de petits voisinages. Les moyennes mobiles traditionnelles sur des voisinages courts produisent normalement beaucoup de faux signaux et de ce fait sont moins rentables. Malgré son caractère volatile, KMA ne génère pas plus de transactions que les moyennes mobiles traditionnelles de même longueur. Par conséquent, il semble que la nature erratique de la courbe de KMA ne conduit pas nécessairement à générer plus de faux signaux pour les stratégies de trend-following.

L'application des stratégies de trading pour les échantillons irréguliers montre que les différents types de moyennes mobiles calculées sur l'échantillon ajusté (pour avoir un échantillon régulier) pourrait conduire à des résultats moins efficaces, que si on calcule la moyenne mobile optimale pour l'échantillon irrégulier par la méthode de krigeage.

La deuxième partie "Une alternative aux bandes de Bollinger: les bandes, basées sur les données transformées", propose une nouvelle approche pour optimiser les bandes de trading.

Bandes de Bollinger ont été proposése au début des années 1980 et restent très populaires parmi les professionnels de nos jours. Bollinger a proposé d'utiliser la moyenne mobile simple comme une ligne centrale:

$$m_t = \frac{\sum_{i=t-n+1}^{t} P_i}{n}$$

avec $\{P_i\}_{0 \le i \le t}, \forall t > 0$ - la prix, n - longeur de la fenetre.

Les bandes sont définies par l' équation suivante :

$$b_t = m_t \pm k\sigma_t,$$

avec $\sigma_t^2 = \frac{\sum_{i=t-n+1}^{t} (P_i - SMA_t)^2}{n}$, $\forall k = const > 0$ - paramètres de Bollinger.

Les conclusions suivantes peuvent être dérivées pour les bandes de Bollinger :

1. Les bandes de Bollinger sont ajustées pour la volatilité des prix, comme la définition des bandes incorpore l'écart-type, en tant que mesure de la volatilité.
2. Les bandes supérieures et inférieures au même moment de temps sont placés à distance égale de la moyenne mobile, c'est-à-dire ces bandes sont symétriques. Toutefois, cette distance peut être différente à différents moments du temps, en raison de l'évolution de la nature de la volatilité des prix.
3. La distance entre les bandes se réduit avec la diminution de la volatilité des prix et s'élargit avec l'augmentation de celle ci.
4. L'usage des SMA comme la ligne centrale est justifiée par le fait que le SMA est la moyenne statistique des prix des sous-échantillons - la même valeur qui est utilisée pour les calculs de l'écart-type de prix. En outre, certaines recherches montrent que la substitution de la SMA par

la moyenne mobile plus rapide ne produit pas de résultats plus élevés (Bollinger, 2002). Nous avons également montré dans la partie 1 qui moyenne mobile optimale krigée (KMA), en moyenne, coïncide avec la SMA : KMA coïncide avec le SMA pour les grandes longueurs de la fenêtre. L'intérêt de l'introduction de la moyenne mobile exponentielle (EMA) au lieu du SMA pourrait consister en méthode récurrente de son calcul (à l'heure actuelle EMA peut être calculée comme une somme pondérée du prix actuel et de la valeur précédente de l' EMA). Toutefois, à cet égard SMA peut également être programmée avec des formules récurrentes, mais il a besoin d'accumuler plus de données à chaque instant que pour le calcul d'EMA.

La méthode traditionnelle de Bollinger est statistiquement justifiée pour les cas de prix au minimum localement stationnaires et avec une distribution symétrique. Les bandes, basées sur les données transformées, fournissent un moyen simple mais puissant pour l'optimisation des bandes. La méthode, est basée sur la transformation des données en une variable normale, qui permet de définir les valeurs extrêmes et en conséquence les valeurs des bandes sans imposition des contraintes sur la fonction de la distribution des résidus. Du point de vue théorique les bandes optimales devraient contenir K% des données (par exemple, K% = 90%); toutes les observations qui se trouvent en dehors des bandes sont considérés comme extrêmes. Les bandes ne sont pas faciles à définir pour la distribution asymétrique ou multimodale et exigent un temps considérable pour la procédure d'optimisation. Notre méthode permet d'obtenir les bandes dans un cadre plus simple et moins intensif en termes de calculs. Pour cette procédure, les données brutes (résidus) sont d'abord transformées en variables normales. Pour la variable normale la distribution est

connue et bien definie ; en conséquence les intervalles qui contiennent *K%* des données sont connus. Ensuite, les bandes pour les données brutes sont obtenues par transformation de l'intervalle (bandes) pour la distribution normale en utilisant d'une fonction d'anamorphose calibrée précédemment. Les DT bandes contiennent le même pourcentage de données que l'intervalle pour les données normales.

Nous avons examiné notamment les résidus $R_i = P_i - SMA_t$, $SMA_t = \frac{1}{n}\sum_i P_i$, $i \in [t-n+1; t]$ pour calibrer la fonction de transformation. Notre objectif principal était de rester dans le contexte de la théorie des bandes de Bollinger qui utilisent ces résidus pour le calcul de l'écart-type des données. En même temps, ces résidus provoquent la forme specifique de ces DT bandes : les bandes forment un escalier qui change de marche si il ya une variation importante de prix. Cependant, les DT bandes sont moins sensibles à des mouvements non significatifs des moyennes mobiles. En plus, il semble que les DT bandes peuvent etre utilisées pour définir d'autres signaux de trading comme les vagues d'Elliot et niveaux Support/Resistance.

On a analysé des stratégies différentes dans la partie II. Les stratégies de bande de Bollinger, comme toutes les stratégies contrariantes envoient de faux signaux au cours de la tendance présente dans les données à cause des erreurs qu'elles font dans la définition de la "vraie" valeur de l'instrument. Pendant les tendances des marchés la «vraie» valeur augmente ou diminue, par conséquent, des signaux de "surévaluation" ou "sous-évaluation" d'instrument sont fausses. Statistiquement, cela implique que les paramètres de nos bandes ne reflètent pas la vraie distribution de probabilité, qui n'est pas constante. En raison de ces

faux signaux les traders ne se basent pas uniquement sur les signes envoyés par les bandes de Bollinger, mais les examinent en combinaison avec d'autres signaux d'analyse technique, dans le but de confirmer la sur-évaluation / sous-évaluation ou de prédire les mouvements futurs des prix (par exemple, inversion de tendance). Dans cette partie nous avons examiné le momentum, comme l'un des signaux de confirmation pour les stratégies basées sur les bandes de Bollinger. Le même signal est utilisé pour la confirmation des stratégies pour les DT bandes. En outre, nous utilisons aussi les signaux de « Elliot » et « Support / Résistance » pour confirmer les signaux des DT bandes. À la suite, quatre stratégies différents sont analysées: (1) les stratégies de base, qui se fondent uniquement sur les signaux envoyés par les bandes, (2) les stratégies confirmées par le momentum, (3) les stratégies confirmées par les signaux d' « Elliot » ; et (4) les stratégies confirmées par les signaux d' « Elliot » et des « Support / Résistance ».

Les résultats de simulations trading pour quatre instruments différents montrent que les DT bandes génèrent plus de profits que les bandes classiques de Bollinger en stratégie confirmée par l'indicateur de momentum; les trajectoires de profil profits/pertes ont une pente plus positive et ascendante. Les majorités des stratégies gagnantes incorporent les DT bandes. En particulier, les stratégies marchent bien pour trois des quatre instruments. En plus, la stratégie de DT bandes était encore rentable en présence des certains coûts de transaction et de slippage. Enfin, les DT bandes pourraient être utiles dans la définition d'autres règles d'analyse techniques - les vagues d'Elliot et les niveaux de support / résistance. En conséquence, les nouvelles DT bandes ne sont pas seulement mieux justifiée d'un point de vue statistique et plus simples

dans leur application, mais elles permettent également générer des profits plus importants.

La partie III "Le krigeage disjonctif en finance: une nouvelle approche pour la construction et l'évaluation des stratégies de trading» présente l'application des méthodes de krigeage disjonctif (DK), pour des décisions plus informative en termes de timing et type de position. Comme beaucoup des stratégies de trading sont basées sur des signaux envoyés par la rupture de certains seuils, ces problème demande plus d'attention. Le krigeage disjonctif, une autre approche géostatistique, permet d'estimer les probabilités, que certain seuils seront atteints dans le futur.

En particulier, nous voulons prédire la probabilité conditionnelle $P\left(Z_{t+\Delta} < z_c \middle| Z_t, Z_{t-1}, ..., Z_{t-m}\right)$ sur la base des dernières observations disponibles dans certains voisinages. Du point de vue statistique, nous avons besoin de connaître la distribution $Z(T)|Z_\alpha$ à (n+1)-dimensions, qui est compliquée, voire impossible à estimer à partir de données empiriques. La méthode de krigeage disjonctif implique seulement la connaissance de la distribution bidimensionnelle $\{Z(t_i), Z(t_j)\}, 0 \le i < j \le n$ et $\{Z(T), Z(t_j)\}, 0 \le i < j \le n$ comme une condition nécessaire pour les calculs de la prédiction d'une fonction non linéaire. Il s'agit d'une hypothèse moins stricte que la connaissance de la distribution à (n+1)-dimensions.

La méthode de krigeage disjonctif est basée sur l'hypothèse qu'une fonction non linéaire de certaines variables aléatoires peut être développée en termes de facteurs d'un polynôme:

$$f[Y(t)] = f_0 + f_1 H_1[Y(t)] + f_2 H_2[Y(t)] + \ldots = \sum_{k=0}^{\infty} f_k H_k[Y(t)]$$

Quand $Y(t)$ est une variable normale et les couples $\{Z(t_i), Z(t_j)\}, 0 \le i < j \le n$ ont une distribution bivariable gaussienne, nous pouvons utiliser des polynômes orthogonaux d'Hermite pour le développement de la function non-linéaire. Grace à l'orthogonalité des polynômes de Hermite, le krigeage disjonctif de la fonction non linéaire est réduit au krigeage des polynômes de Hermite.

Pourtant l'application de la méthode de krigeage disjoncitif à données financières demande quelques ajustements en raison de la particularité de celles ci. Un des problèmes est la non-normalité de la variable analysée. En ce cas la variable et les seuils sont transformés en variables normales. Le principal problème est pourtant la non-stationnarité qui exige la ré-estimation des paramètres de la méthode, notamment de la fonction d'anamorphose. Nous avons proposé la méthode qui permet d'ajuster la fonction de transformation principale (basique) à la volatilité locale des données.

Deux types de probabilités disjonctives peuvent être définis et évalués. Les probabilités disjonctives ponctuelles sont les probabilités estimées par krigeage disjonctif en des points particulier ; ils reflètent la probabilité qu'un certain seuil sera dépassé à un certain moment de temps. Cette probabilité peut être estimée, mais ne peut pas être validée. Le krigeage disjonctif d'un intervalle reflète la probabilité qu'un certain seuil sera dépassé sur un intervalle de temps futur. Ce type de probabilité peut être validé par les fréquences empiriques - la proportion des observations lorsque le prix a été au-dessous d'un certain seuil.

Ces probabilités ont été estimées pour quatre instruments différents. Les résultats sont cohérents. L'intervalle DK probabilité (calculé pour la fonction d'anamorphose constamment ajustée à la volatilité locale) démontre une bonne prédiction en termes de timing et des valeurs en comparaison avec les fréquences empiriques. Nous avons également montré que seule la longueur de l'intervalle pour lequel la prévision a été faite et la longueur de l'échantillon utilisé pour l'ajustement de la fonction d'anamorphose ont un impact sur la prévision par la méthode de krigeage disjonctif.

La pouvoir de la prédiction de la méthode de krigeage disjonctif a été évalué aussi par la comparaison des résultats des stratégies de trading, qui incorporent cette probabilité krigée. Nous avons construit deux types des stratégies: (1) stratégie de krigeage disjonctif, où la décision sur la position d'entrée est faite sur la base des probabilités krigées, et (2) la stratégie aléatoire, où la décision sur la position d'entrée est faite au hasard (avec probabilité de 0.5). Notre étude révèle que la stratégie de krigeage disjonctif produits des résultats positifs pour l'intervalle continu des seuils. La stratégie aléatoire produit les bénéfices à nature aléatoire. La distribution de profit per transaction montre que le krigeage disjonctif permet de diminuer le nombre de transactions avec les pertes et augmenter la nombre de transaction avec les profits, si on compare avec la distribution de la stratégie aléatoire, qui produit une distribution symétrique pour les gains des transactions.

En conséquence, nous avons montré comment cette méthode peut être appliquée / ajustée pour les données financières d'une manière continue que

par rapport à la stratégie aléatoire, la stratégie de krigeage disjonctif améliore le processus de prise de décision.

Dans ce travail, nous sommes concentrés sur des études d'application d'analyse technique et ses stratégies à un seul instrument. Les recherches futures devraient envisager l'optimisation des stratégies pour un portefeuille d'instruments. En particulier, une autre méthode géostatistique multivariable, telle que cokrigeage peut être utilisée pour l'estimation de la moyenne du portefeuille et la prévision de sa valeur.

Afin de séparer les effets de l'amélioration de la ligne centrale (par l'introduction de la KMA) et l'amélioration de bandes (par l'introduction de la DT bandes), nous n'avons pas examiné les DT bandes qui intègrent la KMA comme la ligne centrale. Il serait particulièrement intéressant d'analyser des stratégies, fondées sur les DT bandes et KMA court.

L'approche des DT bandes indique les directions suivantes de recherche seraient à envisager. L'ajustement de la fonction d'anamorphose à la volatilité locale, réalisée dans la partie III pour la méthode de krigeage disjonctif, peut être appliquée à la définition des DT bandes.

L'étude de la relation entre la rentabilité de la stratégie et la valeur de paramètre *K%* utilisée pour la définition des DT bandes permettra augmenter les profits des stratégies définies sur la base de ces bandes.

Cette approche crée des nouvelles possibilités à l'amélioration d'autres instruments et règles d'analyse technique. Par exemple, la stratégie confirmée

par l'indicateur de momentum est fondée sur la définition des seuils optimaux pour cet indicateur. Dans cette analyse les seuils n'étaient pas optimisés, mais l'approche de transformation des donnée utilisée pour la définition des DT bandes peux être utilisée pour définir les seuil de momentum. Cela pourrait conduire à des seuils asymétriques de momentum. L'autre exemple est l'application des DT bandes à la définition d'autres indicateurs techniques, comme les vagues d'Elliot et de support / résistance.

Enfin, l'application de la méthode de krigeage disjonctif aux données financières peut encore être améliorée par un meilleur ajustement de la fonction cumulée de la distribution utilisée pour la transformation de données aux changements de moyenne locale ou à l'asymétrie de distribution.

Les résultats d'analyse prouvent que les techniques proposées dans cette thèse sont prometteuses et peuvent être utilisées en pratique. Ils indiquent aussi de nombreux domaines de recherche pour l'avenir.

Contents

General Introduction ……. 1

Part I : Moving average optimization: kriging approach ……... 12
Introduction …….. 12
1. Moving average as a trading instrument……... 14
1.1. Definition of moving average ……...……. 14
1.2. Types of the moving average ……. 14
2. Kriging method: Theory ……. 16
2.1. **Geostatistic instruments and terminology: short overview** ……..……. 17
2.1.1. Stationary and intrinsic stationary functions ……..…... 18
2.1.2. Variogram …….. 19
2.2. Simple kriging: prediction of the process with zero mean ……... 21
2.3. Simple kriging: prediction of the process with known mean ……..….. 23
2.4. Ordinary kriging: prediction of the process with unknown mean ……..….. 24
2.5. Ordinary kriging: estimation of the unknown mean ……... 25
2.6. Kriging non-stationary variable …….. 26
2.6.1. Universal kriging: trend estimation ……. 27
2.6.2. Kriging intrinsic function IRF-0 ……. 28
3. Peculiarities of the kriging method applications in finance …… 30
3.1. Data non-stationarity ……. 31
3.2. Data sampling peculiarities …… 31
4. Kriging results: non-stationary, evenly spaced time-series data ……. 33
4.1. Variogram analysis and optimal kriging weights: Bund …….. 38
4.2. Trading results: kriged moving average versus simple moving average …….. 38
4.2.1. Bund …….. 46
4.2.2. DAX …….. 47
4.2.3. Brent ……..… 49
4.2.4. X instrument …….. 51
5. Kriging results: bounded, evenly spaced time-series data ……... 53
5.1. MACD and trading strategies ……. 55
5.2. MACD strategy: optimal signal line ……... 55
5.3. Results of the trading strategy, based on the MACD indicator and signal line …… 57
5.3.1. Bund …….. 59
5.3.2. DAX ……. 61
5.3.3. Brent ……. 63
5.3.4. X instrument ……. 65
6. Kriging results: unevenly spaced data ……... 67
6.1. Unevenly spaced time dependent price series ……... 68
6.2. Unevenly spaced price series subordinated to cumulative volume …….. 70
7. Conclusions ……. 75
8. Appendices I ……... 83
86

Part II. An Alternative to Bollinger bands: Data transformed bands ……
Introduction ……... 111
1. **Classical Bollinger bands concepts** ……..….. 111
1.1. Bollinger bands definition …… 114
1.2. Bollinger bands and trading strategies …….. 115
1.3. Theoretical framework for Bollinger bands ……..……..…….. 117
2. **Data transformed bands** ……..……... 121
2.1. Data transformed bands: algorithm description …….. 124
2.2. Data transformed bands: empirical observations ……..…. 125
2.3. Data transformation bands: other technical rules description ……... 127
3. Strategies descriptions …… 128

3.1. Strategies …… 132
3.2. Choice of the streategy parameters ……. 132
3.3. Analysis of the outcomes …… 136
4. Trading outcomes for different instruments ……..…… 138
4.1. Bund …..…… 139
4.2. DAX ……... 139
4.3. Brent ……... 145
4.4. Instrument X ……...…… 152
5. Conclusions ……..……..… 157
6. Appendices ……..……..……... 161
164

Part III. Disjunctive kriging in finance: a new approach to construction and evaluation of trading strategies 180
Introduction …..…… 180
1. Theory ……...…… 182
1.1. Disjunctive kriging ……...…… 183
1.2. Disjunctive kriging: normal random process ……..... 185
1.3. Disjunctive kriging: non-normal random process …… 189
1.4. Disjunctive kriging: case of regularized variable ……... 189
1.5. Disjunctive kriging: particular case of the variable with exponential covariance model …….. 190
2. Peculiarities of the disjunctive kriging application to the financial data ……. 192
2.1. Calibration of the transform function: kernel approach ……... 193
2.2. Data used for the estimation of the CDF: locally adjusted transform function …. 199
3. Examples of the DK probabilities under stationarity assumption…… 202
3.1. Point DK probabilities: DAX case ...…… 202
3.2. Interval DK probabilities for global transformation function: Bund case ……. 205
4. Examples of the DK probabilities under local stationarity assumption …….. 212
4.1. Framework for the estimation and analysis of the interval DK probabilities …… 212
4.2. Interval DK probabilities: DAX instrument ……. 215
4.2.1. Impact of the window length …… 216
4.2.2. Impact of the interval length ..…… 218
4.2.3. Impact of the DCI length ….……. 220
4.2.4. Impact of the bandwidth parameter …… 222
5. Application of the DK method to the strategy construction …… 223
5.1. DK trading strategy …… 224
5.2. Random walk (or benchmark) strategy ……. 224
5.3. Strategies outcomes …….. 225
5.3.1. DAX ……. 226
5.3.2. Bund ...……. 228
5.3.3. Brent ……. 231
5.3.4. X instrument ……. 233
6. Conclusions …… 235
7. Appendices III …… 238

General Conclusions …… 252

Bibliography …….. 255

General introduction

Technical analysis is a hotly debated topic among researchers and traders. It has its devoted supporters, so called technicians or technical analysts, as well as the opponents who do not accept its methods. The debate takes place not only between different groups of traders (fundamentalists vs. technicians[1]), but also between the representatives of academic circles and traders (technicians). The discussion between different types of traders is explained by different principles and relationships that are used for predicting future prices. Fundamentalists base their predictions and, thus, their strategies on the market fundamentals, such as macro-indicators and micro-indicators. Macro-indicators evaluate general market situation; inflation, interest rate, unemployment rate, inventories, consumer confidence index, etc. are the examples of such indicators. Micro-indicators represent the characteristics of an instrument, for which the prediction is made; for example, for the prediction of the price movements of a particular stock, the traders analyze company's revenues, assets, balance sheet, etc.

In their turn, technical analysts believe that prices incorporate all information available in the market (i.e. macro-, micro-indicators, expectations, etc.). Therefore, arguably, it is sufficient to use the existing price observations to make predictions about future price movements. Thus, technical analysis techniques are predominantly based on the price data.

When it comes to the academic audiences, most of them[2] refuse to accept technical analysis as a consistent price forecast method. As noted Lo *et al.* (2000), many academic researchers who easily accept fundamental factors believe that "the difference between fundamental analysis and technical analysis is not unlike the difference between astronomy and astrology"[3]. Taking into account that the technical analysis exists for more than 100 years, such resistance of the researchers is quite puzzling, and we believe that there may be an explanation to this phenomenon.

First, the technical analysis theory is based on the assumption that the past price observations can be used to predict the future price movements. This assumption contradicts the efficient market hypothesis (EMH)[4] that is the cornerstone of the financial theory and on which many financial models are based. At the same time many departures from EMH are observed in the real markets due to over- or under-reaction, certain market anomalies (such as size effect), behavioral effects. Bernard and Thomas (1990), Banz (1981), Roll (1983), Chan, Jegadeesh, Lakonishok (1996), Huberman and Regev (2001) are the examples of such research. Treynor, Ferguson (1984) demonstrated theoretically that past prices, combined with other information, can predict the future price movements. Lo and MacKinaly (1988, 1999), showed that past prices can be used as a forecast for future prices. Finally, Grosman and Stiglitz (1980) argue that mere presence of the trading and investment activity and the possibility to earn profits in financial markets undermines the credibility of EMH. Despite the existence of such anomalies, the supporters of EMH still believe that the investment opportunities occur only in the short-term, and they are eliminated in

[1] Nowadays pure technicians or fundamentalist among traders rarely exist. Technicians generally do follow the financial news and the macro-indicators, while fundamentalists apply some of the technical analysis techniques.

[2] Some researchers though believe in the prediction power of the technical analysis. Further we will provide these works in general literature review.

[3] Lo, A. W., Mamaysky, H. and J.Wang. 2000. "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation", *The Journal of Finance*, Vol. LV, #4 (August, 2000), pp.1705-1765, p.1705.

[4] The EMH states that the more efficient the markets are, the more random the price movements in these markets are. As a result, it is impossible to use the past prices to predict the future movements under this hypothesis. Literature review on the EMH can be found in Lo (2007).

the long run. As a result, at present there is no consensus regarding the validity of the EMH in real markets.

The second point is that most technical analysis techniques have been developed on the basis of the empirical observations, rather than derived or modeled mathematically. For example, the majority of chart patterns, such as "Support/Resistance", "Head-and-Shoulders", etc. were the results of regular observations of the price behavior. Evidently, the experiment and observation laid the foundation for many major inventions in physics, mechanics, chemistry and engineering. The key difference between the scientists and technicians appears to manifest itself in the way they treat the observed results: contrary to the scientists, technicians frequently do not bother to prove or explain their observations, but take them for granted.

The third explanation is driven by the fact that the technical trading rules are often unjustifiably presented as "optimal". We can frequently see the traders making claims about "optimal" values of certain technical parameters (for example, the moving average length) without any additional support or explanations how and for what data type (instrument, data frequency, etc.) these values were obtained. Obviously, such statements raise lots of skepticism from the scientists.

Finally, the "language barriers" created by the usage of the technical jargon on one side and statistical terms and tests names on the other complicate the assimilation of the new ideas by both sides (Lo *et al.*, 2000). In addition, the researchers frequently mistakenly believe that the technical instruments are only about "charting", disregarding the mathematical concepts that are used in building the technical strategies (for example, moving average, momentum, etc.)

Thus, we can conclude that the absence of both the scientific representation of the method and of a formal analysis of the method's prediction power creates a misunderstanding between the technical traders and academic researchers. As defined by Neftci (1991), "technical analysis is a broad class of prediction rules with unknown statistical properties, developed by practitioners without reference to any formalism"[5].

On our part, we believe that technical analysis should be viewed more as a "bank" of empirical observations of the financial markets that can be further used by the researchers to develop models or well-defined statistical trading techniques. We also believe that all the academic research performed to-date in this field, is a necessary input in narrowing the gap between the theoretical and practical finance.

According to some researchers, technical analysis studies can be split into the following groups:

- Trend studies

  This group represents the indicators that identify the trend and the trend breaks. Among the most popular indicators are moving averages, support and resistance levels, etc.
- Directional studies

  This group contains the indicators that define the length and strength of current trend/forecast. Among these indicators are DMI, Parabolics, range oscillator, etc.
- Momentum studies

  This group concentrates on the measurements of the velocity of the price movements. The examples of such indicators are Stochastic, Momentum, Rate of

[5] Neftci, S. N. "Naïve Trading Rules in Financial Markets and Wiener-Kolmogorov Predicition Theory: A study of Technical Analysis", *Journal of Business,* Volume 64, Issue 4 (October, 1991), pp. 549-571, p.549.

change, MACD, Trix and CCI. Momentum instruments are frequently used for the definition of the trend breaks that often follow the slowness of the price velocity.

- Volatility studies

  This group presents the trading rules that incorporate instruments' volatility and the notion of extreme values. The examples of such rules are trading bands, among which the most popular case is the Bollinger bands.

- Volume studies

  In the context of the technical analysis, volume is the second (after price) important data element that measures the trading activity in the markets. Volume itself as well as the indicators that incorporate volume information (for example, volume weighted moving averages) completes this group.

Each group of technical studies contains indicators and chart patterns. Indicators cover all buy/sell rules that are formulated on the basis of the well-defined mathematical expressions (for example, trading rules based on the moving average). In contrast, the charts cannot be explicitly defined by formulas; they are the graphical patterns defined subjectively by a trader. Therefore, one trader can recognize a specific chart as particular price pattern, while the other trader would see no pattern at all. The examples of such charts are Support/Resistance levels, Head and Shoulders and Triangles. It should be noted that researchers try to program the chart patterns by algorithms and estimation methods (see Lo *et al.*, 2000), however, the results of the chart pattern recognition depends on the algorithm itself.

The financial literature that exists in the field, can be split into the following groups:

1. Development of the scientific framework and formalization of the technical analysis.
2. Evaluation of the prediction power of technical trading rules, as well as their comparison with other prediction methods.
3. Analysis of the statistical properties of technical indicators and their trading outcomes.
4. Optimization of the existing indicators/rules/strategies; their improving; development of the new technical trading instruments.

For the first group of studies[6], the cornerstone work is the paper by Neftci (1991) "Naïve Trading Rules in Financial Markets and Wiener-Kolmogorov Prediction Theory: A study of Technical Analysis". It proposes the general approach that defines which technical rules can be formalized and which cannot. According to this framework, a well-defined rule should be a Markov time, i.e. it should use only information available up to current moment for its construction[7]. Most of the prediction technique that is used for financial market forecasts lies in the Wiener-Kolmogorov prediction theory framework, according to which "time-varying vector autoregressions (VARs) should yield the best forecasts of a stochastic process in the least square error (MSE) sense"[8]. However, this framework is not suitable for the forecast of the non-linear series. For example, Neftci (1991) defines at least two cases when linear models cannot produce plausible forecasts such as (1) producing sporadic buy and sell signals (non-linear problem by its nature); and (2) predicting some particular patterns, such as stock exchange crashes. Consequently, any other systems of forecasts that can predict non-linear time series can improve the forecast proposed by the Wiener-Kolmogorov framework. Similar conclusions are obtained in Brock, Lakonishok and Lebaron (1992). According to Neftci (1991), it may be the case that technical analysis informally tries to analyze the information captured by the higher order moments of asset prices. In fact the patterns and rules of the technical analysis can be

[6] See also Rode, Friedman, Parikh, Kane (1995) for formalization of the technical analysis.

[7] The method will be presented further in more details.

[8] Neftci, S. N. "Naïve Trading Rules in Financial Markets and Wiener-Kolmogorov Predicition Theory: A study of Technical Analysis", *Journal of Business,* Volume 64, Issue 4 (October, 1991), pp. 549-571, p.549.

characterized "by appropriate sequences of local minima and/or maxima"[9], that lead to non-linear prediction problems (Neftci, 1991). As the result, he believes that technical analysis can improve the forecasts of the future price movements.

Another part of the formal academic research that can diminish the number of skeptics about the technical analysis is the formalization of the technical indicators and instruments themselves. While there is a lot of literature devoted to the description, definition or calculations of the technical indicators (Murphy, 1999; Achelis, 2000), there are few works that explain or justify the method from theoretical standpoint; the examples are Bollinger (2002) on the Bollinger bands, Ehlers ([38]) on moving average. Group of papers try to explain the predictability of the technical indicators in the context of the microstructure theory through the relationship between technical analysis and liquidity provision. The researchers believe that the technical analysis may indirectly provide information captured in limit-order books to make predictions about future price movements. Osler (2003) provided the explanation of the prediction power of such technical trading rules as Support/Resistance, proving the following hypothesis: the clusters of take-profit and stop-loss orders are the reasons why the rules succeed in predicting future price movements. Kavajecz, Odders-White (2004) related the moving average indicators (price moving averages of different length) to the relative position of depth on the limit order book.

The second group of studies that measure the predictive properties of the technical analysis is best represented in the financial literature. The majority of the papers in this field are devoted to the statistical (econometric) analysis of the prediction power of the technical rules, while comparing them with other (non-technical) predictors or variables. The early works in the field of the technical analysis did not find the superior prediction properties of the technical rules comparing them with the Buy-and-Hold strategy; as the result these works supported the EMH theory (Alexander, 1961, 1964; Fama and Blume, 1966; James, 1968; Van Horne and Parker, 1967; Jensen, Benington, 1970). More recent work by Allen, Karjalainen (1999) and Ratner and Leal (1999) has also found little evidence in favor of the technical analysis. At the same time other research provides the evidence in favor of the technical analysis. Brock, Lakonishok and Lebaron (1992) showed that 26 technical trading rules applied to Dow Jones Industrial Average over 90 years over-perform the strategy of holding cash. Sullivan, Timmermann, White (1999) shows that some of the technical rules considered in Brock *et al.* (1992) are actually profitable even after using the bootstrap method to adjust for the data-snooping biases. Levich, Thomas (1993) found that some moving average and filter rules were profitable in the foreign exchange markets. Osler, Chang (1995) also found the evidence of the profitability of the "head-and-shoulders" patterns in foreign exchange markets. Lo, Mamaysky, Wang (2000) showed that the same technical charts provide incremental information about future price movements by comparing unconditional distribution of the stocks returns with the conditional distribution of the returns (conditional on the presence of the chart pattern). Blume, Easley, O'Hara (1994) demonstrated that the traders who use information contained in the market statistics such as prices and volume do better than the one who do not use it. In this context, technical analysis is a component of trader's learning process. Blanchet-Scallient et al. (2005) compare the results of the technical rules to the strategies, based on the mathematical model. Under certain assumptions (prices follow one-dimensional Brownian motion, trader's wealth utility is represented by the logarithmic function) they show that MA rule can outperform the strategies based on the mathematical models in case of severe misspecifications of the model parameters.

As for the third group of studies, Acar, Satchell (1997) analyzed the statistical properties of the returns from the trading rules, based on the moving averages of the length 2. They showed that

[9] Neftci, S. N. "Naïve Trading Rules in Financial Markets and Wiener-Kolmogorov Predicition Theory: A study of Technical Analysis", *Journal of Business,* Volume 64, Issue 4 (October, 1991), pp. 549-571, p.550.

in the case when the asset price distribution is a Markovian process, the characteristic function (and, therefore, the distribution function too) of the realized returns could be deduced.

The optimization of the technical trading rules has a crucial importance for the traders, who use this approach in the construction of their strategies. Both researchers and traders contribute to this field of studies. For some instruments that are more popular, many optimization approaches exist, while for the other the niche is largely underdeveloped. For example, many works devoted to the optimization of the rules based on the moving averages, momentum (Gray, Thomson, 1997). Certainly, the choice of the optimization techniques largely depends on the type of the technical indicator. However, there is one approach that is applied to many different techniques – a simulation of the trading strategy based on the available historic data samples. According to this approach the optimal parameters correspond to the global/local maximum or minimum of the trading outcomes (profit/losses, Sharpe ratio, number of trades, etc.). The example of such optimization approach can be found in Williams (2006). Although this approach is universal, as it can be applied to all rules that are used in the trading strategies, the method has its drawbacks. The outcomes are dependent on the historical data used for optimization; thus, the parameters optimal for the studied data set might be no longer optimal for a new data sample.

Finally, the development of the new instruments is a very dynamic field that is constantly enlarging both with the new types of the existing instruments and totally new ones. For example, Arm ([11], V.8:3) proposed the volume-weighted moving average, Chande ([28], V.10:3) developed the volatility adjusted moving average, Chaikin and Brogan in their time introduced Bomar bands (Bollinger, 2002).

As we can see, the majority of the papers in the field of technical analysis are devoted to the analysis of its prediction power within the context of EMH. Despite a significant number of papers on this topic, there is still no consensus whether technical analysis has superior prediction power over other prediction methods. Therefore, we will accept the hypothesis, similar to one in Grosman and Stiglitz (1980), that survival of the technical analysis among traders for the past 100 years can be considered a proof that it can be integrated into profitable trading strategies, at least for some particular instruments; otherwise the traders would have stopped using it.

At the same time, fewer researchers concentrate on the development of the theoretical framework for the analysis and optimization of the technical rules, although there is a pool of users (traders, technicians) who create the demand for this type of research. Thus, in this thesis, we want to concentrate on the optimization and development of the new trading techniques based on the existing technical strategies. The technical indicators/rules chosen for the analysis will be formalized and explained from the point of view of the statistical theory. Contrary to many existing works in the field that use trading simulations to define the optimal parameters, we want to use the optimization approaches, based on the statistical characteristics of the data in the first place. We will use the trading simulations in all parts of this thesis, mainly to compare the optimized indicator/strategy with other non-optimal (in the context of this work) indicators/strategies.

It is obvious that an exhaustive analysis of all technical rules is quasi-impossible: the set of trading rules is extremely large and it expands constantly with the development of the new rules and patterns. We decided to concentrate our analysis on such popular technical analysis techniques as trading bands. While being part of the volatility studies, trading bands frequently incorporate (in their constructions or their strategies) other techniques of the technical analysis from such group of studies as trend and momentum (see classification above). Besides, we will prove that the

method itself is well defined within the framework developed by Neftci (1991), briefly presented further.

Trading bands are lines plotted around a measure of central tendency, shifted by some percentage up and down (upper and lower bands) (Bollinger, 2002). The schematic representation of the concept is given in Figure 1.

Trading bands have four key components (see Figure 1):

(1) price (quotes),
(2) mid-line,
(3) upper band, and
(4) lower band.

The way these components are defined implies the existence of different bands types, such as envelopes and channels (Bollinger, 2002, Murphy, 1999).

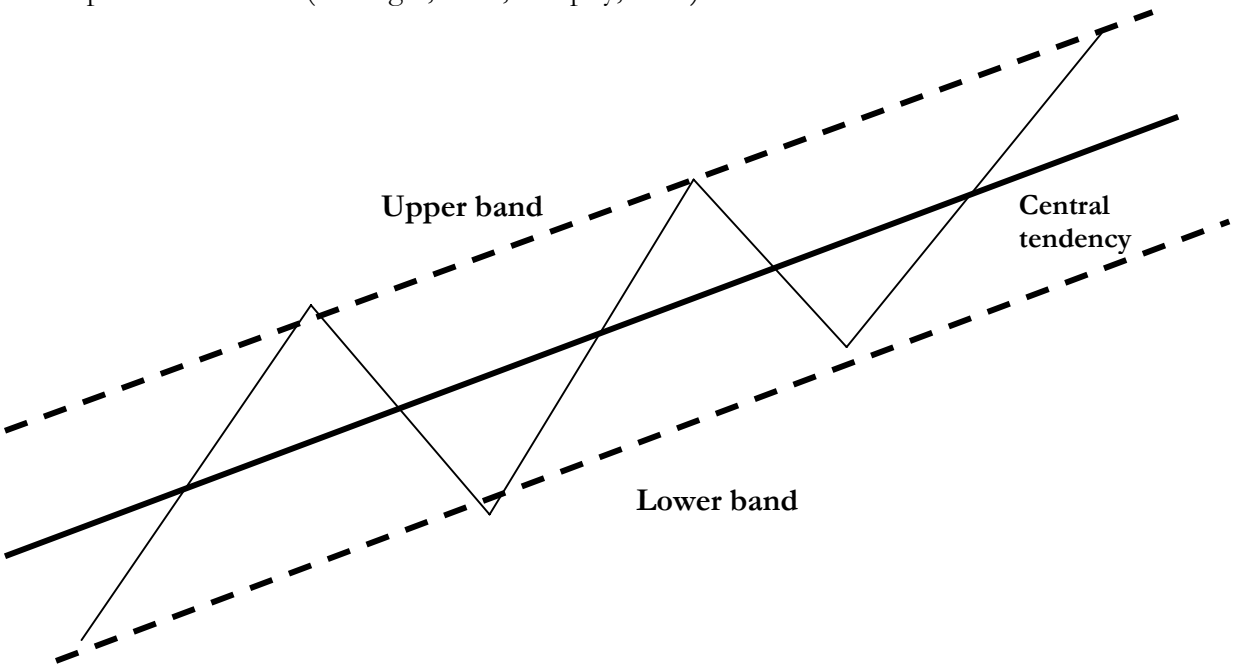


Figure 1. Schematic representation of trading bands

Despite the differences in constructing the bands, the strategies based on them are quite similar. Touching/breaching upper/lower bands give trader information on the direction of price movements or on relative price levels (whether the instrument is oversold or overbought), which are used as signals in strategy construction. As a result, both trend following and contrarian strategies can be defined on the basis of the trading bands.

For example, let some moving average represent the mid-line in the bands. Suppose prices crossed the upper band of the trading bands after continuous fluctuations within the upper/lower bands and the mid-line crossing. For the trend following strategy, the bands are the confirmation of an established trend: the first signal that the upward trend had been established happen at the crossover of the moving average and price curve[10]. Therefore, a breach of the

[10] Breaching the upper bands implies that the price has been previously breaching the moving average line from below.

upper band can be used as a confirmation signal of an upward trend. In this context, trading bands allow to eliminate false signals generated by the moving average trading rule.

Breaching the trading bands in the context of the contrarian strategy confirms that the instrument is mis-priced. Therefore, breaching the upper band means that currently the instrument is overpriced and the price should return to its average (moving average line) in future.

As the same trading bands can be used in the strategies that lead to the opposite trading decisions[11], traders frequently use trading bands in combination with other technical trading signals that confirm the presence or absence of a trend. In case of the trend-following strategy, these extra rules give additional confirmation signals that the trend has been established, while in the case of the contrarian strategies they allow avoiding trending patterns, where contrarian strategy sends false signals.

As have been mentioned already, it can be proven that some types of the trading bands are well defined. According to Neftci (1991), technical trading rule is well defined if it is a Markov time.

Let $\{X_t\}$ be an asset price; I_t - sequences of information sets (sigma-algebras) generated by X_t and other data sets observed up to time *t*.

Definition 1
A random variable τ is a Markov time if the event $A_t = \{\tau < t\}$ is I_t–measurable.

Simply speaking, it means that a rule/indicator is well defined if for making a decision or its calculation it uses only information available up to the current moment, but not the one that anticipates the future. For example, the first moment of time when prices increase 20% from the initial level at $t = 0$ is Markov time: on the basis of available history of prices up to moment t, we can determine whether this event has happened or not. As a result, the Markov time approach eliminates many technical rules that anticipate the future, among which many chart patterns.

The definition of a technical rule as a Markov time implies (1) possibility to quantify the rule, (2) feasibility of the rule, (3) possibility to investigate rule's predictive power. However, the fact that the rule is well defined cannot justify its usage. In order to be used, a rule should produce (buy/sell) signal at least once, i.e. the probability that the signal is generated at least once should be equal to one (see Definition 2).

Definition 2
A Markov time τ is finite if

$$P(\tau < \infty) = 1.$$

In addition, a rule should have at least the same predictive power as other well-defined forecasting techniques. As a result, a trading rule gives a consistent forecast of the future price movements if it is a finite Markov time that has at least the same predictive power as more formalized forecasting methods. For example, Neftci (1991) showed that the moving average trading rules are finite Markov times and in some cases have higher predictive power than the linear forecast methods, such as AR or ARMA models.

[11] The trading positions taken within each strategy would be the opposite: for the trend-following strategy a long position is appropriate, while for the contrarian strategy – a short position should be taken.

Further we provide the proof that the trading bands are Markov times.

Lets define $\{X_t\}$ as some random process that represents price time series; I_t as a sequences of information sets (sigma-algebras) generated by X_t and other data sets observed up to time *t*. Trading band is defined as following:

- middle line: m_t;
- upper bands: $m_t + \delta_t^{(1)}$, $\delta_t^{(1)} > 0$;
- lower bands: $m_t + \delta_t^{(2)}$, $\delta_t^{(2)} < 0$.

Proposition 1

Let functions m_t, $\delta_t^{(1)}$, $\delta_t^{(2)}$ be I_t-measurable. Lets define the following variables:

$$Z_t = X_t - m_t \quad (1)$$
$$Y_t^{(1)} = Z_t - \delta_t^{(1)} \quad (2)$$
$$Y_t^{(2)} = \delta_t^{(2)} - Z_t \quad (3)$$

Then,

- generated times $\{\tau_i^{entry}\}$:
$$\tau_i^{entry} = \inf_t \left[t > \tau_{i-1}^{entry} : Y_t^{(1)} \cdot Y_t^{(2)} \le 0 \right], \ \tau_0^{entry} = 0, \quad (4)$$
are Markov times.
- generated times $\{\tau_i^{exit}\}$:
$$\tau_i^{exit} = \inf_t \left[t > \tau_i^{entry} : Z_t \cdot Z_{t-1} \le 0 \right], \quad (5)$$
are Markov times.

Proof.

Note that Z_t, $Y_t^{(1)}$, $Y_t^{(2)}$ are I_t-measurable. This implies that the products $Z_t \cdot Z_{t-1}$ and $Y_t^{(1)} \cdot Y_t^{(2)}$ are also I_t-measurable. τ_i^{entry} and τ_i^{exit} are defined as the first entry of $Y_t^{(1)} \cdot Y_t^{(2)}$ and $Z_t \cdot Z_{t-1}$ in the interval $(-\infty;0] \in \Re$ respectively. Then, according to the Theorem (Shiryayev, 1985), which states that the first entry of the process $\{X_t\}$ in some defined interval is always Markov time, the τ_i^{entry} and τ_i^{exit} are Markov times.

Proposition 1 states that if estimates of the middle line and trading bands are defined on the basis of the information available up to moment t, the trading rules based on these trading bands is well-defined. For example, moving average $MA_t = \sum_{i=1}^{n} w_i X_{t-n+i}$ that is often used as the middle line is I_t-measurable. Constant scalars $\delta_t^{(1)}$, $\delta_t^{(2)}$ together with the I_t-measurable middle line define bands, which are I_t-measurable. Finally, Bollinger bands, with the middle line $MA_t = \frac{1}{n}\sum_{i=1}^{n} X_{t-n+i}$ and $\left|\delta_t^{(1)}\right| = \left|\delta_t^{(2)}\right| = k\sigma_t$, where $k > 0$ some scalar, σ_t - experimental standard deviation of the $\{X_i\}_{t-n+1 \le i \le t}$, are also I_t-measurable.

Note that the trading strategy defined in Proposition 1 is as follows: a position is opened when the price breaches one of the bands; this position is closed, when the price crosses the middle line.

Proposition 1 demonstrates that there are trading bands strategies that can be considered as well defined. The question is now whether the signals they send are finite, i.e. $P\left(\tau^{entry} < \infty\right) = 1$, $P\left(\tau^{exit} < \infty\right) = 1$. It is obvious that the distance between the bands defines whether these trading bands will send finite signals. In particular, if the distance between the two bands is extremely large the entry signal might never be generated. The entry signals will be generated for the stationary process $\{X_t\}$ and unbiased estimator of its mean m_t, if the following inequality holds:

$$0 < P\left(m_t + \delta_t^{(2)} < X_t < m_t + \delta_t^{(1)}\right) < 1 \qquad (6)$$

The expression (6) can be considered as criteria for the choice of the distance $\left|\delta_t^{(1)}\right| + \left|\delta_t^{(2)}\right|$ between the bands that can generate any entry signals. As for the exit signals, Neftci (1991) showed that the trading rules, based on the cross-over of price and MA curves, generate finite signals in the case when price process $\{X_t\}$ is stationary and *m*-dependent price process.

The majority of the trading bands used by traders incorporate moving average as middle line. Moving average allows constructing well-defined trading bands; therefore, we will narrow our analysis to this particular type of bands. Such trading bands are the function of four different parameters or vectors of parameters:

$$TB = f\left(n, \{w_i\}_{1 \le i \le n}, d_U, d_L\right),$$

where $n, \{w_i\}_{1 \le i \le n}$ - parameters of the moving average: length of the moving window and weights attached to the price in the window (see Ch.1.1 for more details);
d_U, d_L - distance between moving average and upper and lower bands respectively. Note that upper and lower distances can be different, as well as each distance itself can be a function of time: $d_U \neq d_L$, $d_{(\,)} = d(t)$.

Despite the large number of the bands parameters, their optimization can be simplified:

1. Search for the optimal parameters for middle line and bands can be separated due to the different roles that they play in the definition of the trading bands strategy. For example, within the context of the contrarian strategy moving average represents the mean value of the instrument, while bands define the extreme values of the prices conditionally on the current mean. We do not need to know the bands value in order to evaluate price mean, while we need a current mean value (and probability distribution function) to define the extreme values, and thus, bands. As a result, the band optimization problem can be split: first, best middle line estimator (its parameters) is obtained, then optimal bands parameters are searched for.
2. According to Proposition 3, the optimal moving average should be the best estimator to the local mean (or trend). This allows choosing clearer and more objective optimization criteria - minimization of the mean squared error:

   $$E\left(\hat{M}_t - m_t\right)^2 \to \min,$$

   where m_t - true trend/mean, $\hat{M}_t$ - estimator of the true mean.

As a result, the first part of the thesis will be devoted to optimization of the moving average, while the second part will concentrate more on the development of the optimal bands (distances).

Finally, it should be noted that optimization of the strategies, based on the technical analysis, is frequently substituted by the problem of the parameter optimization. However, the decision-making around taking or exiting a position is a non-linear problem, while the parameter optimization frequently involves linear methods. Therefore, we will consider some non-linear approaches to optimize the decision-making in addition to the optimization of the parameters of the technical indicators.

Within the analysis of the trading bands, in addition to moving average we will also consider some other technical analysis techniques such as Momentum, Moving Average Convergence Divergence (MACD), Bollinger bands, Support and Resistance patterns and Elliot waves.

This thesis is split into three parts. Part I concentrates on the search of the optimal trend estimator. Part II proposes the approach to optimize bands. Finally, Part III is devoted to the optimal decision-making.

The first part "Optimization of the moving average indicator: kriging method" considers kriging as a method to estimate the mid-line in the bands. Kriging approach, a geostatistical technique, defines the optimal mean estimator as a weighted sum of the observations in some close neighborhood; in this respect the kriged estimator coincides with the definition of the moving average. Unlike other linear method usually used in finance, this method can be applied to both equally spaced data (in our context, traditional time series) and data sampled at unequal intervals of time or other axis variable. The latter is the case of the instruments that are not regularly traded, or subordinated price processes obtained by changing the time coordinate to other random variable coordinate.

The kriging method is based on the statistical characteristics of data such as covariance (autocovariance) function. This allows obtaining optimal estimates that depend on the statistical characteristics of the data rather than on the historical data itself as in the case of the simulation studies[12]. This method optimizes the weights structure for a given length of the moving window.

We will see that the optimal kriged moving average (KMA) estimated on the equally spaced data sample has a specific weight structure for certain covariance models: the largest weights are attached to the first and last observation, while all the other weights are low. As a result, this KMA coincides in lag with the simple moving average with all equal weights (further referred to as SMA), but is more volatile. Moreover, these "border" weights values depend only on the covariance model. The weights structure for the subordinated (unequally spaced) samples exhibits non-stable patterns that largely depend on the discrepancy in the values of the variable used to subordinate the price curve or the distance between the time-observations of price: the larger the discrepancy – the more volatile the weights structure is.

The comparison of KMA with the traditional types of moving averages returned interesting results. The results of applying trading strategies based on the crossovers of moving average and price curves show that for the majority of considered instruments KMA generates higher results than simple or exponential moving average. Moreover, the global maxima of the KMA-based

[12] Certainly, in real-life applications the results will still depend on the historical data used for the evaluation of the statistical properties of the variable. However, the results are more dependent on the estimation accuracy than on the data itself.

strategies are achieved at short lengths of the moving window, where traditional moving averages normally generate more false signals, and thus, less profitable. Despite its volatile nature, KMA does not generate more transactions than traditional moving average of the same length, and therefore, does not seems to generate more false signals.

The second part "An Alternative to Bollinger bands: Data transformed bands" describes the new approach to optimizing the bands. We propose a method that enables the definition of the extreme values and therefore bands values, without constraining assumptions about the distribution function of the residuals. From theoretical standpoint the optimal bands should contain K% of the price data (for example, K%=90%); all data points that lie outside of the bands are considered extreme. However, such an interval and, thus, bands are not easy to define for asymmetrical or multimodal distribution and require time-consuming optimization procedures. Our method allows obtaining the bands in a more straightforward and less-intensive procedure. For this purpose the raw data (residuals) are transformed first into standard normal variable. For this distribution the intervals that contain K% of the data are known. Afterwards the data-transformed bands (DT bands) are obtained by backward transformation of the interval for normal distribution by the means of a previously calibrated transformation function. The obtained DT bands exhibit a peculiar stair-like pattern; the bands change the level only if there is a significant price change. As for the trading outcomes, confirmed DT bands strategies generate more profits than the classical Bollinger bands that are more monotonous and upward sloping. As a result, the new DT bands are not only more justified from the statistical point of view and straightforward in their application, but also they allow generating higher profits.

The third part "Disjunctive kriging in finance: a new approach to construction and evaluation of trading strategies" presents the disjunctive kriging method for more informative decision making about the timing and the value of a position. Frequently the traders would like to know in advance that certain thresholds/bands would be breached. Disjunctive kriging (DK), another geostatistical method, allows estimating the probability that some thresholds will be reached in the future. We demonstrate how this method can be applied/adjusted to the financial data on a continuous basis and show that in comparison to the random-walk hypothesis, DK improves the trading decision-making.

The general conclusions can be found in the last section of this thesis.

Part I. Moving average optimization: kriging approach

Introduction

Prediction of future instruments value movements, as well as estimation of a trend plays an important role in the analysis of financial data. Traditional approaches to trend estimations are linear filters[1] that can identify such features as trend, seasonality, noise, etc. A moving average (MA) is an example of such filters, which is used for the identification and extractions of series' trends.

As for the trading applications, a MA is the most widely used technical analysis techniques. Many research works found that some moving average and filter rules are profitable (Brock, Laconishok, Lebaron, 1992; Sullivan, Timmermann, White, 1999; Levich, Thomas,1993).

By its construction method, the MA is a weighted average. Two principal parameters should be identified before MA calculations: (1) the length of a sub-sample, for which the MA is estimated; (2) the MA weights, attached to each observation in the sub-sample. In addition, the manner in which the sub-sample is chosen should be decided in advance (for example, it can precede the estimation point, or it can include the estimation point, etc.). These parameters are responsible for two characteristics of the MA: (1) its smoothness; and (2) a lag, by which the MA is late in the prediction of the price movements. As many MA trading rules are based on the relative positioning of the price and MA curves[2], the smoothness of the MA curve is considered to be have direct impact on the number of false signals generated by the rule: smoother MA sends less false signals. For the same reason, the lag is responsible for the speed of the trading signal: for smaller lag the signal about trend reversion is sent more rapidly. The dilemma is that the smoother MA implies larger lag between the price and MA curves and vice versa. The trade-off between these characteristics lays in the basis of many MA optimization procedures.

The research in the field of the MA optimization can be split in the following groups:

(1) theoretical studies;
(2) simulation studies.

The first group of studies involves the search of optimal parameters, which is based on some theoretical approach or existing relationships in the market. For example, Achelis (2000) believed that the length of the MA should fit peak-to-peak cycle of a security price movement[3]. Gray, Thomson (1997) used the compromise criterion between the smoothness and lag to optimize the MA parameters. Ehlers (http://www.mesasoftware.com/papers) proposed a method to calculate the optimal weights for the MA as a function of the lag that a trader can tolerate[4]. Di Lorenzo, Sciarretta (1996) also defined the optimal MA parameters as a function of MA lag, which they define at the level that minimizes the number of false signals generated by MA. As the result they tried to develop adaptive moving average that takes into account the transaction costs and price volatility. Chande ([28], V.10:3) developed a Variable index dynamic average that also incorporate the notion of stochastic volatility in the definition of the exponential MA weights structure. Arms ([11], V.8:3) exploited the relationship between price and volumes in the construction of a volume-adjusted moving average. Arrington, [16], V.10:6) tried to incorporate data statistical characteristics, such as extreme values, into definition of the optimal MA length for a Variable length moving average.

---

[1] Detailed analysis of the linear filters can be found in Gençay, Selçuk, Whitcher (2002).
[2] Some rules are based on the relative positioning of the two MA of different lengths.
[3] See Appendix A1 for more details.
[4] See Appendix A2 for more details.

The simulation studies search for the optimal MA parameters through the simulation of a trading strategy or a rule for the historic data and analysis of the trading outcomes. Then the MA parameters that maximize these outcomes (or at least generate profits) are considered as optimal. The examples of such works are Williams (2006), Brock, Laconishok, Lebaron (1992), Sullivan, Timmermann, White (1999). Besides, all the "expert judgments" of traders regarding optimal MA types or parameters most likely are based on the experimental applications of these indicators to the historic data. The disadvantage of the approach is that the optimal parameters are conditional on the trading rule and strategy for which the simulations were performed.

The objective of our analysis is to introduce an optimisation method that takes into account the statistical characteristics of the data. We would also like to concentrate more on the optimisation of the MA weights, as many existing papers in both groups of studies are devoted to the optimisation of the MA length. Optimisation of the MA weights is more complicated task, as it involves the search of n interrelated values. We propose a kriging method, a geostatistical approach, to optimise the moving average weights.

The goal of geostatistics is to provide "quantitative descriptions of the natural variables distributed in space or in time and space"[5]. Soil properties or ore grades in a mineral deposit are the examples of such natural variables. The principal objective of the geostatistics is "the reconstruction of a phenomenon over domain on the basis of values observed at limited number of points"[6]. Kriging method is used for the trend estimation and can be applied to the problem of MA optimisation. The kriged estimator of a trend at some point is a weighted sum of the values in the near neighbourhood to this point, which evolves the direct comparison with the classical MA indicator. One of the objectives of the kriging estimation procedure is to find the optimal weights of the linear estimator of a variable mean.

There are several differences between financial and geostatistical data. Financial data is mainly sampled in time. Financial data samples are much larger than the geostatistical data sets. The objective of the financial analysis is mainly not a reconstruction of the phenomenon, but a prediction of the future price movements, for which the filtering of the trend is done.

For the time series data a formal analogy of the method that uses past values to predict future one are AR, ARMA, ARIMA models. The references in the domain are Greene (2007), Box, Jenkins, Reinsel (2008).

Financial data is usually treated as time-series data with values sampled at the at regular time intervals as when data is unevenly spaced, most of the methods used in the time signal processing cannot be applied. However, in reality, financial data is documented only at the moments when a transaction takes place. As the result, for less-actively traded instruments the data is not equally spaced. The data sampled at very high frequency (for example, 1 second) will also most likely be unequally spaced. Finally, when the time coordinate is changed to another variable coordinate (for example, volume), the data subordinated to another process would most likely be unevenly spaced. In the case of the unevenly spaced financial data, the geostatistic methods will bring better result than the classical time-series methods.

However, even for the equally spaced data the difference exists between kriging method and time-series models: the kriging approach does not demand completely specified model of the

[5] Chiles, J.-P. and P. Delfiner. "Geostatistics. Modeling Spatial Uncertainty", *John Wiley and Sons, Inc.*, 1999, p. 1.
[6] Chiles, J.-P. and P. Delfiner. "Geostatistics. Modeling Spatial Uncertainty", *John Wiley and Sons, Inc.*, 1999, p. 150.

process as in the case of the time series model; Only second-order properties are modelled for linear kriging.
The analysis of the kriging method is made in the following way. First, we present a definition and brief description of the MA as a trading instrument. Chapter 2 presents the kriging method. Chapter 3 discusses the peculiarities of method application to the financial data. Chapter 4, 5 and 6 analyse the results of the kriging method application to the equally and unequally spaced historic data. Chapter 7 summarizes the obtained results.

1 Moving average as a trading instrument

Moving averages (MAs) are one of the most widely used technical indicators by the traders. MAs lay in the basis of the many technical rules and strategies. They are also used to construct new technical indicators, such as Moving Average Convergence Divergence (MACD). Further in the Chapter 1 we present the definition and types of the MA. The trading strategies, which incorporate MA, are discussed in more details in Ch.4, 5 and 6 when applied to the historic data.

1.1 Definition of the MA

Let's $\{x_t\}_{-\infty\le t\le\infty}$ represent discrete time series (sampled at equal time intervals); while $\{x_t\}_{0\le t\le N-1}$ are the observed time series with N observations: x_0 represents the first observation and x_{N-1} - the last observation. A linear filter converts time-series $\{x_t\}_{0\le t\le N-1}$ into $\{y_t\}_{0\le t\le N-1}$ by some linear transformation (Gençay, Selçuk, Whitcher, 2002). The output $\{y_t\}$ is the result of the convolution of the vector x_t with a coefficient vector w_t:

$$y_t = w_t \circ x_t = \sum_{i=-\infty}^{\infty} w_i x_{t-i} \qquad \text{(I.1.1)}$$

Many applications are not feasible for $i < 0$ as it implies the usage of future x_t values. Therefore some restrictions might be imposed with respect to *i*-parameter:

$$y_t = \sum_{i=0}^{\infty} w_i x_{t-i} \qquad \text{(I.1.2)}$$

The filter (I.1.2) is called a *causal filter*, while (I.1.1) – a *non-causal filter*.

The other classifications are based on the impulse response of the filters:

1. Finite impulse response (FIR) filter
2. Infinite impulse response (IIR) filter

Further, we wil consider only FIR filters, defined as following:

$$y_t = \sum_{i=1}^{n} w_i x_{t-n+i} \qquad \text{(I.1.3)}$$

The majority of the moving averages used in finance are the representatives of the group of causal FIR filters. MAs can be constructed for any financial series. However, most frequently

time-series $\{x_t\}_{0\le t\le N-1}$ are either instrument prices or the indicators, derived from price (for example, Momentum, Moving average convergence divergence indicator, etc.). The example of the simple MA, a particular type of the MA with all equal weights ($w_i = 1/n$, where n is the length of the rolling window used for the MA calculations as in (I.1.3)) is given in Figure 1.1.

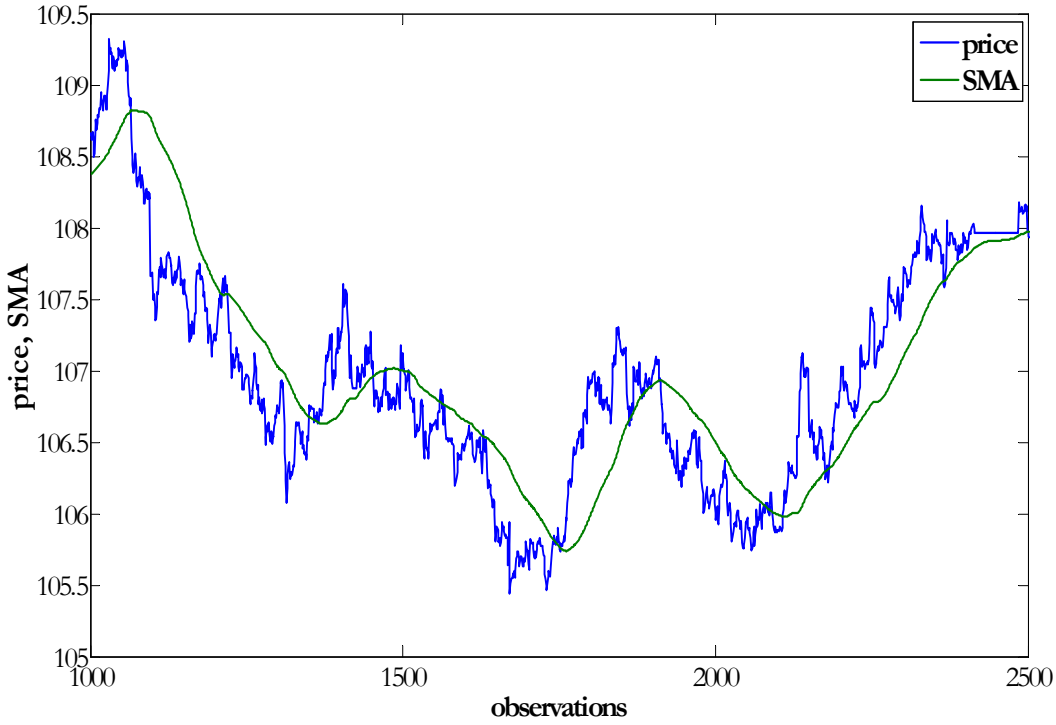


Figure 1.1. Bund quotes (30/7/2003-7/12/2006, frequency – 30 minutes) and simple moving average of the length n=115 observations

From the expression (I.1.3), we can see that MA has two parameters involved in its definition:

- n – size of the rolling window[7] (sub-sample) used for MA calculations: $[t-n+1;t]$
- $\{w_i\}_{1\le i\le n}$ - vector of the weights for each observation in rolling window;

In all technical analysis applications the following two constraints are imposed on weights:

$$\sum_{i=1}^{n} w_i = 1 \qquad \text{(I.1.4)}$$

$$w_i > 0 \qquad \text{(I.1.5)}$$

None of these constraints are general filters requirements, but they can be justified in some of the cases. The constraint (I.1.4), or universality condition, assures that the MA can be considered as an unbiased estimator of an instrument mean, which is important when mean is unknown. Positive weights (constraint (I.1.5)) assure that the MA, as an estimator of the mean of some always positive variable (e.g. price), does not reach negative values; this might happen if negative weights are attached to some extreme observations. Note that while the universality condition (I.1.4) is frequently applied by geostaticians, the constraint (I.1.5) is not used in the geostatistical applications.

[7] Further, the term "window" indicates the interval of data used for a moving average calculations.

MA defines a trend by smoothening the data, i.e. removing the higher frequency components from price data. Data smoothening of higher degree allows defining more clearly the existing (long-term) trend in the data. However, the negative side of the strong data smoothening is the elimination of short-term trends together with noise. As the results, the filtered data lag the original data. In financial applications the lag mean that MA does not anticipate the turning point in the trend - the signal about the change in the trend comes after the trend reversal has actually taken place. MA cannot predict local maximum or minimum of the price function, but only confirm it. The smoother is the data, the larger is the lag between MA and price curves. As have been mentioned in the introduction, this trade-off lies in the basis of some optimization approaches.

1.2 Types of the moving average

The weights vector defines the type of MA. Many types of MAs exist in the financial applications. We can split them in two large groups:

1. MA with fixed (constant) weights
2. MA with variable weights

MA with fixed weights assigns the same weights vector to all moving windows; weight value is only a function of the position within moving window: $w_{i,t} = f(i)$, where i – relative position of the weight within estimation (rolling) window; $i = 1$ corresponds to first observations in the window, $i = n$ - to the last observations. Taking into account that the interval $i \in [1:n]$ is the same for all time moments, the weights are considered as fixed. The examples of the MAs with fixed weights are given in Appendix B. The most widely used examples of this MA group are:

- simple moving average (SMA):

$$MA_t = \sum_{i=0}^{n-1} w_{n-i} P_{t-i} = \sum_{i=0}^{n-1} \frac{1}{n} P_{t-i} = \frac{1}{n} \sum_{i=0}^{n-1} P_{t-i}$$

  where $\{P_i\}_{i \geq 0}$ - price series; n - length of the window;

- exponential moving average (EMA):

$$EMA_t = \alpha P_t + (1-\alpha) EMA_{t-1}$$

  where $\{P_i\}_{i \geq 0}$ - price series; α - parameter of the EMA: $\alpha = \frac{2}{2+n}$, n - length of the SMA that has the same lag.

MA with variable weights assigns different weights vector in different time moments; weight is a function of other variables (indicators) at moment t: $w_{i,t} = f(T_t)$, where T_t - is the vector of indicators dependent on moment *t*. The traders try to adjust the MA to volatility, trading activity, etc. Among these MAs are Volume Adjusted Moving Average, Variable index dynamic average (VIDYA), Variable-length moving average (VLMA) (see Appendix C for more details on these MAs).

2 Kriging method: Theory

Kriging is part of the geostatistics used to estimate values of a random variable at some point, where it is not observed. Kriging can also be used to define a trend in the data, deterministic or random.

Matheron introduced the term "kriging" in 1970. The method emerge as the improvement of the moving average technique developed by D.G. Krige, a South African mining engineer, for estimating of gold grades; it was named after him (Armstrong, 2004). Kriging is a linear prediction method used to obtain unbiased and efficient (in terms of a variance) spatial estimate/predictor $Z(x_0)$ or the mean of the random process Z, from the available observations: $(Z(x_1), Z(x_2), ..., Z(x_n))$.

For example, in the mining context, the kriging is used to find the grade at some chosen point of space, taking into account the information about the grade available at the other points were the sampling has been done (Matheron, 1970). Kriging method attributes the weights to the grades of available sample deposit points, creating the weighted average estimate. The weights are chosen in a way that minimizes the estimation variance. The optimal weights take into account the geometric form of the deposit, the positioning of the available samples. Intuitively, we might assume that the sample points distant from the prediction point should have less weight than the closer one. However, there are more complex phenomena in mining that in some cases contradicts the intuition. Therefore, kriging method takes into account not only geometrical form of the deposit and how the sampling points are positioned, but also the statistical characteristics of a random variable Z, such as covariance (or variogram) functions.

The distinctive features of the kriging approach are the following:

1. Kriging method is used for spatial estimation. This implies the following data peculiarities:
   - data frequently cannot be defined as "past" and "future";
   - data is frequently unevenly sampled;
   - data has "continuous rather than discrete location indexing space"[8].
2. Contrary to trend estimation methods that need the predefined deterministic function, kriging estimates of the trend is based on the statistical characteristics of the data.
3. Besides deterministic trends, kriging can also estimate the random trend in the data.
4. The kriged estimates avoid systematic error, caused by the difference between samples empirical and "true" statistical characteristics. Matheron (1970) explains that the histogram of the real grade of the deposit contains less extreme values and more intermediate values than the experimental histogram built on the analysed sample, which often cause underestimation of a mean of the deposit grade. The kriging procedure allows avoiding the underestimation error.

The kriging approach is developed within the scope of second-order statistical models that use only mean and covariance (or variogram[9]) model. Contrary to other trend estimation methods, kriging approach is subject to fewer constraints:

1. None assumption about distribution properties of the random variables are made.
2. Kriging does not define a priori the function that represents the trend.

[8] Chiles, J.-P. and P. Delfiner. "Geostatistics. Modeling Spatial Uncertainty", *John Wiley and Sons, Inc.*, 1999, p. 151.
[9] The term is presented in more details further in Ch.2.1.

3. Some of the kriging types (universal kriging) allows addressing the problem of random trends.
4. Kriging allows representing trend in the form of a weighted average that can be easily translated into a MA.
5. Almost all kriging types provide not only the estimates of the trend, but also the variance of the estimators that allows building the confidence intervals for the estimators.

The following types of the kriging exist:

1. Simple kriging (SK).
   Simple kriging is applied when the mean of a process is constant and known. The process should be stationary; covariance model is used to derive the kriging estimates.

$$Z(x) = m_{known} + Y(x)$$

2. Ordinary kriging (OK).
   Ordinary kriging is applied when the mean is constant, but unknown. In this case the process is intrinsic[10]; variogram model is used to derive the kriging estimates.

$$Z(x) = m + Y(x)$$

3. Universal kriging (UK)
   Universal kriging is applied when mean is variable and unknown. The variability of trend can have either deterministic or random nature.

$$Z(x) = m(x) + Y(x),\ m(x) \text{ - deterministic trend}$$
$$Z(x,w) = m(x,w) + Y(x,w),\ m(x,w) \text{ - random trend}$$

The application of the simple kriging in our context is limited due to a small amount of financial instruments, which can be considered as stationary with known mean. The exceptions are the technical indicators derived from the price data, such as Momentum or MACD that are oscillators by nature and thus fluctuates around 0-line.

Ordinary and universal kriging has much more applications possibilities as they can be applied to both second-order stationary and non-stationary data.

Further we present each of the methods in more details[11]. We start, however, this chapter with the description of some of the geostatistical terminology and concepts, used further in kriging applications.

2.1 Geostatistic instruments and terminology: short overview

Kriging involves some common statistical definition like stationarity, covariances, as well as some instruments less known for a wider statistical community, such as intrinsic stationarity and variograms.

[10] The term is presented in more details further in Ch.2.1.

[11] Further the description of the theoretical approach is based on the G.Matheron's book "La théorie de variables régionalisées, et ses applications", Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fascicule 5, 1970, pp.117-186, adjusted for notation to our particular case. We will avoid additional citations to simplify the reading of the text.

2.1.1 Stationary and intrinsic stationary functions

Any random variable Z can be characterized by its probability distribution. However, an estimation of the probability distribution function is sometimes complicated and time-consuming. Thus, statisticians frequently substitute probability analysis by the calculations of the statistical moments, such as mean and variance-covariance matrix.
Random functions are more complicated statistical entities, as they represented by the sets of random variables: $Z(t_1), Z(t_2),..., Z(t_n)$. The main problem with the estimation of their statistical characteristics lies in the lack of observations: in real life we often have a single realization for each variable: $z(t_1), z(t_2),..., z(t_n)$. Thus, in order to make some valuable statistical inferences about the random function from these realizations, it is often assumed that variable is stationary.

Stationarity imposes the invariance of a joint probability density function and, therefore, all its moments under temporal (or spatial) shift (translation), i.e. for any value *h*, probability distribution of the variables $Z(t_1), Z(t_2),..., Z(t_n)$ is the same as of the variables $Z(t_1 + h), Z(t_2 + h),..., Z(t_n + h)$. For the temporal data it means that all moments of the random variable does not depend on the time at which the variables are observed, but only on the distance between them. Simply speaking, "information about the process [is] the same no matter where it is obtained"[12].

However, even under stationarity assumptions a joint probability distribution is often difficult to estimate; therefore, the statisticians has introduced second-order stationarity - the invariance of the first and second moments under translation (see Definition 1.1).

Definition 1.1
A random function $\{Z(t)\}$ is stationary of order two if its mean and covariance do not depend on time, but only on the distant between the variable, i.e. $\forall t, h$:

$$\mathrm{E}[Z(t)] = m \qquad \text{(I.2.1)}$$

$$Cov[Z(t), Z(t+h)] = \mathrm{E}[Z(t)Z(t+h)] - m^2 = C(h) \qquad \text{(I.2.2)}$$

The covariance (I.2.2) has the following properties:

1. $C(0) = \sigma^2$ - variance of the random variable
2. $C(h) = C(-h)$
3. $|C(h)| \le C(0)$

Note that the covariance for stationary variable or order two is bounded.

Unfortunately, the group of the second-order stationary processes is not very large. In financial applications these processes are even more rare. For example, standard Brownian motion does not belong to the group of second-order stationary process. Therefore, the larger group of the random functions was introduced to enlarge the group of random processes for which many geostatistical methods, including kriging, can be applied. This group consists of *intrinsic functions* (see Definition 1.2).

Definition 1.2
A random function $\{Z(t)\}$ is intrinsic if its increments are stationary; i.e. $\forall t, h$

$$\mathrm{E}[Z(t+h) - Z(t)] = 0 \qquad \text{(I.2.3)}$$

[12] Anselin L. "Variogram analysis", presentation

$$Var[Z(t+h)-Z(t)] = \mathrm{E}\left[(Z(t+h)-Z(t))^2\right] = 2\gamma(h) \qquad \text{(I.2.4)}$$

The group of the intrinsic processes is significantly larger. In particular, it contains the group of the stationary processes of order two, i.e. each stationary process of order two is intrinsic stationary (see Box 1.1. for the proof); but not every intrinsic stationary process is stationary of order two (see Box 1.2. for the example).

Box 1.1
The proof that each process, stationary of order two is intrinsic stationary process

Suppose some process $\{X_t, t \in [0,T]\}$ is stationary of order two. Then this implies the following equalities:

$$\mathrm{E}(X_t) = m$$
$$Var(X_t) = v^2$$
$$Cov(X_t, X_{t+h}) = C(h).$$

Lets consider its increments: $\forall t > 0, \forall h > 0 : X_{t+h} - X_t$. Then the following equalities hold:

$$\mathrm{E}(X_{t+h} - X_t) = 0$$
$$Var(X_{t+h} - X_t) = Var(X_{t+h}) + Var(X_t) - 2Cov(X_{t+h}, X_t) = 2v^2 - 2C(h) = \gamma(h)$$

Therefore, $\{X_t, t \in [0,T]\}$ is intrinsic.

Box 1.2
Intrinsic stationary process: Standard Brownian motion

A random process $\{W_t, t \in [0,T]\}$ is a standard Brownian motion if:

1. $\forall t > 0, \forall s > 0 : W_{t+s} - W_t \propto N(0,s)$ (normally distributed with 0-mean an s - variance)
2. $\forall t > 0, \forall s > 0 : W_{t+s} - W_t$ is independent of W_t.
3. W_t is a continuous function of time and $W_0 = 0$.

As we can see from definition the increments of the standard Brownian motion $\forall t > 0, \forall s > 0 : W_{t+s} - W_t$ are stationary of order 2:

1. $\mathrm{E}[W_{t+s} - W_t] = 0$
2. $Var[W_{t+s} - W_t] = s$
3. $Cov[W_{t+s} - W_t, W_t] = 0$

Therefore, $\{W_t, t \in [0,T]\}$ is intrinsic stationary. However, we know that covariance of the process $\{W_t\}$, $E[W_t W_\tau] = \min(t, \tau)$ is time-dependent. Thus, the process itself $\{W_t, t \in [0,T]\}$ is not stationary of order 2. As the result, we have showed that an intrinsic process is not necessary a stationary process of order two.

Intrinsic random functions of order k (IRF-k) are the generalization of the intrinsic functions. "The IRF-k is a random function with stationary increments of order k"[13]. IRF-0 is the intrinsic function with stationary increments. IRF-k enlarge the group of the processes for which the kriging method can be applied. The analogy of these functions in the time series is the ARIMA

[13] Chiles, J.-P. and P. Delfiner. "Geostatistics. Modeling Spatial Uncertainty", *John Wiley and Sons, Inc.*, 1999, p. 245.

processes: ARIMA of order d is, in fact, IRF-$(d-1)$. There are still differences between ARIMA and IRF-k models. Firstly, ARIMA models are discrete, while IRF can be both continuous and discrete. Secondly, ARIMA models are completely specified, while IRF should only be second-order model. Finally, ARIMA are one-dimensional, while IFR can be defined in R^n.

In this work we concentrate mainly on stationary and IRF-0 models. Although IRF-k models allow working with non-stationary data, the experience has shown that in the practical applications too much information about variable is lost when the estimation procedures use its stationary increments. Similar conclusions were the incentive for the development of cointegration.

2.1.2 Variogram

Definition of the intrinsic functions is based on the variance of increments. This variance is called variogram, a concept widely used in geostatistics.

Definition 1.3
The following function $\gamma(h)$ is called semi-variogram, or less formally variogram:

$$\gamma(h) = \frac{1}{2} Var[Z(t+h) - Z(t)] = \frac{1}{2} \mathrm{E}\left[(Z(t+h) - Z(t))^2\right] \qquad \text{(I.2.5)}$$

Expression (I.2.5) indicates a significant advantage of the variogram over the covariance: its definition does not involve variable mean that is usually unknown and should be estimated.

For stationary variables the following relationship exists[14] between variogram and covariance:

$$\gamma(h) = C(0) - C(h) = \sigma^2 - C(h) \qquad \text{(I.2.6)}$$

This relationship is represented graphically in Figure 1.2.

A variogram has the following characteristics:

1. $\gamma(0) = 0$.
2. It can be discontinuous just after the origin (so called nugget effect).
3. Variogram is bounded for the stationary variables of order two and tend to be increasing for non-stationary variable.

Several parameters characterize a variogram (see Figure 1.3). Sill is a level, by which the variogram is bounded. A lag, at which the variogram is stabilized around the sill level, called range. The range indicates the lag at which there is no more correlation between samples (no autocorrelation for the time dependent random functions). If discontinuity is present at the origin, it is called nugget effect.

[14] The formulae (I.2.5) can be rewritten as following:

$$\gamma(h) = \frac{1}{2} \mathrm{E}\left[(Z(t+h) - Z(t))^2\right] = \frac{1}{2} \mathrm{E}\left[(Z(t+h) - m)^2 + (Z(t) - m)^2 - 2(Z(t+h) - m)(Z(t) - m)\right] =$$

$$= \frac{1}{2}\left[2C(0) - 2C(h)\right] = \sigma^2 - C(h)$$

Appendix D presents the most frequently used variograms $\gamma(h)$ and covariance $C(h)$ models.

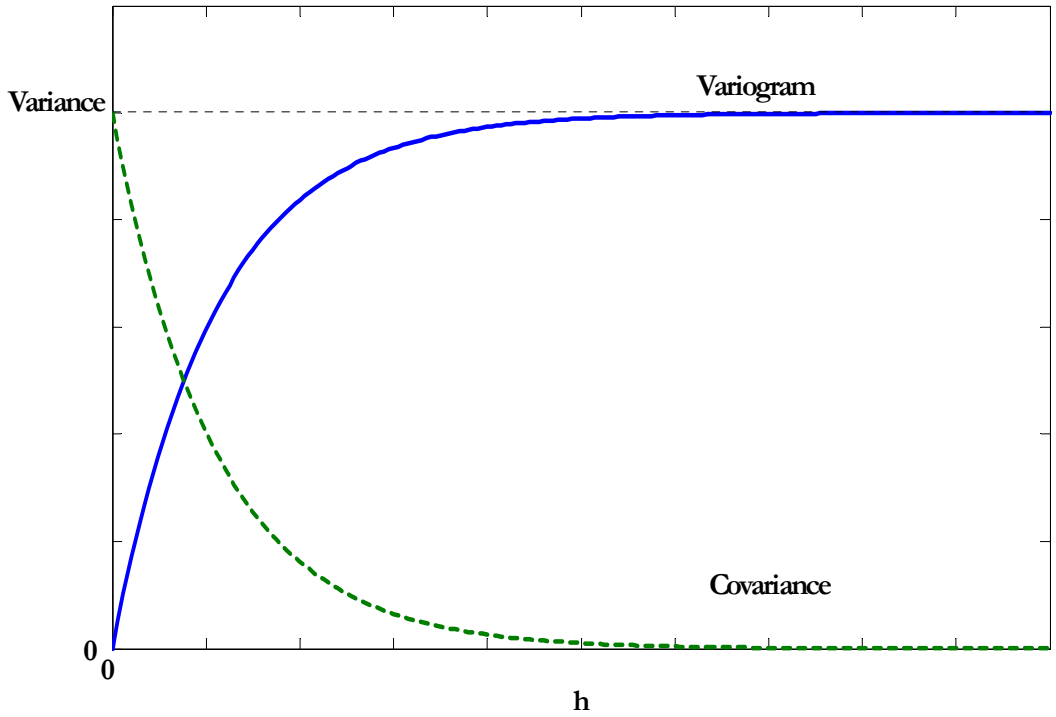


Figure 1.2. Relationship between variogram and covariance for stationary variable.

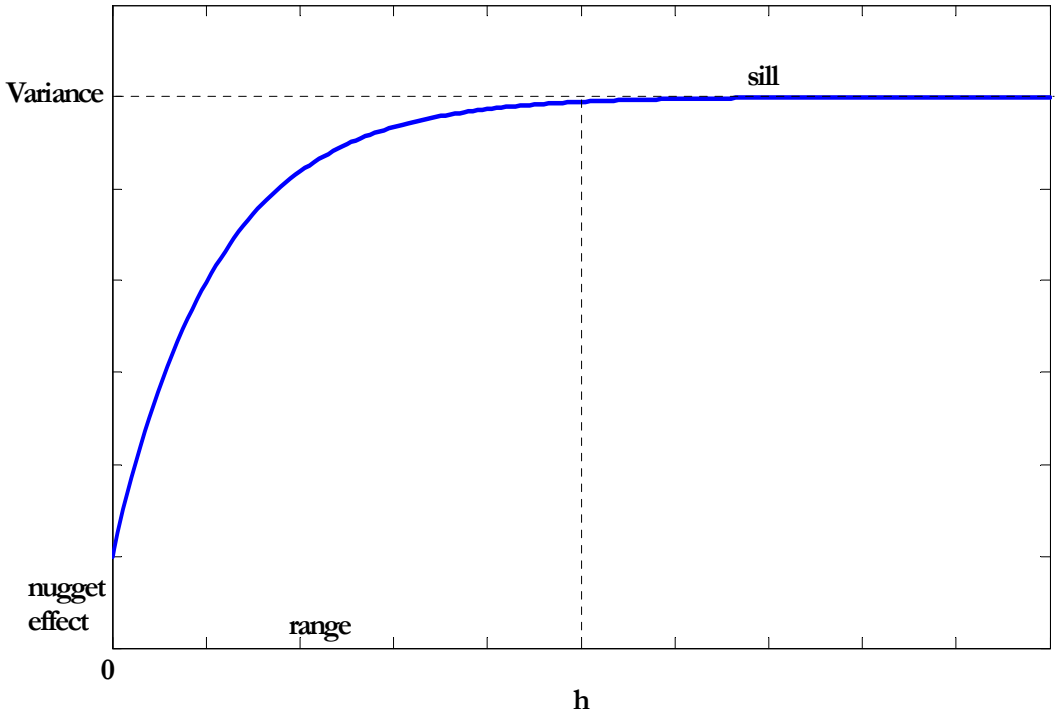


Figure 1.3. Variogram parameters

According to Chilès, Delfiner (1999), there are two ways of fitting model to the empirical variogram:

1. Manual fitting
2. Automatic fitting

Frequently geostatistician fit the model manually as variograms are non-linear in their parameters, such as range (Matheron, 1968). Automatic fitting can be performed by the least square technique: ordinary least square, generalized least square or weighted least square (Cressie, 1991). As many observations are available, further in the applications we will use manual approach to variogram fitting.

It should be stressed that missing observations cannot be ignored in a variogram estimations. If data is missed due to a regular absence of the trading activity (for example, week-ends or night hours) this missing data can be ignored. In order to avoid an overnight effect (the absence of the overnight data) an empirical variogram can be substituted by the average daily variogram. However, if data is missing due to the irregular activity (for example, holiday), the missing data should be treated as non-available and differences in the variogram formulae that incorporate these observations should not be taken into account.

2.2 Simple Kriging: prediction of the process with zero mean

The kriged estimator Y_{n+1}^K is the orthogonal projection of the Y_{n+1} on a Hilbert space $H(s)$, i.e. Y_{n+1}^K is a unique element of the $H(s)$, such that $Y_{n+1}^K - Y_{n+1}$ is orthogonal to all other $Y \in H(s)$. In our case, the Hilbert space $H(s)$ represents the linear span of the available known points.

Let's consider some random function $Y(t)$ with zero expectation ($\mathrm{E}[Y(t)] = 0$) and variance-covariance matrix $\sigma_{uv} = \mathrm{cov}[Y(u), Y(v)]$[15]. Let Y_α be realizations of the random variable $Y(t)$ at some experimental points $t_\alpha = \{t_i\}_{1 \le i \le n}$. Lets represent the kriged estimator of the $Y(t)$ at point t_{n+1} as a linear combination of Y_α:

$$Y_K(t_{n+1}) = \lambda_K^\alpha Y_\alpha = \sum_{i=1}^{n} \lambda_{K,i} Y_i \qquad \text{(I.2.7)}$$

The weights vector λ_K^α minimizes the estimation variance $\mathrm{var}[(Y_{n+1} - Y_K)] = \mathrm{E}[(Y_{n+1} - Y_K)^2]$[16]:

---

[15] The variable can be a function of any variable (time, volume, etc.). We use index *t* that corresponds to the time coordinate, as this is the case most often addressed by the technical analysis. However all kriging methods can be applied to other cases.

[16] Note that $\mathrm{E}[(Y(t) - Y_K(t))] = 0, \forall t$ and

$$\mathrm{E}[(Y_{n+1} - Y_K)^2] = \mathrm{E}(Y_{n+1}^2) - 2E(Y_{n+1} Y_K) + \mathrm{E}(Y_K^2) = \sigma_{n+1}^2 - 2\sum_{i=1}^{n} \lambda_i \mathrm{E}(Y_{n+1} Y_i) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \mathrm{E}(Y_i Y_j) =$$

$$= \sigma_{n+1}^2 - 2\sum_{i=1}^{n} \lambda_i \sigma_{n+1,i} + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \sigma_{i,j}$$

$$\mathrm{E}\left[\left(Y_{n+1}-Y_K\right)^2\right]=\sigma_{n+1}^2-2\sum_{i=1}^{n}\lambda_i\sigma_{n+1,i}+\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\sigma_{i,j} \qquad \text{(I.2.8)}$$

Taking partial derivatives of the function (I.2.8) with respect to the weights vector λ_K^α gives the following system of *n*-equations with *n*-unknowns:

$$\lambda_K^\beta\sigma_{\alpha\beta}=\sigma_{\alpha,n+1}, \qquad \text{(I.2.9)}$$

where *i*-th equation is:

$$\sum_{j=1}^{n}\lambda_j\sigma_{i,j}=\sigma_{n+1,i}$$

The system (I.2.9) is regular and have unique solution if matrix $\sigma_{\alpha\beta}$ is strictly positive definite, which is usually assumed for theoretical variance-covariance matrix.

Kriging variance, defined in (I.2.8) takes the following form[17]:

$$\sigma_K^2=\mathrm{E}\left[\left(Y_{n+1}-Y_K\right)^2\right]=\sigma_{n+1}^2-\sum_{i=1}^{n}\lambda_i\sigma_{n+1,i} \qquad \text{(I.2.10)}$$

2.3 Simple kriging: prediction for the process with known mean

Let Z be some random function with known mean $\mathrm{E}(Z)=m$ and variance-covariance matrix $\sigma_{ij}=\mathrm{cov}\left[Z_i,Z_j\right]$.

Lets define some new random variable Y as following:

$$Y=Z-\mathrm{E}(Z)=Z-m$$

Y has zero mean and σ_{ij} variance-covariance matrix.

Then the optimal estimator of $Z(t_{n+1})$ is the kriged estimator Z_{n+1}^K of the following form:

$$Z_{n+1}^K=m+Y_{n+1}^K=m+\lambda_K^\alpha\left(Z_\alpha-m\right) \qquad \text{(I.2.11)}$$

where λ_K^α are the kriged weights that satisfies kriging system (I.2.9) for Y.

The variance of kriging is the same as in (I.2.10)[18]:

---

[17] The equation (I.2.9) implies that: $\lambda_K^\alpha\lambda_K^\beta\sigma_{\alpha\beta}=\lambda_K^\alpha\sigma_{\alpha,n+1}$, or $\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\sigma_{i,j}=\sum_{i=1}^{n}\lambda_i\sigma_{n+1,i}$. Then substituting this expression in (I.2.8) leads to (I.2.10).

$$\sigma_K^2 = \mathrm{E}\left[\left(Z_{n+1} - Z_{n+1}^K\right)^2\right] = \mathrm{E}\left[\left(Y_{n+1} - Y_K\right)^2\right] = \sigma_{n+1}^2 - \sum_{i=1}^{n} \lambda_i \sigma_{n+1,i} \qquad \text{(I.2.12)}$$

2.4 Ordinary kriging: prediction of the process with unknown mean

Ordinary kriging is applied to the random function Z, when its mean is unknown. Suppose $Z(t)$ is stationary, i.e. its mean m is unknown, but constant: $\mathrm{E}[Z(t)] = m$. $\sigma_{uv} = \mathrm{cov}[Z(u), Z(v)]$ represents its variance-covariance matrix. Let Z_α be realizations of the random variable $Z(t)$ at some experimental points $t_\alpha = \{t_i\}_{1 \le i \le n}$. A kriged estimator of the $Z(t)$ at point t_{n+1} is a linear combination of Z_α:

$$Z_K(t_{n+1}) = \lambda_K^\alpha Z_\alpha = \sum_{i=1}^{n} \lambda_i Z_i \qquad \text{(I.2.13)}$$

The optimal weights λ_K^α should minimize the estimation variance $\mathrm{var}[(Z(t_{n+1}) - Z_K(t_{n+1}))]$. As mean is unknown the kriged value Z_{n+1}^K should be an unbiased estimator of the $Z(t_{n+1})$, i.e. the mean of the estimation error should be zero:

$$\mathrm{E}[(Z(t_{n+1}) - Z_K(t_{n+1}))] = 0 \qquad \text{(I.2.14)}$$

Taking into account that $\mathrm{E}(Z(t_{n+1})) = m$ and $\mathrm{E}(Z_K(t_{n+1})) = m\sum_{i=1}^{n} \lambda_i$ condition (I.2.14) takes the following form:

$$\mathrm{E}[(Z(t_{n+1}) - Z_K(t_{n+1}))] = m\left(1 - \sum_{i=1}^{n} \lambda_i\right) = 0 \qquad \text{(I.2.15)}$$

The equality (I.2.15) holds only if $m = 0$ (the case of simple kriging), or

$$\sum_{i=1}^{n} \lambda_i = 1 \qquad \text{(I.2.16)}$$

Condition (I.2.16) is called an universality condition.

As the result, the kriging problem can be reformulated as the following optimization problem:

---

[18] The equation (I.2.11) implies that $\mathrm{E}\left(Z_{n+1}^K\right) = m$; therefore $\mathrm{E}\left[\left(Z_{n+1} - Z_{n+1}^K\right)\right] = 0$ and $\mathrm{var}\left[\left(Z_{n+1} - Z_{n+1}^K\right)\right] = \mathrm{E}\left[\left(Z_{n+1} - Z_{n+1}^K\right)^2\right]$.

$$\begin{cases} \min\limits_{\lambda}\left(\mathrm{var}\left[\left(Z(t_{n+1})-Z_K(t_{n+1})\right)\right]\right) \\ \sum\limits_{i=1}^{n}\lambda_i = 1 \end{cases} \qquad \text{(I.2.17)}$$

If universality condition (2.16) holds then $\mathrm{E}\left[\left(Z(t_{n+1})-Z_K(t_{n+1})\right)\right]=0$ and $\mathrm{var}\left[\left(Z(t_{n+1})-Z_K(t_{n+1})\right)\right]=\mathrm{E}\left[\left(Z(t_{n+1})-Z_K(t_{n+1})\right)^2\right]$.

With respect to this we can rewrite (I.2.17) as following:

$$\begin{cases} \min\limits_{\lambda}\left(\sigma_{n+1}^2-2\sum\limits_{i=1}^{n}\lambda_i\sigma_{n+1,i}+\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}\lambda_i\lambda_j\sigma_{i,j}\right) \\ \sum\limits_{i=1}^{n}\lambda_i = 1 \end{cases} \qquad \text{(I.2.18)}$$

The solution to the problem (I.2.18) should satisfy the following system of equations:

$$\begin{cases} \sum\limits_{j=1}^{n}\lambda_j\sigma_{ij} = \sigma_{i,n+1}+\mu \\ \sum\limits_{j=1}^{n}\lambda_j = 1 \end{cases} \qquad \text{(I.2.19)}$$

where μ - is a Lagrange multiplier.

The variance of kriging is[19]:

$$\sigma_K^2 = \sigma_{n+1}^2+\mu-\sum_{i=1}^{n}\lambda_i\sigma_{n+1,i} \qquad \text{(I.2.20)}$$

2.5 Ordinary Kriging: Estimation of the unknown mean

Let $Z(t)$ be some random function, stationary of order two. Its mean m is unknown, but constant: $\mathrm{E}\left[Z(t)\right]=m$. $\sigma_{uv}=\mathrm{cov}\left[Z(u),Z(v)\right]$ represents its variance-covariance matrix. Let Z_α be realizations of the random variable $Z(t)$ at some experimental points $t_\alpha=\{t_i\}_{1\le i\le n}$. According to Matheron (1970) the estimator of the m as a linear combination of the available observations Z_α:

$$m^* = \lambda_m^\alpha Z_\alpha = \sum_{i=1}^{n}\lambda_i^m Z_i \qquad \text{(I.2.21)}$$

[19] The expression follows from (I.2.19): $\lambda^\alpha\lambda^\beta\sigma_{\alpha\beta} = \lambda^\alpha\sigma_{\alpha,\,n+1}+\mu\sum\limits_{\alpha}\lambda_\alpha = \lambda^\alpha\sigma_{\alpha,\,n+1}+\mu$.

Note that the weights vector λ_m^α is not the same as vector weights λ_K^α, used to define the predictor of the function $Z(t)$.

In order to assure the unbiased mean estimator, we impose the universality condition:

$$\sum_{i=1}^{n} \lambda_i^m = 1.$$

To obtain the efficient estimator, we minimize the variance of the estimation error:

$$\mathrm{var}\left[\left(m - m^*\right)\right] = \mathrm{var}\left[\left(m^*\right)^2\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i^m \lambda_j^m \sigma_{ij} \qquad \text{(I.2.22)}$$

As the result, the search of the optimal estimator (I.2.21) is reduced to the following optimization problem:

$$\begin{cases} \min\limits_{\lambda}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} \lambda_i^m \lambda_j^m \sigma_{i,j}\right) \\ \sum\limits_{i=1}^{n} \lambda_i^m = 1 \end{cases} \qquad \text{(I.2.23)}$$

The optimal weights λ_m^α should satisfy the following system of equations:

$$\begin{cases} \sum\limits_{j=1}^{n} \lambda_j^m \sigma_{ij} = \mu_m \\ \sum\limits_{j=1}^{n} \lambda_j^m = 1 \end{cases} \qquad \text{(I.2.24)}$$

where μ_m - is a Lagrange multiplier.

The corresponding kriging variance is[20]:

$$\sigma_K = \mu_m \qquad \text{(I.2.25)}$$

2.6 Kriging non-stationary variable

There are several approaches in the geostatistics to the kriging of a non-stationary variable:

1. Universal kriging
2. Kriging an intrinsic function IRF-0.
3. Kriging an intrinsic function IRF-k.

---

[20] Note that the following expression follows from (I.2.24): $\lambda^\alpha \lambda^\beta \sigma_{\alpha\beta} = \mu_m \sum_\alpha \lambda_\alpha = \mu_m$.

We will further present only first two approaches. As have been mentioned already in Ch.2.1.1, the kriging of the IRF-k leads to the loss of some information about the principal non-stationary variable. In addition, the main objective of out study is to define a drift of our (non-stationary process). The problem is that in general IRF-k has "no uniquely definable drift"[21], for the exception of the case when the analyzed process $Z(x)$ can be represented as the stationary process $Y_{st}(x)$ plus some polynomial drift:

$$Z(x) = Y_{st}(x) + \sum_l A_l x^l \qquad \text{(I.2.26)}$$

This representation (I.2.26) coincides with the universal kriging model with random coefficients (Chilès, Delphiner, 1999).

2.6.1 Universal kriging: trend estimation

Let $Z(t)$[22] be some random function, which is non-stationary with unknown mean m. $\sigma_{uv} = \mathrm{cov}[Z(u), Z(v)]$ represents its variance-covariance matrix:

$$\begin{cases} \mathrm{E}[Z(t)] = m(t) \\ \mathrm{E}[Z(t_1)Z(t_2)] = m(t_1)m(t_2) + \sigma(t_1, t_2) \end{cases} \qquad \text{(I.2.27)}$$

The universal kriging (UK) provide the best linear estimator of the trend $m_t = \mathrm{E}[Z(t)]$. We can consider the function $m(t)$ to be regular and continues or irregular and a random variable itself.

There are the following hypothesis on which the estimation method is based:

1. The $m(t)$ function is estimated locally and it can be approximated by the following expression:

   $$m(t) = \sum_{l=0}^{k} a_l f^l(t),$$

   where $f^l(t)$ a function, chosen a priori and fixed through all applications (for example, time polynomial); a_l - unknown coefficient corresponding to *l*-function (might be a function of *t*.

2. We suppose that the covariance (variogram) between two points of time can be estimated locally as $\gamma(t_1, t_2) = \omega\gamma_0(|t_2 - t_1|)$ and it is deforming slowly. The parameters of the covariance (variogram) function should be estimated and further controlled, which is quite complicated task. However, to simplify things, we suppose that the parameters are known a priori.
3. $Z(t)$ is some random non-stationary function that satisfy conditions (I.2.27). Let Z_α be the realization of the random variable $Z(t)$ at some experimental points $t_\alpha = \{t_i\}_{1 \le i \le n}$.

Suppose that $Z(t)$ can be represented as the following process:

---

[21] Chiles, J.-P. and P. Delfiner. "Geostatistics. Modeling Spatial Uncertainty", *John Wiley and Sons, Inc.*, 1999, p. 270.

[22] We define the universal method for the time coordinate, though as in the previous defintion of the kriging t-variable can be substituted by the spatial *x*-coordinates.

$$Z(t,w) = Y(t,w) + m(t,w), \qquad \text{(I.2.28)}$$

where $m(t,w)$ - trend of the series, random function; $Y(t,w)$ - random function, such that $E[Y(t)] = 0$, some *Y* are correlated.

Although it is not necessary, lets assume that $Y(t,w)$ and $m(t,w)$ are independent, in order to divide the structural effects on process $Z(t)$ from each of them. Furthermore, $m(t,w)$ should by its nature have much more larger effect on the $Z(t)$ and much more smaller volatility than $Y(t,w)$. The regularity and continuity of the $m(t,w)$ is attained by the following assumption:

$$m(t,w) = a_l(w) f^l(t) = \sum_{i=0}^{l} a_i f^i(t), t \in V,\ f^0(t) = 1 \qquad \text{(I.2.29)}$$

where $a_l(w)$ are random variables that reflect the coefficients of the trend; $f^l(t)$ are polynomials of the order *l* that defines the form of the trend. Lets assume for simplicity reason that the trend is described by the linear function:

$$m(t,w) = a_o(w) + a_1(w)t, t \in V \qquad \text{(I.2.30)}$$

For the trend estimation, lets assume that Z_α are the values of $Z(t)$ at the experimental points $t_\alpha \in S \subset V$, with *S* bounded.

Lets assume that the estimator of the trend $m(t)$ at one point $t \in V$ is the following:

$$M^*(t) = \lambda_{UK}^{\alpha} Z_{\alpha(n)} = \sum_{i \in \alpha} \lambda_i Z(t_i) = \sum_{i \in \alpha} \lambda_i Y(t_i) + \sum_{i \in \alpha} \lambda_i m(t_i), \qquad \text{(I.2.31)}$$

The optimal weights vector λ_{UK}^{α} minimizes the estimation variance:

$$E\left(M_t^* - m_t\right)^2 = \lambda^\alpha \lambda^\beta \sigma_{\alpha\beta} = \sum_{i \in \alpha(n)} \sum_{j \in \alpha(n)} \lambda_i \lambda_j \sigma_{ij}\ , \qquad \text{(I.2.32)}$$

with $\sigma_{ij} = \mathrm{cov}(Y(t_i), Y(t_j)) = E(Y(t_i) \cdot Y(t_j))$

The universality constrain for the minimization problem (I.2.32) is:

$$\sum_{\alpha} \lambda_\alpha f_\alpha^l = f_0^l \qquad \text{(I.2.33)}$$

Let Σ represent variance covariance matrix of Y, with the elements σ_{ij}; F - the vector of simple functions in the form: $f = \begin{bmatrix} 1 & t & t^2 \ldots \end{bmatrix}$, $F = \begin{bmatrix} 1 & t_1 & t_1^2 & \ldots \\ 1 & t_2 & t_2^2 & \ldots \\ & \ldots & & \ldots \\ 1 & t_n & t_n^2 & \ldots \end{bmatrix}$.

Then the minimization problem (I.2.32)- (I.2.33) in matrix representation is:

$$\begin{bmatrix} \Sigma & F \\ F' & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ f_0 \end{bmatrix} \qquad \text{(I.2.34)}$$

The UK variance of the mean estimator:

$$\sigma_{UK} = \mathrm{E}\left[\left(M^*_{t_0} - m_{t_0}\right)^2\right] = -\mu' f_0 = f_0{}' \left(F' \Sigma^{-1} F\right)^{-1} f_0 \qquad \text{(I.2.35)}$$

It should be mentioned here that for a finite case the UK model could be regarded as a linear regression model with the correlated residuals of the following (matrix) form (Chilès, Delphiner, 1999):

$$Z = Fa + Y$$

Then the generalized least square (GLS) solution for the optimal (unbiased, efficient) estimator of the coefficients vector a will be:

$$a^* = \left(F' \Sigma^{-1} F\right)^{-1} F' \Sigma^{-1} Z$$

The covariance of the estimated residuals is:

$$\mathrm{E}\left[\left(Z - Fa^*\right)\left(Z - Fa^*\right)'\right] = \Sigma - F\left(F' \Sigma^{-1} F\right)^{-1} F' \qquad \text{(I.2.36)}$$

The expression (I.2.36) shows that the covariance of the estimated residuals is a biased estimate of the covariance of the true residuals Y.

2.6.2 Kriging intrinsic function IRF-0

The problem with the intrinsic function IRF-0 is that the constant term of the drift coefficient a_0 cannot be determined, since the random function is determined by its increments (variogram). In order to avoid this problem, we can suppose that over some limited domain (for example, time-interval) for some very large $A > 0$, function $A - \gamma(h)$ represent the covariance. Substituting the new covariance into the optimization problem (2.32)-(2.33) we will obtain similar equation system that does not depend on A:

$$\begin{bmatrix} \Gamma & F \\ F' & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ f_0 \end{bmatrix} \qquad \text{(I.2.37)}$$

where $\Gamma = \lfloor \gamma_{\alpha\beta} \rfloor$ is the variogram matrix., which is strictly conditionally negative definite. This condition is met for a valid variogram models.

Contrary to the UK approach with covariance, the kriging variance of the mean estimate cannot be defined.

3 Peculiarities of the kriging method applications in finance

As have been mentioned already kriging method is developed to confront the spatial data. In this chapter we discuss how the difference in the financial and spatial data might have impact on the financial applications of the kriging method.

Mean value of geostatistic data is estimated for some sample that forms a close neighbourhood to the estimation point. For the financial data, only past observations are available, therefore the close neighbourhood is formed by the sub-sample that precedes the estimation point. If we choose some n –value for the length of the sub-sample that precedes and contain the estimation point and consider it as the "close neighbourhood", then the kriged estimator of a variable mean at each moment of time will coincide with the definition of the weighted MA in the (I.1.3).

There are two principal peculiarities of the financial data:

1. The majority of financial samples are non-stationary due to the presence of trends in the data.
2. Many financial variables are sampled at equal distances. This is true in particular for a low frequency data (for example, daily, monthly, annual observations).

3.1 Data non-stationarity

As have been shown in Chapter 2, all kriging approaches are based on the second-order moments of the process Z or residuals Y that supposed to be known. In real life, however, the covariance (variogram) should be estimated first and a valid model should be fit to the estimates.

In the case of the stationary instrument, the raw variogram can be easily estimated and fitted. However, the presence of a drift in the data introduces a bias into the estimates of the raw variogram. Therefore, in order to obtain the best possible estimate of the true variogram the drift should be removed from data and the variogram of the residuals should be estimated (Chilès, Delphiner, 1999).

The problem is that the trend is usually unknown; thus, its estimates should be used to define residuals. Let say the realizations of a random function $\{Z\}_{i\geq 0}$ are available at the experimental points x_α. This process is represented as following:

$$Z(x) = m(x) + Y(x) \qquad \text{(I.3.1)}$$

Lets define the residuals $R(x_\alpha) = Z(x_\alpha) - \hat{m}(x_\alpha)$, where $\hat{m}(x_\alpha)$ is the trend estimator at points x_α. Lets consider a variogram of the residuals:

$$\gamma_R(x_\alpha, x_\beta) = \frac{1}{2}\mathrm{E}(x_\beta - x_\alpha)^2 = \gamma(x_\beta - x_\alpha) - Cov(Z_\beta - Z_\alpha, \hat{m}_\beta - \hat{m}_\alpha) + \frac{1}{2}Var(\hat{m}_\beta - \hat{m}_\alpha) \quad \text{(I.3.2)}$$

The expression (I.3.2) can be simplified if only $\hat{m}(x)$ is an optimal linear trend estimator $m^*(x)$. Then:

$$\gamma_R(x_\alpha, x_\beta) = \gamma(x_\beta - x_\alpha) - \frac{1}{2}Var(m^*_\beta - m^*_\alpha) \quad \text{(I.3.3)}$$

Expression (I.3.3) shows that even if the optimal trend estimator is chosen the residual variogram might still underestimate the true variogram. The good news is that this bias is small at short distances, but can be significantly increased at large distances (Chiles, Delfiner, 1999).

Chiles, Delphiner (1999) believe that despite the presence of a trend it is always possible to return to the standard structural analysis of the stationary case. For example, if trend $m(x)$ is mild, than the estimated empirical variograms of Z on several data sub-samples will not differ significantly at the short distances. As kriging often is applied to rather close neighbourhood of data, the empirical variogram can be accepted as a good estimate of the true variogram.

At the situation when we can assume the stationarity of the residual term $Y(x)$ and at some sufficiently large sub-sample of the available observations the trend is equal to zero: $m(x_\nu) = 0, x_\nu \in x_\alpha$, then the empirical variogram estimated on this sub-sample can again be accepted as a good estimate of the true variogram.

As the result, we have the following solutions to address data non-stationarity:

1. Assume that a process Z can be represented as in (I.3.1) with stationary residuals Y. Then the covariance/variogram can be estimated by using one of the two approaches:
    a. Estimate and eliminate present trend in the data; then use the variogram of the residuals as an estimate of the true variogram.
    b. Estimate the empirical raw variogram on the sub-sample, where no trend is observed.
2. Assume that process Z is non-stationary; use intrinsic model to fit the raw variograms and apply approach 2.6.1 to obtain the optimal weights estimates.

As for the solution (1a), the estimation of a trend should be done before the application of the kriging method. As have been shown in (I.3.3), in order to minimize the error when accepting the empirical residuals variogram for the true variogram, we should apply the linear method to the trend estimation. For example, line or polynomial can be fit to the data by least square methods. However, curve fit will demand the subjective choice of the polynomial. We propose to estimate trend as a moving average of predefined length. In particular, we propose to subtract EMA of predefined (medium) length ($EMA_{t,T}$) and evaluate the variogram model for the price residuals ($\{R_i : R_i = P_i - EMA_i\}_{t-n+1 \le i \le t}$). The method is not optimal in statistical term, but it has the following advantages:

1. EMA is the Markov time indicator in the sense that only available historic data is used for its calculations.
2. EMA, as most popular technical indicator, is introduced in many trading software making them very easy to use.
3. EMA has only one parameter to choose – its length. Our main criterion is to choose such length, which guarantees the convergence of the residuals variogram at some not very large range values. From the trading point of view, EMA should reflect a medium-term

trend. The EMA length certainly depends on the data frequency and traders time horizon. For example, for 1-second frequency and intraday trader medium term can correspond to the two hours (2*60*60=7200) window length; while for the 30-minutes frequency it might corresponds to one week MA length.

We also propose to use only one simple function $f^0 = 1$ in the trend, implying constant (within moving window) unknown trend. In the empirical geostatistical applications the number of functions in the trend is usually limited to one or two, as more functions were not improving significantly the results.

3.2 Data sampling peculiarities

The way the financial data is sampled has direct effect on the results of the kriging method application. On one hand, non-regular sampling typical for instruments that are not traded frequently, or for subordinated processes, justifies the usage of the kriging approach at the place of the usual filtering methods. On the other hand equal-space sampling has an interesting impact on the structure of the optimal kriging weights.

In fact, Castelier, Laurenge (1993) showed that in the case of regular sampling the optimal weights for the mean estimator has quite similar behaviour. Under assumption of one simple function in the trend, we have relatively high weights for first and last observations in the window, and relatively low in the absolute terms (sometimes negative) the rest of the weights. They have derived the following close-form solution for the case of the exponential variogram model:

$$\begin{gathered} \lambda_1 = \lambda_N = \frac{1}{N-(N-2)b} \\ \lambda_i = \frac{1-b}{N-(N-2)b}, \forall i : 1 < i < N, \\ b = e^{-\frac{1}{a}} \end{gathered} \qquad \text{(I.3.4)}$$

where N is a length of the window (length of the sample of observations used in the kriging procedure), a - range parameter of the variogram $\gamma(h) = \sigma^2\left(1 - e^{-\left|\frac{h}{a}\right|}\right)$.

The expression (I.3.4) shows that optimal weights have the same structure independently on the length of the sample *N* or the variogram parameters; they have impact only on the absolute values of the weights.

The examples of the weights structures for different types of the variogram models are proposed in Figures 1.4-1.8. The following models are considered:

- Stable model: $\gamma(h) = \sigma^2\left(1 - e^{-\left|\frac{h}{a}\right|^\alpha}\right)$;
- Fractal model: $\gamma(h) = \sigma^2 |h|^\alpha$;

– Spherical model: $\gamma(h) = \begin{cases} \sigma^2\left(\frac{3h}{2a} - \frac{h^3}{2a^3}\right), h \le a \\ \sigma^2, otherwise \end{cases}$;

All figures 1.4-1.8 show similar results: concentration of the principal weights for the first and last observations in the sub-sample window, with comparatively lower and relatively stable other weights.

The weights structures for the stable model for different parameters are given in Figures 1.4-1.6. Figures 1.4-1.5 represent the impact of the range parameter a on the weights values, while fixing α at two levels: $\alpha_1 < 1$ and $\alpha_2 > 1$. Figure 1.4 shows the increased weights volatility with increase in the range parameter for $\alpha = 0.5$: very high range values corresponds to very high "border" weights (their sum is close to 2), which are compensated with relatively large negative weights (#2 and #9) and more or less stable and small (negative or positive) other weights. Figure 1.5 presents much smoother parabolic weights structure, convexity of which increases with the increase in the range parameters (the "border" weights also increases with range parameter). Figure 1.6 show the transformation of the weights structure from the volatile type (as in Figure 1.4) to the parabolic type (as in Figure 1.5) with the increase of α-parameter from 0 to 2[23].

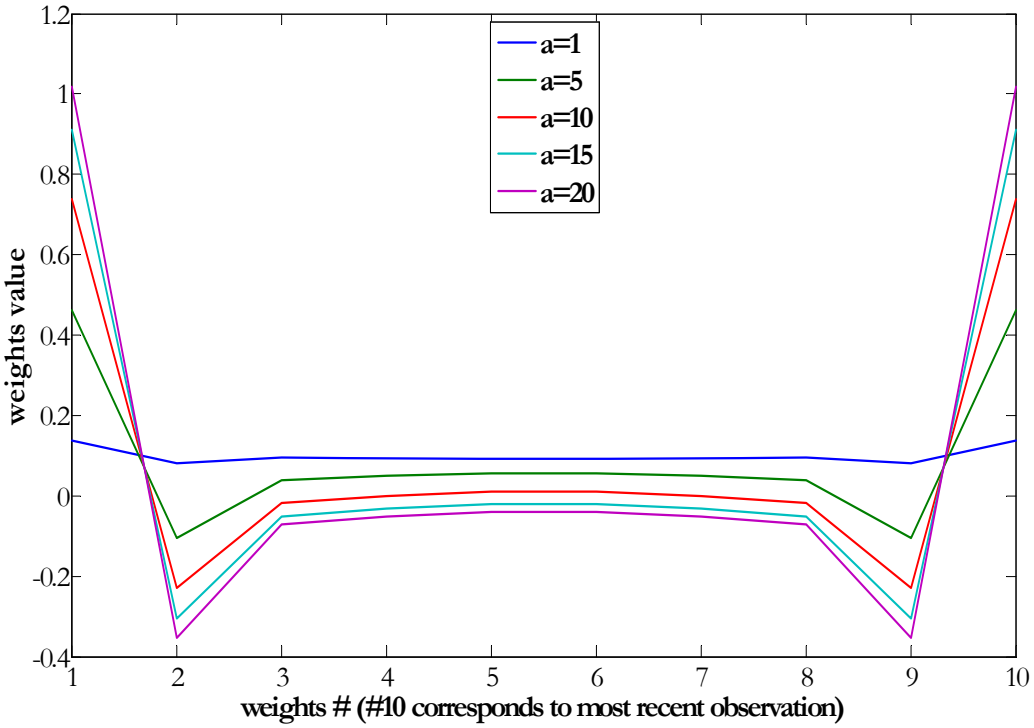


Figure 1.4. Weights, which correspond to different a variable in stable model: window length=10, $\alpha = 1.5$.

[23] For the range parameter $a > 1$ for the unit sampling frequency, the oscillation of the weights for the gaussian model or $\alpha = 2$ is much more severe. See Castelier, Laurenge (1993) for more examples.

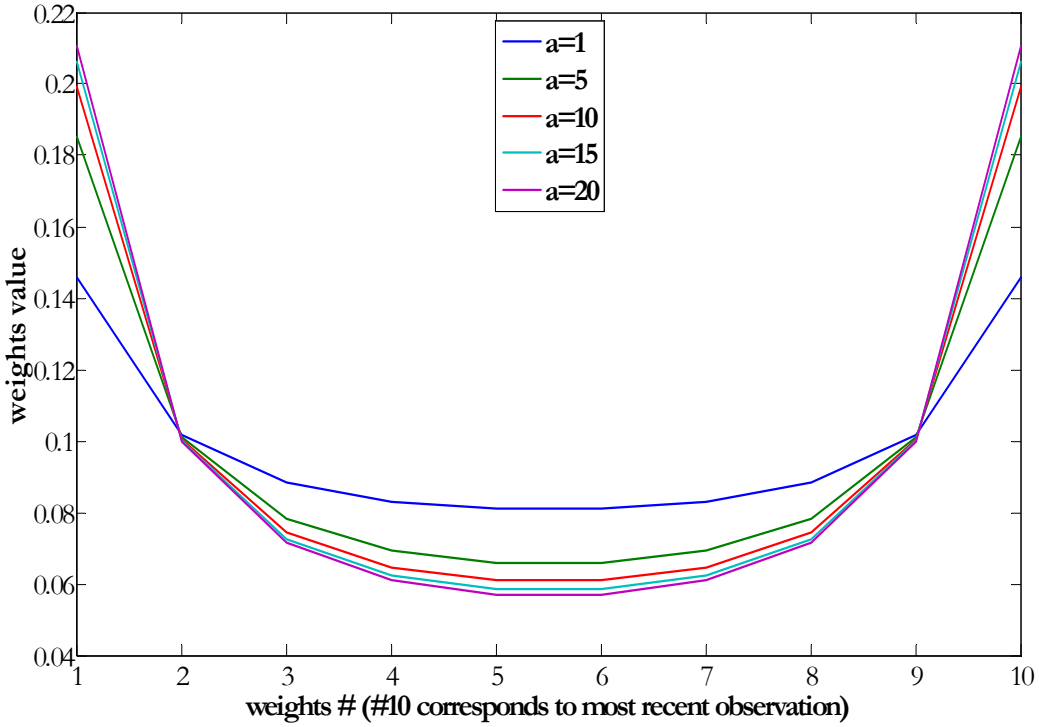


Figure 1.5. Weights, which correspond to different a variable in stable model: window length=10, $\alpha = 0.5$.

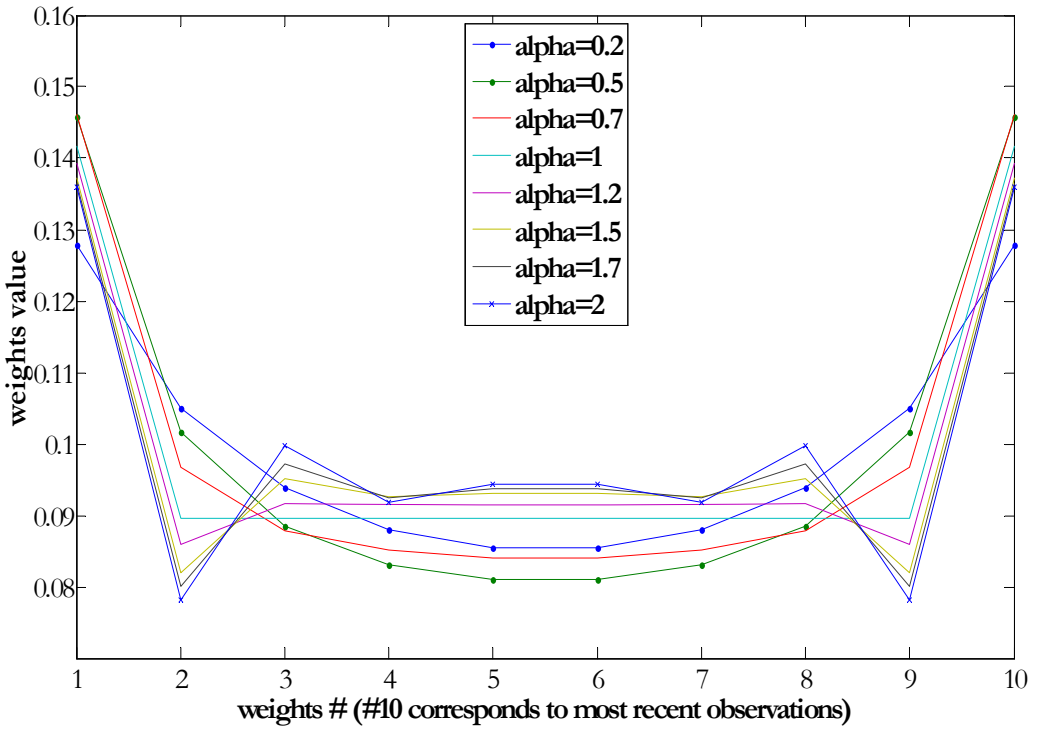


Figure 1.6. Weights, which correspond to different α variable in stable model: window length=10, $a = 1$.

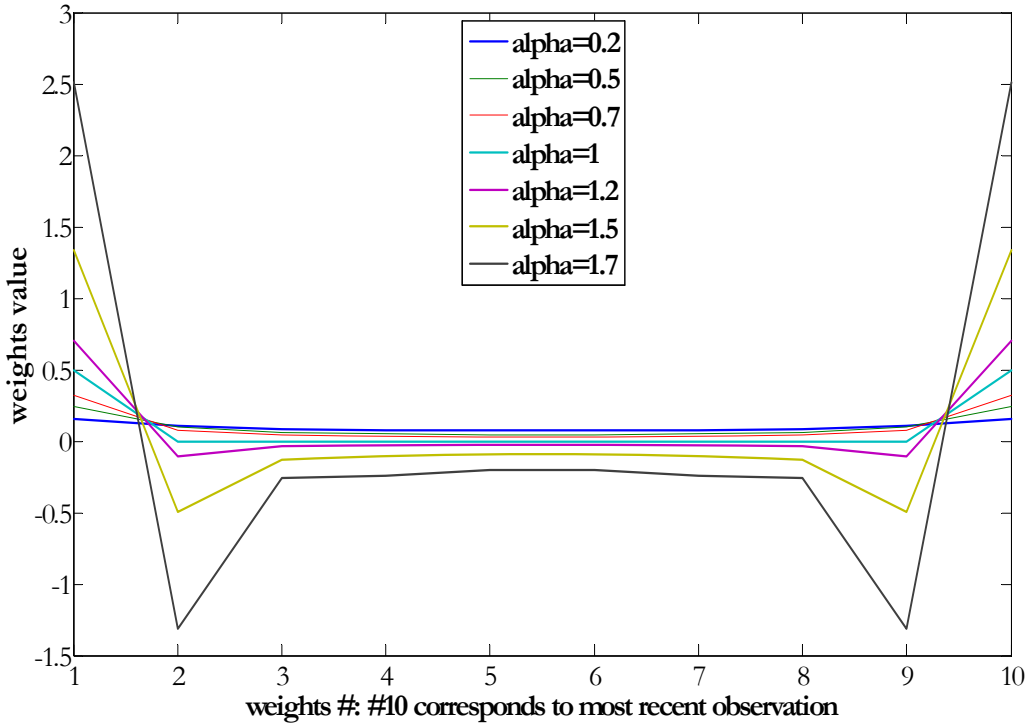


Figure 1.7. Weights, which correspond to different α variable in fractal model: window length=10, $a = 1$.

Figure 1.7 presents the weights structure for the fractal variogram model. Fractal model corresponds to the IRF-k functions. The weights structure resembles the case of stable model for $\alpha_1 < 1$.

Figure 1.8 considers the case of the spherical variogram model. The case when $a = 1$ implies the optimal MA in the form of the simple moving average; as data is considered to be sampled at unit distance $a = 1$ implies the absence of the correlation and the closeness of the data to the white noise. For the other range values, the same persistence of the large weights for first and last observation in window is observed.

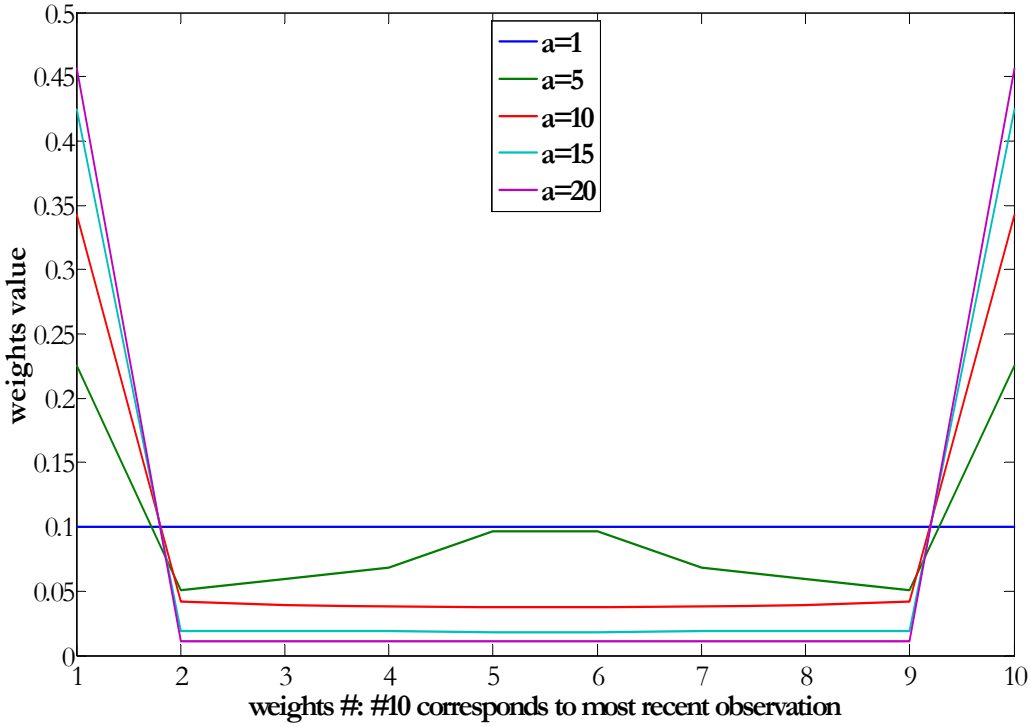


Figure 1.8. Weights, which correspond to different a variable in spherical model: window length=10.

As the result, we can see that independently of the window size and variogram models[24], kriged MAs are not significantly different from SMA in the term of lag, while some of them are less smooth (more volatile). In fact relative volatility of the weights depends only on the variogram model (not on the length of the window), in particular on the regularity of the variogram model at the origin. Varying parameter α (in stable or fractal models) has impact on the form of the weights curve and adds significantly to the volatility of the kriged MA. We can expect that kriged MA for stable and spherical models are more or less smooth; at the same time the kriged MA for fractal model supposed to be very volatile and unstable (both "borders" weights are larger than 1).

As the result, we can conclude that in the case of regular sampling (one-dimension), kriging optimal weights have the similar structure: for the exceptions of the border weights (first and last observation), other weights are quite low in the absolute terms. The «border» behavior of the weights in its turn is defined by the covariance structure of the variable and depends on the regularity of the variogram at the origin, i.e. by the presence of the range. Pure nugget model gives arithmetic average; spheric and exponential models produced more important oscillations around arithmetic weights, while gaussian model produce violent oscillations (Castelier, Laurenge, 1993).

Further we present examples of the kriging method applications to the different set of data. In Ch.4 and 5 we present the case of the instruments, sampled at equal intervals. Chapter 4 consider non-stationary examples of prices, while Chapter 5 concentrates on the MACD indicators with bounded paths without distinctive trends. Chapter 6 presents the examples of the application of the kriging method to the case of the data, sampled at uneven intervals.

[24] At least three classical model considered above and in Castelier, Laurenge (1993).

4 Kriging results: Non-stationary, evenly spaced time-series data

This chapter analyses the application of the kriging to four different instruments: (1) Bund; (2) DAX; (3) Brent; (4) X instrument[25]. The analysis of the Bund is presented in details, while only trading outcomes are presented for other instruments.

4.1 Variogram analysis and optimal kriging weights: Bund

The Bund sample represents the quotes[26] for three different contracts[27] due March 8, June 8 and September 8, 2006, sampled at one-second frequency. The data is sampled at very high frequency, therefore, even for Bund, an actively trading instrument, the data is not available at each point of time. In order to obtain equally spaced data, missed data is interpolated at the levels of the last available data. Example of the Bund price path for the contract due on March 8, 2006 is given in Figure 1.8. Other contracts can be find in Figures E1, E2 in the Appendix E. All figures indicate the presence of trends in the data. Moreover the trends are clearly non-linear. Figure 1.9 supports the hypothesis of the non-stationarity with the unbounded price variograms evaluated for the Bund contracts.

Linear variogram implies the price process could be modelled as an IRF-0. We have fit the following linear model to the variogram for the March 8, 2006 Bund contract (h in seconds):

$$\gamma(h) = 0.93 \cdot 10^{-6} |h|$$

The optimal weights estimated for n=3600s under assumption that local covariance is $A - \gamma(h)$ for some very large $A > 0$ have similar structure than the weights considered in the Chapter 3 (Figure 1.7): $\lambda_1 = \lambda_{3600} = 0.5$, $\lambda_2 = ... = \lambda_{2999} = 0$.

The analysis of the variograms for the residuals is more complicated.

[25] Due to the confidentiality reason we cannot present detailed description of the instrument.
[26] The Bund quotes, used for vaiogram and kriging calculations, are in fact the index built on the basis of different Bund prices (quotes). Due to the confidentiality reason we cannot provide the formulae.
[27] Bund is a futures contract.

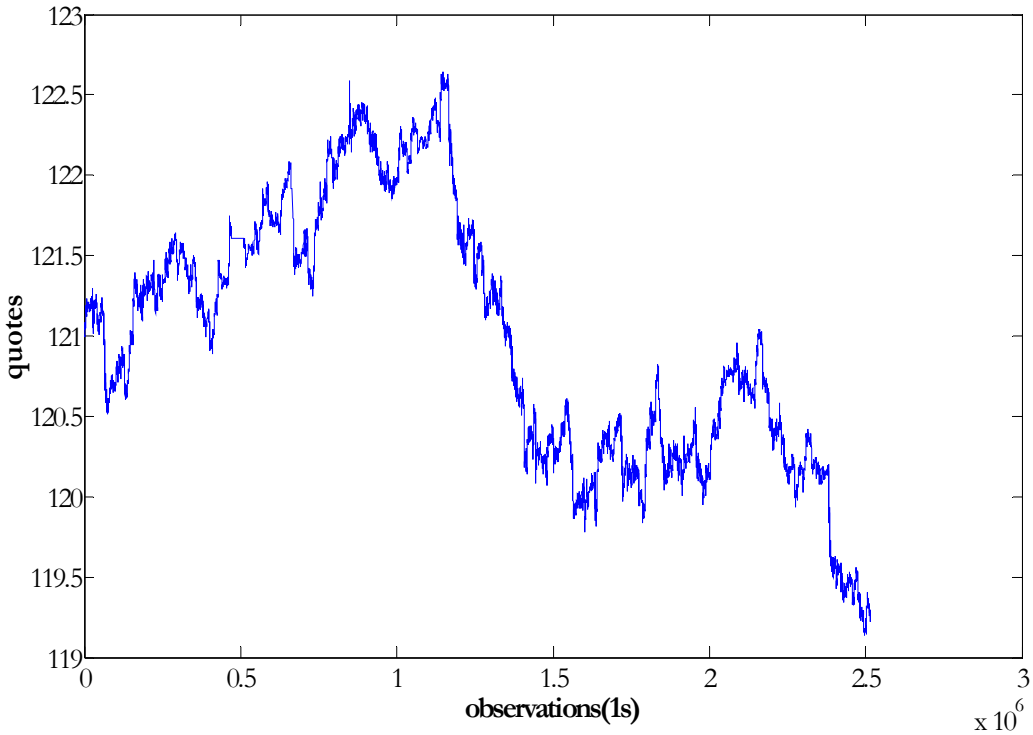


Figure 1.8. Bund (December 9, 2005 - March 8, 2006, frequency – 1 sec)

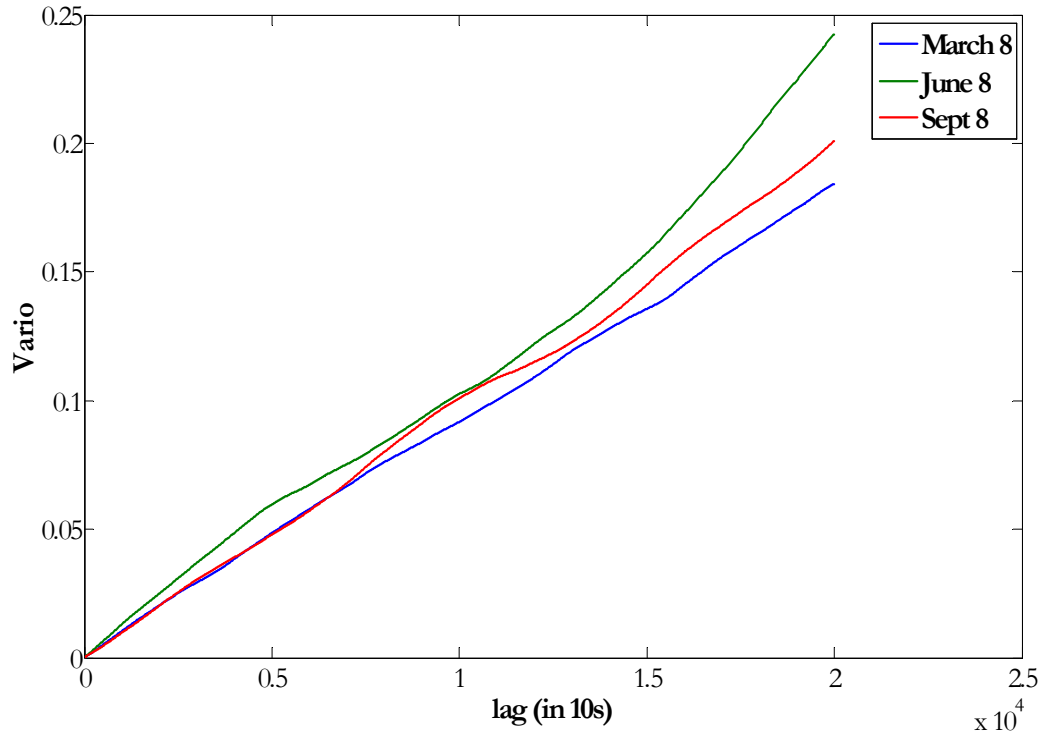


Figure 1.9. Variograms for different Bund contracts due at different dates in 2006 (December 9, 2005 - March 8, 2006, March 9 – June 8, 2006, June9-September 8, 2006, frequency – 1 sec)

As have been discussed already in Chapter 3, in order to obtain the estimates of true covariance, the trend should be eliminated from the data. Further we provide the analysis of how different EMAs (the trend estimator) can affect the form of the variogram/covariance used in kriging applications. Moreover, we would like to analyse whether covariance structure of the residuals is stable over different contracts.

It is obvious that different lengths of the EMAs, used for trend subtraction, will have impact on the parameters of the variogram (at least its sill). Effective length of the EMA defines how close the EMA is approaching the price curve and how smooth it is; longer EMA length implies smaller distance between two curves and smoother EMA curve. As the result, the variance of the residuals will increase with the increase of the EMAs length. Figure 1.10 presents the variograms, estimated for September 8, 2006 contract, which correspond to seven different EMA lengths. As expected, the sill of the variograms is an increasing function of the EMA length.

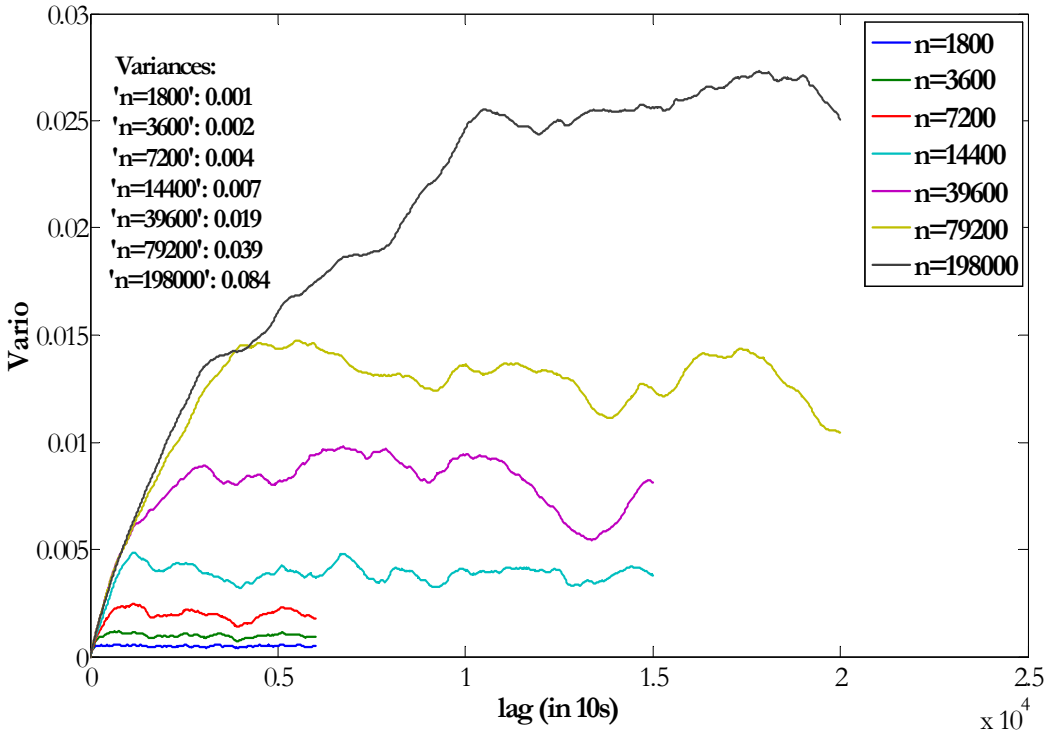


Figure 1.10. Bund (June9-September 8, 2006, frequency – 10 sec): Variograms for Bund (8 September contract) residuals, which corresponds to EMAs of different length

The impact of the EMA lengths on the range of the variograms is though unclear from the Figure 1.10. Therefore, we propose to consider the variograms, normalized by their respective sill (see Figure 1.11). From Figure 1.11 we can see the ranges of the variograms also depend on the length of EMA, used for the trend subtraction: the longer is the EMA – the larger is the variogram range.

The analysis of the residuals over different contracts shows whether the covariance structure is the same over the time. Figures 1.12 and E3-E4 in the Appendix E show the examples of the variograms for the residuals for different Bund contracts, obtained by the subtraction of the EMA of the following lengths: $n_1 = 1800$ sec (30 minutes), $n_2 = 7200$ sec (2 hours), $n_3 = 79200$ sec (2 days) respectively. We cannot conclude from the observed variograms about

residuals stationarity. All the variograms stabilized around some sills. However, these sills are all well below the estimated variances.

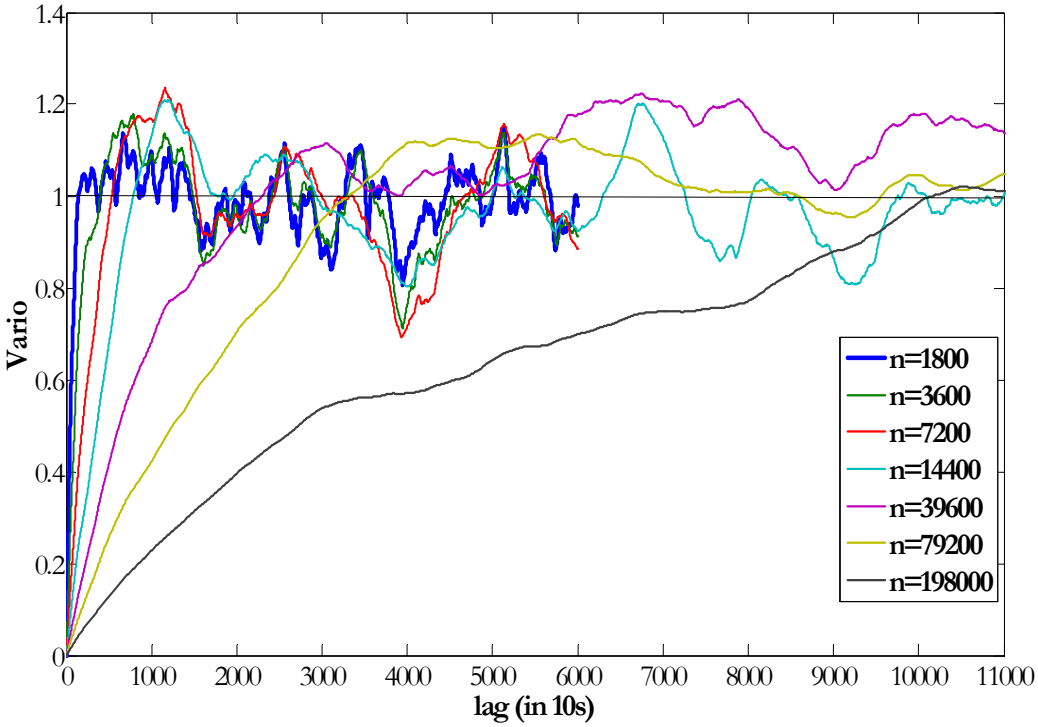


Figure 1.11. Bund (June9-September 8, 2006, frequency – 1 sec): Normalized variograms for Bund residuals, which corresponds to EMAs of different length

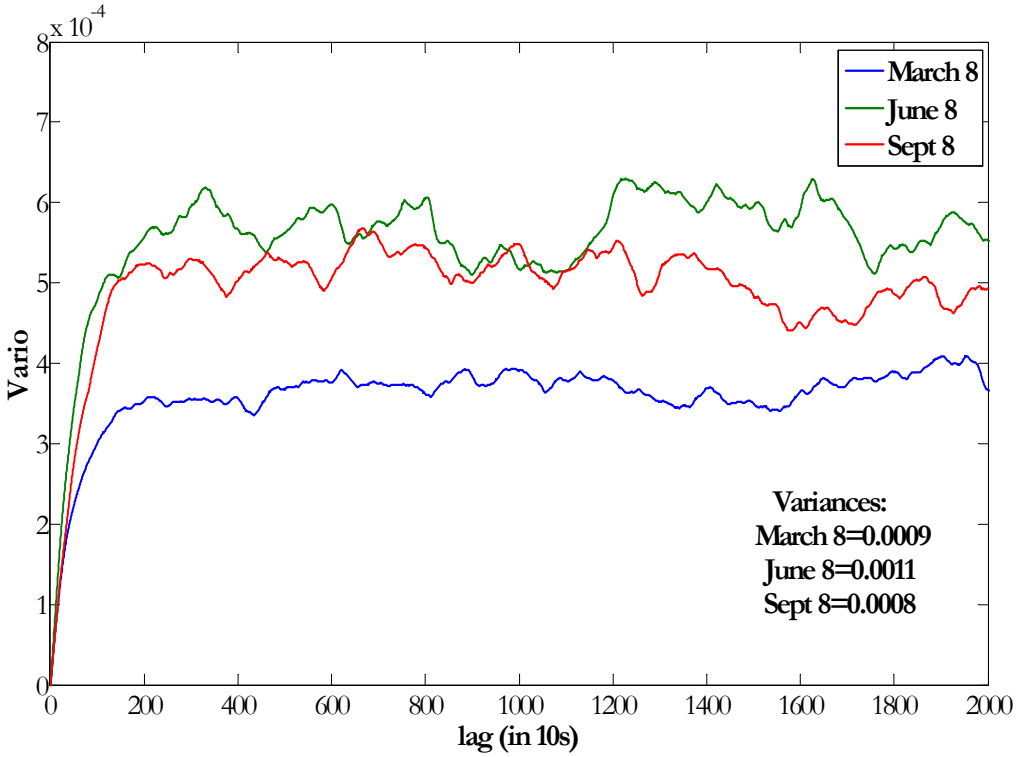


Figure 1.12. Bund (December 9, 2005 - March 8, 2006, March 9 – June 8, 2006, June9-September 8, 2006, September 9 – December 8, 2006, frequency – 1 sec): Variograms for Bund residuals (n(EMA)=1800 sec) for different contracts

Stabilization of the variograms below variance implies persistent autocorrelation (correlation is no longer = 0 at large lags). However, explanations for this phenomena can lay also in the data peculiarities such as non-constant volatility caused by such events as overnight jump of the prices, high volatility «after 14:30 hours» etc. These events can cause an overestimation of the true variance. Therefore, calculation of the average variogram over the periods that does not include these effects can be one of the solutions for this problem. For example, the estimation of the average «daily» variogram can help to define whether overnight effect can be a cause for these types of the variograms that we have observed.

Estimation of the average variogram (and average variance) over all available (non-0) days for three different contracts can be done according to the following formulas:

$$\mathrm{E}(\gamma(h)) = \frac{\sum_{i=1}^{N(days)} \gamma_i(h)}{N} \qquad \text{(I.4.1)}$$

$$\mathrm{E}(\mathrm{var}(Bund)) = \frac{\sum_{i=1}^{N(days)} \mathrm{var}_i(Bund)}{N} \qquad \text{(I.4.2)}$$

For these estimations we take only days with complete series of prices. Our maximum lag is certainly constrained by 1 day (in reality, by more lower value). Figure 1.13 shows the estimated average variograms for three Bund contracts. The variograms show that daily variograms have stabilized around estimated variance. New estimated variance is lower than the variances estimated over the 3 months sub-sample. This means that the presence of the overnight effects can be the cause for variance overestimation. The form of the average variogram is smooth and «model-like», which will facilitate significantly the variogram modeling.

The variogram range is approximately equal for all contracts, while the sill is different. Therefore, we can expect the same weights for the optimal MA forecast, but different variance of the estimator for different Bund contracts.

As the result of the analysis, we can see that the range parameter of the residuals variogram is constant over time (does not depend on the contract), however its value depends directly on the EMA length used for the residuals construction. The residuals variance is time-dependent and also depends on the EMA length.

In order to obtain the kriging equation system that is non-singular, a model should be fitted to the empiric covariance estimates, which will guarantee positive definiteness of the variance-covariance matrix. We choose to model the average "daily" variogram in Figure 1.13 for the March 8, 2006 contract. It means that we use the EMA of the 1-hour (3600 seconds) length as a trend estimator. Figure 1.14 proposes the fit of the exponential model $\gamma(h) = \sigma^2\left(1 - e^{-\left|\frac{h}{a}\right|}\right)$ with the parameters $a = 1900$, $\sigma^2 = 0.002$.

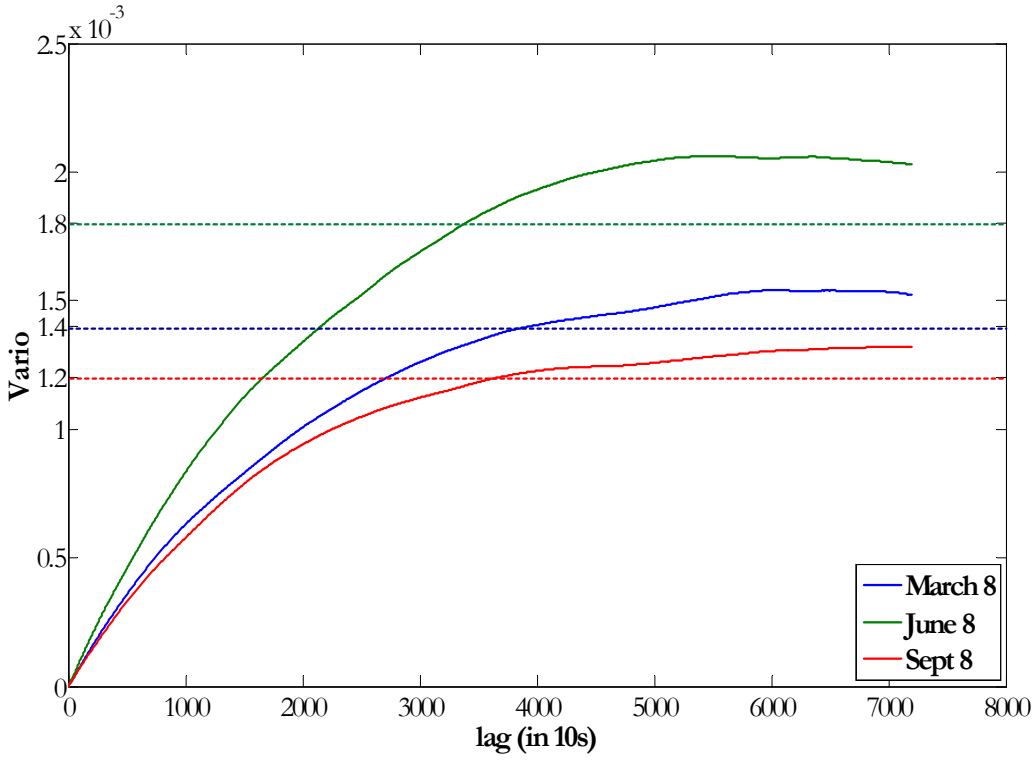


Figure 1.13. Bund (December 9, 2005 - March 8, 2006, March 9 – June 8, 2006, June9-September 8, 2006, frequency – 1 sec): Average variograms over one day for Bund residuals (n(EMA)=3600s (1 hour))

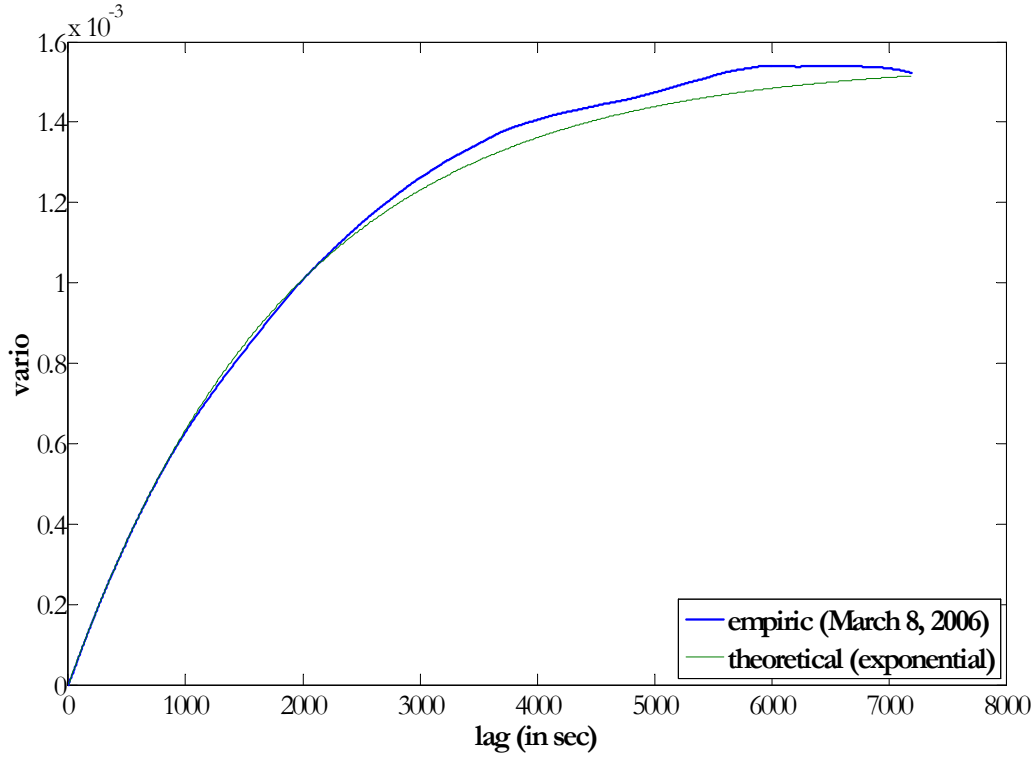


Figure 1.14. Bund (December 9, 2005 - March 8, 2006, frequency – 1 sec): Empirical average "daily" variogram for Bund residuals (n(EMA)=3600s (1 hour)) and theoretical model $\gamma(h) = 0.002\left(1 - e^{-\left|\frac{h}{1900}\right|}\right)$

From Chapter 3 we know that equally spaced samples will produce the optimal kriging weights of the particular form. In particular, for the exponential variogram model, these weights values can be even calculated analytically according to the formula (I.3.4) (n=7200s):

$$\lambda_1 = \lambda_{7200} = 0.1728,$$
$$\lambda_i = 9.09 * 10^{-5} \approx 0, i = 2,3,...,7199$$

The optimal weights are presented in Figure 1.15. The kriged moving average (KMA), which corresponds to these weights is given in Figure 1.16. As can be observed at this length KMA is very close to SMA of the same length, but less smooth. More volatile nature of the KMA is demonstrated in Figure 1.17 for a shorter SMA and KMA. We can notice that KMA oscillates around SMA and has much less smooth nature.

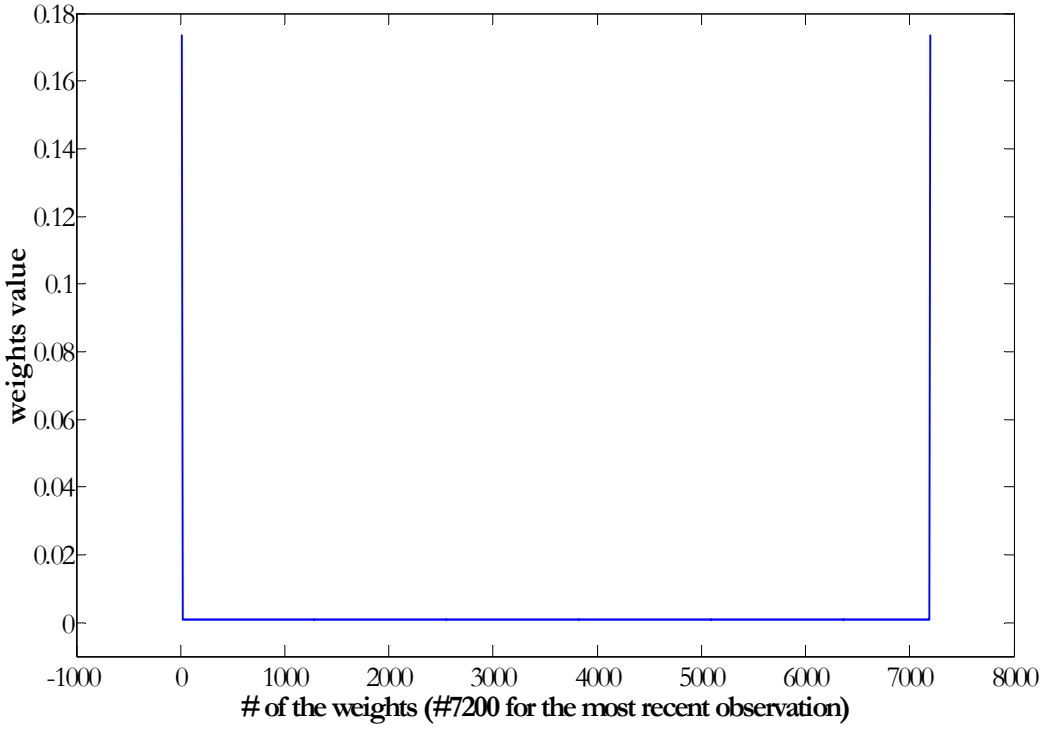


Figure 1.15. Bund (December 9, 2005 - March 8, 2006, frequency – 1 sec): Optimal weights (mean estimator) calculated for the theoretical exponential variogram $\gamma(h) = 0.002\left(1 - e^{-\left|\frac{h}{1900}\right|}\right)$: window=7200s.

As we can see the weights structure is the same whether we use residuals covariance or IRF-0 model variogram. As the result we propose further the analysis within stationary framework (residuals covariance model), as we can expect the trading outcomes to be approximately the same.

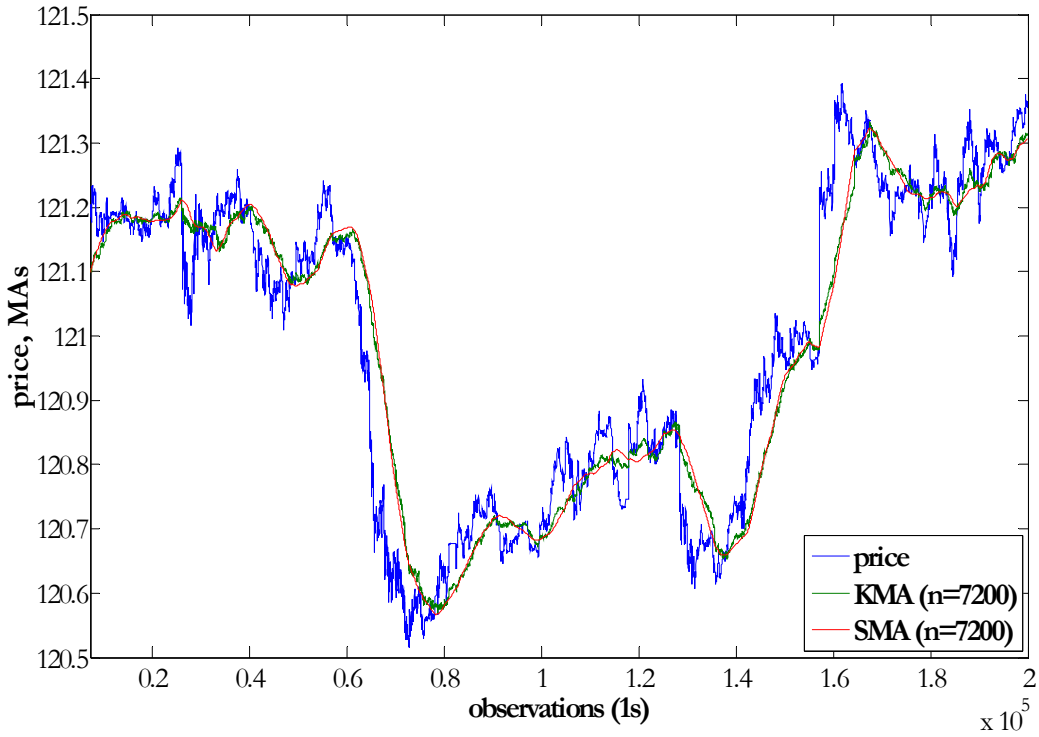


Figure 1.16. Bund (December 9, 2005 - March 8, 2006, frequency – 1 sec, observations aprox. 7201-200000): Price, KMA and SMA (window length =7200s)

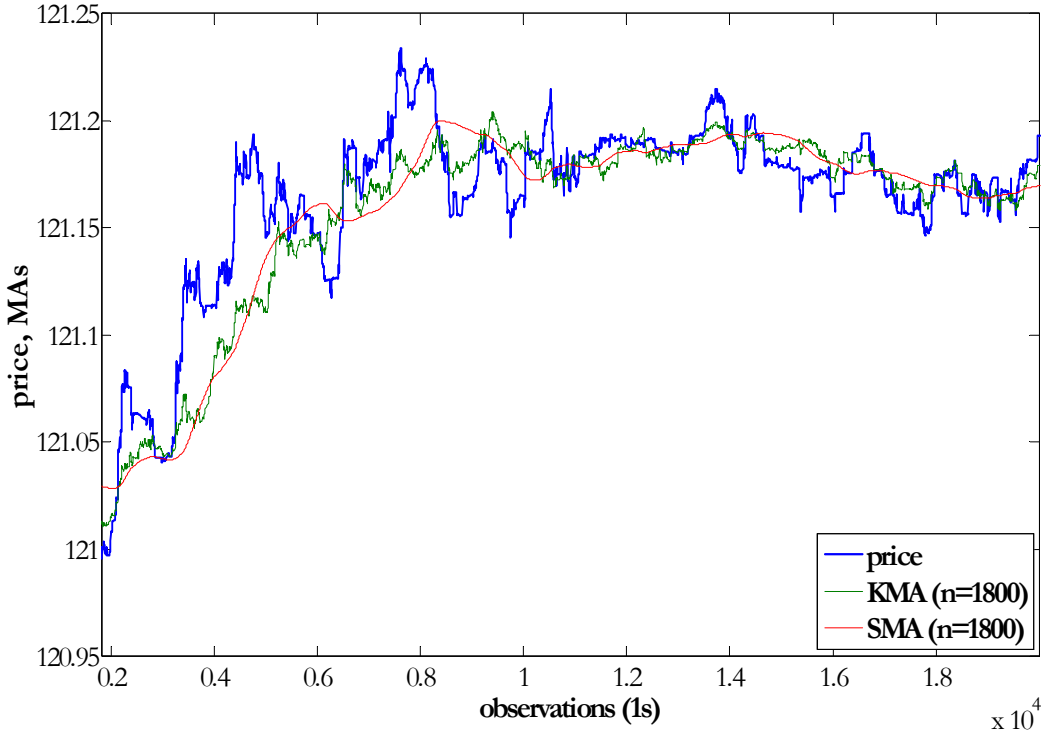


Figure 1.17. Bund (December 9, 2005 - March 8, 2006, frequency – 1 sec, observations aprox. 1801-20000): Price, KMA and SMA (window length =1800s)

As we can see, the optimal MA weights, which take into account the auto-covariance of the instruments value results in the MA, which is close by the lag to the SMA, but has more volatile

structure. The question is whether trading strategies based on KMA can bring higher outcomes than SMA.

4.2 Trading results: KMA versus SMA

The following sub-chapter 4.2 analyzes trading results, obtained for the simulation of strategies, based on KMA and SMA for such instruments as Bund, DAX, Brent and X instrument.

We apply simple trading strategy, based on the crossovers of the price and MA lines (in this case, SMA and KMA). The long position should be taken (and the short position should be closed) when the price crosses the MA curve from below, which is a confirmation of the upward trend; the short position should be taken (and the long position should be closed) when the price crosses the MA curve from above, which is a confirmation of the downward trend. Thus, the strategy is defined as following:

Trend-following strategy (I.4.3)

1. Trading costs are 0. Profits are defined in quotes units[28].
2. Lets define $R_i = P_i - MA_i$.
3. The initial trading position $Pos_0 = 0$; trading outcome $\Pi_0 = 0$.
4. The first trade ($Pos_i, i > 0$) is undertaken at the first crossovers of the MA and price lines, i.e. under following condition:
   if $(R_i R_{i-1} \le 0)$ and $(R_{i-1} < 0)$: $Pos_i = 1,\ P_{entry} = P_i,\ \Pi_i = 0$
   if $(R_i R_{i-1} \le 0)$ and $(R_{i-1} > 0)$: $Pos_i = -1,\ P_{entry} = P_i,\ \Pi_i = 0$
   otherwise, $Pos_i = 0,\ \Pi_i = 0$
5. Afterwards, at the new trading signals (curves crossovers) the following trades are executed:
   if $(R_t R_{t-1} \le 0)$ and $(R_{t-1} < 0)$:
   - exit (previously taken) position *Short* ($Pos_{t-1} = -1$): $Pos_t = 0, P_{exit} = P_t$
   - entry position *Long*: $Pos_t = 1,\ P_{entry} = P_t$
   - cumulative trading outcome: $\Pi_t = \Pi_{t-1} + Pos_{t-1}(P_t - P_{t-1})$

   if $(R_t R_{t-1} \le 0)$ and $(R_{t-1} > 0)$:
   - exit (previously taken) position *Long* ($Pos_{t-1} = 1$): $Pos_t = 0, P_{exit} = P_t$
   - entry position *Short*: $Pos_t = -1, P_{entry} = P_t$
   - trading outcome for this operation: $\Pi_t = \Pi_{t-1} + Pos_{t-1}(P_t - P_{t-1})$

   otherwise, $Pos_t = Pos_{t-1}, \Pi_t = \Pi_{t-1} + Pos_{t-1}(P_t - P_{t-1})$

The results of the strategy (I.4.3) simulations for different instruments and different lengths of KMA and SMA are presented in Table 1.1. Except for the Bund case, we use different samples for the variogram estimation and trading simulations. Each sample (DAX, Brent, X instrument) is split into two sub-samples of approximately the same length; the first sub-sample is used for the variogram estimation, and second sub-sample - for the trade simulations.

[28] Instruments values are usually quoted in ticks, not in currency equivalents.

Note that the trading results can be compared only within particular instrument due to the difference in the length of the sample, its frequency and value of the quotes for different instruments: we cannot compare Bund to DAX, but only the DAX results for the strategies, based on KMA and SMA.

Table 1.1
The outcomes of the simulated trading strategies

Instrument	Frequency	Optimal length (obs)		Max profit (in quotes)		Number of trades	
		KMA	SMA	KMA	SMA	KMA	SMA
Bund	1 sec	1740	660	1.71	1.89	689	1037
DAX	30 min	45	118	1736	1525	614	350
Brent	30 min	45	49	14.13	9.45	124	140
X instrument	1 hour	20	11	-10.52	-6.32	254	404

Table 1.1 shows that KMA are steadily more effective at short lengths, while the optimal length of the SMA is varying from short to long. In general at short MA lengths, KMA generates fewer trades than the SMA (see Brent case). Besides the profit per trade was higher for the KMA for Bund and Brent instruments. Further we analyze how results of the strategies, based on the different MAs depend on their length, as well as have a look at the P&L paths for the optimal MA lengths for KMA and SMA to see whether they exhibit monotone and positive trend.

4.2.1 Bund

Figure 1.18 gives the end-of-period cumulative outcomes for the trading strategies based on the KMA and SMA of the different lengths. KMA does not show better results than the SMA, though it still works better at short lengths than in long. Similar outcomes for long SMAs and KMAs are explained by the fact that at long lengths both curves almost coincides. Shorter KMA are more volatile than SMA and oscillates with larger amplitudes around SMA leading to the difference in the trading outcomes.

SMA-based trading strategy accumulates fewer trades than KMA (see Figure 1.19).

The KMA and SMA paths, which correspond to their respective optimal lengths, are given in Figure 1.20. We can see that both paths exhibit positive trend. Moreover, after 60000 observations their behavior is synchronized, in fact the end-of-period difference in the outcomes for both MA is caused by better performance of the SMA periods at the beginning of the sample. From Figure 1.16 we can see that the price pattern up to the 60000 observations is characterized by the trendless period; so it seems like SMA due to their more smooth nature perform better during time, when markets are not trending.

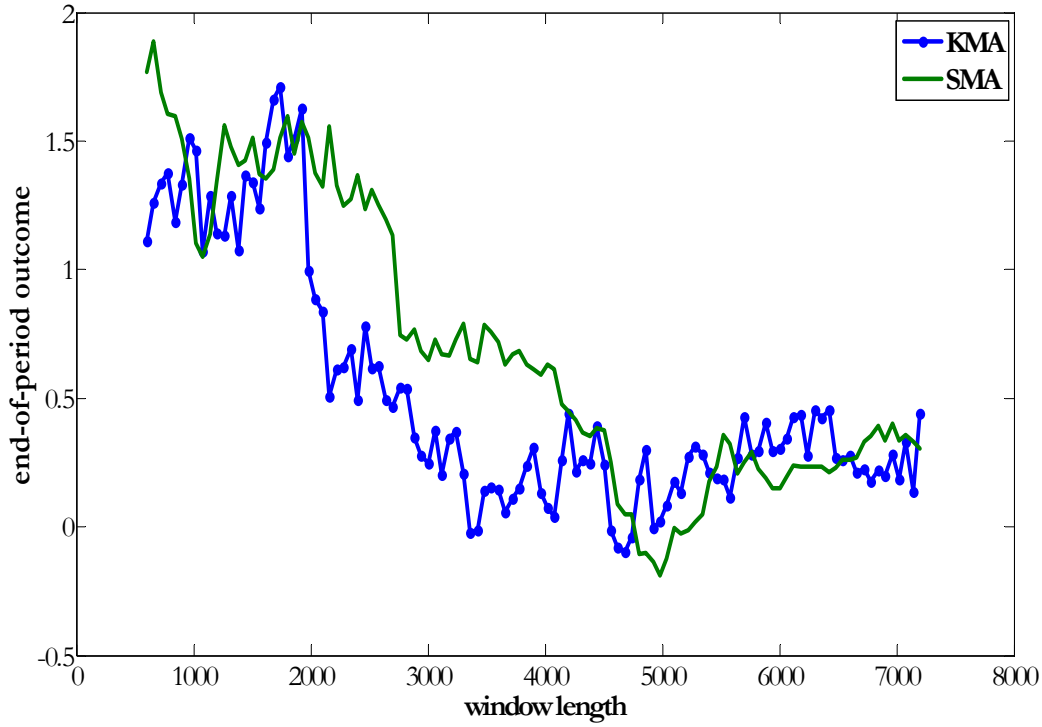


Figure 1.18. Bund (December 9, 2005 - March 8, 2006, frequency – 1 sec, observations 1-200000): End-of-period outcomes for the strategies, based on KMA and SMA of the different lengths

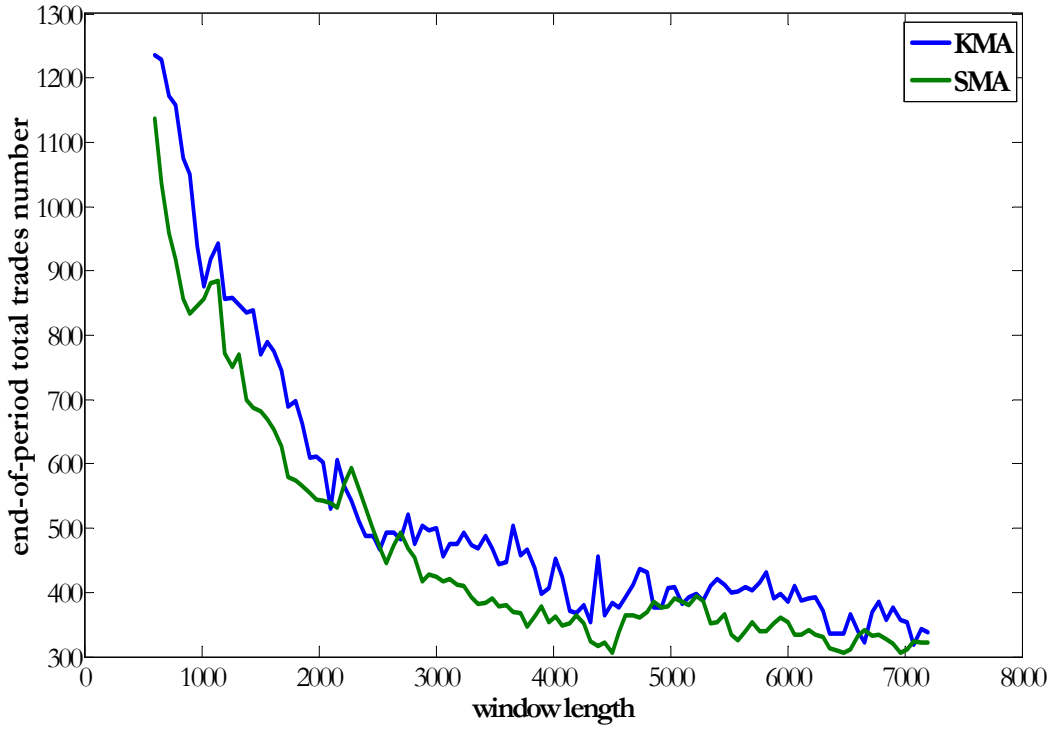


Figure 1.19. Bund (December 9, 2005 - March 8, 2006, frequency – 1 sec, observations 1-200000): End-of-period total trade number for the strategies, based on KMA and SMA of the different lengths

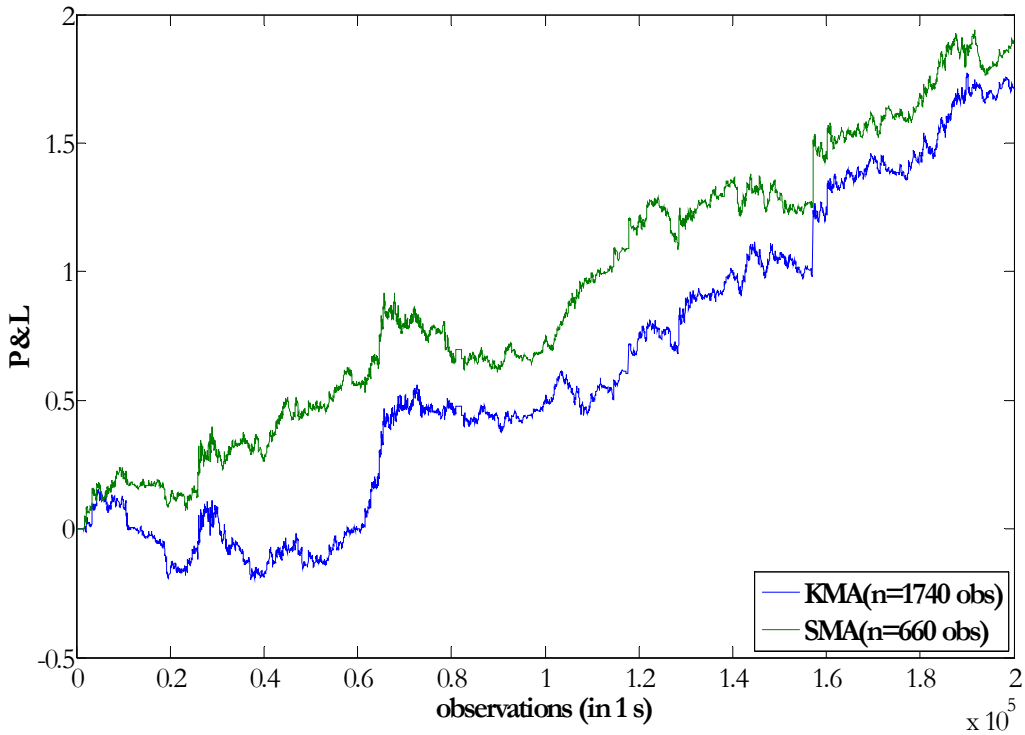


Figure 1.20. Bund (December 9, 2005 - March 8, 2006, frequency – 1 sec, observations 1-200000): Optimal P&L paths for the strategies, based on KMA (length=1740 observations) and SMA (length=660 observations)

4.2.2 DAX

Short description of the DAX sample, as well as variogram used in the kriging applications is given in Appendix F.

Figure 1.21 presents the end-of-period cumulative outcomes for the trading strategies, based on the KMA and SMA of different lengths. It seems like KMA is more effective than SMA at short lengths (between approx. 30-60 observations). At long lengths SMA leads to higher outcomes, though the results for both MAs are comparable. Similar trading outcomes for the long SMAs and KMAs are explained by the fact that two curves almost coincide at these lengths. Shorter KMAs are more volatile than respective SMAs: they oscillate with larger amplitudes around SMA leading to the difference in trading outcomes.

Contrary to Bund case, the KMA accumulates fewer trades than the SMA almost for all window length (see Figure 1.22). It means that even having more erratic nature, and therefore, higher probability of sending false signals, KMA crosses price curve less frequently than the SMA curve.

The KMA and SMA paths that correspond to their respective optimal lengths[29], are given in Figure 1.23. We can see that both paths exhibit positive trend.

[29] We choose the KMA optimal length at the level that does not generate the global P&L maximum, but fells within the interval of optimal KMA lengths.

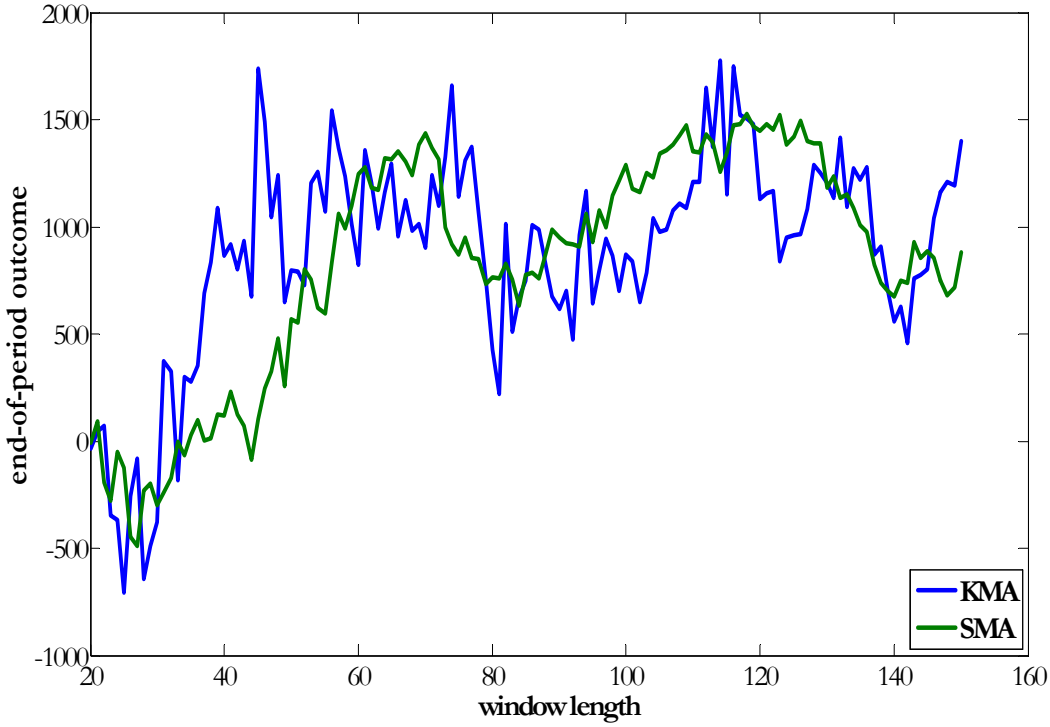


Figure 1.21. DAX (30/7/2003-7/12/2006, frequency 30 minutes, observations 10086-20171): End-of-period outcomes for the strategies, based on KMA and SMA of the different lengths

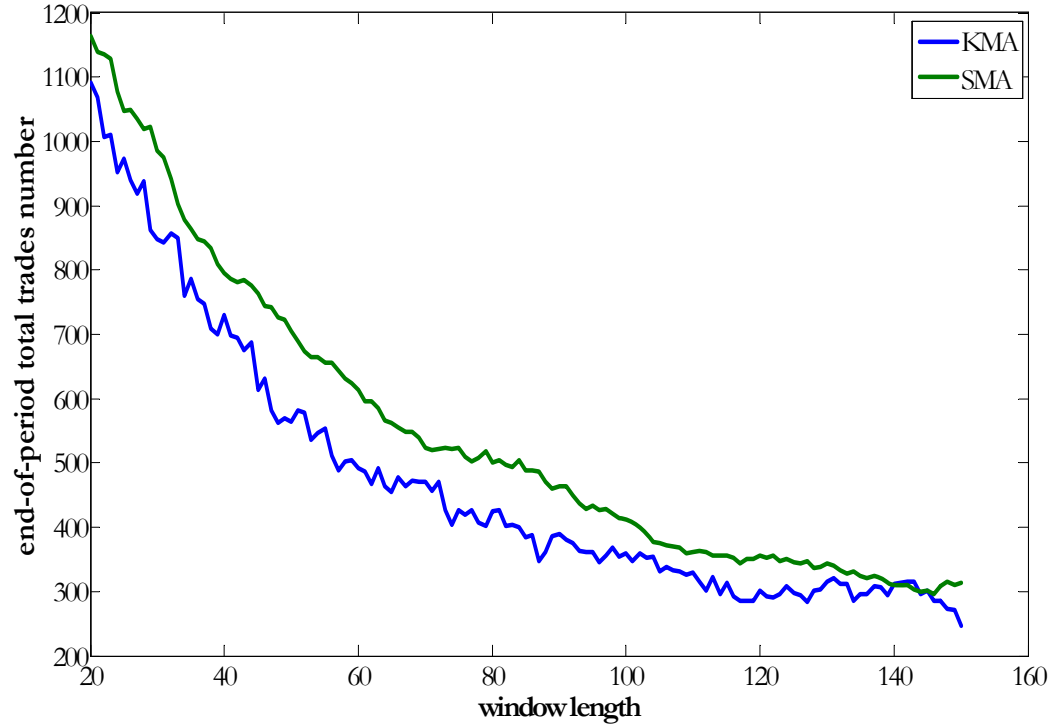


Figure 1.22. DAX (30/7/2003-7/12/2006, frequency 30 minutes, observations 10086-20171): End-of-period total trades number for the strategies, based on KMA and SMA of the different lengths

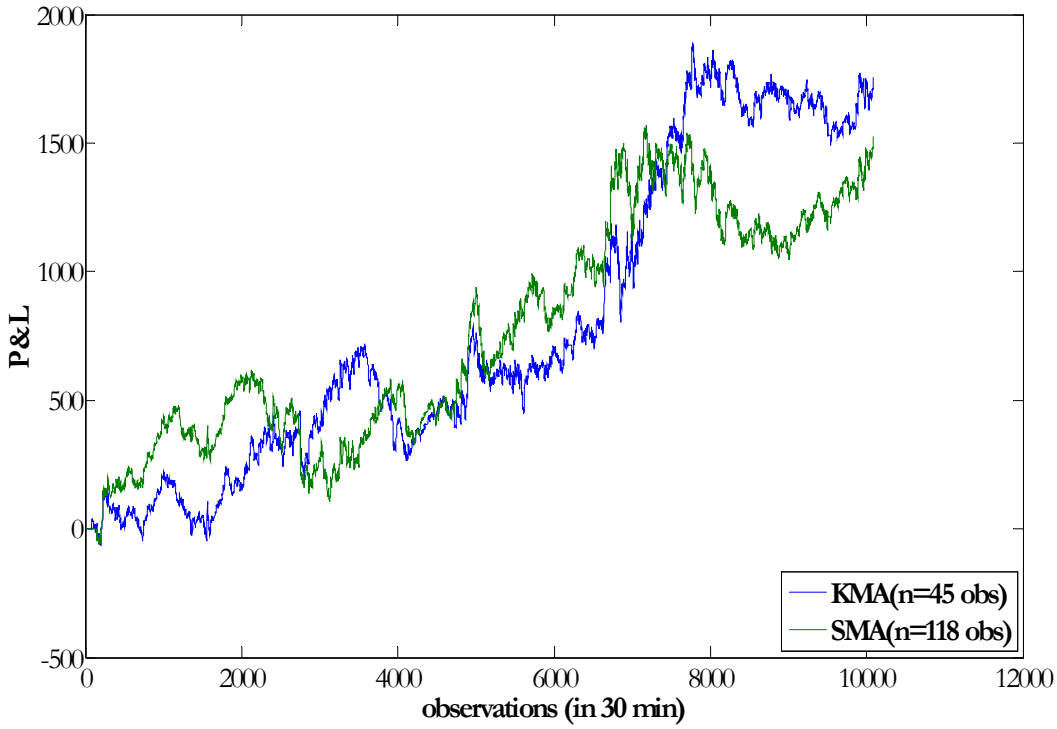


Figure 1.23. DAX (30/7/2003-7/12/2006, frequency 30 minutes, observations 10086-20171): Optimal P&L paths for the strategies, based on KMA (length=45 observations) and SMA (length=118 observations)

4.2.3 Brent

Short description of the Brent sample, as well as variogram used in the kriging applications are given in Appendix F.

Figure 1.24 presents the end-of-period cumulative outcomes for the trading strategies, based on the KMA and SMA of different lengths. Again, the shorter length of the KMA leads to higher outcomes than the longer length, though contrary to the DAX case no superiority over SMA results are observed.

As in the case of DAX instrument, the KMA accumulates fewer trades than the SMA for short and medium lengths, but slightly higher number of trades for the long length (see Figure 1.25). For shorter lengths KMA sends less false signals than SMA.

The KMA and SMA paths, which correspond to their respective optimal lengths, are given in Figure 1.26. Contrary to the DAX case, KMA path exhibit steeper trend than the optimal SMA path that is more random.

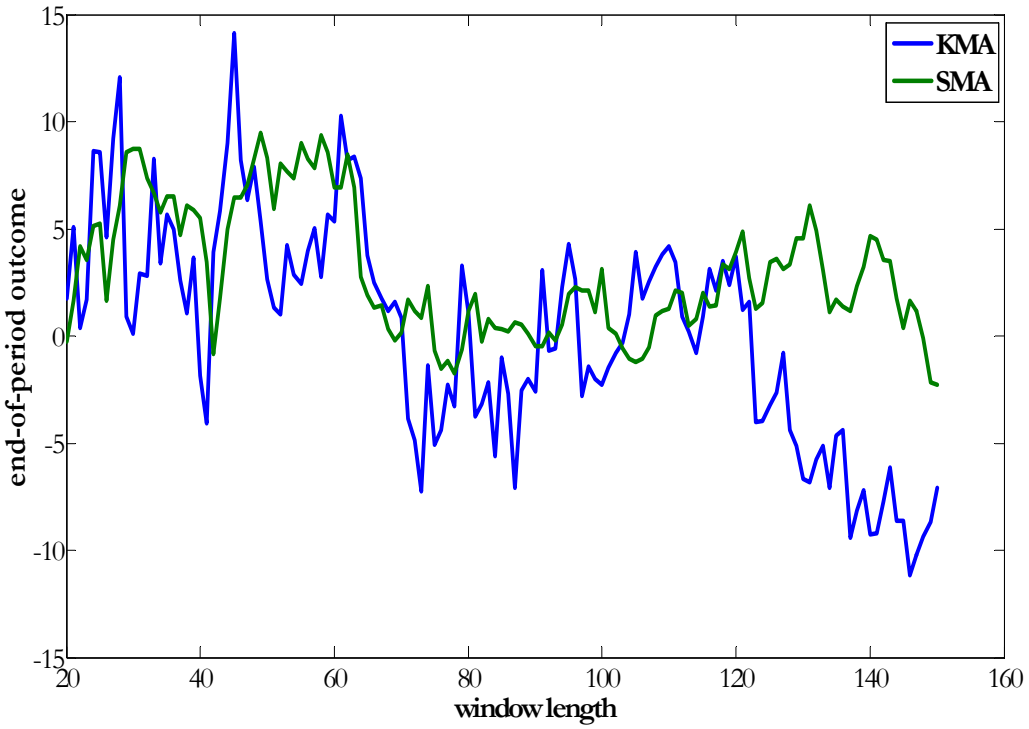


Figure 1.24. Brent (17/12/04-27/01/06, frequency 30 minutes, observations 2515-5029): End-of-period outcomes for the strategies, based on KMA and SMA of the different lengths

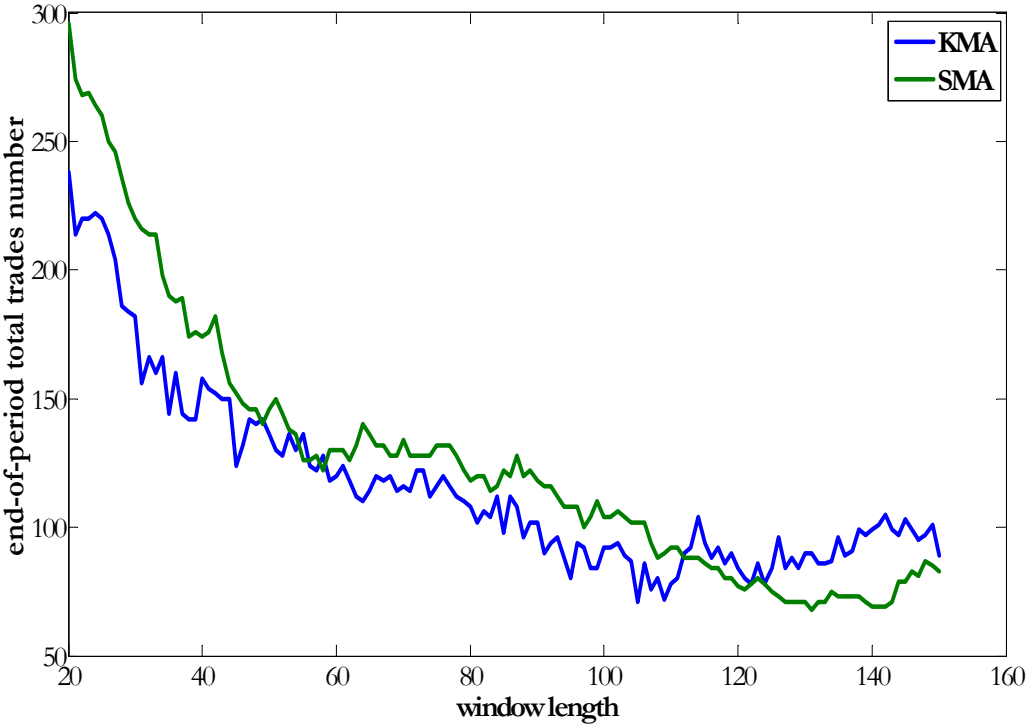


Figure 1.25. Brent (17/12/04-27/01/06, frequency 30 minutes, observations 2515-5029): End-of-period total trades number for the strategies, based on KMA and SMA of the different lengths

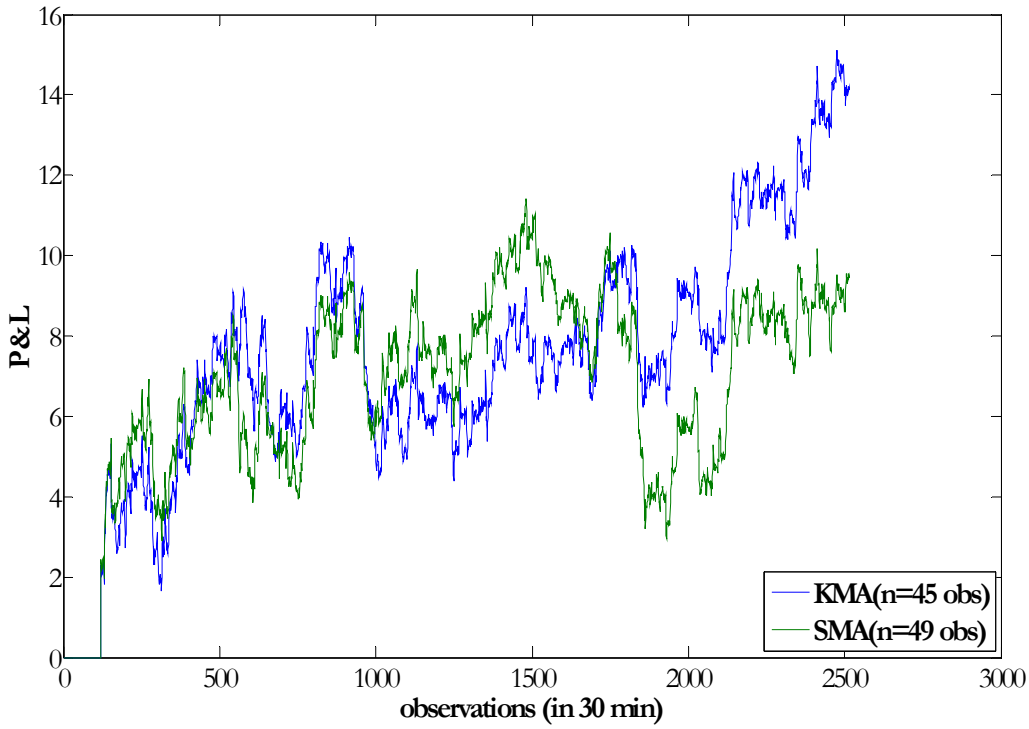


Figure 1.26. Brent (17/12/04-27/01/06, frequency 30 minutes, observations 2515-5029): Optimal P&L paths for the strategies, based on KMA (length=45 observations) and SMA (length=49 observations)

4.2.4 X instrument

Short description of the X instrument sample as well as variogram, used in the kriging applications are given in Appendix G.

Figure 1.27 presents the end-of-period cumulative outcomes for the trading strategies, based on the KMA and SMA of different lengths. Contrary to the other instruments, none of the MAs provides the profitable trading strategy. One of the possible explanations might be that the X instrument has mean-reverting nature in the long run, therefore, trend-following strategies might not work for this instrument.

As for the DAX and Brent cases, KMA generates fewer trades than SMA (see Figure 1.28).

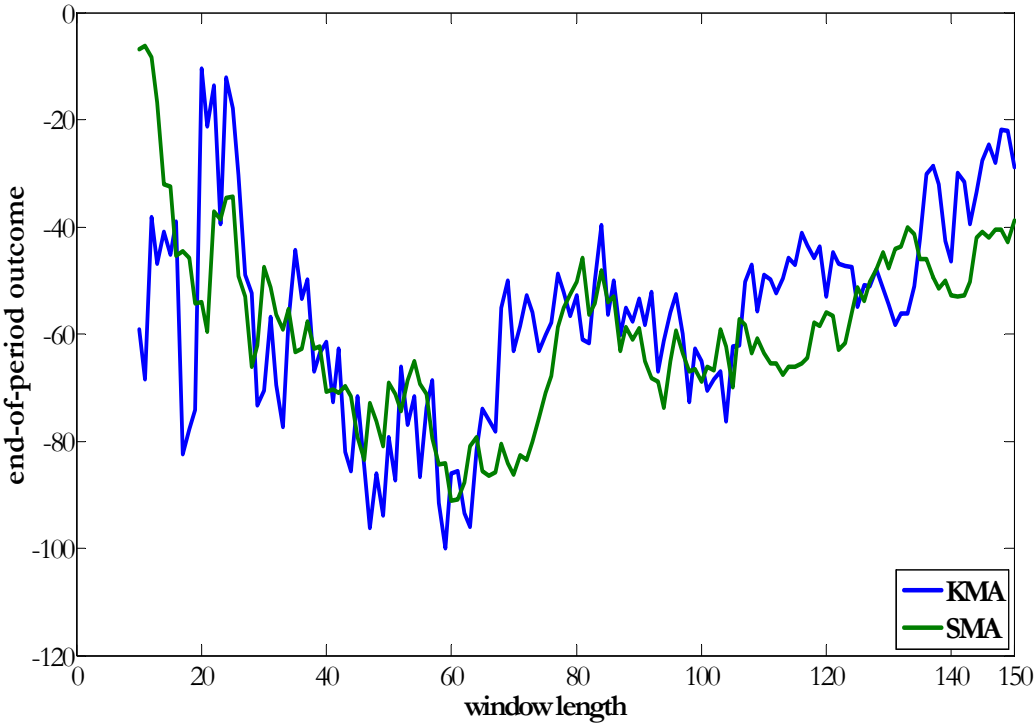


Figure 1.27. X instrument (frequency – 1 hour, observations 2372-4743): End-of-period outcomes for the strategies, based on KMA and SMA of the different lengths

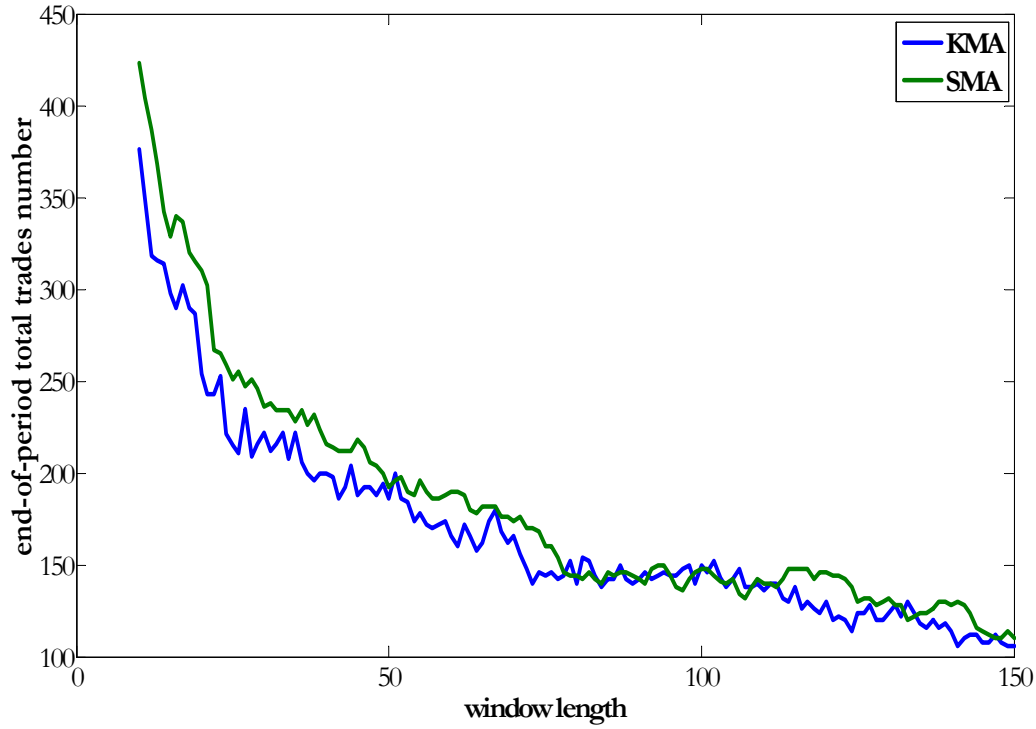


Figure 1.28. X instrument (frequency – 1 hour, observations 2372-4743): End-of-period total trades number for the strategies, based on KMA and SMA of the different lengths

Comparing the performance of the KMA and SMA of the same length for the considered instruments, we can conclude that at short lengths KMA performs better than SMA, producing fewer trades; at long lengths the difference in the performance is less pronounced. The explanation to this phenomenon lays in the behavior of both curves. KMA is more volatile than the SMA and it oscillates around the SMA curve. The volatility and amplitude of the oscillations is the indirect function of the KMA length: the longer the KMA the less it is volatile and it coincides more with the SMA curve.

KMA seems to perform better during trending periods and worse during trendless periods than the SMA of the same length. The explanation to this lays in the weight structure of the MAs. As markets are trending, KMA is more sensitive to the last large price changes, than SMA as it assigns larger weight to the last available observation[30]. During the trendless periods such structure makes the KMA more erratic than the SMA, causing more false signals, and thus, less profit. This might be an explanation why the trading results for the X instruments are so poor.

For trending price patterns KMA produces fewer trades than the SMA of the same length even having a more erratic nature. Taking into account non-zero transaction costs in real-life applications, fewer trades end up at lower trading costs.

5 Kriging results: Bounded, evenly spaced time-series data

The previous Chapter 4 shows the results of the kriging method application to the estimation of the price mean for series that are normally non-stationary due to the presence of trend. This chapter analyzes the case of the data, which is bounded and has mean-reverting nature.

5.1 Moving Average Convergence/Divergence indicator (MACD) and trading strategies

Moving Average Convergence/Divergence indicator (MACD) is a technical indicator, developed by G. Appel (Murphy, 1999).

Definition 1.4
MACD is the difference between two exponential moving averages of different length:

$$MACD_t = EMA_{1,t}\left(\alpha\left(n_1\right)\right) - EMA_{2,t}\left(\alpha\left(n_2\right)\right), \qquad \text{(I.5.1)}$$

where $EMA_{i,t}$ - exponential moving average at the moment of time t;
$\alpha\left(n_i\right)$ - parameter of the EMA, as a function of its effective length n_i.

Figure 1.29 presents the example of the MACD indicator for the effective lengths of $n_1 = 12$ and $n_2 = 26$ days estimated for the sample of Bund (1991-2006) with daily frequency.

[30] KMA optimal structure also assigns more weight to the first observation in the window, but due to the presence of the trend this value, depending on the trend direction, is much smaller or much bigger than the last available observation; therefore its weighted impact is smaller on the mean value than for the last observation in the window.

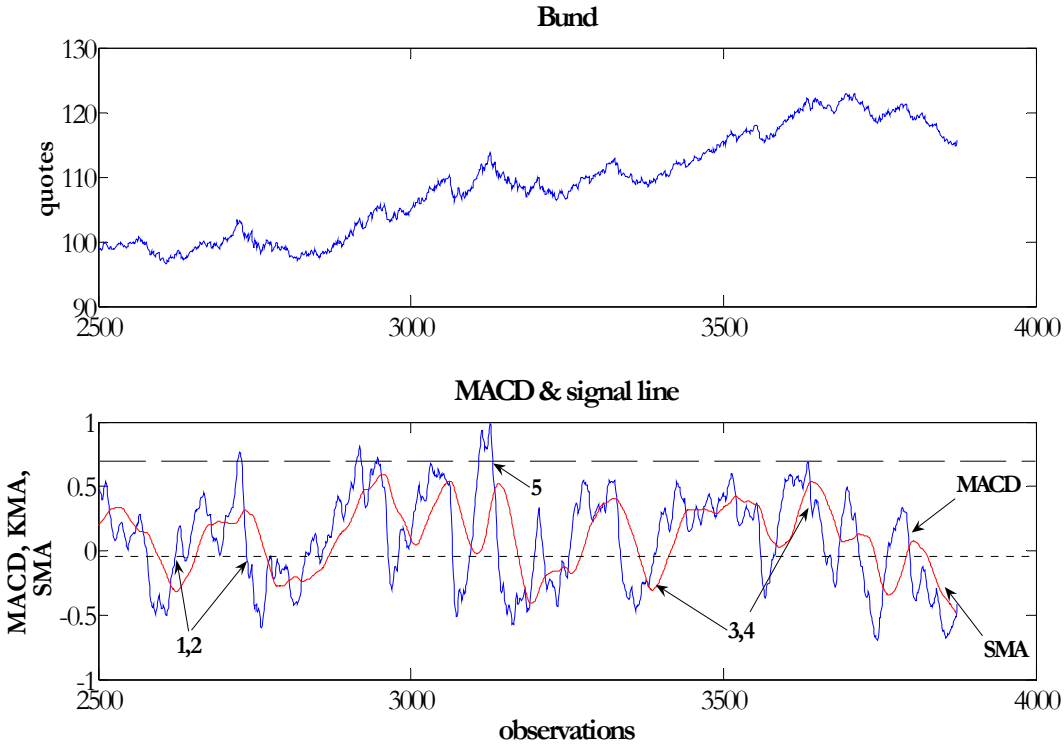


Figure 1.29. Bund (1991-2006, frequency=1day), MACD ($n_1 = 12$, $n_2 = 26$ days)

Moving Average Convergence/Divergence indicator (MACD) belongs to the group of trading oscillators. As long as the choice of the EMAs effective lengths in (I.5.1) implies the crossover of both EMA curves, the MACD oscillates around 0-level. The amplitude and frequency of the oscillation depend on the choice of the n_1, n_2 values (see Figure I1 in the Appendix I for the examples of the MACD indicator calculated for the Bund sample).

From the statistical point of view, the major quality of the MACD indicator is its mean-reverting nature; therefore it is more likely to be stationary than the price series.

Different trading rules are defined on the basis of the MACD values (Murphy, 1999). One strategy coincides with the technical rule, based on the crossovers of the two MAs of different length[31]: when MACD line crosses the zero-line above - the buy signal is generated (point #1 in Figure 1.29), when MACD line crosses the zero-line below - the sell signal is generated (point #2 in Figure 1.29).

Another rule is based on the crossovers of the MACD line with its signal line. Signal line is defined by MA (for example EMA), which is constructed on the basis of the MACD values. The crossing of the MACD line above the signal line is the buy signal (point #3 in Figure 1.29); the crossing of the MACD line below the signal line generates the sell signal (point #4 in Figure 1.29).

[31] The trading rule based on the two MA of different length states the following:

- Buy signal: Shorter moving average rises above the longer moving average;
- Sell signal: Shorter moving average fall below the longer moving average.

A contrarian strategy is constructed on the basis of the extreme MACD values, which suggest that prices have gone too far too fast and therefore, are subjects for some corrections: an overbought conditions are present when the MACD is too far above the zero line (point #5 in Figure 1.29), while the oversold conditions are present when the MACD is too far away below the zero line.

Finally, a rule is constructed on the convergence/divergence of the price and MACD trends. Negative divergence takes place when the MACD line is well above the zero line and starts the negative trend at the time when prices exhibit positive trend; in this case, the sell signal is generated. The positive divergence happens when the MACD line is well below the zero-line and starts exhibiting the positive trend earlier than the price line does; then, the buy signal is generated.

We choose for evaluation and optimisation the strategy, based on the crossovers of MACD and signal lines. This will allow defining the optimal signal line as a kriged MA.

5.2 MACD strategy: optimal signal line

A strategy, based on the crossovers of MACD and signal lines has at least three parameters to optimise: n_1, n_2 - the lengths of the EMAs and n_S - the length of the MA chosen as a signal line. Taking into account that the goal of this sub-chapter is to find the optimal MA for the stationary data, we choose to optimise only the value n_S, while accepting some default values for the lengths n_1, n_2 involved in MACD calculations. From the empirical variogram in Figure H2 (Appendix I) we have chosen default parameters $n_1 = 12$ and $n_2 = 26$ for MACD calculations, which represent some average variogram for the MACD indicator for the Bund instrument (see Figure 1.30).

Figure 1.30 presents the empirical variogram for MACD(12-26) and the combination of two theoretical models fitted to these values. The model that is fitted to the data is a sum of the gaussian model $\gamma(h) = \sigma^2\left(1 - e^{-\left|\frac{h}{a}\right|^2}\right)$ and damped cosines model $\gamma(h) = \sigma^2\left(1 - e^{-\alpha\left|\frac{h}{a}\right|}\cos\left|\frac{h}{a}\right|\right)$:

$$.\gamma(h) = 0.05\left(1 - e^{-\left|\frac{h}{17}\right|^2}\right) + 0.04\left(1 - e^{-0.4\left|\frac{h}{18}\right|}\cos\left|\frac{h}{18}\right|\right). \qquad \text{(I.5.2)}$$

The optimal weights[32] for the KMA estimates are presented in Figure 1.31 (window=50 observations). Figure 1.32 represents the MACD and two signal lines - KMA and SMA of the same length. We can see again that KMA is more volatile than the SMA and oscillates around it.

[32] The kriging weights that correspond to the variogram model (I.5.2).

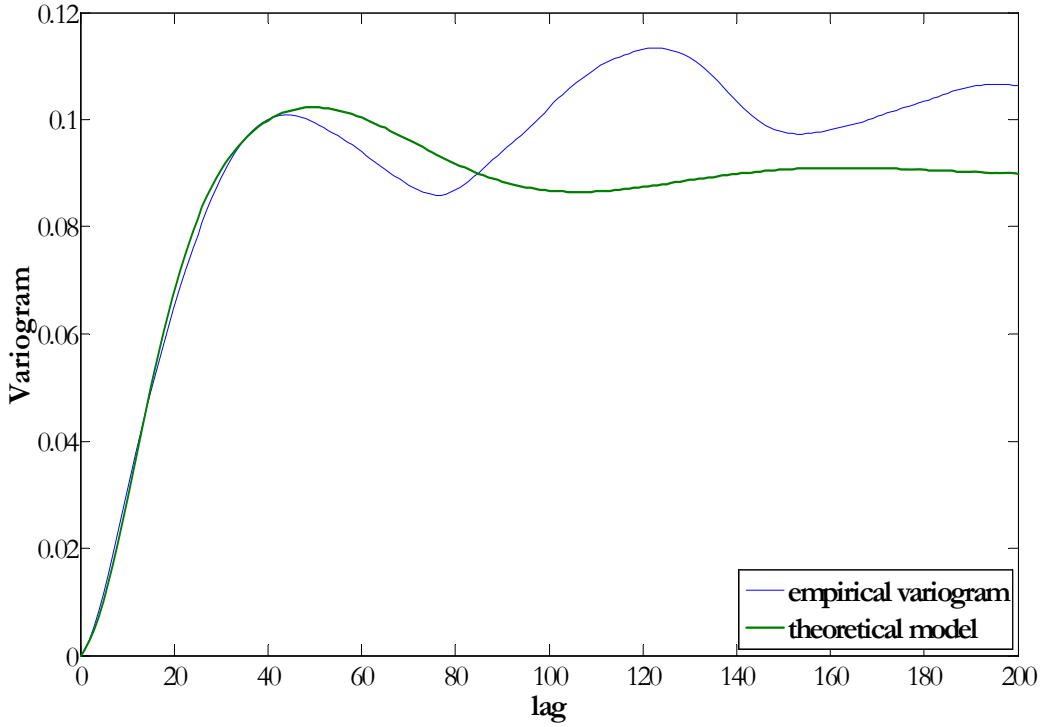


Figure 1.30. Bund (1991-2006, frequency = 1 day): Empirical variogram of the MACD indicator ($n_1 = 12$ and $n_2 = 26$) and theoretical model $\gamma(h) = 0.05\left(1 - e^{-\left|\frac{h}{17}\right|^2}\right) + 0.04\left(1 - e^{-0.4\left|\frac{h}{18}\right|}\cos\left|\frac{h}{18}\right|\right)$

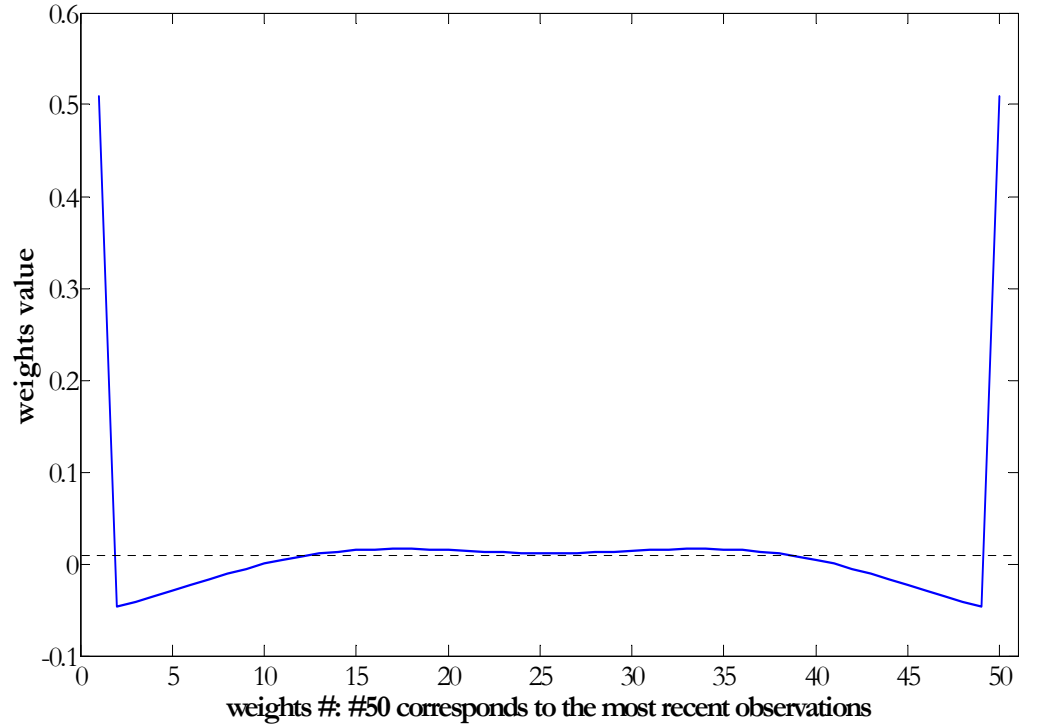


Figure 1.31. Bund (1991-2006, frequency = 1 day): Optimal weights for the MA estimate for MACD indicator ($n_1 = 12$ and $n_2 = 26$), variogram model

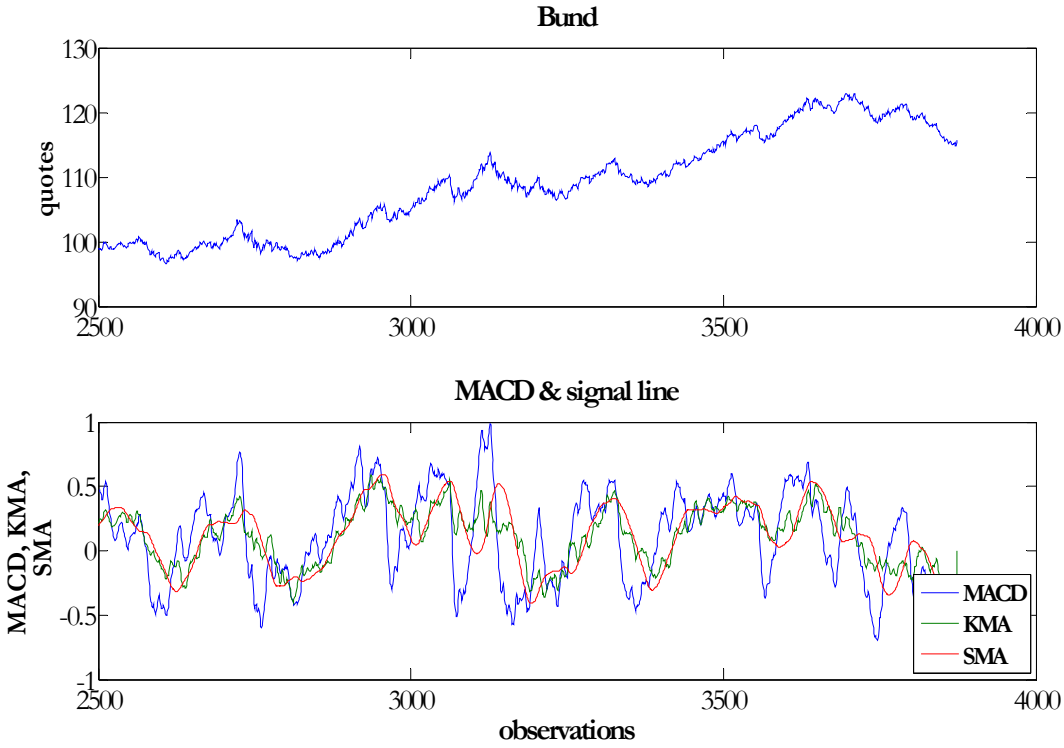


Figure 1.32. Bund (1991-2006, frequency = 1 day), MACD and its signal lines represented by KMA and SMA of the same length (50 observations)

Further in Chapter 5.3 we compare the trading outcomes for the trading strategies, based on the crossovers of MACD curve and the signal lines such as KMA, SMA and EMA of the same length. The analysis will be conducted for four instruments[33]: (1) Bund; (2) DAX; (3) Brent; (4) X instrument. The default parameters of the MACD indicators are chosen at levels specific for each instrument taken into account its frequency.

5.3 Results of the trading strategy, based on the MACD indicator and its signal lines

Suppose $\{P_i\}_{i>0}$ is an instrument's prices and $EMA_{1,t}, EMA_{2,t}$ are the exponential MAs of the length n_1, n_2:

$$EMA_{i,t} = \alpha_i P_t + (1-\alpha_i) EMA_{t-1},\ \alpha_i = \frac{2}{2+n_i}.$$

Then MACD indicator is $\{MACD_i\}_{0<i\le t}$: $MACD_t = EMA_{1,t} - EMA_{2,t}$.

The trading strategy based on MACD can be formulated as following:

1. Trading costs are 0. Profits are defined in quotes units[34].
2. On the basis of the history of estimated MACD indicators $\{MACD_i\}_{0<i\le t}$ for each moment t we construct the signal line in the form of MA of the length n_S: $\{MA_{S,i}\}_{i>0}$. The following are the types of the MA used as a signal line:

[33] Note that DAX, Brent and X instrument samples are the same as in Chapter 4, while we consider new sample for Bund instrulent.

[34] Instruments values are usually quoted in ticks, not in currency equivalents.

– Exponential moving average:

$$EMA_{S,t} = \alpha_S MACD_t + (1-\alpha_S) EMA_{t-1},\ \alpha_S = \frac{2}{2+n_S};$$

– Kriged moving average:

$$KMA_{S,t} = \sum_{i=1}^{n_S} \lambda_i MACD_{t-n_S+i}$$

– Simple moving average:

$$SMA_{S,t} = \frac{\sum_{i=1}^{n_S} P_{t-i+1}}{n_S}.$$

3. Lets define $R_i = MACD_i - MA_{S,i}$.
4. The initial trading position $Pos_0 = 0$; trading outcome $\Pi_0 = 0$.
5. The first trade ($Pos_i, i > 0$) is undertaken at the first crossovers of the MA and MACD lines, i.e. under following condition:
   if $(R_i R_{i-1} \le 0)$ and $(R_{i-1} < 0)$: $Pos_i = 1,\ P_{entry} = P_i,\ \Pi_i = 0$
   if $(R_i R_{i-1} \le 0)$ and $(R_{i-1} > 0)$: $Pos_i = -1,\ P_{entry} = P_i,\ \Pi_i = 0$
   otherwise, $Pos_i = 0,\ \Pi_i = 0$
6. Afterwards, for the new trading signals (curves crossovers) the following trades are executed:
   if $(R_t R_{t-1} \le 0)$ and $(R_{t-1} < 0)$:
   - exit (previously taken) position *Short* ($Pos_{t-1} = -1$): $Pos_t = 0, P_{exit} = P_t$
   - entry position *Long*: $Pos_t = 1,\ P_{entry} = P_t$
   - cumulative trading outcome: $\Pi_t = \Pi_{t-1} + Pos_{t-1}(P_t - P_{t-1})$

   if $(R_t R_{t-1} \le 0)$ and $(R_{t-1} > 0)$:
   - exit (previously taken) position *Long* ($Pos_{t-1} = 1$): $Pos_t = 0, P_{exit} = P_t$
   - entry position *Short*: $Pos_t = -1, P_{entry} = P_t$
   - trading outcome for this operation: $\Pi_t = \Pi_{t-1} + Pos_{t-1}(P_t - P_{t-1})$

   otherwise, $Pos_t = Pos_{t-1}, \Pi_t = \Pi_{t-1} + Pos_{t-1}(P_t - P_{t-1})$

This trading strategy was applied to four different data samples of different frequency. For DAX, Brent and X instrument the same data samples are used as in the Chapter 4. For Bund instrument we chose the new sample of the daily frequency. For Bund case the variogram model (I.5.2) estimated on the whole sample, that is also used for kriging applications. For the DAX, Brent and X instrument we have divided each sample in two sub-samples: the first sub-sample is used for the variogram estimations and the second sub-sample – for the simulation of the trading activity.

The results of the application of the trading strategy, described above are summarized in Table 1.2. The following general conclusions can be made:

1. The optimal signal line defined by KMA has much shorter length than the optimal SMA and EMA. (The only exception is the X instrument, but these results are not representative as optimal SMA and EMA in fact bring minimal losses not maximum profits).
2. KMA leads to higher absolute profits for all four instruments.

3. For Bund and Brent instruments, the efficiency of trades (profits per trade) is the lowest for KMA.

Table 1.2
The outcomes of the simulated trading strategies, based on MACD indicator

Instrument	Frequency	Optimal length (obs)			Max profit (in quotes)			Number of trades		
		KMA	SMA	EMA	KMA	SMA	EMA	KMA	SMA	EMA
Bund	1 day	18	37	103	45.65	39.48	43.79	195	127	112
DAX	30 min	36	145	102	1587	623	680	264	134	161
Brent	30 min	49	97	85	32.63	28.93	29.45	45	29	31
X instrument	1 hour	37	35	21	9.18	-19.71	-40.39	57	53	73

Further we consider the results of the strategy simulation for each instrument in more details.

5.3.1 Bund

For the simulation of the trading strategy we use the sample of daily observations for Bund (1991-2006) analyzed in Chapters 5.1 and 5.2. MACD is defined as the difference of the EMAs of the lengths $n_1 = 12$ and $n_2 = 26$; the estimated variogram follows model (I.5.2).

Figure 1.30 and 1.31 compare the profits and trade number for different length and types of the signal line. The MACD strategy leads to the positive results for all types and length of the signal lines (Figure 1.30). The usage of the KMA as a signal line leads to the highest possible profits, although the trading outcomes are more volatile than for other signal lines. The optimal lengths of the KMA lay between 10 – 25 observations (days). EMA shows more consistent results: for longer EMAs the strategy brings high and less volatile outcomes than the other signal lines. Figure 1.31 shows that except for very short lengths 10-20 observations, KMA leads to relatively lower trades number than the signal lines defined by SMA and EMA of the same lengths.

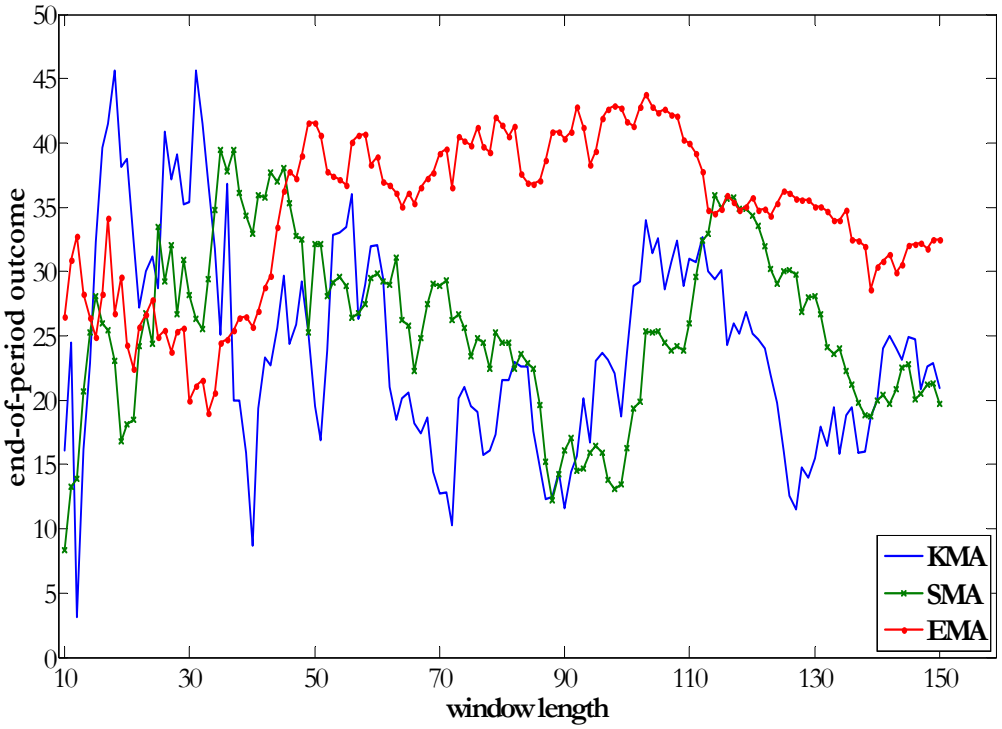


Figure 1.30. Bund (1991-2006, frequency = 1 day): End-of-period outcomes for the strategies, based on MACD ($n_1 = 12$, $n_2 = 26$) and signal lines (KMA, SMA and EMA) of the different lengths.

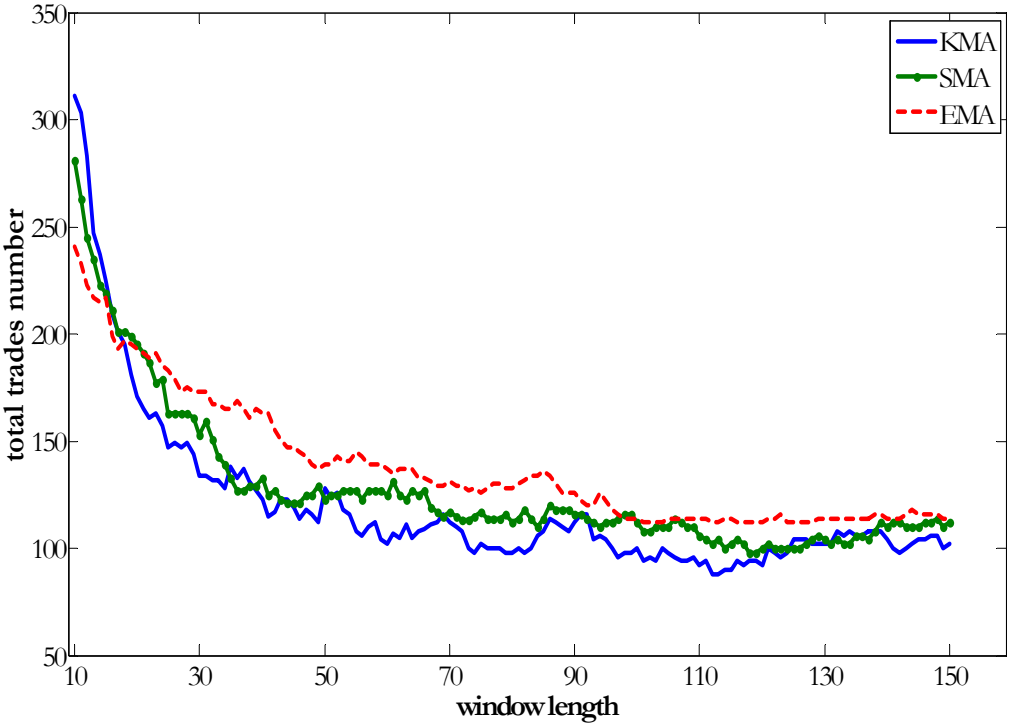


Figure 1.31. Bund (1991-2006, frequency = 1 day): End-of-period total trades number for the strategies, based on KMA, SMA and EMA of the different lengths

As for the optimal P&L paths (see Figure 1.32), all signal lines types leads to the paths that exhibits positive trends.

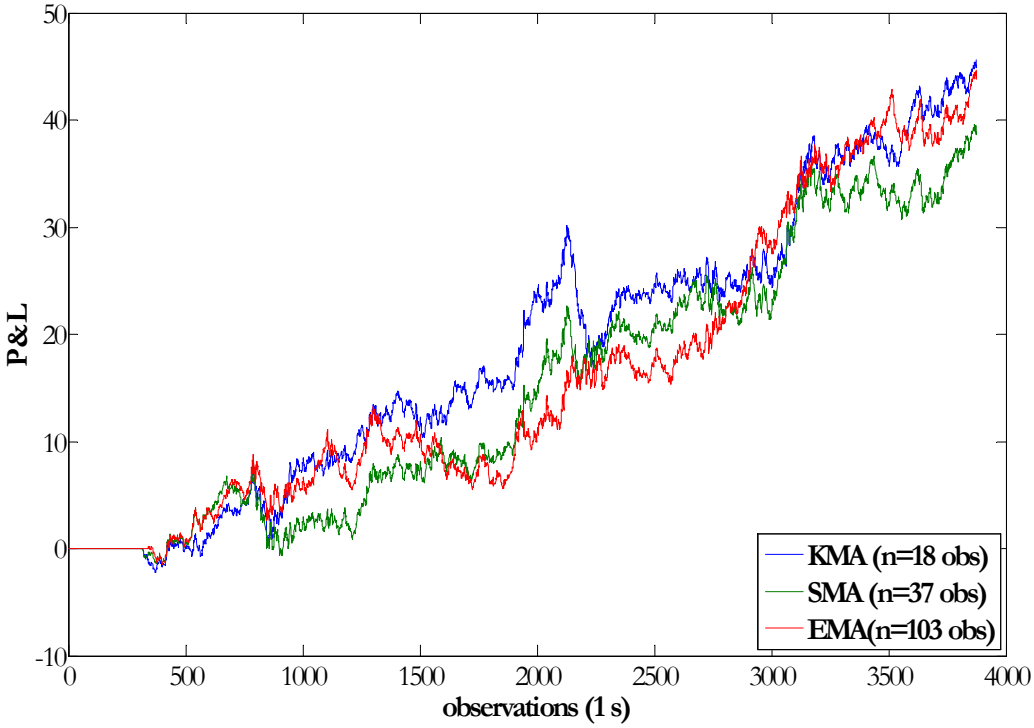


Figure 1.32. Bund (1991-2006, frequency = 1 day): Optimal P&L paths for the strategies, based on KMA (length=18 observations), SMA (length=37 observations) and EMA (length=103 observations)

5.3.2 DAX

The description of the DAX data sample is given in Appendix F. MACD indicator was constructed as the difference of the EMAs of the lengths $n_1 = 40$ and $n_2 = 80$ (see Figure E4). The MACD variogram model (see Figure F5 in the Appendix F) that is used in kriging applications is:

$$\gamma(h) = 123\left(1 - e^{-\left|\frac{h}{50}\right|^2}\right) + 123\left(1 - e^{-0.3\left|\frac{h}{80}\right|}\cos\left|\frac{h}{80}\right|\right).$$

The trading outcomes generated for different lengths of the signal line defined by KMA, SMA and EMA are presented in Figure 1.33. Contrary to Bund case, some signal lines lead to losses. The strategy based on the KMA signal line generates the highest possible profits; however, its outcomes are volatile. The optimal KMA lengths belong to the interval of low values 20-40 observations; EMA optimal lengths lay between 80-150 observations.

The number of the total trades for the KMA-based strategy is higher for short lengths; at long lengths it is similar to the number of trades generated by other types of the signal lines (see Figure 1.34).

As for the monotonicity of the optimal P&L paths in Figure 1.35, they do not exhibit constantly the patterns with positive trend and are quite volatile for ally types of the signal line.

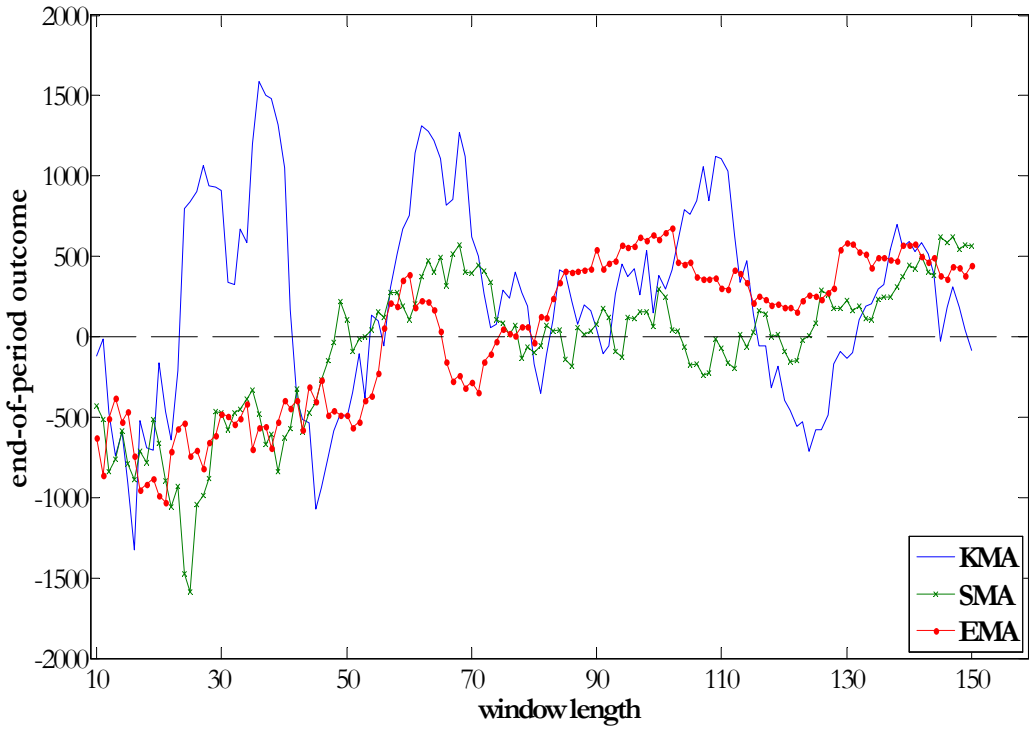


Figure 1.33. DAX (30/7/2003-7/12/2006, frequency 30 minutes, observations 10085-20170): End-of-period outcomes for the strategies, based on MACD ($n_1 = 40$, $n_2 = 80$) and signal lines (KMA, SMA and EMA) of the different lengths.

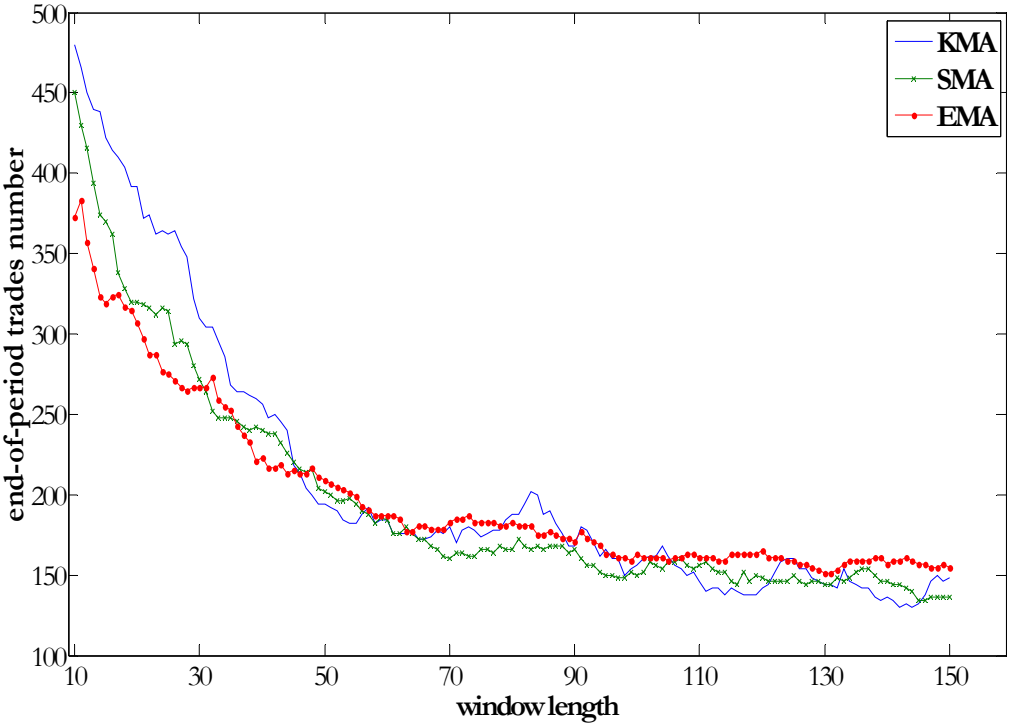


Figure 1.34. DAX (30/7/2003-7/12/2006, frequency 30 minutes, observations 10085-20170): End-of-period total trades number for the strategies, based on MACD ($n_1 = 40$, $n_2 = 80$) and signal lines (KMA, SMA and EMA) of the different lengths.

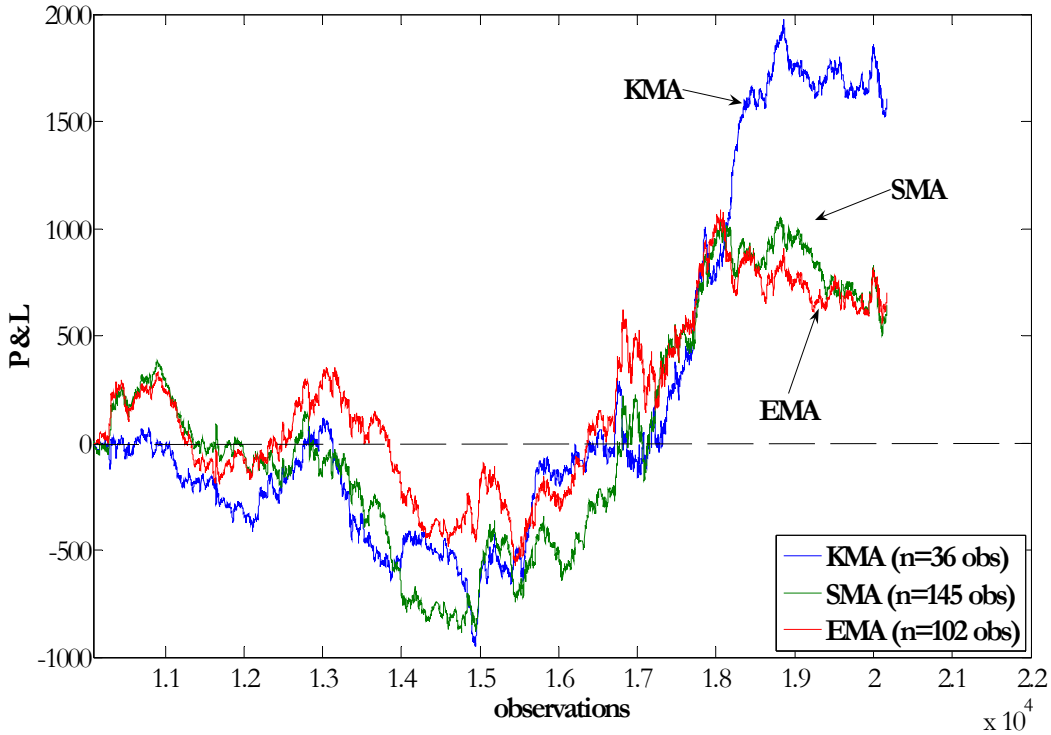


Figure 1.35. DAX (30/7/2003-7/12/2006, frequency 30 minutes, observations 10085-20170): optimal P&L for the strategies, based on MACD ($n_1 = 40$, $n_2 = 80$) and optimal signal lines (KMA, SMA and EMA)

5.3.3 Brent

Brent data sample is briefly presented in Appendix G. The MACD indicator was constructed as the difference of the EMAs of the lengths $n_1 = 40$, $n_2 = 80$ (see Figure G4 in the Appendix G). The following variogram model (see also Figure G5) is used in the kriging applications:

$$\gamma(h) = 0.12\left(1 - e^{-\left|\frac{h}{50}\right|^2}\right) + 0.1\left(1 - e^{-0.3\left|\frac{h}{50}\right|}\cos\left|\frac{h}{50}\right|\right).$$

Trading profits/losses and trades numbers generated for different lengths of the signal lines are presented in Figure 1.36 and Figure 1.37 respectively. As in the previous cases the KMA generates the highest possible profits, while its outcomes are quite volatile (see Figure 1.36). The optimal lengths of the KMA belong to the interval of the low values between 20-70 observations. Again SMA and EMA perform better as the signal line at the long distances.

As for the total number of trades the KMA generates fewer trades than respective EMA, except for the short lengths up to approximately 30 observations (see Figure 1.37).

Finally the optimal P&L path for the KMA signal line exhibits the steepest positive trend comparatively to the SMA or EMA signal lines (see Figure 1.38).

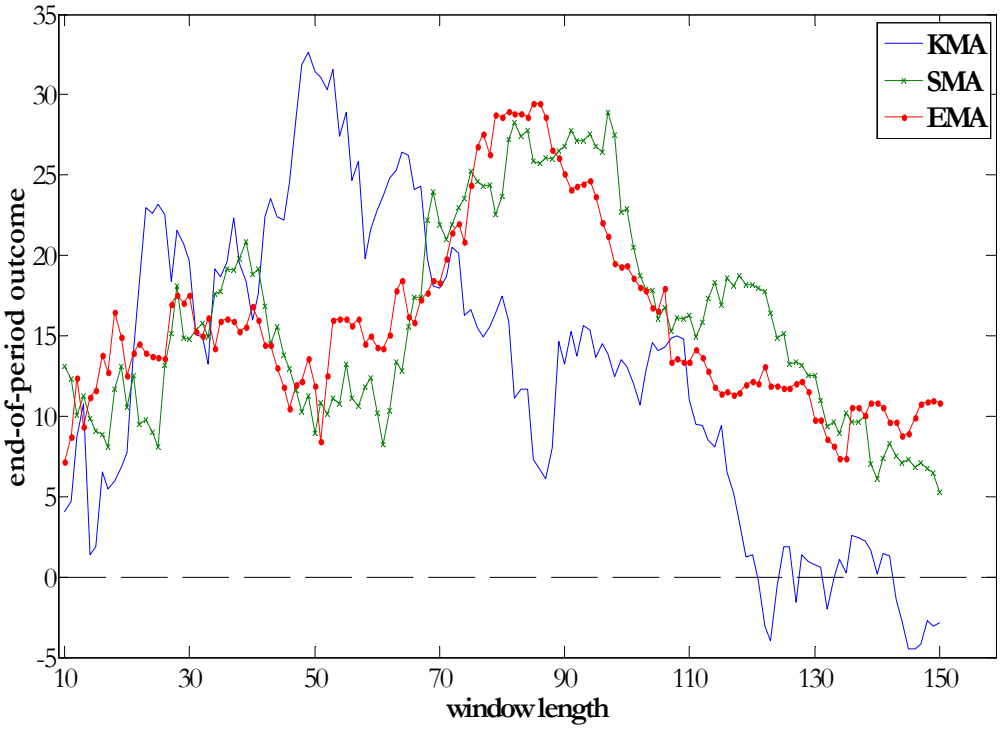


Figure 1.36. Brent (17/12/04-27/01/06, frequency 30 minutes, observations 2515-5029): End-of-period outcomes for the strategies, based on MACD ($n_1 = 40$, $n_2 = 80$) and signal lines (KMA, SMA and EMA) of the different lengths.

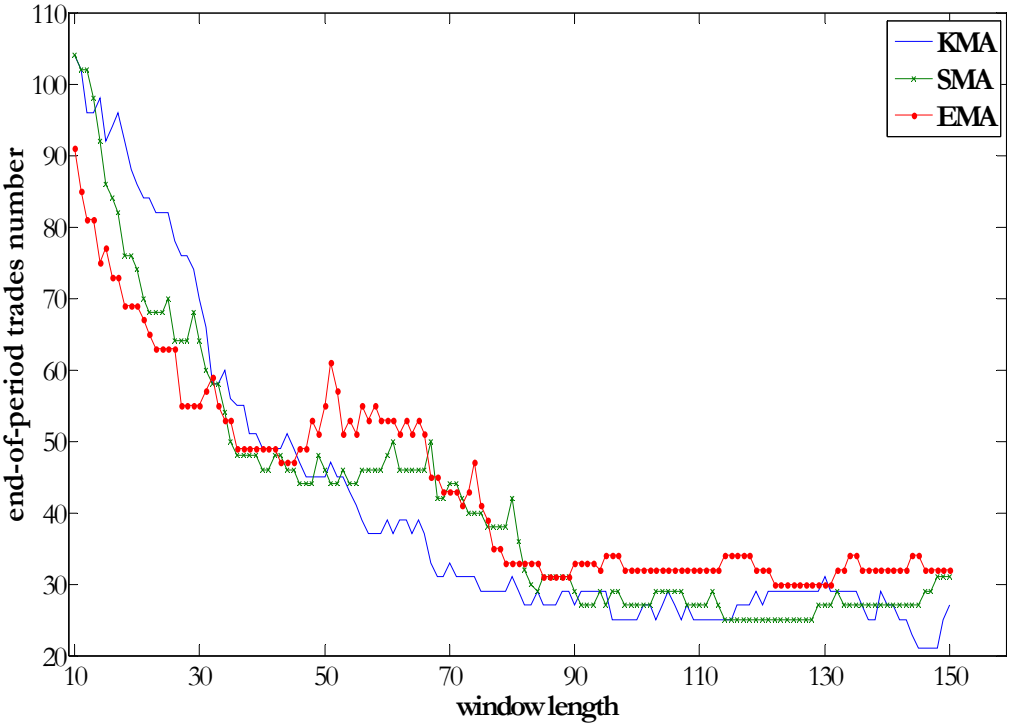


Figure 1.37. Brent (17/12/04-27/01/06, frequency 30 minutes, observations 2515-5029): End-of-period total trades number for the strategies, based on MACD ($n_1 = 40$, $n_2 = 80$) and signal lines (KMA, SMA and EMA) of the different lengths.

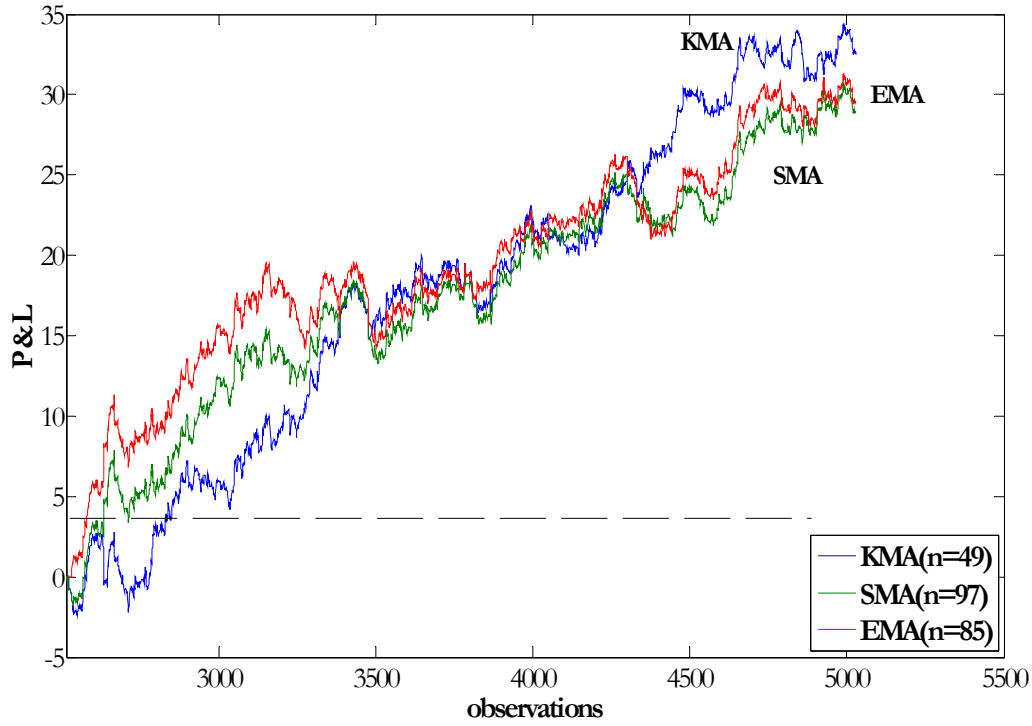


Figure 1.38. Brent (17/12/04-27/01/06, frequency 30 minutes, observations 2515-5029): optimal P&L for the strategies, based on MACD ($n_1 = 40$, $n_2 = 80$) and optimal signal lines (KMA, SMA and EMA).

5.3.4 X instrument

The data sample used for the analysis of the X instrument is presented in Appendix H. The MACD indicator was constructed as the difference between EMAs of the following lengths: $n_1 = 40$, $n_2 = 80$ (see Figure H4 in the Appendix H). The following model is fit to the empirical variogram estimates (see Figure H5 in the appendix H) and used further in the kriging applications:

$$\gamma(h) = 0.7\left(1 - e^{-\left|\frac{h}{45}\right|^2}\right) + 0.3\left(1 - e^{-\left|\frac{h}{80}\right|}\right).$$

Figure 1.39 shows that only one strategy based on the KMA of the length 37 observations as the signal line generates some positive profits. All the other MACD strategies generate losses. Due to this fact we do not provide the optimal P&L paths for this instrument.

As in the previous cases, KMA generates higher trades number at short lengths and relatively lower trades number at long lengths comparatively to the SMA and EMA.

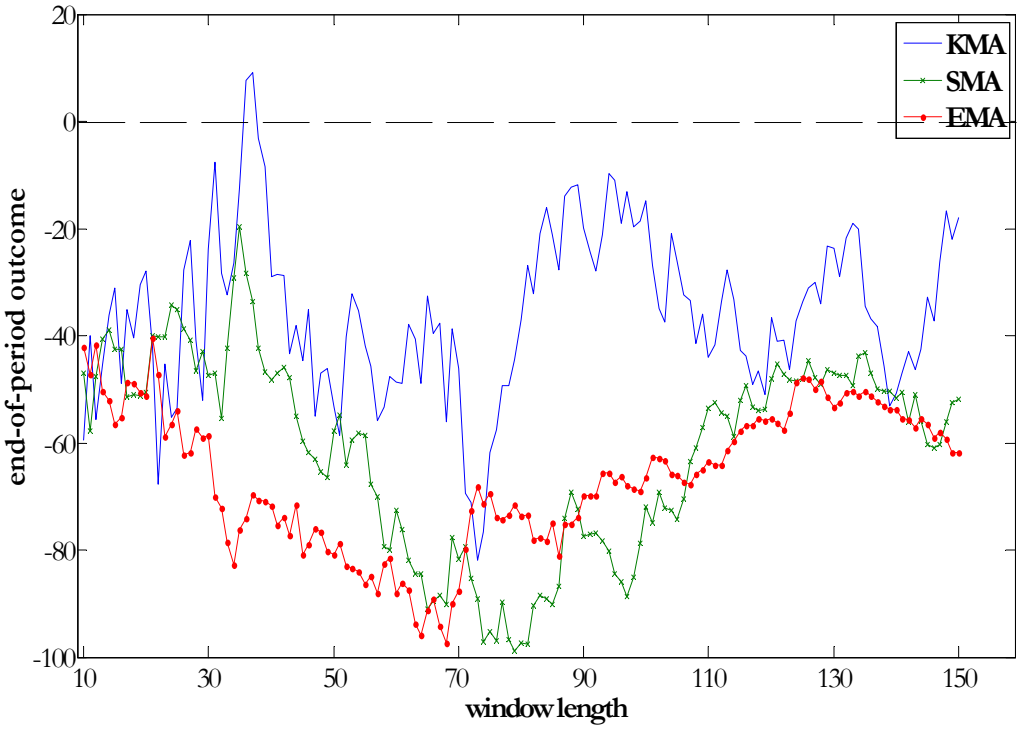


Figure 1.39. X instrument (frequency - 1 hour, $n_1 = 40$, $n_2 = 80$, obs. 2372-4743): End-of-period outcomes for the strategies, based on MACD ($n_1 = 40$, $n_2 = 80$) and signal lines (KMA, SMA and EMA) of the different lengths.

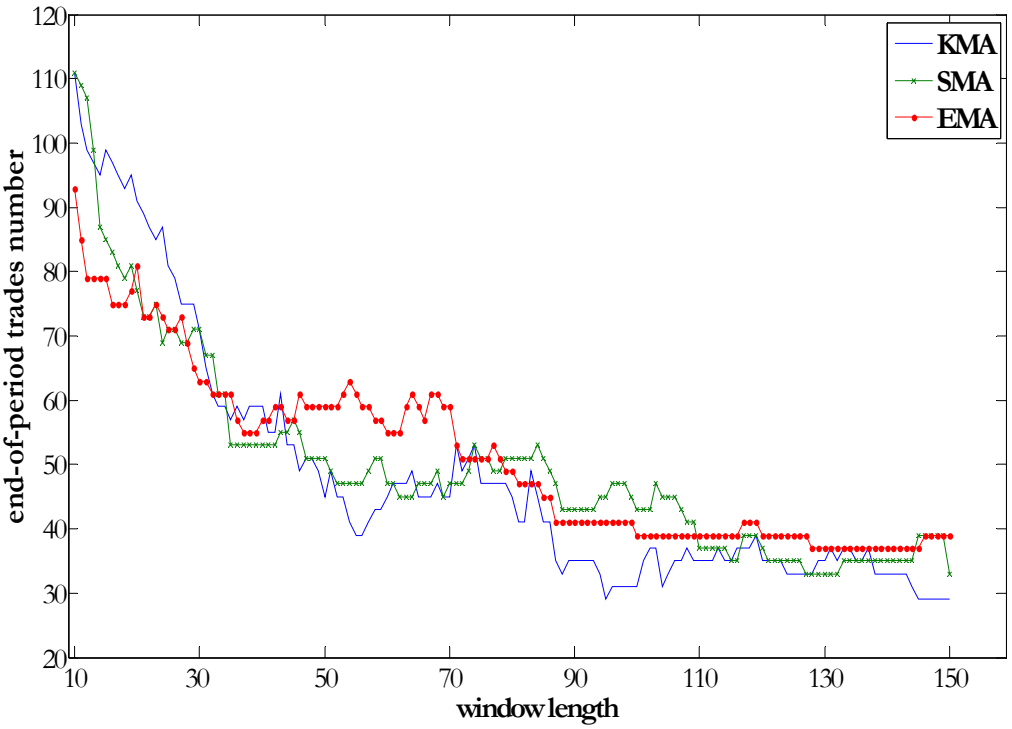


Figure 1.40. X instrument (frequency - 1 hour, $n_1 = 40$, $n_2 = 80$, obs. 2372-4743): End-of-period total trades number for the strategies, based on MACD ($n_1 = 40$, $n_2 = 80$) and signal lines (KMA, SMA and EMA) of the different lengths.

As the result, we can see that application of the kriging approach to the parameter optimization of the trading strategy, based on MACD indicator leads to similar results as in Chapter 4. KMA also leads to the absolute maximum profits for the majority of the instruments. The optimal lengths of the KMA coincide with the low values contrary to optimal EMA and SMA signal lines. KMA generates higher trades number at short lengths, but lower trades number at long lengths relatively to the EMA and SMA of the same length.

We can conclude that KMA improves the MACD trading strategy under zero-transaction-costs hypothesis. Moreover it works better at short distances, which is unlikely for the majority of other MAs that generates more false signals at these lengths.

6 Kriging results: unevenly spaced data

As have been shown in Chapter 3, for the equally spaced samples of the financial data the optimal KMA, defined by the kriging method, has a specific weight structure; KMA is close in the lag to the SMA, calculated on the same window, but might have higher volatility. Moreover the optimal weights are the same throughout the time, as the same distances $(0,1,2,...,n-1)$ between the moment of time t, where the MA is calculated and observations in the moving window of the length n that precedes this moment are used to define these weights.

However, frequently the sampling of the financial data is not done at equal distances, due to the fact that the price/quote is documented if the transaction is undertaken. Other case of the unequal sampling appears when an instrument's price is considered as a function of other than

time variable, for example volume. Such approach can be presented as a change of the coordinates. In this case the price can be considered as a subordinated process[35]. The change of coordinates helps to smooth the jumps in the price process, frequently associated with the large changes in the traded volume.

The incorporation of the volume into the process of decision-making might improve the trading results. Blume, Easley, O'Hara (1994) demonstrated that the traders who use information contained in market statistics such as prices and volume do better than the one that do not. The other articles that analyzed the importance of volume for the price prediction are Lo, Wang (2000), Campbell, Grossman and Wang (1991), Harris and Raviv (1991), Wang (1991).

Taking into account that the kriging method accounts for the difference in the distance between the points through the covariance/variogram model, it can be used to construct the optimal MA for the unevenly spaced data. In this chapter we present the examples of such method applications.

In order to analyse the KMAs in more details we have also simulated trend-following strategy, based on the crossovers of the price and MA curves. The same strategy, based on the SMA curve, is considered as a benchmark for the results comparison. The crossovers of the price and MA lines (SMA and KMA) define the strategy entry/exit signals. Lets define variable $R_i = P_i - MA_i$. Then the strategy is formulated in the following way.

Trend-following strategy

1. Trading costs are 0.
2. The initial trading position $Pos_0 = 0$; trading outcome $\Pi_0 = 0$.
3. The first trade ($Pos_i, i > 0$) is undertaken under following condition:
   if $(R_i R_{i-1} \le 0)$ and $(R_{i-1} < 0)$: $Pos_i = 1,\ P_{entry} = P_i,\ \Pi_i = 0$
   if $(R_i R_{i-1} \le 0)$ and $(R_{i-1} > 0)$: $Pos_i = -1,\ P_{entry} = P_i,\ \Pi_i = 0$
   otherwise, $Pos_i = 0,\ \Pi_i = 0$
4. Afterwards, if the trading signals are generated the following trades are executed:
   if $(R_t R_{t-1} \le 0)$ and $(R_{t-1} < 0)$:
   - exit (previously taken) position *Short* ($Pos_{t-1} = -1$): $Pos_t = 0, P_{exit} = P_t$
   - entry position *Long*: $Pos_t = 1,\ P_{entry} = P_t$
   - cumulative trading outcome: $\Pi_t = \Pi_{t-1} + Pos_{t-1}(P_t - P_{t-1})$

   if $(R_t R_{t-1} \le 0)$ and $(R_{t-1} > 0)$:
   - exit (previously taken) position *Long* ($Pos_{t-1} = 1$): $Pos_t = 0, P_{exit} = P_t$
   - entry position *Short*: $Pos_t = -1, P_{entry} = P_t$
   - trading outcome for this operation: $\Pi_t = \Pi_{t-1} + Pos_{t-1}(P_t - P_{t-1})$

   otherwise, $Pos_t = Pos_{t-1}, \Pi_t = \Pi_{t-1} + Pos_{t-1}(P_t - P_{t-1})$

As the result Chapter 6 is organized in the following way. Chapter 6.1 presents the case of unevenly spaced time-dependent data caused by the missed observations when no transaction takes place. We show how to construct KMA and compare its results with the traditional SMA,

[35] From the statistical point of view the change of the time coordinates to the volume-based (or other variable-based) axe introduce the subordinated processes for price.

estimated on the same sample. We also analyse whether the adjustment of the sample to the regular spaced one by an interpolation of missed observations can produce better KMA. Chapter 6.2 presents the kriged volume weighted moving average (KVWMA) constructed for the irregular spaced sample due to the change of the coordinate from time to volume. We also analyse the trading outcomes of the trend-following strategy, based on this KVWMA.

6.1 Unevenly spaced time dependent price series

In this chapter we present application of the kriging method to the Bund data sampled at 1 second-frequency for the day of April 18, 2006. The frequency is very high, therefore even for Bund that is traded very actively, many gaps are present in the data. Some of these gaps count up to minutes. Figure 1.41 presents the sample at real time coordinate (in seconds); 0 correspond to 9:00:00, #300 – to 9:05:00, etc.

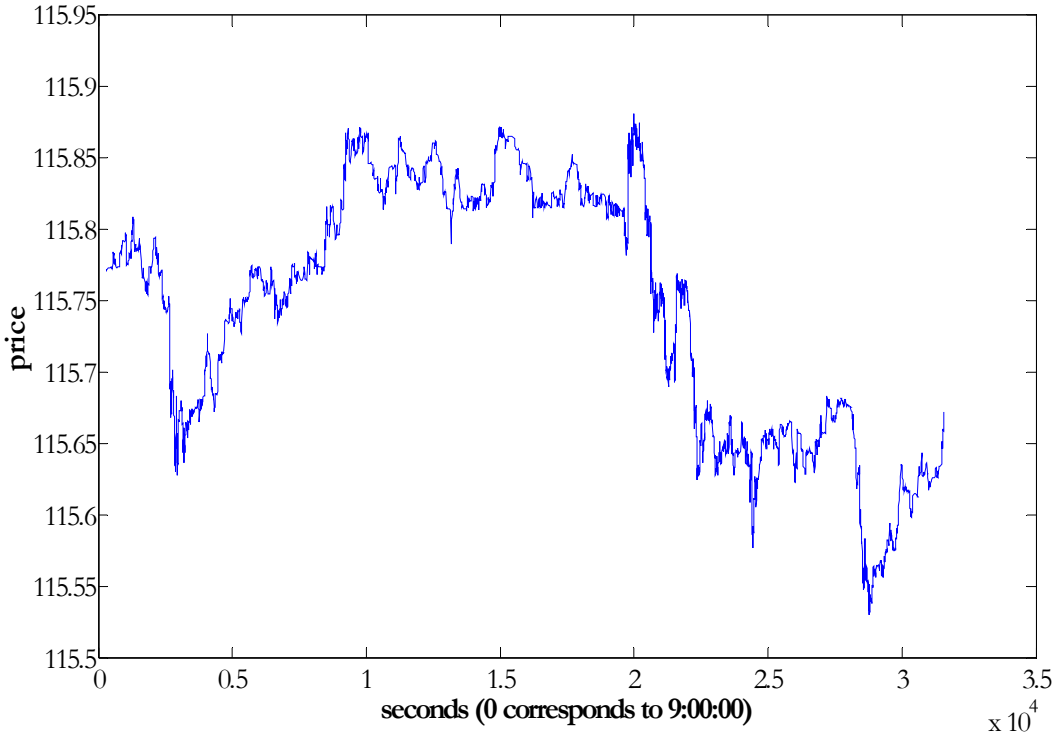


Figure 1.41. Bund (April 18, 2006, frequency 1 sec): quotes (unevenly spaced data)

In order to analyse how these gaps might impact the estimation of the optimal MAs, we propose to consider two samples:

(1) irregular sample: raw data at available time points of time
(2) regular sample: the sample obtained by filling the gaps in the data by the same values available in the previous moments of time. Such approach is based on the assumption that the when there is no trading the price stays at the level corresponding to the last transaction.

Figure 1.42 presents two variograms estimated on these samples. As we can see, the adjustment of sample to the regular sampling changes significantly the form of the variogram.

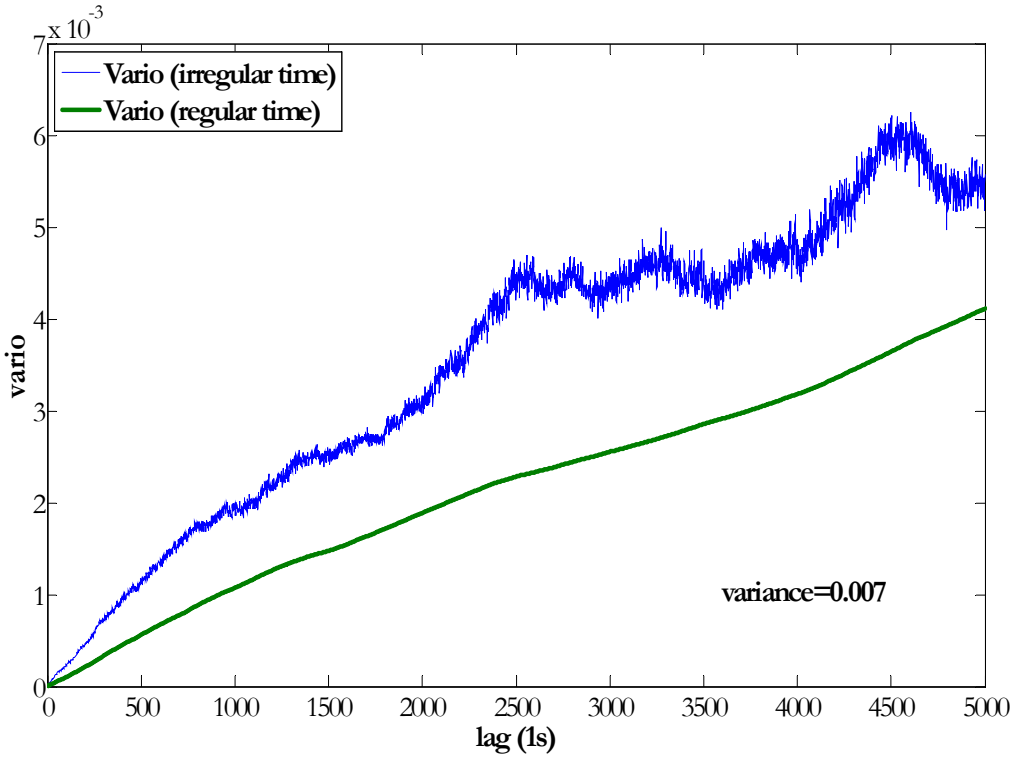


Figure 1.42. Bund (April 18, 2006, frequency 1 sec): variograms for unevenly and evenly spaced price samples

Two different models were fit to the empirical variograms:

(1) irregular sample: $\gamma(h) = 0.007\left(1 - e^{-\left|\frac{h}{3000}\right|}\right)$.

(2) regular sample : $\gamma(h) = 1.1 \cdot 10^{-6}|h|$.

The weights structure corresponding to these variogram models are given in Figures 1.43 and 1.45. Figure 1.43 presents the example of the weights for the irregular sample. It should be stressed that although the distribution of the weights is the same (the largest weights are assigned to the first and last observation), the weights will differ from window to window as they are defined also by the distance from the estimation point to other point in the window, which are irregular. This observation is support by the Figure 1.44 that presents the first weight estimated for each window. We can see that while the first weight (and the last) values contained in the interval between 0.4 and 0.5, they almost never constant.

As for the regular sample, the same weight structure that corresponds to the linear model preserves through time: $\lambda_1 = \lambda_N \approx 0.5, \lambda_{1<i<N} \approx 0$ (see Figure 1.45).

The examples of the KMA and SMA, estimated on the basis of irregular sample and KMA, estimated on the basis of regular sample and re-sampled at the points of irregular sample, are presented in Figure 1.46. Note that the effective lengths of the MAs estimated on the regular and irregular samples are different.

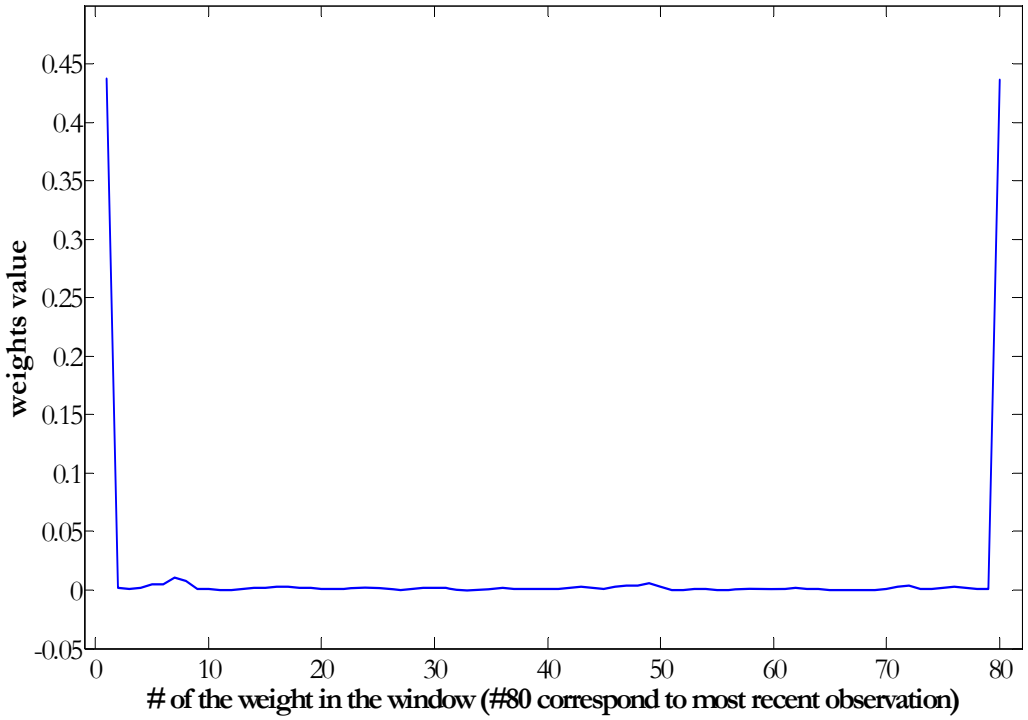


Figure 1.43. Bund (April 18, 2006, frequency 1 sec): Example of the optimal weights for the KMA for the data sampled irregular (window beginning at 3050 observation)

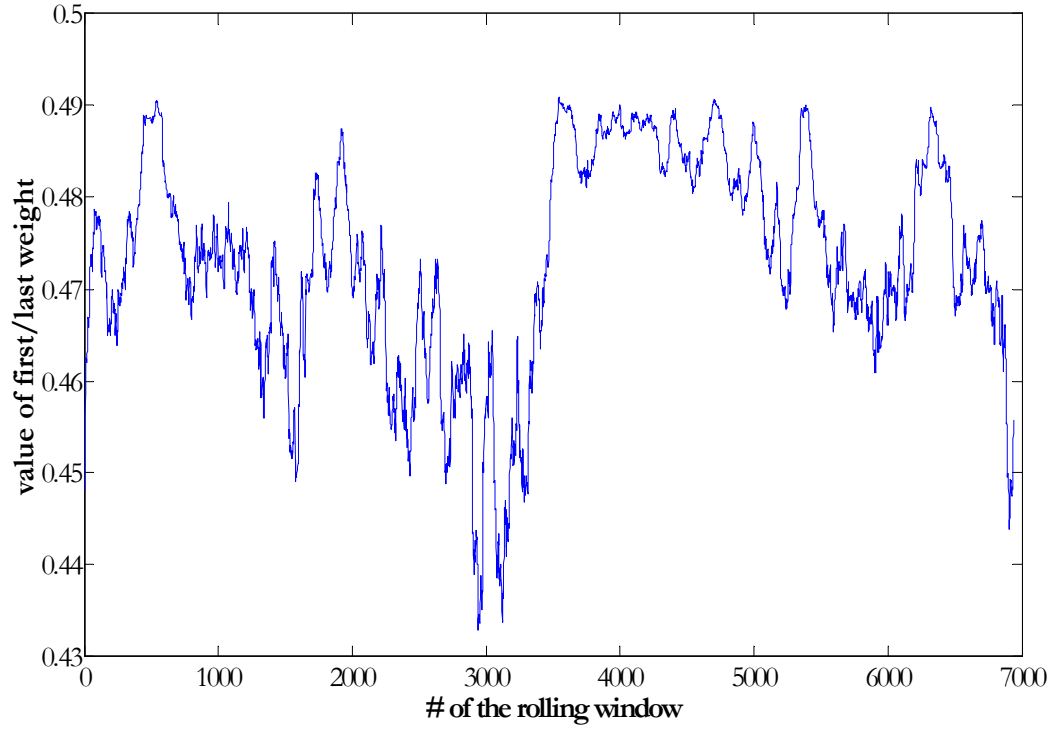


Figure 1.44. Bund (April 18, 2006, frequency 1 sec): Value of the first/last weights for the optimal KMA as a function of the window used for its estimations

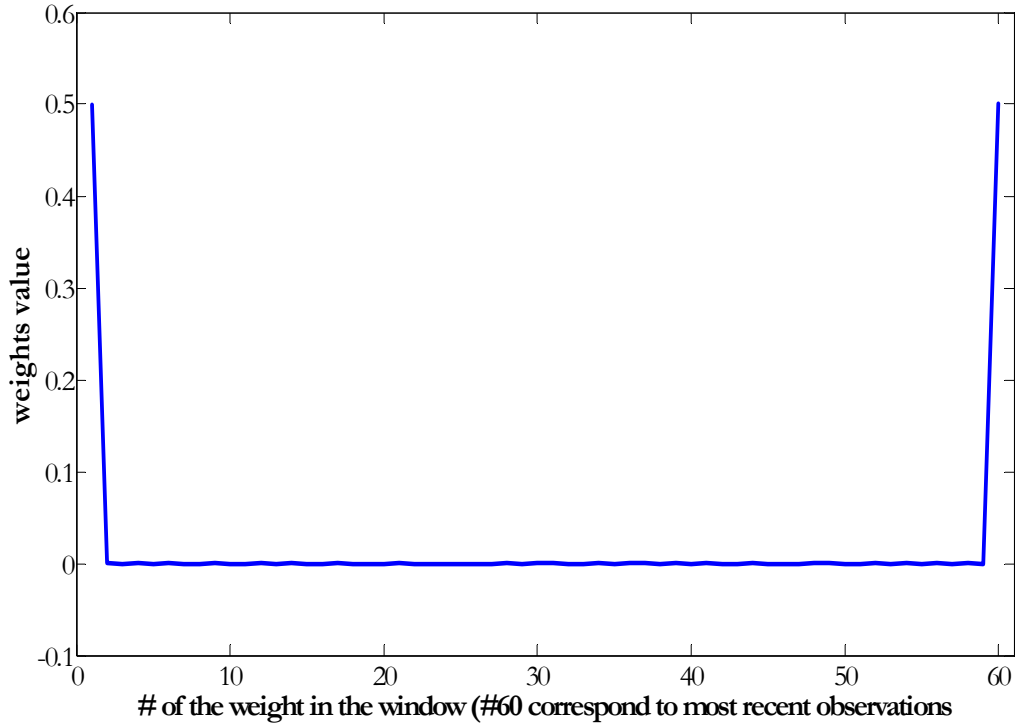


Figure 1.45. Bund (April 18, 2006, frequency 1 sec): The optimal weights for the KMA for the data re-sampled regularly

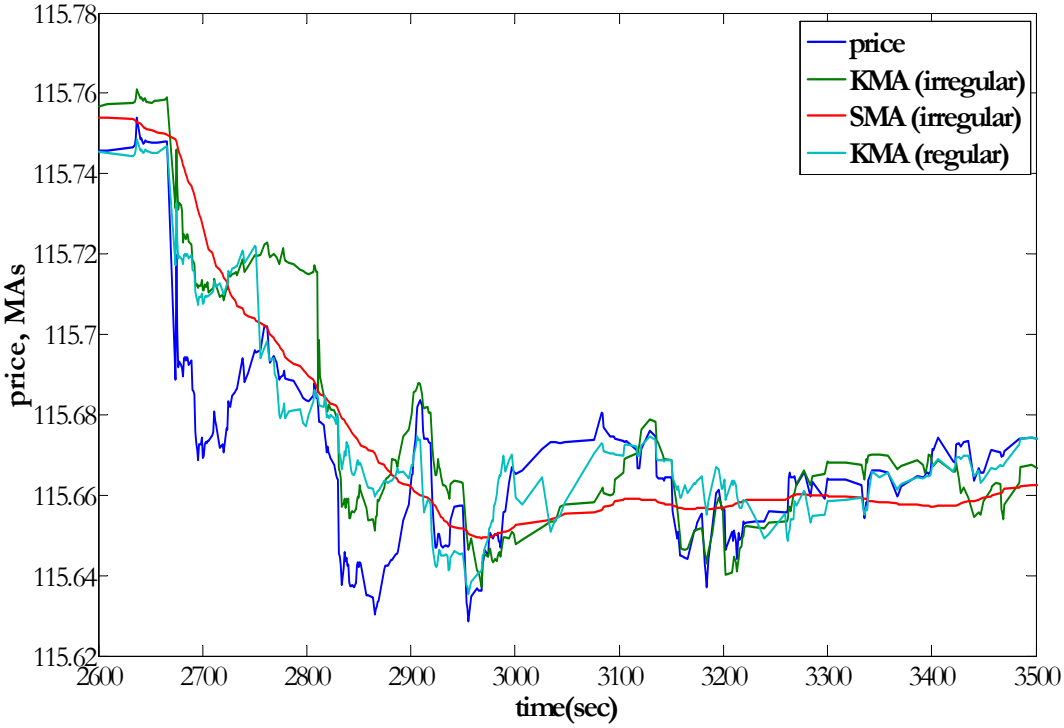


Figure 1.46. Bund (April 18, 2006, frequency 1 sec, obs. 2600-3500s): Price and different MA types (window=80 observations)

As we can see for this particular length KMA estimated on the irregular sample generate less false signals than the SMA and KMA estimated on the regular sample. For example, between 2900-3000 seconds the SMA, estimated on the irregular sample and KMA, estimated on the regular sample send false signals by crossing several times the price curve, while the KMA estimated on the irregular sample crosses it only once.

Table 1.2 and Figure 1.47 summarize the maximum profits generated by the trend-following strategy, based on the different MA types[36].

Table 1.2
Optimal outcomes of the trend-following strategy, based on the different MA types

Outcomes	KMA (irregular)	SMA (irregular)	KMA (regular)
MA length	80	300	260
Cumulative value	0.4078	0.4218	0.395
Number of trades	256	103	279

As we can see the highest outcome is achieved for the strategy, based on the SMA. The KMA, estimated for irregular samples generates comparable profits at much shorter lengths. At the same time, we can notice that the KMA calculated on the adjusted regular sample and resample for the irregular points, generate more false signals than the KMA estimated on the irregular sample; at the same time it generates some stable profits for different MA lengths.

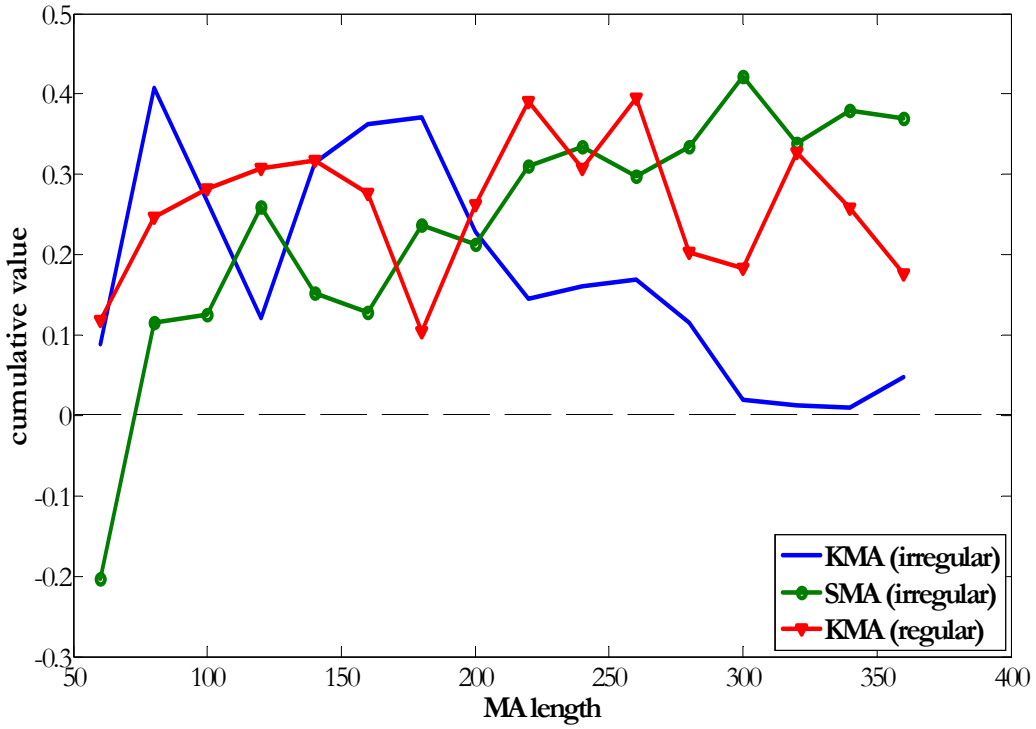


Figure 1.47. Bund (April 18, 2006, frequency 1 sec): End-of-period cumulative value of the strategy for different types of the MAs and data samples

[36] It should be noted that the MA of the length considered in the analysis are too short and less likely to be used for the data with 1-second frequency in the real-life applications. We though propose to consider these values and data as some general example.

The paths generated by the optimal MAs are presented in Figure 1.48. As we can see, the SMA does not generate consistently large profits than the other KMA. In fact the KMA, calculated on the irregular sample generate the most consistent profit path.

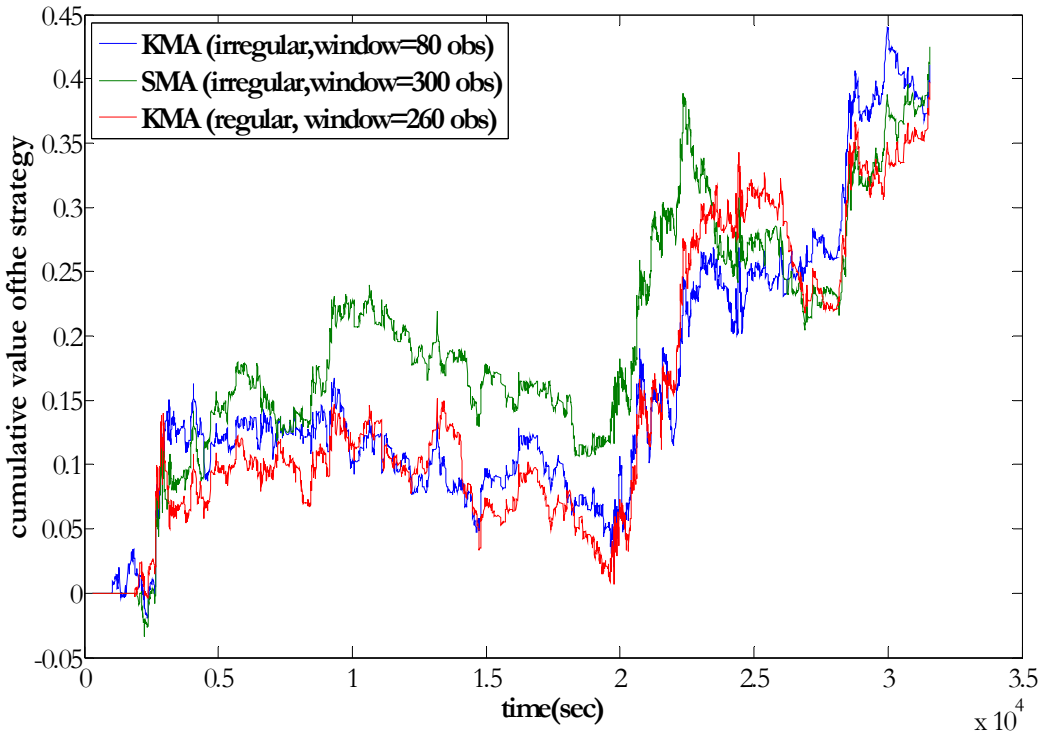


Figure 1.48. Bund (April 18, 2006, frequency 1 sec): Cumulative value of the strategy (paths) for different types of the optimal MAs

As the result we can see that KMA, calculated for the irregular samples generate consistent and the highest results at the short lengths of the MA, where traditional MAs do not work well. The KMA calculated on the adjusted regular sample generates more false signals.

6.2 Unevenly spaced price series subordinated to cumulative volume

Another case when data is unevenly sampled appears when the coordinate is changed from time to another variable, for example cumulative volume. The trading literature proposes to construct the MA that incorporate information accumulated in the trading volume. Volume-adjusted moving average (VAMA) can be represented as following:

$$VAMA_t = \sum_{i=1}^{n} w_i^* P_{t-n+i}\ ,\ w_i^* = w^*(V_i),$$

where $\{P_i\}_{i>0}$ are prices; n is MA length; and $\{w_i^*\}_{1\le i\le n}$ are MA weights as a function of the cumulative volume at the i th moment of time $\{V_i\}_{i>0}$.

The idea behind VAMA is to use some additional information about the trading activity in the definition of the MA. Traditional time-based (sampled) MA suppose that all moments of times are equal with respect to the information that they bring. VAMAs treat the activity of trading as

the measure of the price importance in the particular moments of time. For example, a sharp increase in trading activity might indicate the change in the current trend, therefore the price at this moment of time have much higher importance for trader than the prices when the market is not traded actively.

Some examples of such volume weighted MA exist. Richard W. Arms Jr ([11], V.8:3) proposed to change the system of coordinates from time- to volume based. The new curve has "prolonged" price values for heavily traded moments of time and "shorten" (or even neglected) the price values for lightly traded moments. The new VAMA can be calculated on the basis of the new price curve. Arms ([11], V.8:3) used graphical approach to the construction of VAMA, in particular equivolume charting, which can be translated into programming algorithm. Suppose that you have daily observations of price (high, low) and volume for a particular instrument. According to the equivolume charting each price observation is represented in the form of box with high being the upper side of the box, low – being the lower side of the box, and volume being the width of the box. The mean price calculated as sum of high and low divided by two, is posted one or several times in the box depending on its width. The trading rule then involves the new price curve and VAMA. As the result of the construction, VAMA approaches price curve during the periods of heavy trading more rapidly than the time-based MA. With this approach the days with heavy trading have higher weights and, thus, more importance for the definition of the MA.

At our opinion the method presented above has several drawbacks. Although the VAMA takes into account the trading volumes at each moment of time, it cannot incorporate the information about possible price (auto-) correlation at the volume points. Besides, building the price-volume blocks of different width implies price interpolation for the non-available cumulative volumes values; the interpolation of the price at the same level as previously available data can be questioned. Finally, while the MA calculation procedure can be programmed, it involves many subjective judgements, among which the definition of the volume frequency and the points at which the MA is calculated.

With respect to this critique, the kriging method can be used to define the volume weighted moving average. In this case we use a variogram $\gamma(v)$ with v as the cumulative volume lags/increments to define the optimal weights for the kriged volume-weighted moving average (KVWMA). Comparing to the VAMA the method is more straightforward. The KVWMA has even more advantages than the VAMA: it not only incorporates the information about the trading activity (volume), but also about the price autocorrelation.

Lets consider the same Bund sample to present the implementation of the method. Figure 1.49 presents the Bund quotes and trading volumes sampled at the frequency not higher than 1 second April 18, 2006. We can see from the volume curve that the trading activity varies during the day; moreover, the average activity increases after 14:00 for the European markets with the opening of the US exchanges (between 4000-5000 observations in Figure 1.49).

Figure 1.50 presents the same price series, but plotted as a function of the volume. We can notice that although some of the price jumps are reduced, the medium-term trends are still present in the price path.

Taking into account that the variability of volume can be significant, in order to calculate the empirical variogram $\gamma(v)$ we need to scale the cumulative volume by dividing each value by some Δ. Otherwise the variogram calculations might be very time consuming. We use the scaling factor of 1000 for the Bund volume values (now the Bund volume is counted in 1000^{th} units).

The price variogram calculated on the whole sample supports the non-stationarity assumption: the variogram breaks the variance of the sample (see Figure 1.51). We known already what type of the weights the linear variogram provides $\lambda_1 = \lambda_N \approx 0.5, \lambda_{0 \le i \le N} \approx 0$

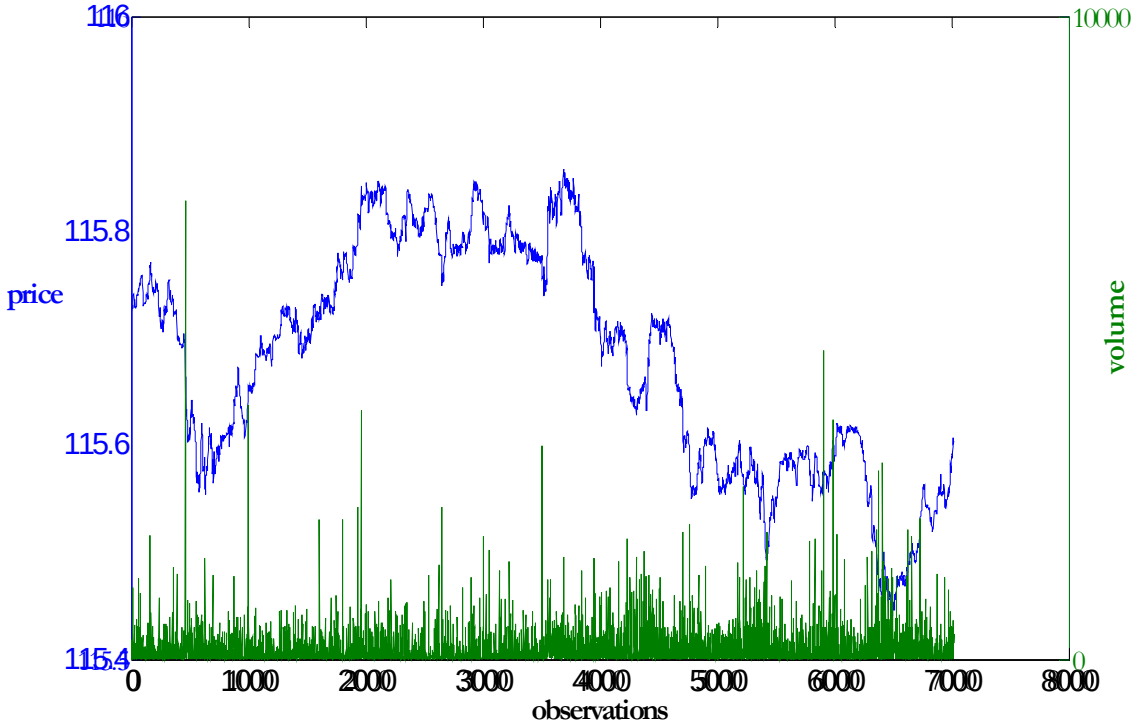


Figure 1.49. Bund (April 18, 2006, frequency 1 sec): quotes and volume.

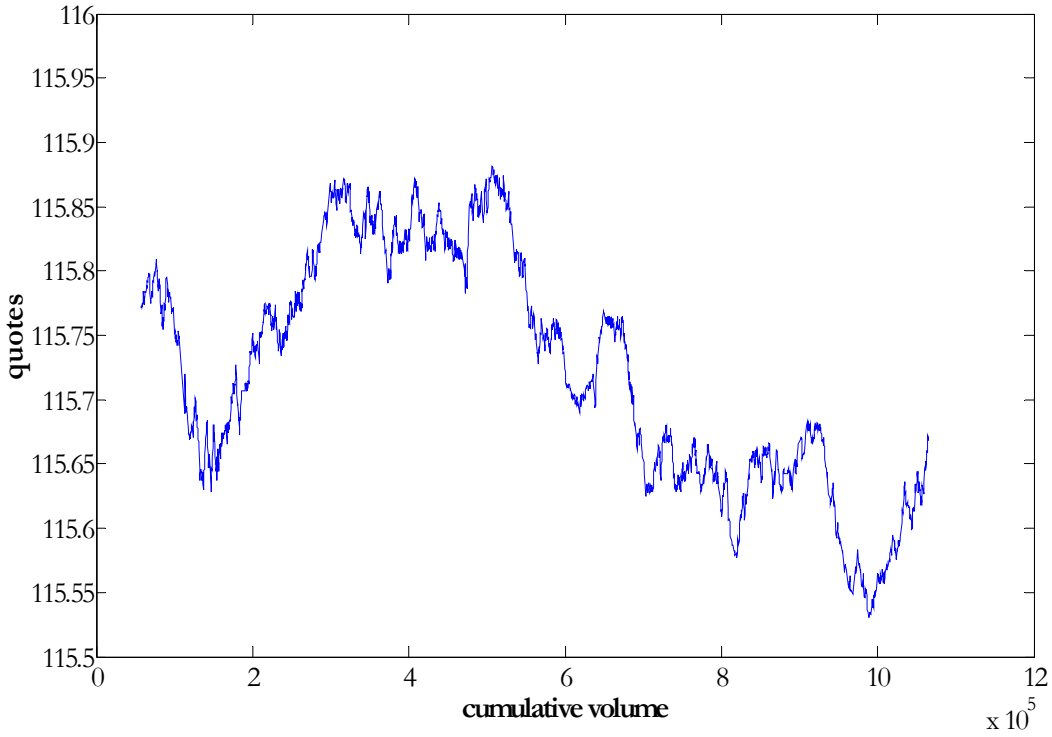


Figure 1.50. Bund (April 18, 2006, frequency 1 sec): Price as a function of the cumulative volume

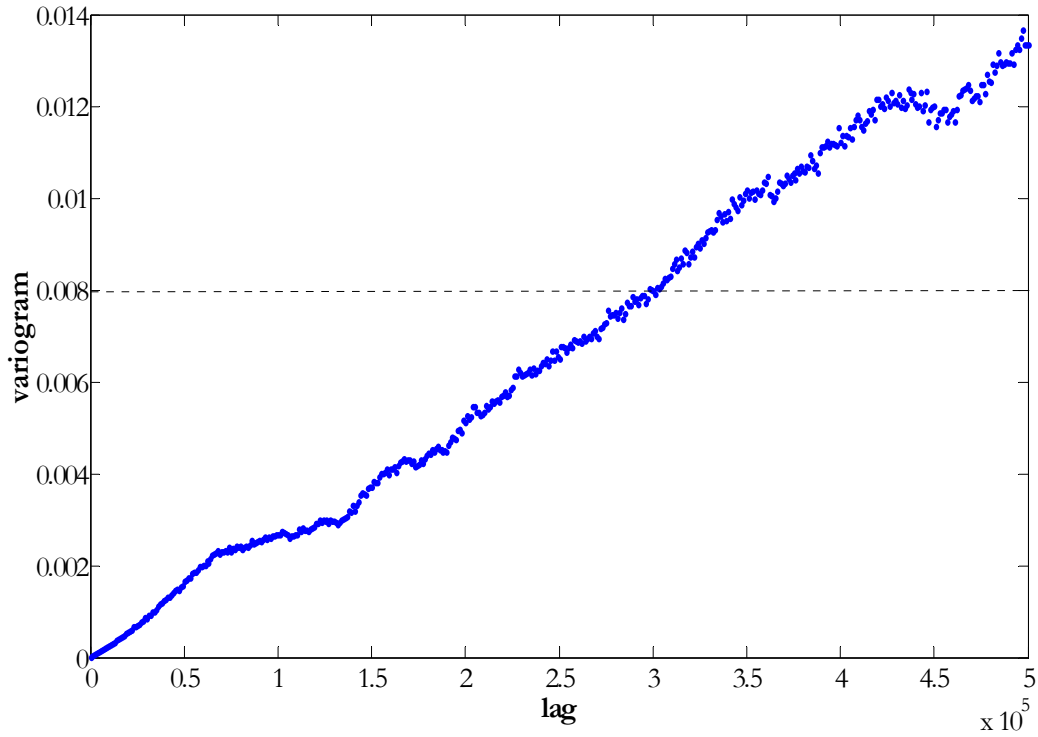


Figure 1.51. Bund (April 18, 2006, frequency 1 sec): Variogram of the price in the cumulative volume coordinate, calculated on the whole sample, lag is in 1000 volume units. (variance=0.008)

We propose to consider the problem in the stationary context. We can notice at Figure 1.49 some trend-less periods, for example, sub-samples (1900; 4000) or (4500; 7017). Therefore, we propose to make the following assumptions: the trendless periods of time allow estimating the non-corrupted covariance structure of the residuals, which is stable throughout the whole sample. We choose sub-sample (4500; 7017) (see Figure 1.52) to estimate the empirical variogram. The following model is fit to the variogram in Figure 1.53:

$$\gamma(v)=\begin{cases}0.0005\left(1-e^{-\left(\frac{v}{40}\right)^2}\right)+0.0015\left(\frac{3}{2}\cdot\frac{v}{80}-\frac{1}{2}\cdot\left(\frac{v}{80}\right)^3\right), v<80\\ 0.0005\left(1-e^{-\left(\frac{v}{40}\right)^2}\right)+0.0015, v\geq 80\end{cases}$$

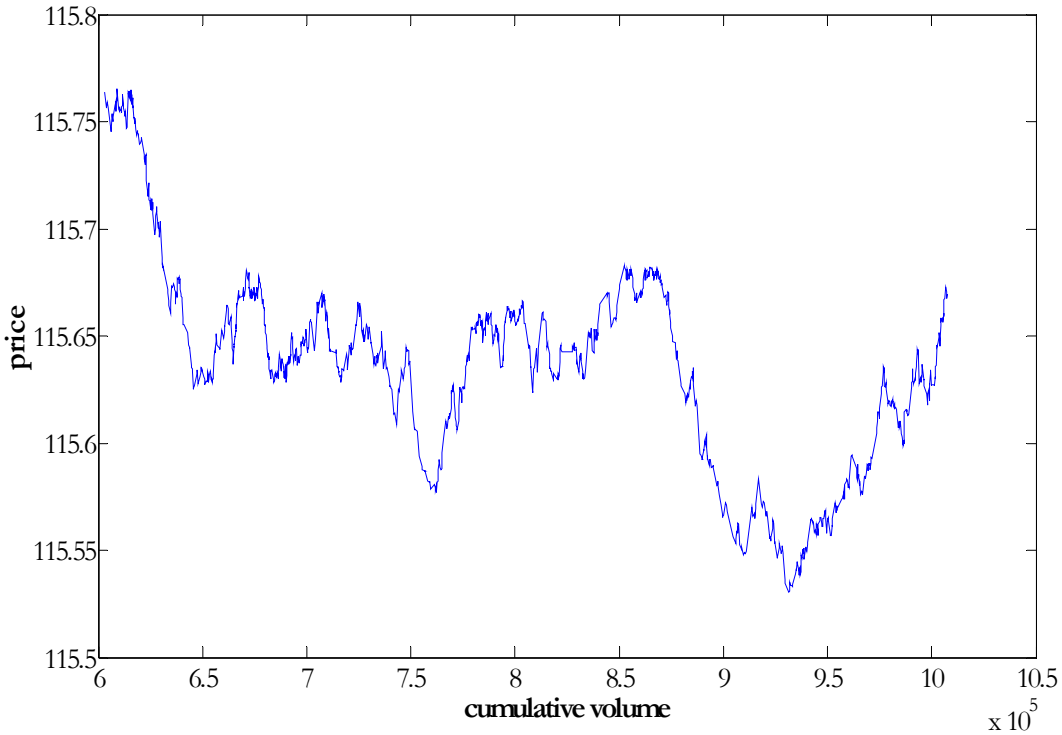


Figure 1.52. Bund (April 18, 2006, frequency 1 sec): Price as a function of the cumulative volume, observations 4500-7017

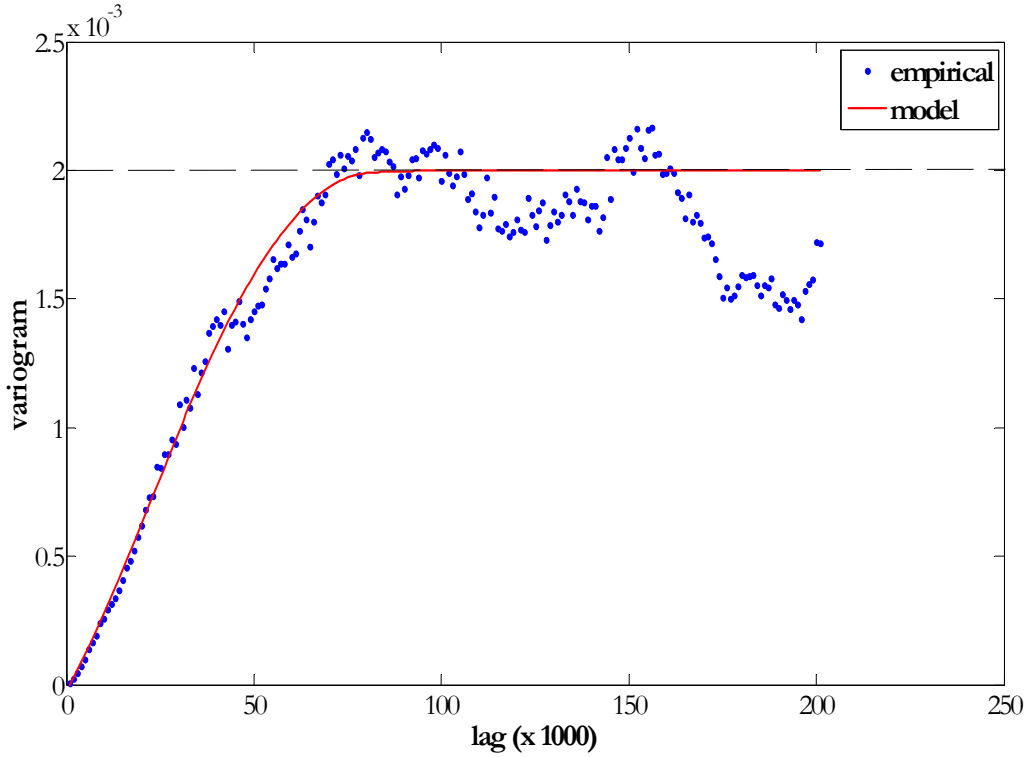


Figure 1.53. Bund (April 18, 2006, frequency 1 sec): Variogram of the price in the cumulative volume coordinate, calculated on the sub-sample for observations 4500-7017 (variance=0.002)

The following figure 1.54 presents some examples of the optimal weights for the length of the moving window: $n_1 = 20$. The structure of the optimal KVWMA is similar to the one obtained for the equally spaced samples: higher weights for first and last observations and relatively low weights for other observations. Although some "middle" weights are different from 0. Some similarities with the equally spaced data can be explained by the following factors. One of the possible explanations for such pattern is that the trading volumes within the day are lower than the trading volumes for data of lower frequency (for example, daily or monthly data). Besides, the scale parameter $\Delta = 1000$ might also smooth the volume discrepancy; thus, changing the scale factor can also change the structure of the weights. For the examples of weights structure different from Figure 1.54 at the presence of the jump in volume see Appendix J.

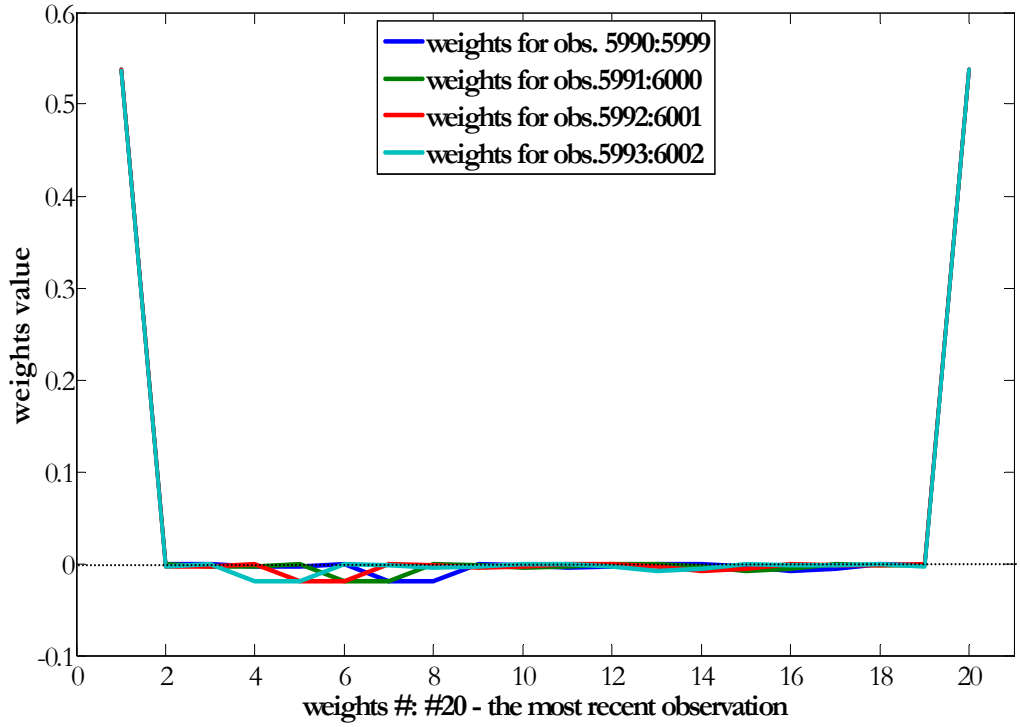


Figure 1.54. Bund (April 18, 2006, frequency 1 sec): Some examples of the optimal weights (window length=20 observations)

Figures 1.55-1.56 present the KVWMA and SMA for different moving windows ($n_1 = 20$ and $n_2 = 200$) for the Bund sub-samples [201; 1500] and [201; 3000] respectively. The following peculiarities of the kriged MA can be noticed:

1. The KVWMA is much more volatile than the SMA of the same length; it is oscillating around the SMA curve.
2. The shorter KVWMA follows the price curve more closely than the longer one.
3. The behaviours of the KVWMA and SMA around price curve are different, in particular for the trending parts of the price curve. During the trending periods, SMA seems to send more false signals than the KVWMA (see Figure 1.56) for the trend following strategy described above. We can notice that for the decreasing trend between approximately 201-700 observations, the SMA sends three false signals (#2,3,4) versus no false signals for the KVWMA. For the upward trend between the observations 700-1500 the SMA sends four false signals (#6,7,8,9) versus two false signals for the KVWMA (#3*, 4*).

4. During the trendless periods both MAs seems brings bad results; this again support the hypothesis that the trend-following strategies work only during the trending periods.

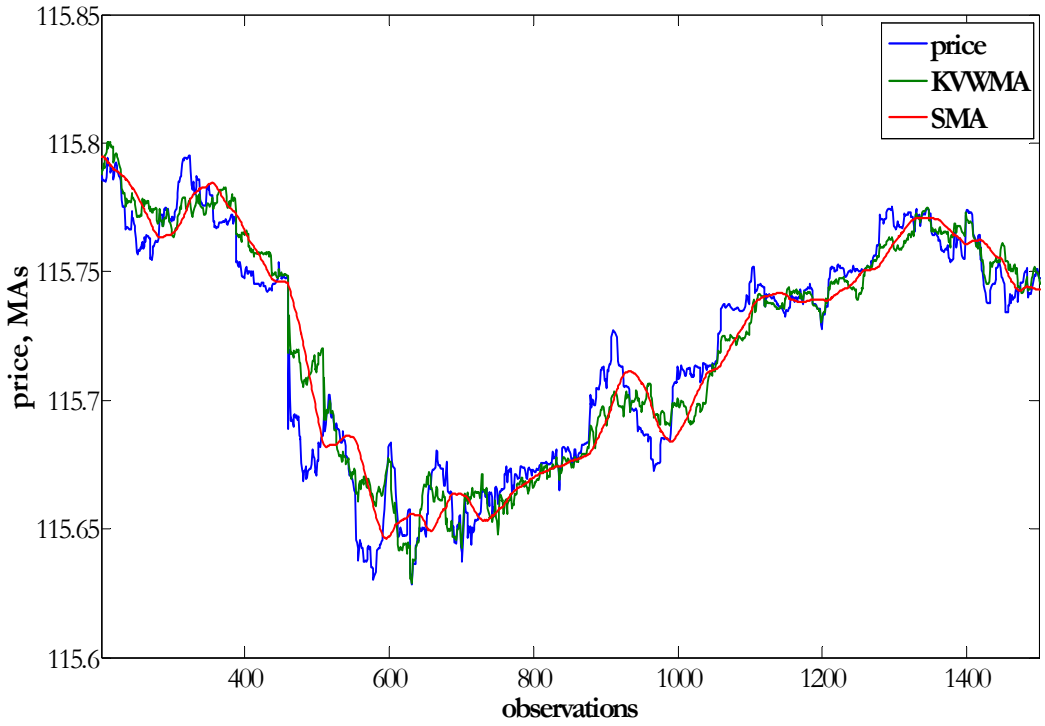


Figure 1.55. Bund (April 18, 2006, frequency 1 sec, observations 201-3000): Price, KVWMA and SMA (window length=20 observations)

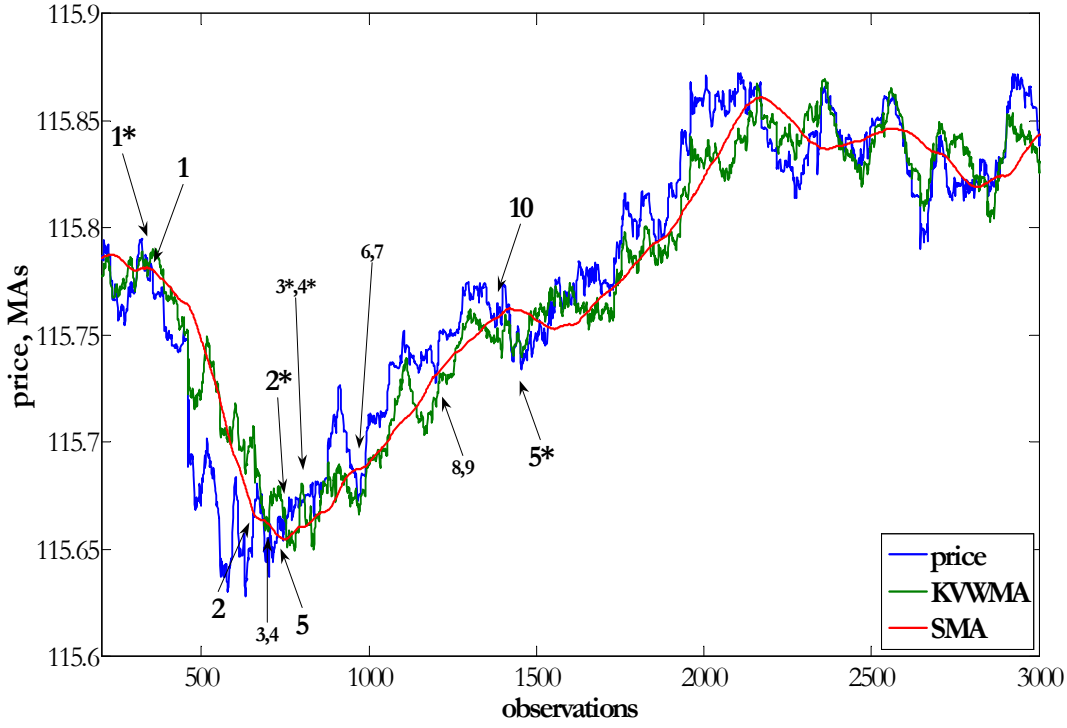


Figure 1.56. Bund (April 18, 2006, frequency 1 sec, observations 201-3000): Price, KVWMA and SMA (window length=200 observations)

The trend-following trading strategy, applied to whole sample (1; 7017) generates the following trading outcomes. Figures 1.57 and 1.58 summarize the end-of-period trading outcomes, simulated for SMA and KVWMA of different lengths.

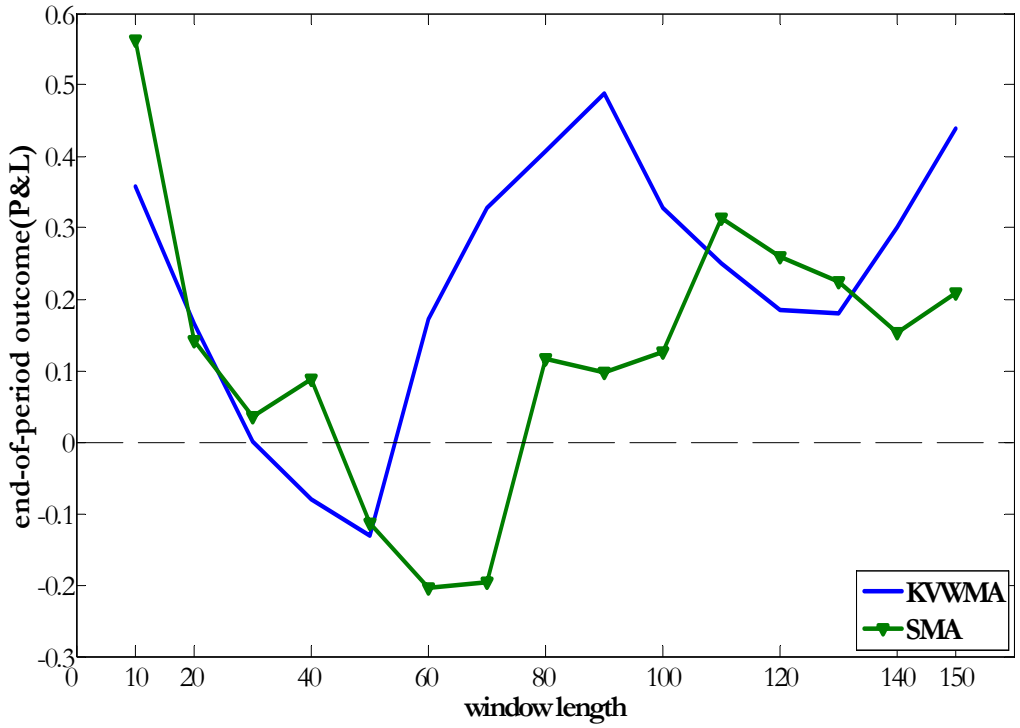


Figure 1.57. Bund (April 18, 2006, frequency 1 sec, observations 201-3000): end-of-period outcomes (P&L) for KVWMA and SMA of different length

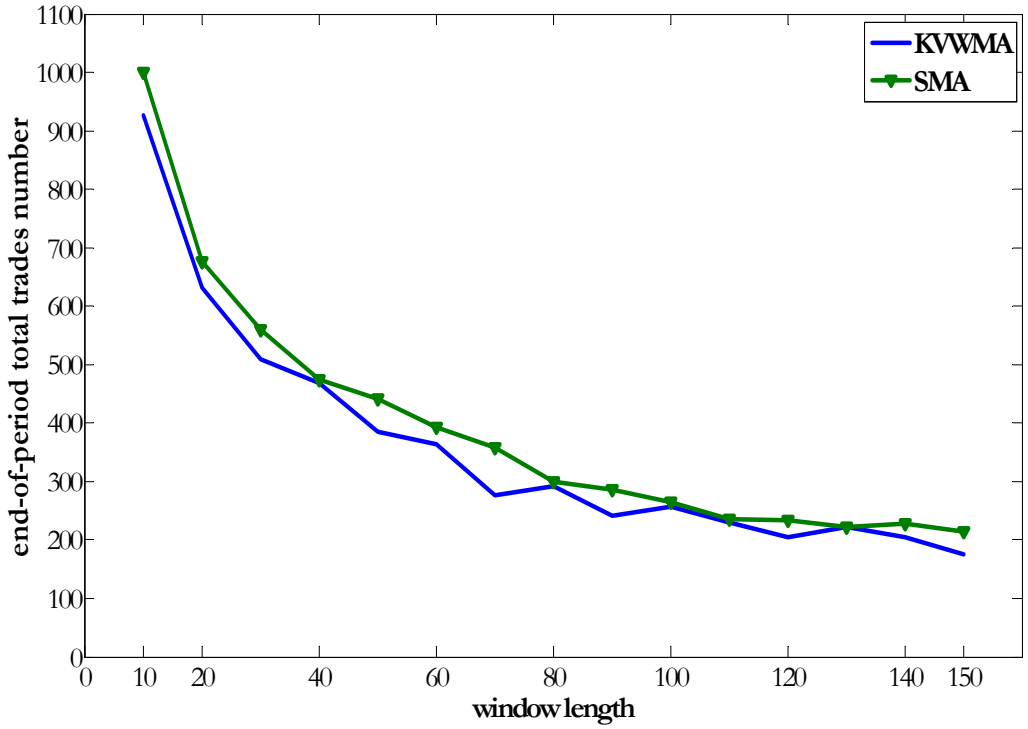


Figure 1.58. Bund (April 18, 2006, frequency 1 sec, observations 201-3000): end-of-period total trades number for KVWMA and SMA of different length

The highest possible cumulative profits are generated by the strategy based on the SMA. At the same time, KWVMA demonstrates higher profits at the window lengths from the interval [60; 100][37]. The outcomes of the KVWMA based strategy are less volatile. Figure 1.58 also shows that the strategy, based on the KVWMA generates less transaction than the strategy, based on the SMA of the same length.

Figure 1.59 presents the optimal P&L paths, generated for the KVWMA (length=90 observations) and SMA (length=10 observations). We can see that SMA path is more volatile than the KVWMA path; it also generates less efficient results (profits per trade).

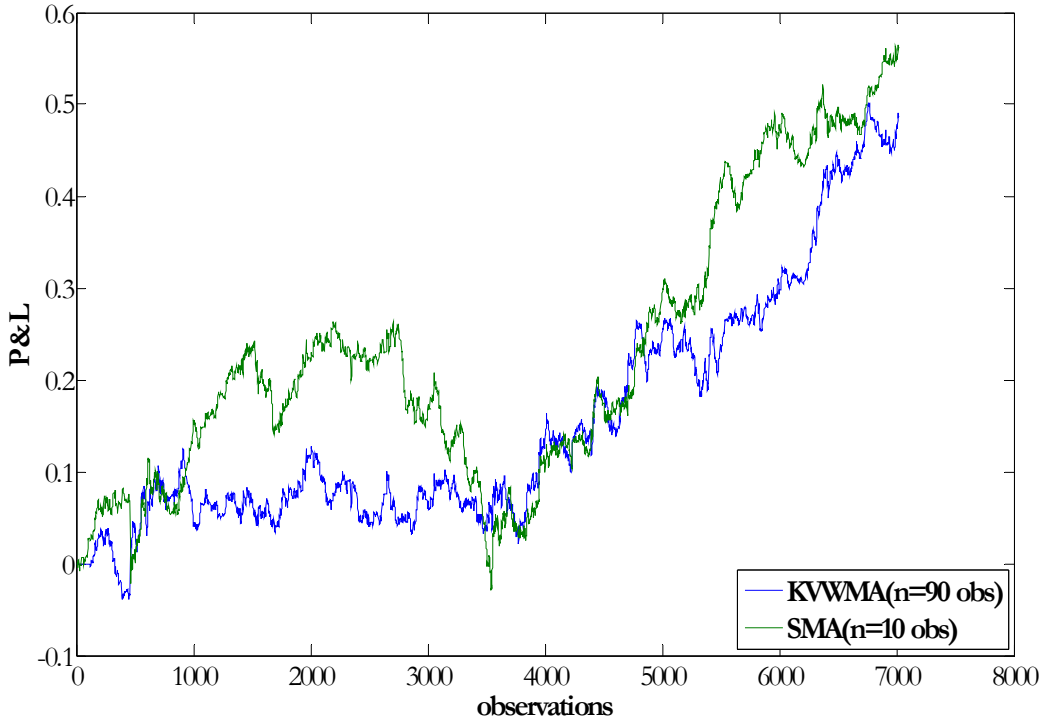


Figure 1.59. Bund (April 18, 2006, frequency 1 sec, observations 201-3000): optimal P&L paths for the KVWMA (length=90observations) and SMA (length=10 observations)

As the result, in this chapter we have demonstrated how construct the KVWMA and showed that it can generate consistent profits at the lengths where traditional MA does not work well.

7 Conclusions

The main objective of the part 1 was to present a statistical method for optimization of one of the most popular rule of the technical analysis – moving average. Part one presented the kriging method, used in geostatistics for the mean estimation and prediction of random variables. According to the method a predictor/estimator is a weighted average of some neighborhood of data, which takes into account the correlation of a random variable at this neighborhood. Such estimator coincides with the definition of the MA, making this approach easy to use for the MA

[37] Note that for the strategy simulations the window length was tabulated with the step of 10 observations.

optimization. The main advantage of the method is that it can be applied to both equally and unequally sampled data. In reality, the financial data, sampled at high frequencies is irregular, as prices and other information about transaction are recorded at the moment of transaction. Another transformation, which makes the price data irregular is the change of the coordinates from time to another variable, such as volume.

As the result, we have considered the two cases of kriging method application, with respect to the sample used in these applications: (1) equally sampled data; (2) unequally sampled data.

The first approach leads us to the definition of the optimal kriged moving average (KMA), based on the instrument prices or some indicator values. We have seen that for the equally spaced data, the optimal weights follow some specific structure, similar to different variogram models and window lengths. In such optimal weight structure the largest weights in absolute value has the first and the last observation in the window, with relatively negligible weights for all other observations. As the result, KMA has approximately the same lag as simple moving average, but is less smooth than SMA. KMA oscillates around the SMA curve. The volatility and amplitude of the oscillations is the indirect function of the KMA length: the longer the KMA the less it is volatile and it coincides more with the SMA curve. Therefore, the trading strategies based on KMA and SMA will take different positions at short window lengths and the same position at long lengths.

As for the trading results, the trend-following strategies, based on KMA generate higher profits at short lengths than SMA, and in some cases produce even fewer trades. At longer lengths the difference in the performance is less pronounced. KMA seems to perform better during the periods that contain price jumps and worse during trendless periods; it also performs better than SMA during the period with jumps and worse than SMA during the trendless period. The explanation to this lays in the weight structure of the MAs. As markets are jumping, KMA is more sensitive to the last large price changes than SMA as it assigns larger weight to the last available observation[38]. During the trendless periods such structure make the KMA more erratic than the SMA, causing more false signals, and thus, less profit. It should be noted that these conclusions should be considered within a context of a particular trading strategy – trend-following, based on the cross-overs of MA and price curves.

The particular, characteristic of the KMA is that frequently it produces fewer trades than the SMA even having a more erratic nature. Taking into account non-zero transaction costs in real-life applications this factor can increase even further the discrepancy between the outcomes for KMA and SMA (when KMA is more profitable). It also implies that erratic nature of the KMA curve does not necessary leads to more false signals generated by the trading strategy.

Finally, KMA seems perform worse for the data sampled at higher frequency due to more volatility present in this data; this might explain poorer results for Bund case, compare to DAX and Brent data.

As for the unequally spaced data we have considered two examples of the application of the kriging method. For the time-dependent data we have shown that KMA, calculated for the irregular samples generate consistent and the highest results at the short lengths of the MA, where traditional MAs do not work well. We also showed that an adjustment of sample to the

[38] KMA optimal structure also assigns more weight to the first observation in the window, but due to the presence of the trend this value, depending on the trend direction, is much smaller or much bigger than the last available observation; therefore its weighted impact is smaller on the mean value than for the last observation in the window.

regular spaced one by the interpolation of the missed observations does not produce optimal MA: the KMA calculated on the adjusted regular sample generates more false signals.

As for the unequally sampled data due to the change of coordinate, the estimation of the volume weighted MA lead to the following conclusion: kriging results largely depend on the variability in the volume. If the variability is small due to the high frequency of the data the structure of the weights resembles the one for the evenly spaced data set: the structure of the weights emphasized the first and the last observations. At the same time the weights are not constant in values and are changing with time. As the result the KVWMA behaves similar to the KMA: it oscillates around the SMA and its smoothness depends on the MA length. It also generates good trading results (although not the highest possible, but more stable). KVWMA also generates fewer transactions than the SMA of the same length. For the case, when the variability of the volume is significant, the weights structure will be more time dependent and volatile.

In general, we can conclude that kriging method allows obtaining quite interesting MA. At the framework of the trend-following strategies considered in this part, optimal KMA or KVWMA has short length, contrary to classic MAs, which are more profitable at long lengths. It also generates fewer transactions than other MAs at the medium length of the moving window.

9 Appendices I

Appendix A

Some MA optimization techniques

A1. Optimisation technique, based on the price series cyclic nature

According to Achelis (www.equis.com/Education/TAAZ/?page=74) the length of the MA should fit the peak-to-peak cycle of a security price movement. The ideal MA length should follow the formula:

$$n_{opt} = \frac{C_l}{2} + 1, \qquad \text{(A1.1)}$$

where C_l - is the length of the cycle.

Achelis proposes the suitable MAs lengths with respect to the trends that they have to filter (see Table A1.1).

Table A1.1
Optimal lengths of the moving averages

Trend	Moving Average length
Very short term	5-13 days
Short term	14-25 days
Minor Intermediate	26-49 days
Intermediate	50-100 days
Long-term	100-200 days

Source: Achelis, S.B. "Technical Analysis from A to Z", www.equis.com/Education/TAAZ/?page=74

One of the principal critiques of the method lies in the necessity to define parameter C_l before the MA optimization. The analysis of the cycles often demands preliminary analysis of the market nature. Each market has its own characteristics, which might define its cyclic nature; therefore, application of the same lengths parameter (see Table A1.1) can be questioned. Besides, these optimal lengths for daily observations are irrelevant for the intraday trader that works on much higher data frequencies.

The other important drawback of the method is its concentration only on the MA's length, although weigths values have an impact on the form and positioning of the MA curve.

A2. Moving average as a center of gravity

Ehlers (www.mesasoftware.com/papers.htm) proposed simple and didactic representation of the trade-off problem[1]. He shows that as long as a trader knows the lag she/he can tolerate, she/he can calculate the optimal weights of the MA. He proposes to consider MA as a center of gravity (c.g.) of the window with price series on its diagonal. The weights of the prices are the weights of the MA, while the lag is the distance between this center and its projection on the right vertical side of the window. Picture A2.1 demonstrates this approach in the simplified version for the case of the simple moving average (all weights are equal). For SMA the center of gravity lies in the middle of the diagonal; thus, the lag (L) of this MA is half of the window length (n): $L = n/2$.

As we can see, the value of the lag L is determined by the position of the center of gravity on the price curve, which in its turn is determined by the price weights and size of the window.

[1] See Ehlers, J. "Signal analysis concepts", http://www.mesasoftware.com/papers.htm for more detail analysis.

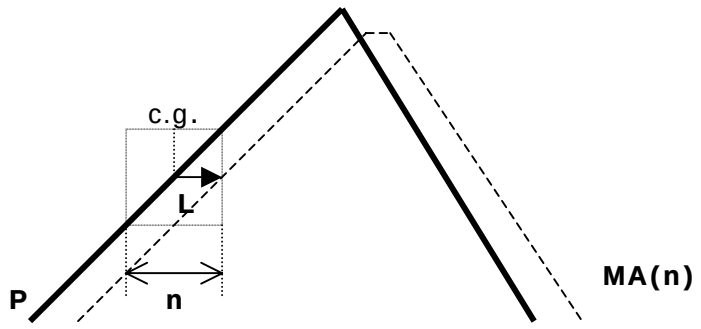


Figure A2.1. Formation of the simple moving average (Ehlers, [38])

The example of the exponential moving average (EMA) can be used to demonstrate the application of the Ehlers's method. Lets assume that price $\{P(t)\}_{t>0}$ is a continuous trend with slope μ and has value $T(t)$ at moment *t*. EMA at moment *t* is defined as following:

$$EMA(t) = \alpha P(t) + (1-\alpha) EMA(t-1). \qquad \text{(A2.1)}$$

Figure A2.2 presents the schematic representation of the definition of the optimal weight α.

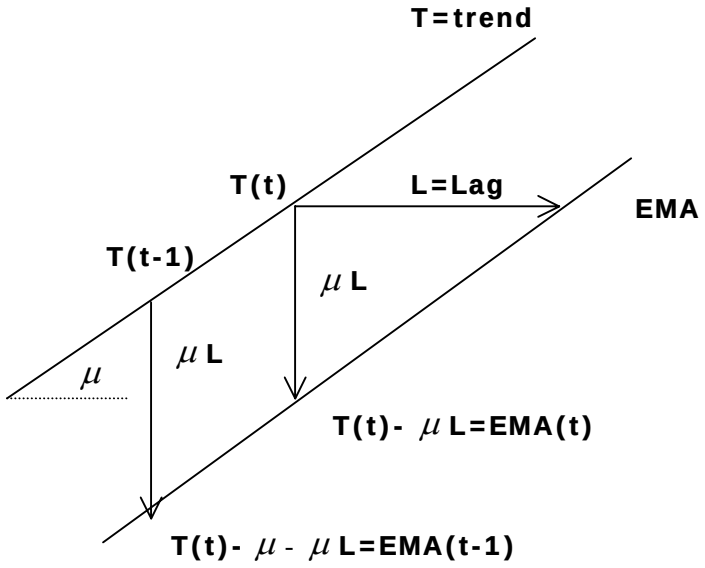


Figure A2.2. Calculation of the optimal weights for the exponential moving average (Ehlers, [38])

According to the figure A2.2 the values of the EMA at moments t and $t-1$ can be defined as following: $EMA(t) = T(t) - L\mu$, $EMA(t-1) = T(t) - L\mu - \mu$. Then expression (A2.1) takes the form:

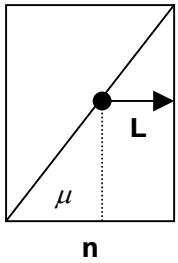
$$T(t) - L\mu = \alpha T(t) + (1-\alpha)(T(t) - L\mu - \mu)$$

We can derive that $\mu = \alpha\mu(1+L)$, or

$$\alpha = \frac{1}{1+L} \qquad \text{(A2.2)}$$

As the result, Ehlers gets (A2.2) for the weight of the EMA as a function of the MA lag regarding price curve.

One of the principal critiques of the method is that knowing only the lag of the MA is not sufficient to define the optimal MA parameters. Either one of the parameter (length or weights) should be predefined or the smoothness criteria should be introduced. From the box presentation of the MA as a gravity center in Figure A2.3 it is obvious that different sets of the optimal weights can generate the same lag L that the trader can tolerate. For example, the moving window $n = 5$ with weights $\left\{\frac{1}{5};\frac{1}{5};\frac{1}{5};\frac{1}{5};\frac{1}{5}\right\}$ and $\left\{\frac{1}{2};0;0;0;\frac{1}{2}\right\}$ will have the same center of gravity and, thus, lag L. At the same time, different weights set will generate different smoothness of the MA curve; therefore, the smoothness criteria in addition to lag can help to define the optimal weights.



Picture A2.3. Moving average as a center of gravity

The other important critique is that the prices never form a straight line making the optimization procedure quite approximate.

Finally, it is not evident how to define the level of lag *L* that the trader can accept.

A3. Adaptive MA

The problem of lag choice is important as it defines how close a MA is to prices curve. This determines the sensibility of the MA rule to false signals. One of the definitions of the false signals is related to the transaction costs. Di Lorenzo (1996) tried to develop adaptive moving average that takes into account the transaction costs and volatility of the price data.

They assume that the price is a linear trend, namely $P(t) = \mu t$. Consequently, moving average at moment t for window n:

$$M_t(n) = \frac{P_t + P_{t-1} + ... + P_{t-n+1}}{n} = \frac{\mu t + \mu(t-1) + ... + \mu(t-n+1)}{n} = \frac{\mu}{n}\sum_{i=0}^{n-1}(t-i) =$$

$$= \frac{\mu}{n}tn - \frac{\mu}{n}\sum_{i=0}^{n-1} i = \mu t - \frac{\mu}{n}\cdot\frac{(n-1)n}{2} = \mu t - \frac{(n-1)\mu}{2}$$

Distance between price trend and MA:

$$d = P_t - M_t(n) = \mu t - \mu t - \frac{(n-1)\mu}{2} = \frac{\mu(n-1)}{2} \qquad \text{(A3.1)}$$

From equation (A3.1), Di Lorenzo (1996) received the formula for rolling window:

$$n = \frac{2d}{\mu + 1} \qquad \text{(A3.2)}$$

According to Di Lorenzo (1996), the distance should be the function of the short-term volatility of the price series and the transaction costs.

The expression (A3.2) is received under assumption that prices follow some linear trend, which is quite simplified version of the reality. Taking into account that the trend factor μ is generally unknown it should be estimated first. Finally, the trend is frequently non-constant value, therefore, should be re-estimated from time to time. All these critique makes the application of the method on the routine basis less likely.

Appendix B

Examples of the MAs with fixed weights

Notation:

$\{P_i\}_{i\geq 0}$ - price series;

$\{w_i\}_{1\leq i\leq n}$ - weights of MA

n - length of moving window (length of MA)

α - length parameter of the EMA: $\alpha = \frac{2}{2+n}$, with n - the length of the SMA that has the same lag.

1. Simple moving average (SMA):

$$MA_t = \sum_{i=0}^{n-1} w_{n-i} P_{t-i} = \sum_{i=0}^{n-1} \frac{1}{n} P_{t-i} = \frac{1}{n} \sum_{i=0}^{n-1} P_{t-i}$$

2. Exponential moving average (EMA):

$$EMA_t = \alpha P_t + (1-\alpha) EMA_{t-1}$$

3. Weighted moving average (WMA):

$$MA_t = \sum_{i=1}^{n} w_i P_{t-n+i}$$

3.1. Linearly weighted moving average (LWMA):

$$MA_t = \sum_{i=0}^{n-1} w_{n-i} P_{t-i} \quad w_j = f(j) = a + bj$$

Widely used case: $w_i = \frac{2i}{n(n+1)}$ (Hutchinson, Zhang, [55], V.11:12).

3.2. Triangular moving average (TMA)

For this MA higher weights are assigned to the middle period observations:

$w_1 < w_2 < \ldots < w_{n/2} = w_{n/2+1} > \ldots > w_{n-1} > w_n$ for *n* – even

$w_1 < w_2 < \ldots < w_{n+1/2} > \ldots > w_{n-1} > w_n$ for *n* – odd

3.3. *Generally weighted moving average (GWMA)* (Hutchinson, Zhang, [55], V.11:12):

$$MA_t = \sum_{i=0}^{n-1} w_{n-i}(\alpha) P_{t-i} \text{ with } w_j(\alpha, n) = \frac{j^\alpha}{\sum_{i=1}^{n} j^\alpha} \qquad \text{(B.1).}$$

Lets consider the following cases for generally weighted moving average $GWMA(t, n, \alpha)$:

(1) $\alpha = 0, w_j(0) = \frac{1}{n}$, i.e. $GWMA(t, n, 0) = SMA(t, n)$

(2) $\alpha = 1, w_j(1) = \frac{j}{\sum_{j=1}^{n} j} = \frac{2j}{n(n+1)}$, i.e. $GWMA(t, n, 1) = LWMA(t, n)$

(4) $\alpha = 2, w_j(2) = \dfrac{j^2}{\sum_{i=1}^{n} j^2} = \dfrac{6j^2}{n(n+1)(2n+1)}$, i.e. $GWMA(t,n,2)$ is a square-weighted MA

(5) $\alpha = 0.5, w_j(0.5) = \dfrac{j^{\frac{1}{2}}}{\sum_{i=1}^{n} j^{\frac{1}{2}}}$ (no closed form), i.e. $GWMA(t,n,0.5)$ is a square-root-weighted MA

Note that the weights constructed for the GWMA according to the formulae (B.1) are always higher for more recent observations: as α increases larger weights are assigned to the recent observations.

Appendix C

Examples of the MAs with variable weights

C1. Volume-adjusted moving average

Volume-adjusted moving average (VAMA) can be represented as following:

$$VAMA_t = \sum_{i=1}^{n} w^*_{n-i+1} P_{t-i+1}\ ,\ w^*_i = w^*(V_i),$$

where $\{P_i\}_{i>0}$ is a price series; n - is MA length; and $\{w^*_i\}_{1\le i\le n}$ - are MA weights as a function of the cumulative volume at the i th moment of time $\{V_i\}_{i>0}$.

The idea behind VAMA is to use some additional information about the trading activity at some point t. Traditional time-based (sampled) MA suppose that all moments of times are equal with respect to the information they brings. VAMAs treat the activity of trading as the measure of the price importance in the particular moments of time: for example, the sharp increase in the trading activity might indicate the change in the trend, therefore these price observations have much higher importance for trader than the prices when the market is not traded actively.

Different approaches to define the function w^* can be used. Richard W. Arms Jr ([11], V.8:3) proposes to change the system of coordinates from time- to volume based. The new curve will have "prolonged" price values for heavily traded moments of time and "shorten" (or even neglected) the price values for lightly traded moments. The new VAMA can be calculated on the basis of the new price curve. Arms ([11], V.8:3) used graphical approach to construct VAMA, in particular equivolume charting, which can be translated into programming algorithm. Suppose that you have daily observations of price (high, low) and volume for a particular instrument. According to the equivolume charting each price observations is represented in the form of box with high being the upper side of the box, low – being the lower side of the box, and volume being the width of the box. The mean price calculated as sum of high and low devided by two, is posted one or several times in the box depending on its width. Then x-unit-volume MA of some pre-chosen weight structure (for example, SMA- or EMA-type) is calculated. The trading rule then involves the new price curve and VAMA. Then VAMA approaches price curve during the periods of heavy trading more rapidly than the time-based MA. With this approach the days with heavy trading have higher weights and, thus, more importance for the definition of the MA.

At our opinion the method presented above has several drawbacks. Although the VAMA takes into account the trading volumes at each moment of time, it cannot incorporate the information about volume possible (auto-) correlation. Besides, building the price-volume blocks of different width implies price interpolation for the non-available cumulative volumes values; the interpolation of the price at the same level can be questioned. Finally, while the MA calculation procedure can be programmed, it involves many subjective judgements, among which the definition of the volume frequency and the points at which the MA is calculated.

C2. Variable index dynamic average (VIDYA)

Traditional MA is based on the assumption of constant data volatility: trend is filtered by assuming that the noise has constant volatility. That is why drastic price movements caused by the change in its volatility would be mistakenly treated as trend enforcing or trend-breaking events. Therefore, adapting a MA to the market volatility should improve its prediction power.

Chande ([23], V.10:3) proposes the following adaptation of EMA to the market dynamic. He has substituted a smoothing constant α in the formulae of the EMA by the following coefficient:

$$\alpha^*_t = kV_t, \qquad \text{(C2.1)}$$

where $k = const$, V_t - dimensionless market-related variable.

Chande ([28], V.10:3) represents V_t as a ratio of two volatility measures: the standard deviation of the instrument's closing price and the market standard deviation defined as a reference measure. Such MA in trading literature is called variable index dynamic average ($VIDYA$):

$$VIDYA_t = k\frac{\sigma_n}{\sigma_{ref}}P_t + \left(1 - k\frac{\sigma_n}{\sigma_{ref}}\right)VIDYA_{t-1}, \qquad \text{(C2.2)}$$

where $VIDYA_t, VIDYA_{t-1}$ - variable index dynamic average for the periods *t* and *t-1*;

$k = const$, $0 < k \le 1$ - smoothing coefficient when $\sigma_n = \sigma_{ref}$ [2];

P_t - the last available closing price;

σ_n, σ_{ref} - rolling standard deviation of the analysed instrument and market (rolling) standard deviation (reference value) (for ex. CAC40); the length of the periods over which the rolling standard deviations are calculated, are *n*- and *ref*-periods; they can be different[3].

As we can see from (C2.1) and (C2.2), if the analysed instrument's volatility is the same as the reference market volatility, then $\alpha_t^* = k$ became a constant; if instrument's volatility increases during some period of time Δ, then $\alpha_t^* > \alpha_{t-\Delta}^*$ and adapted MA approaches closer the price curve; otherwise, if instrument's volatility diminishes during some period of time Δ, then $\alpha_t^* < \alpha_{t-\Delta}^*$ and adapted MA diverges from the price curve. The mathematical exercise can prove this VIDYA's property.

Suppose some instrument Y and reference index has the same volatility change, i.e. $\frac{\sigma_n(t_1)}{\sigma_{ref}(t_1)} \approx \frac{\sigma_n(t_2)}{\sigma_{ref}(t_2)}$.

We can always find some $EMA(\alpha_{ema})$ that corresponds to VIDYA at some precise moment of time *T*:

$$VIDYA_T = EMA_T$$

Then the effective lengths of the EMA and VIDYA coincides, and $\alpha_{vidya} = \alpha_{ema}$, or

$$k\frac{\sigma_n(T)}{\sigma_{ref}(T)} = \frac{2}{n+2}, \qquad \text{(C2.3)}$$

where *n* – is the effective length of EMA., which coincides with VIDYA effective length.

Then

---

[2] It seems like author defines the coefficient as $k = \frac{2}{n+2}$, where *n* is the length of the SMA which has the same lag as EMA with smoothing coefficient *k*, when $\sigma_n = \sigma_{ref}$

[3] In Chande's example ([28], V.10:3) the reference standard deviation defines the long-term volatility, contrary to short-term instrument's volatility: $n < ref$

$$n = \frac{2\left(\sigma_{ref}(T) - k\sigma_n(T)\right)}{k\sigma_n(T)}. \qquad \text{(C2.4)}$$

The equation (C2.4) confirms the observations made above: the higher is the instrument's volatility relatively to market volatility the smaller the effective length of the VIDYA, i.e. the VIDYA converges towards the price curve; and vice versa. As the result the *VIDYA* follows the price curve more closely when volatility increases.

Other measures of relative volatility V_t exists, such as Chande momentum oscillator (CMO)[4]:

$$CMO = \frac{100\left(S_u - S_d\right)}{S_u + S_d} \qquad \text{(C2.3)}$$

where S_u, S_d - sum (for n-days period) of the respectively upward/downward closing price change:

$$S_u = \sum_{i=t-n+\Delta+1}^{t} \Delta_i^+ ,$$

$$S_d = \sum_{i=t-n+\Delta+1}^{t} \Delta_i^- ,$$

where $\Delta_i^+ = \left(P(i) - P(i-\delta)\right) \cdot I\left(P(i) - P(i-\delta) > 0\right)$; $\Delta_i^- = \left(P(i) - P(i-\delta)\right) \cdot I\left(P(i) - P(i-\delta) \le 0\right)$;

$P(i)$ - price at moment i;

$I(\cdot)$ - indicator function: $I\left(V(x) \le i\right) = \begin{cases} 1, \forall V(x) \le i \\ 0, otherwise \end{cases}$

C3. Variable-length moving average (VLMA)

Arrington ([16], V.10:6) introduced variable-length moving average. The idea behind this MA is to define its length as a function of relative magnitude of recent price increments: "if recent price changes are "unusually" large the length of the moving average is shortened and average automatically becomes more sensitive to the emerging trends"[5]. Conversely, when the trend is stable the length of the MA is increasing. The "usual" price change is proposed to define by the means of the price changes distribution:

$$R_u \in \left[\mu - 2\sigma; \mu + 2\sigma\right], \qquad \text{(C3.1)}$$

where R_u, R_e - usual and extreme price increments: $R = P(t) - P\left(t - \delta\right)$

μ, σ - mean and standard deviation of the price increments: $R \sim L\left(\mu, \sigma^2\right)$

We believe that the hypothesis behind VLMA is that the larger price movements precede trend change. However, Arrington ([16], V.10:6) does not explain in his article the assumptions on which his MA rests (for example, what is the probability distribution of the price increments). He also does not precise how to define the change in the MA length. Therefore we can see, that although conceptually the idea is interesting, this VLMA should be better defined.

---

[4] Note that CMO is very similar to the other technical indicator – relative strength index (RSI):

$$RSI = \frac{100 S_u}{S_u + S_d},$$

where S_u, S_d - sum (for n-days period) of the respectively upward/downward closing price change.

[5] Arrington, G.R. "The basics of moving averages" Stock & Commodities, V.10:6, pp.275-278, Copyright © Technical Analysis Inc.

Appendix D

Some examples of the variogram $\gamma(h)$ and correlation $C(h)$ models

Let h - lag of the variogram (distance between the observations);

$\gamma(\cdot)$ - variogram;

$C(\cdot)$ - covariance;

a - range parameter;

σ^2 - variance (sill);

α - parameter.

I. Stationary models

1. **Stable model**: $0 < \alpha \le 2$

$$\gamma(h) = \sigma^2\left(1 - e^{-\left|\frac{h}{a}\right|^{\alpha}}\right)$$

$$C(h) = \sigma^2 e^{-\left|\frac{h}{a}\right|^{\alpha}}$$

Particular cases:

(1) $\alpha = 1$ - exponential model

(2) $\alpha = 2$ - gaussian model

2. **Spheric model**:

$$\gamma\ (h) = \begin{cases} \sigma^2\left(\frac{3h}{2a} - \frac{h^3}{2a^3}\right), h \le a \\ \sigma^2, otherwise \end{cases}$$

$$C(h) = \begin{cases} \sigma^2\left(1 - \frac{3h}{2a} + \frac{h^3}{2a^3}\right), h \le a \\ 0, otherwise \end{cases}$$

3. **Dumped cosines model:**

$$\gamma(h) = \sigma^2\left(1 - e^{-\frac{h}{a\alpha}} \cdot \cos\left(\frac{2\pi h}{a}\right)\right)$$

$$C(h) = \sigma^2 e^{-\frac{h}{a\alpha}} \cdot \cos\left(\frac{2\pi h}{a}\right)$$

II. Non-stationary models

1. **Fractal model**: $0 < \alpha < 2$

$$\gamma(h) = \sigma^2 |h|^{\alpha}$$

Particular case:

(1) $\alpha = 1$ - linear model

Appendix E

Short description of the Bund instrument

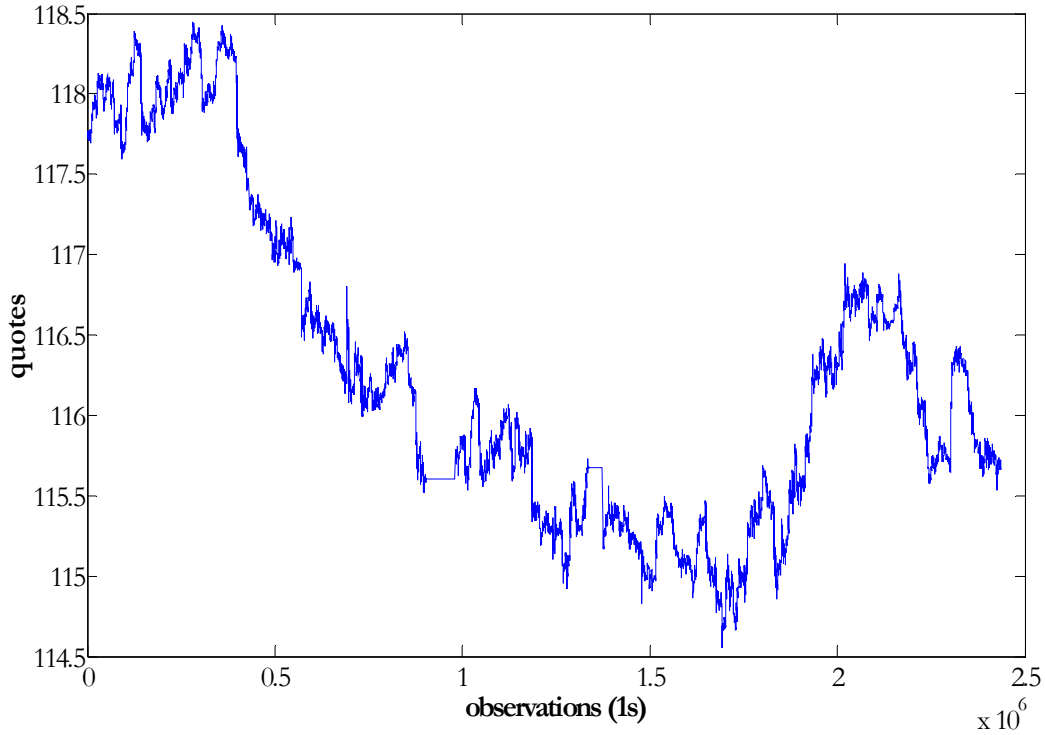


Figure E1. Bund (March 9 - June 8, 2006, frequency – 1 sec)

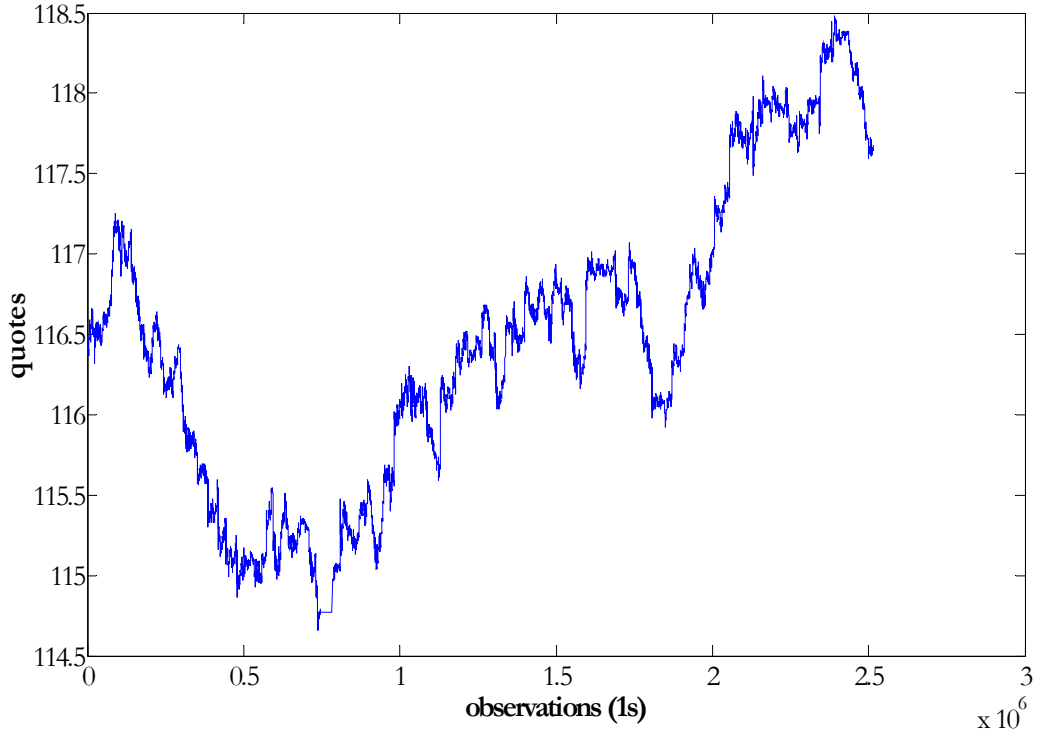


Figure E2. Bund (June 9 – September 8, 2006, frequency – 1 sec)

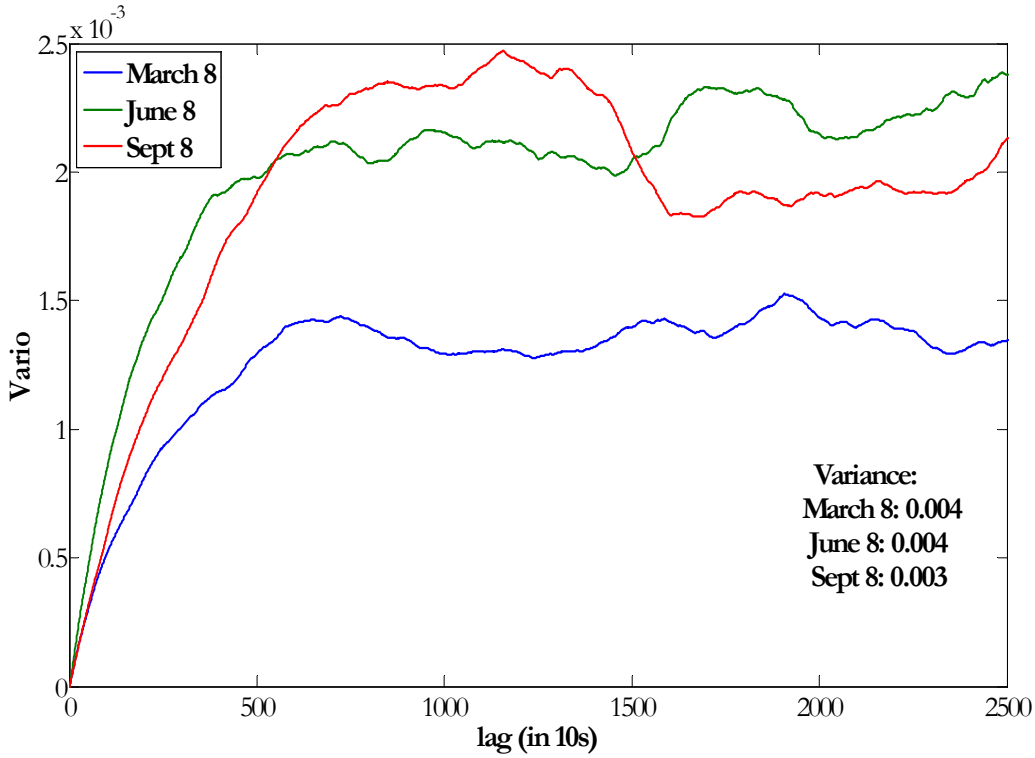


Figure E3. Bund (December 9, 2005 - March 8, 2006, March 9 – June 8, 2006, June9-September 8, 2006, September 9 – December 8, 2006, frequency – 1 sec): Variograms for Bund residuals (n(EMA)=7200 sec) for different contracts

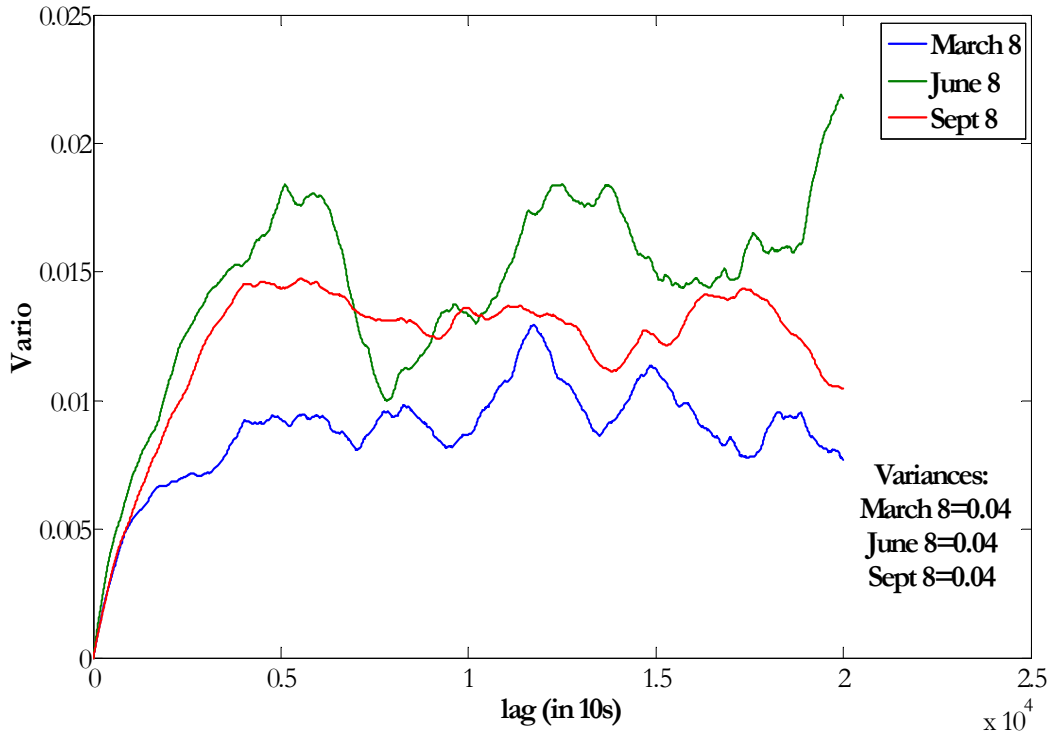


Figure E4. Bund (December 9, 2005 - March 8, 2006, March 9 – June 8, 2006, June9-September 8, 2006, September 9 – December 8, 2006, frequency – 1 sec): Variograms for Bund residuals (n(EMA)=79200 sec) for different contract Short description of the Brent instrument

Appendix F

Short description of the DAX instrument

DAX index represents the stocks market; its historic quotes are given in Figure F1. DAX data corresponds to the time interval of 30/7/2003-7/12/2006 and the frequency of 30 minutes.

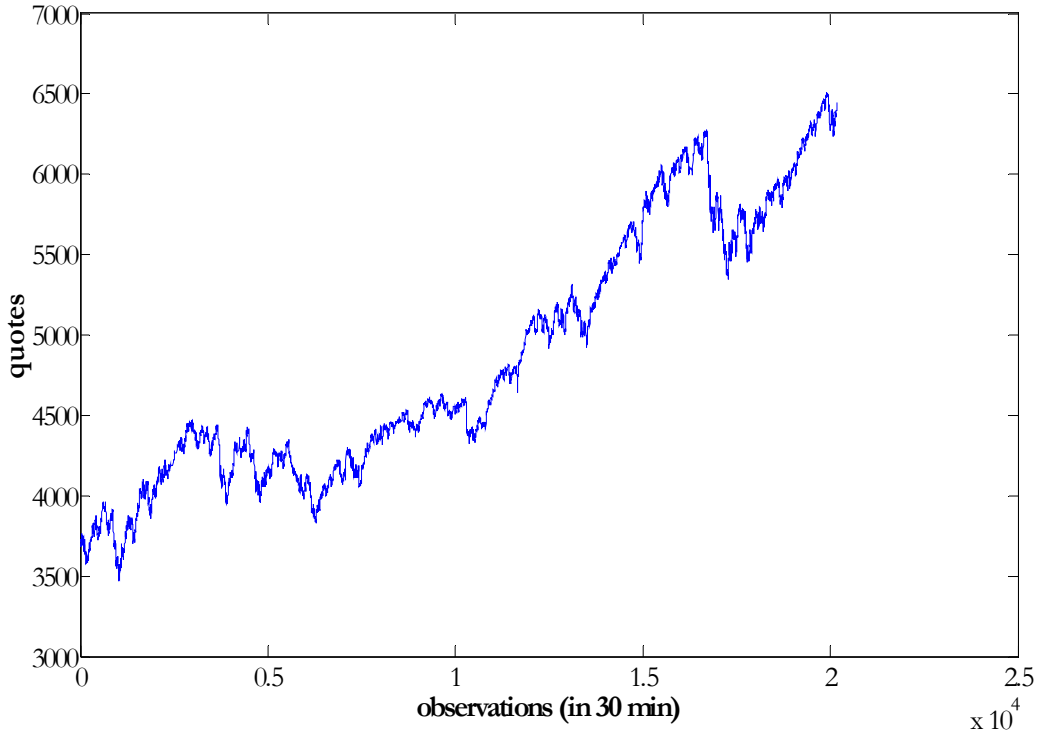


Figure F1. DAX quotes for period 30/7/2003-7/12/2006 (frequency – 30 minutes).

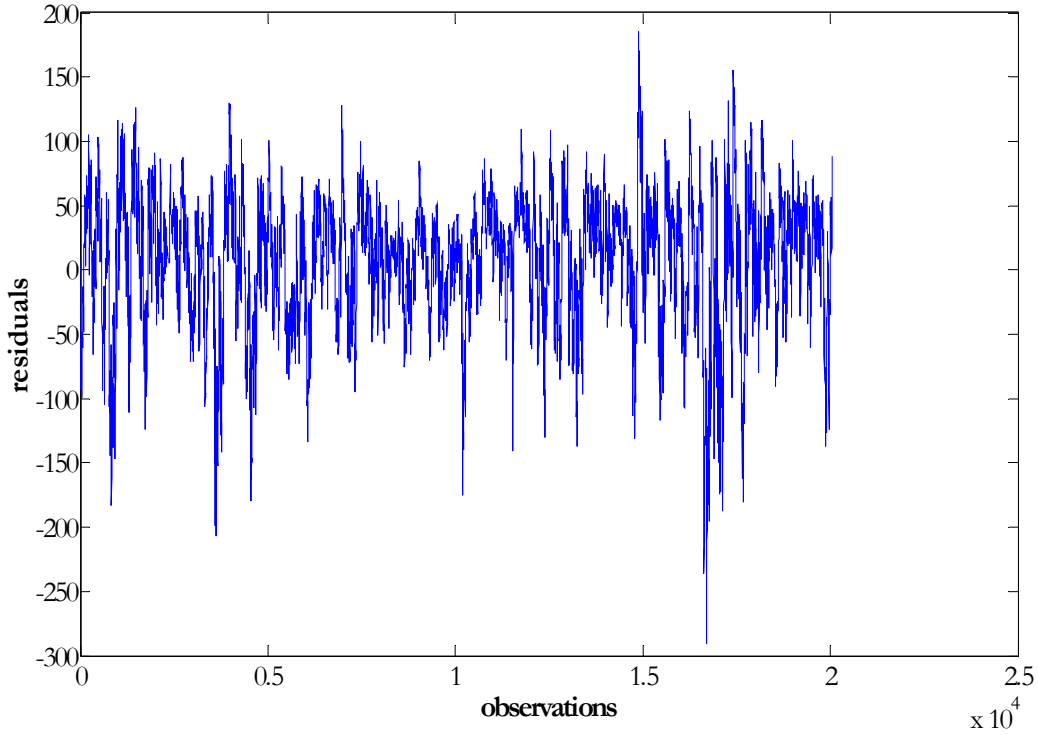


Figure F2. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 observations)

As we can see from Figure F1, the data is positively trending, and thus, non-stationary. Therefore, the EMA of the length of 115 observations is extracted from the price data to estimate the covariance of the residuals. As we can see from figure F2, the obtained residuals do not exhibit any trend.

The estimated variogram of the residuals for the sub-sample [1;10085] is presented in Figure F3. The following theoretical variogram model (the sum of exponential and damped-cosines model) was fit to data:

$$\gamma(h) = 1100\left(1 - e^{-\frac{h}{30}}\right) + 900\left(1 - e^{-\frac{h}{75 \cdot 0.2}} \cdot \cos\left(\frac{2\pi h}{75}\right)\right).$$

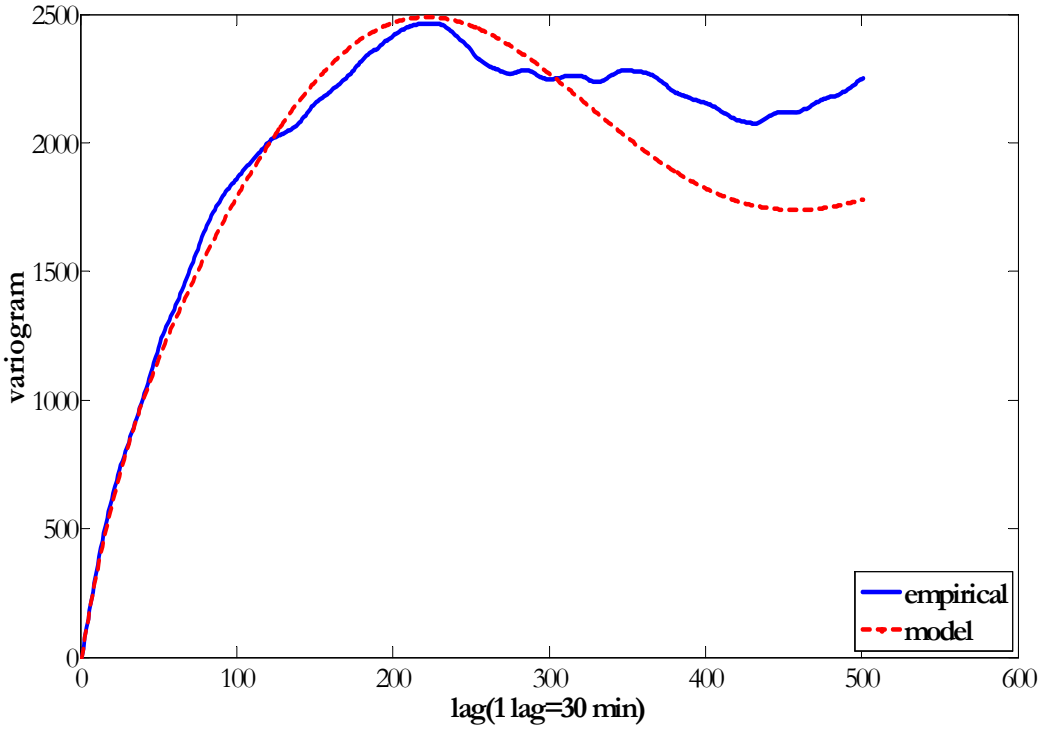


Figure F3. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 observations, observations 1-10085): empirical variogram of the residuals and its theoretical fit

The MACD indicator estimated as the difference of the EMAs of the lengths $n_1 = 40$, $n_2 = 80$ is presented in Figure F4, while their variogram estimated on the sub-sample [1;10085] is given in Figure F5. The following model was fir to the empirical variogram:

$$\gamma(h) = 123\left(1 - e^{-\left|\frac{h}{50}\right|^2}\right) + 123\left(1 - e^{-0.3\left|\frac{h}{80}\right|} \cos\left|\frac{h}{80}\right|\right)$$

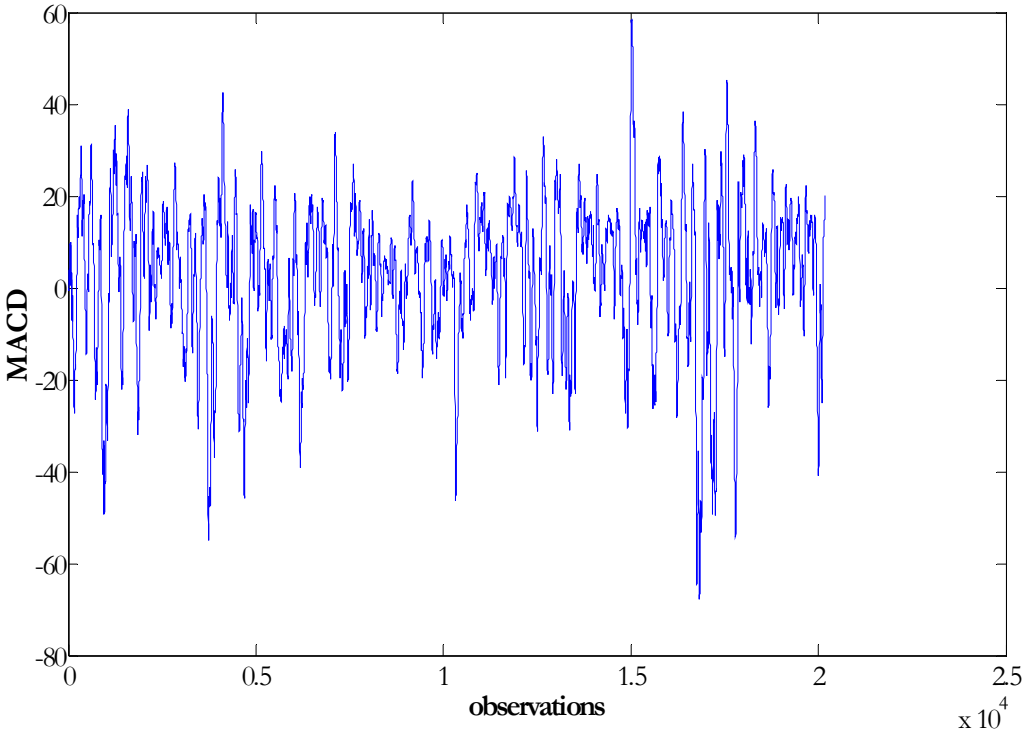


Figure F4. DAX (30/7/2003-7/12/2006, frequency 30 minutes, $n_1 = 40$, $n_2 = 80$): MACD indicator ($n_1 = 40$, $n_2 = 80$)

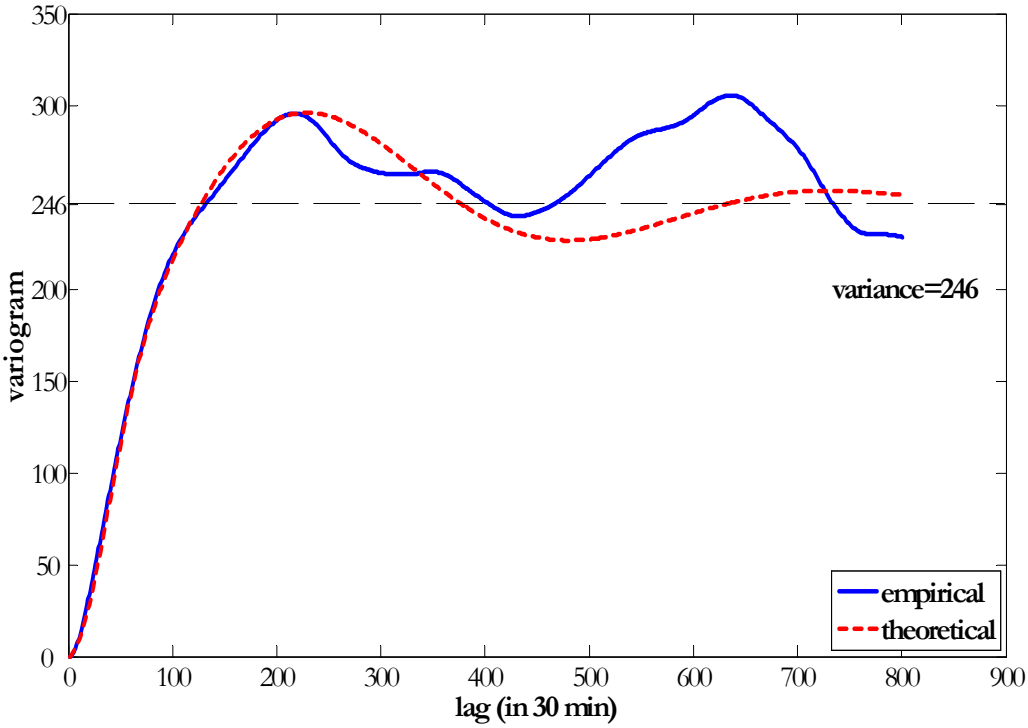


Figure F5. MACD indicator for DAX (30/7/2003-7/12/2006, frequency 30 minutes, $n_1 = 40$, $n_2 = 80$, observations 1-10085): empirical variogram of the residuals and its theoretical fit

Appendix G

Short description of the Brent instrument

Brent is a futures on crude oil. Therefore, we consider this instrument as a representative of commodity markets. The peculiarity of this instrument is that it is highly volatile. Our data represents the period of 17/12/04-27/01/06 with 30 min frequency (see Figure G1).

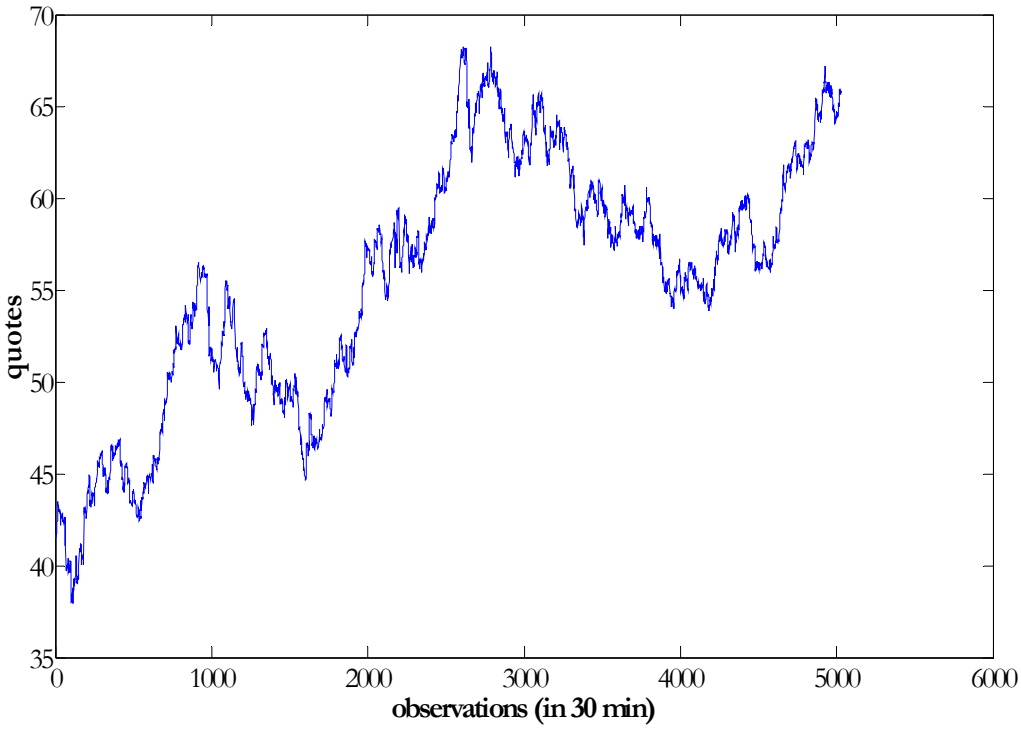


Figure G1. Brent quotes for period 17/12/04-27/01/06 (frequency 30 minutes).

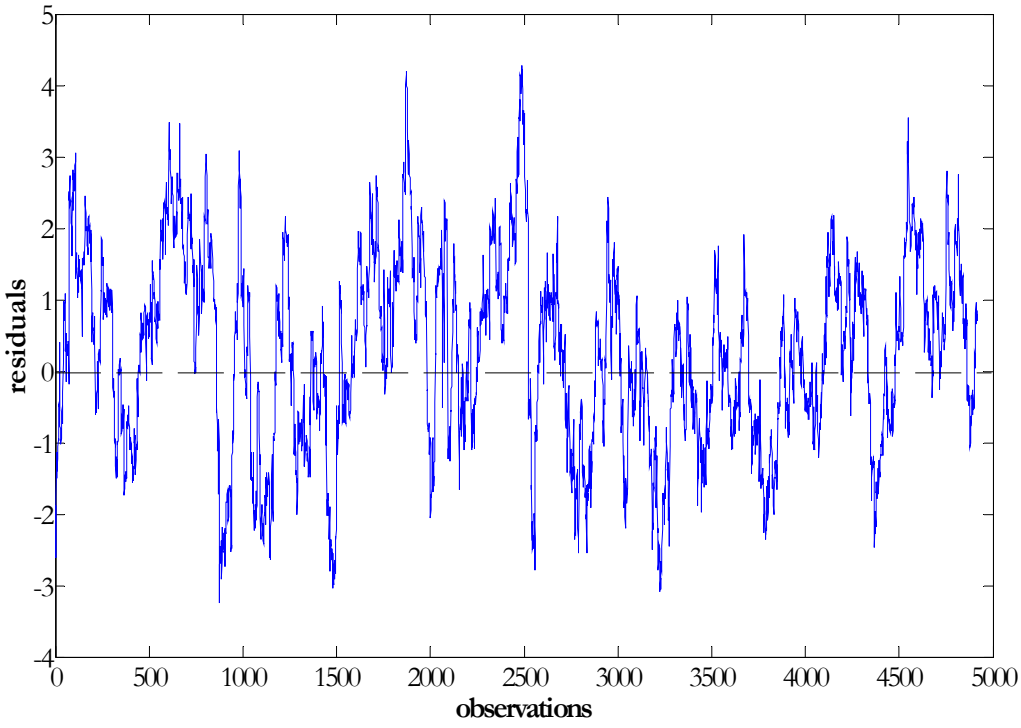


Figure G2. Brent residuals (17/12/04-27/01/06, frequency 30 minutes, n(EMA)=110 observations)

The data is trending, and thus, non-stationary. We eliminate the EMA of the length 110 observations to calculate residuals.

The variogram of the transformed residuals (see Figure G3) is evaluated on the sub-sample $[1;2514]$:

$$\gamma(h) = 2\left(1 - e^{-\left|\frac{h}{50}\right|}\right).$$

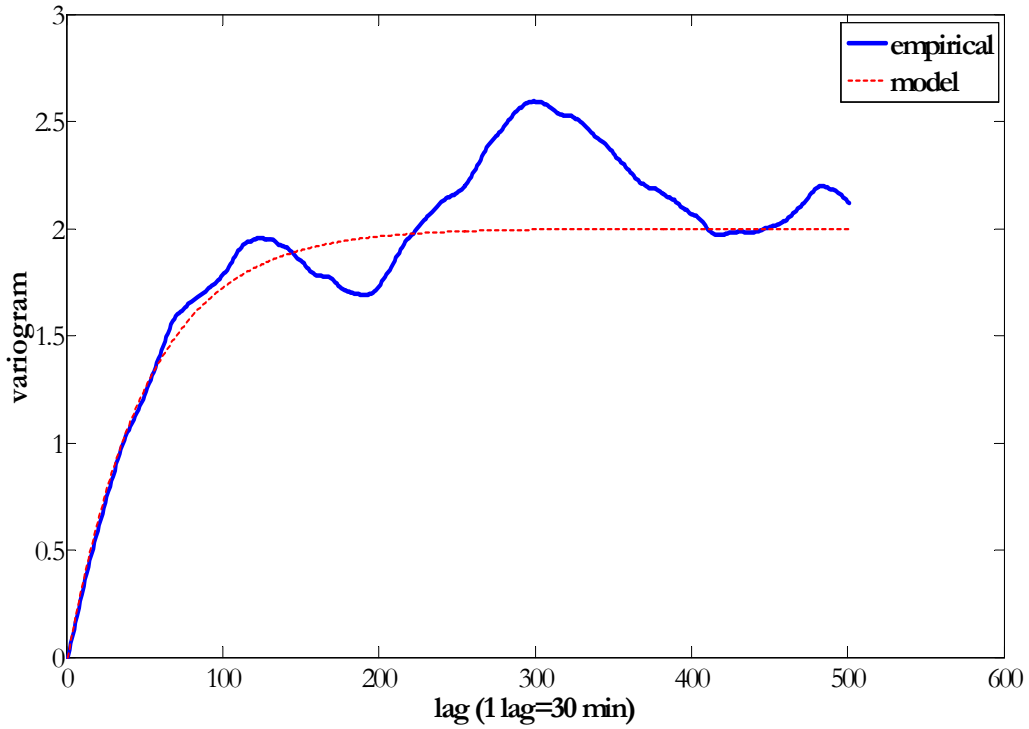


Figure G3. Brent residuals (17/12/04-27/01/06, frequency 30 minutes, n(EMA)=110 observations, observations 1:2514). Variogram of the residuals and theoretical model fit to the data $\gamma(h) = 2\left(1 - e^{-\left|\frac{h}{50}\right|}\right)$

The MACD indicator was constructed for the parameters $n_1 = 40$, $n_2 = 80$ (see Figure G4). The variogram estimated on the first half of the sample in presented in Figure G5. The following theoretical model is fit to MACD data:

$$\gamma(h) = 0.12\left(1 - e^{-\left|\frac{h}{50}\right|^2}\right) + 0.1\left(1 - e^{-0.3\left|\frac{h}{50}\right|}\cos\left|\frac{h}{50}\right|\right).$$

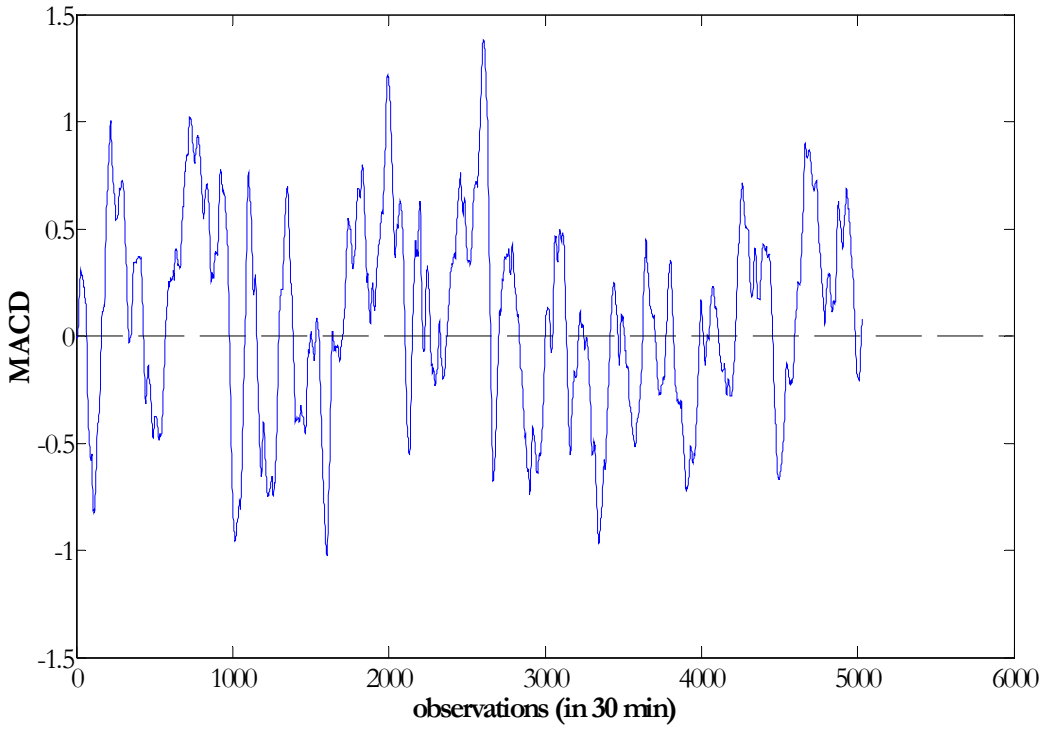


Figure G4. Brent (17/12/04-27/01/06, frequency 30 minutes). MACD indicator, constructed for EMA lengths $n_1 = 40$, $n_2 = 80$.

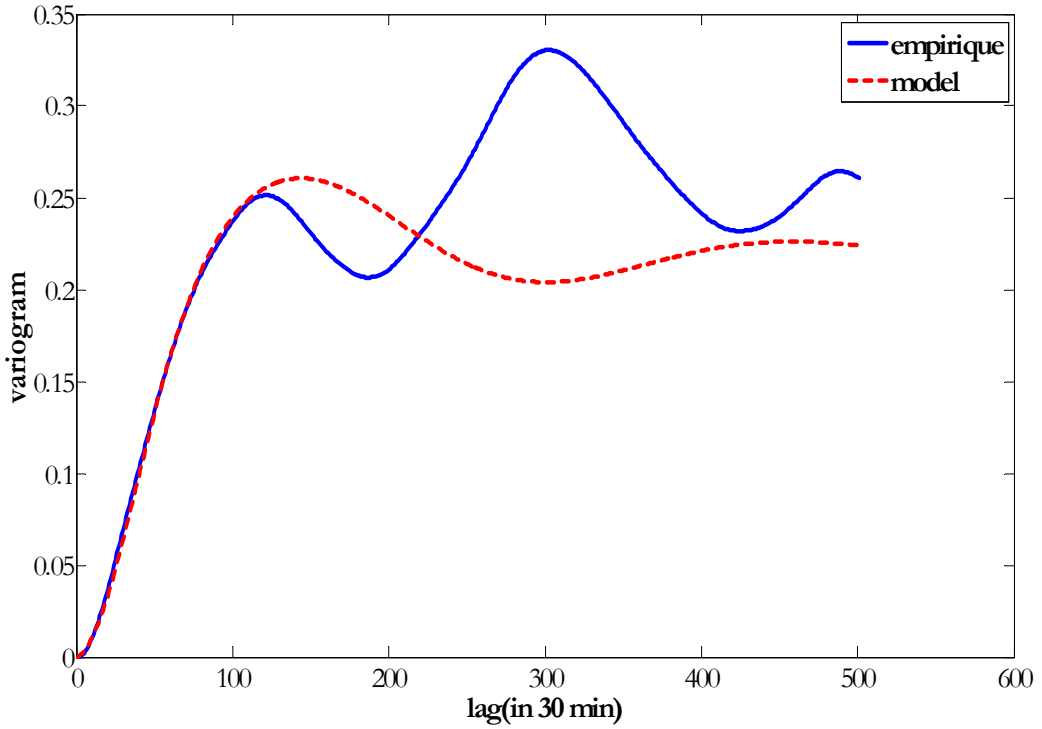


Figure G5. MACD indicator for Brent (17/12/04-27/01/06, frequency 30 minutes, $n_1 = 40$, $n_2 = 80$, obs. 1:2514): Variogram empirical and theoretical

Appendix H

Short description of the X instrument

Instrument X represents an artificially created index, used by one bank for strategy constructions. Due to the confidentiality reason we cannot neither present its detail description, nor provide the information on its real quotes. That is why on Figure H1, which presents the quotes path during some period of time, there are no ticks on the Y-coordinate. The only information we can provide is that data frequency is 1 hour and that index has mean-reverting nature in long run.

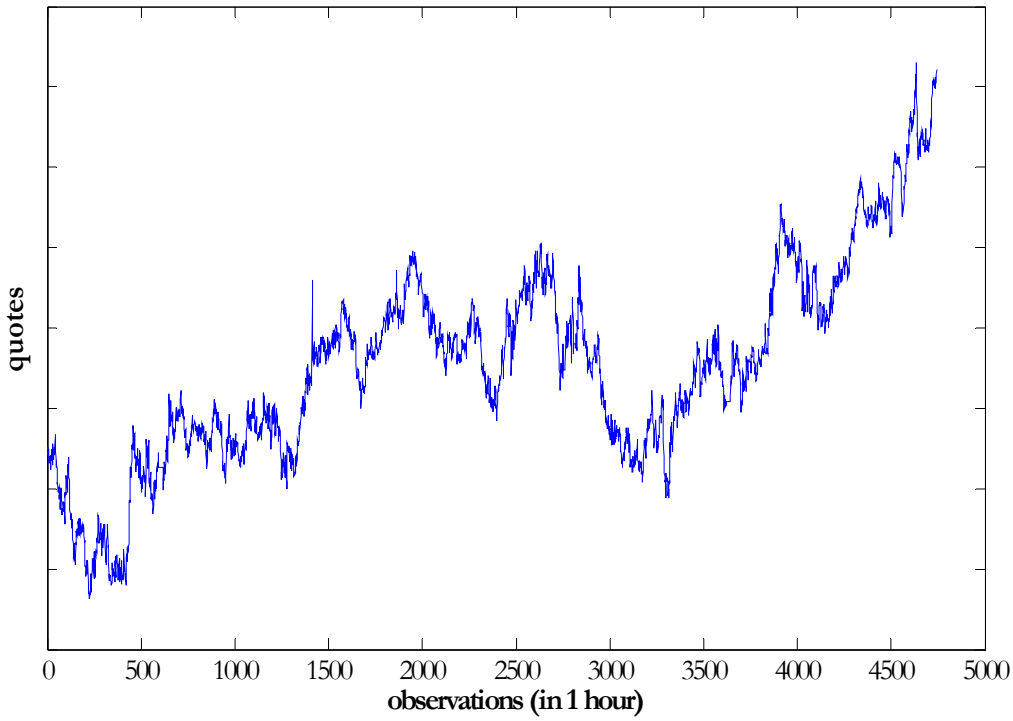


Figure H1. X instrument quotes (frequency – 1 hour)

As we can see from Figure H1, some trends are presented in the data. We eliminate the EMA of the length of 50 observations to obtain residuals used for the covariance calculations. As we can see from figure H2, the residuals do not exhibit any trend.

Empirical variogram of the residuals in the sub-sample [1;2371] is given in Figure H3. The following theoretical model is fit to empirical variogram:

$$\gamma(h) = 5.2\left(1 - e^{-\left|\frac{h}{18}\right|}\right).$$

The MACD indicator was constructed as the difference between EMAs of the following lengths $n_1 = 40$, $n_2 = 80$ (see Figure H4). The empirical variogram of this indicator estimated on the sub-sample [1; 2371] is presented in Figure H5. The following model is fit to the data:

$$\gamma(h) = 0.7\left(1 - e^{-\left|\frac{h}{45}\right|^2}\right) + 0.3\left(1 - e^{-\left|\frac{h}{80}\right|}\right).$$

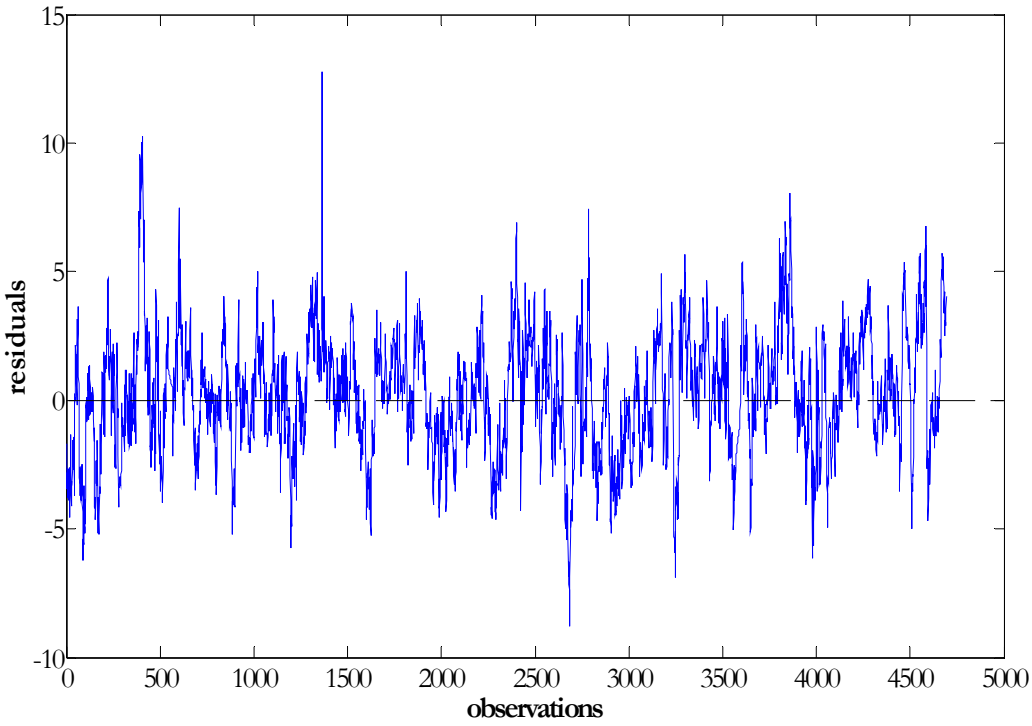


Figure H2. X instrument residuals (frequency – 1 hour, n(EMA)=50 observations)

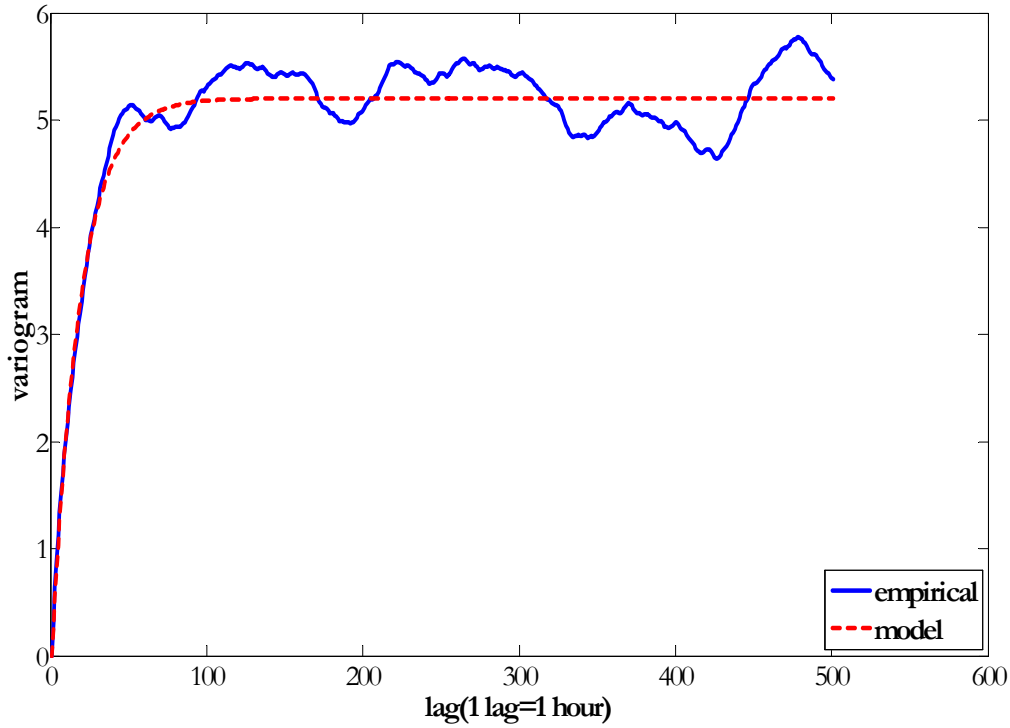


Figure H3. X instrument residuals (frequency - 1 hour, n(EMA)=50 observations): Variogram of the transformed residuals for sub-sample [1:2347] and theoretical model fit to the data.

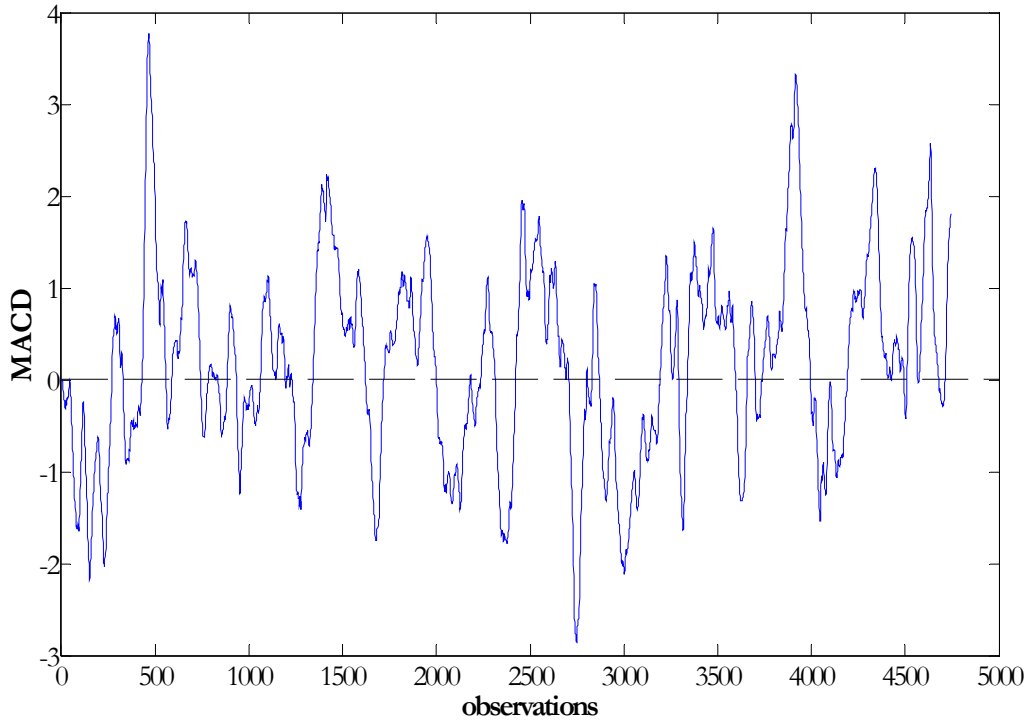


Figure H4. X instrument (frequency - 1 hour,): MACD indicator ($n_1 = 40$, $n_2 = 80$)

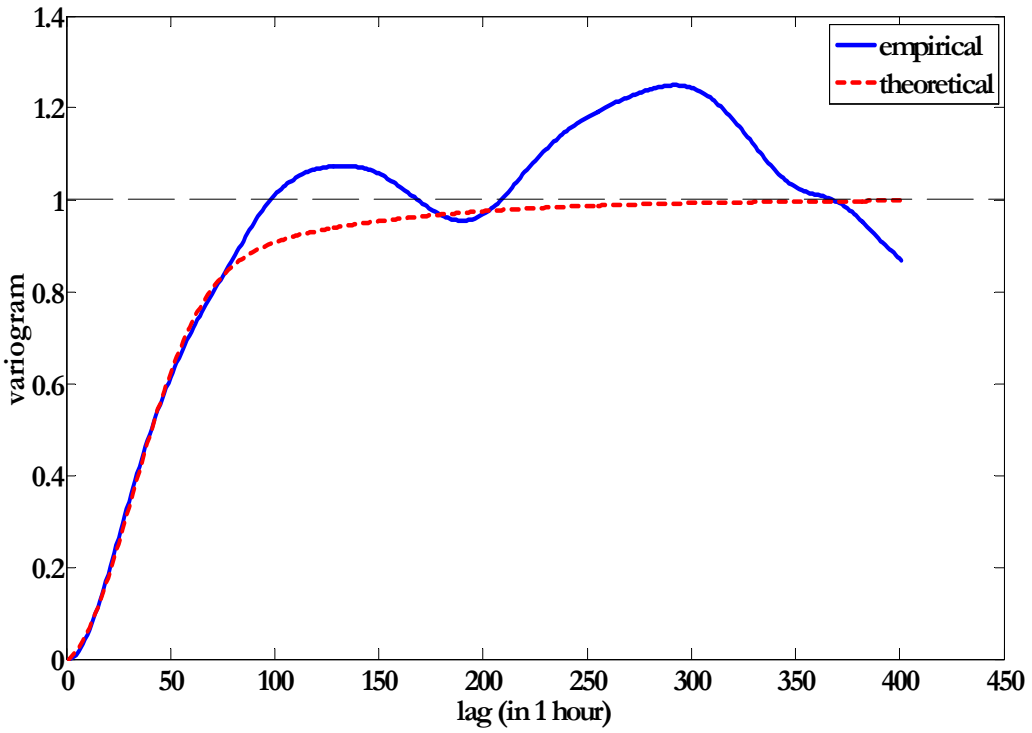


Figure H5. MACD indicator for X instrument (frequency - 1 hour, $n_1 = 40$, $n_2 = 80$, obs. 1:2371): Variogram empirical and theoretical (variance=1)

.

Appendix I
Examples of the MACD for the Bund data

The mean-reverting nature of the MACD indicator allows obtaining bounded variograms for these indicators. Further we consider how the choice of the EMAs lengths has impact on the MACD indicator and its variograms. Couples of EMAs lengths, for which MACD and its variograms are calculated, are given in Table I1.

Table I1
Effective lengths of the EMAs used for MACD calculations

Case #	1	2	3	4	5
n_1	12	6	18	12	6
n_2	26	13	40	52	26

Figure I1 presents examples of the MACD indicators calculated for the EMA lengths in Table I1. As we can see MACD indicators have similar forms. Some MACD curves have more smoothed form, but oscillate with higher amplitudes (case 4), some MACD curves are less smoothed but oscillate with lower amplitudes (case 5). As the MACD indicator exhibit mean-reverting pattern, we might expect the variogram to converge versus some level. However, taking into account significant discrepancy between the variances of different MACD, we propose to consider the variograms normalized by their respective variances (see Figure I2). Although the variograms curves do not coincide, the theoretic models used for their modeling most likely would be the same, which will differ only in parameters.

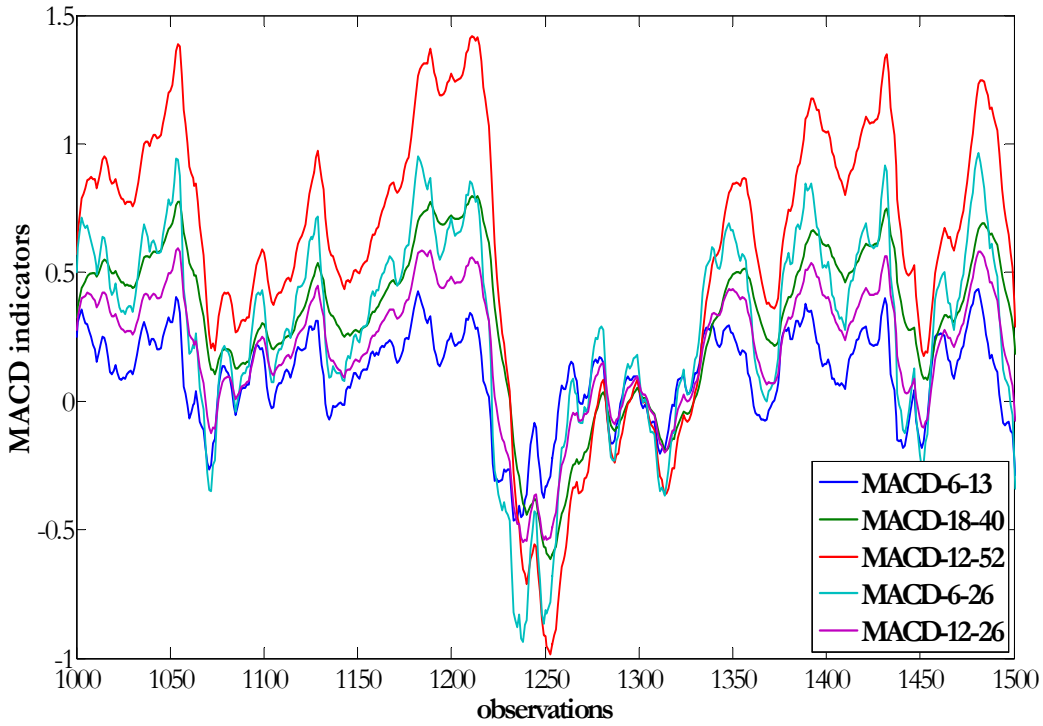


Figure I1. Bund (1991-2006, frequency = 1 day, observations 1000-1500): MACD, constructed for different EMAs

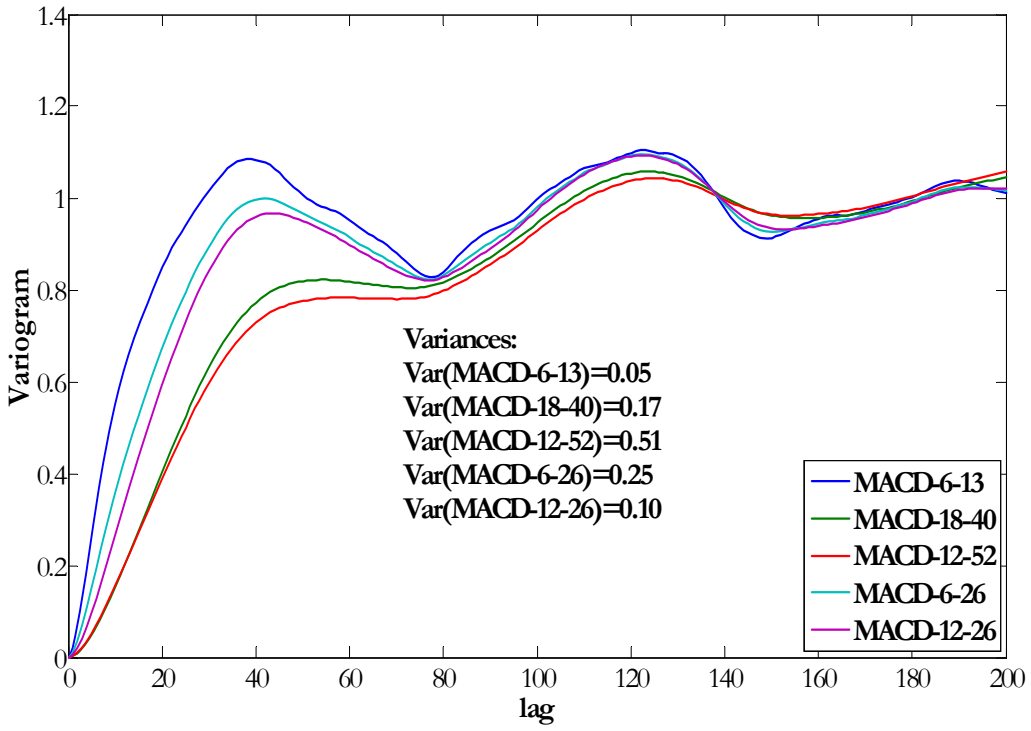


Figure I2. Bund (1991-2006, frequency = 1 day): Normalized variograms for the different MACD indicators, constructed for different EMA lengths

Appendix J

The examples of the weights for the kriged volume weighted moving average in the case of large variability in the cumulative volume

As have been discussed the presence of greater variability in the volume will have impact on the weight structure. We propose the following example to illustrate this hypothesis. Suppose the jump is present in the cumulative volume (see Figure J1). The kriged weights based on the observations in the interval [455;464] presents totally different pattern from what we have observed before. The weights are much more volatile.

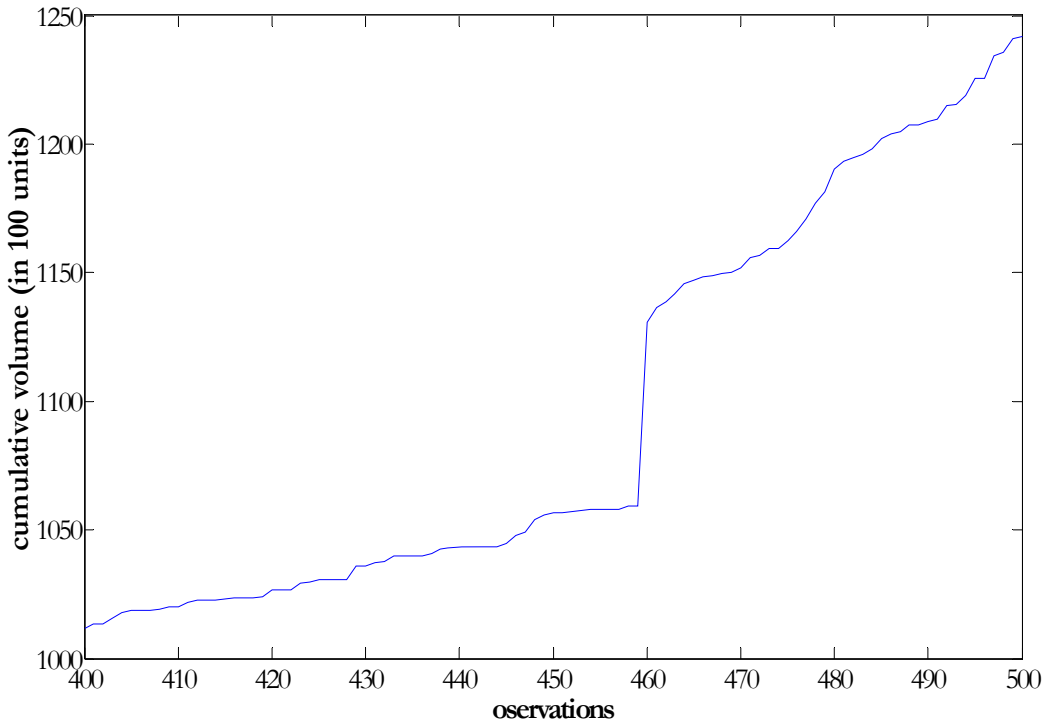


Figure J1. Bund (April 18, 2006, frequency 1 sec): cumulative volume (scale factor=100) at the observations 400-500.

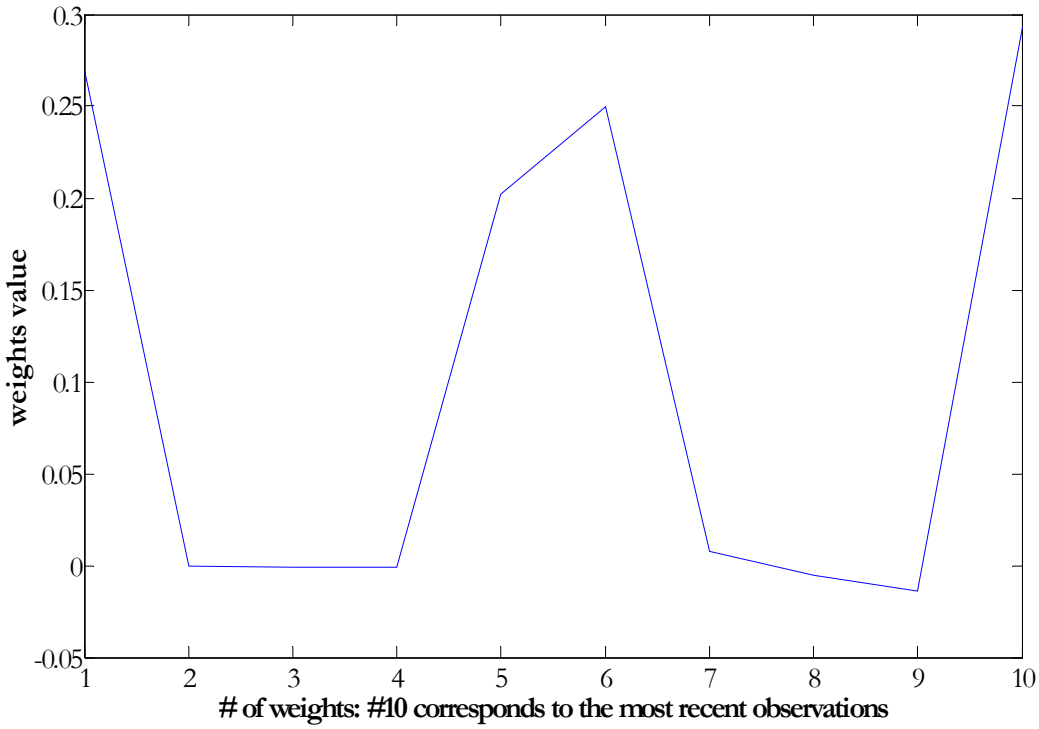


Figure J2. Bund (April 18, 2006, frequency 1 sec): weights of the kriged volume weighted MA for the price interval [455;464].

Part II. An Alternative to Bollinger bands: Data transformed bands

Introduction

Trading bands are one of the most frequently used tools of technical analysis. They are lines plotted around a measure of central tendency, such as the moving average, shifted by some percentage up and down (upper and lower bands) (Bollinger, 2002). The schematic representation of the concept is given in Figure 2.1. Different types of such bands are defined, such as envelopes, price channels (Bollinger, 2002; Chande, 2001; Murphy, 1999). While the difference between them lies only in the way that they are defined and constructed; the strategies based on them are quite similar. Touching/breaching these bands gives the trader information on the direction of price movements, or about relative price levels (whether the instrument is oversold or overbought), which are used as signals in strategy construction.

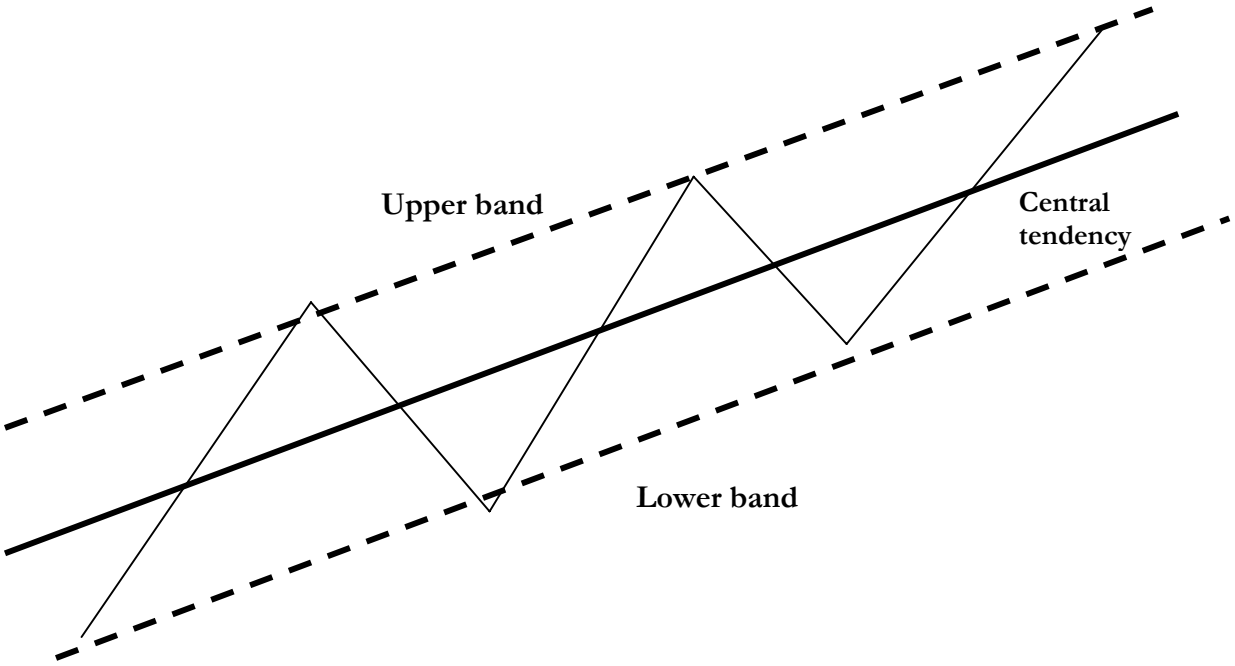


Figure 2.1. Schematic representation of trading bands

According to Bollinger (2002) trading bands were first used more than 50 years ago[1]. Some papers analyze the profitability of the bands, as well as compare their prediction power with other economic methods (Williams, 2006). Unfortunately, less work is dedicated to the analysis of the concept from the theoretical and statistical point of view. Later we will provide a short description of the bands without their deep analysis and explanations. An expanded overview of the different bands is given in the book "Bollinger on Bollinger bands" (2002), to which the interested reader can refer for more information. In this paper we prefer to look at the conceptual similarities and differences between the methods as we could not find this type of overview in any other papers.

From our point of view, the methods differ in two aspects:

[1] See Bollinger (2002) for more detail insight into history of the trading bands.

(1) the way that they are constructed, i.e. the indicators used to construct middle, upper and lower bands;
(2) the conceptual ideas, based on the data statistical characteristics, that are incorporated in these bands.

The difference in the construction method is mainly justified by the definition of different trading strategies. For example, if the original Bollinger bands use simple moving average (SMA) as the middle line, some internet trading sources propose to use exponential moving average (EMA) as the middle line justifying it by the possibility to generate higher trading profits. The problem with such justification is that any obtained profits are conditional on many trading factors such as strategy itself, traded instrument, its frequency, etc. Generalization of such results on larger samples of instruments without their preliminary testing is quite dangerous. Therefore, we found this approach more subjective.

Four groups of the conceptual ideas can be distinguished:
(1) constant volatility;
(2) non-constant volatility;
(3) distributional characteristics of the price or residuals (Price-MA);
(4) others.

The fourth group includes bands which incorporate either techniques, based on the concepts to define the bandwidth other than data statistical characteristics (for example, dividends); or techniques that were not described in the finance literature; or techniques that are very subjective in their presentation and usage. These among others include valuation envelopes based on the dividend yield, that were introduced in 1966 by the *Investment Quality Trends* newsletter; Le Doux's Twin-Line technique; and envelopes defined by Hurst in 1970 on the basis of (hand-made) cycles identification. This group of techniques will not be discussed any further in this paper. Our interest lies within first three groups of bands.

All the bands equally distanced from the middle line throughout the time are the representative of the first group of studies. The most widespread example of such bands is a parallel shift of the MA up and down by some fixed percentage (Katz, McCormick, 2000, Lien, 2006).

The desire to incorporate the stochastic (non-constant) price volatility led to the development of other band types. This boosted the development of the second group of the studies. Keltner's method was among the first approaches, used in the 1960s, which incorporated a rolling estimate of the instrument volatility (Keltner, 1960 in Bollinger 2002). In fact it used a 10-day MA, built on the typical price[2], as a center line, and (moving) 10-day average range[3] as trading bands. Technical traders are aware that range is one of the volatility measurements. The Bollinger bands, developed in early 1980s by John Bollinger, became the most popular representative of this conceptual group. Bollinger used standard deviation to measure an instrument's volatility and to construct bands around some SMA. We will discuss this method in more detail later.

Examples of the third conceptual groups include Bomar bands and the Donchian channel. In early 1980s, asymmetrical Bomar bands were proposed by Chaikin and Brogan (presented by Bollinger 2002), based on the following construction formula: upper/lower bands contain the 85% of data above/below the 21-day MA for the past 250 periods. As we can see, the developers should use the distributional properties of the instrument in order to define the confidence intervals that contain the 85% of data. We will discuss the problems linked to this method further

[2] Typical price is the sum of the high, low and close prices, divided by 3.
[3] Range is a difference between high and low prices at given moments of time.

in more general framework. Here we just want to cite Bollinger, who explained the main advantages and drawbacks of the method[4]:

"The major benefit conferred by Bomar Bands was that analysts were no longer forced to **provide their own guesses about what the proper values for the bands** were. Instead, they were free to focus on decision making and let their PCs set the bandwidth for them. Unfortunately, Bomar Bands were **extremely computationally intensive** for their time, and to this day are not readily available beyond Instinet's research and analytics (R&A) platform. Thus, they **have not achieved the broad acceptance** they deserve."

John Bollinger, 2002, "Bollinger on Bollinger Bands", p.46.

The first conclusion is that Chaikin and Brogan's study was the first attempt to formulate a mathematical optimisation problem to find optimal bandwidth values. It was based on probabilistic concepts such as the distributional characteristics of the random variable under study. It is clear, that all the studies and their conclusions produced before Bomar bands were subjective and data sensitive (based on traders' guesses or analysis and provided for concrete instruments, i.e. data samples). The second conclusion is that due to their computational intensity these bands did not become very popular.

Donchian's "four-week rule" is older (defined in earlier 1960s), but no less interesting: buy when the four-week high is exceeded and sell when the four-week low was breached (Bollinger, 2002). This rule became a prototype for the Donchian channel - the concept that sets the upper band at the *n*-period high and lower band at the *n*-period low. We classify these band types in the third group, as we can see in the basis of the rule the statistical concept of extreme values that were either consciously applied or was empirically observed and used by the developers. Roughly speaking, the traders define the maximum/minimum (extreme) price values for some past periods as a threshold for the definition of the overbought/oversold markets in some future periods. Again, we would like to cite Bollinger on a Donchian channel:

"This concept is rumored to be at the heart of one of the more[5] successful trading approaches in wide use today, that employed by the Turtles[6]."

Bollinger J, 2002, "Bollinger on Bollinger Bands", p.40

Another unusual fact about these bands is their "stair"-like form. This form, however, can be easily explained by the fact that data within some time frames are locally stationary with the same distributional properties and thus, extreme values. Simply speaking, consider a subsample of n values on the interval $[t-n+1;t]$ to define the local maximum and minimum of the random data. For locally stationary data it is unlikely that rolling the subsample forward by 1 observation (or a small Δ number of observations) to the interval $[t-n+2;t+1]$ (or $[t-n+1+\Delta;t+\Delta]$) would significantly change the local maximum/minimum. That is why we can obtain these, let's call them, Δ-steps of the "stair"-like bands.

To be "statistically" correct, we should say that the first and second groups are just special cases of the third group in which the volatility is one of the statistical characteristics that are parts of more general distributional analysis of the data.

In this paper, we propose an alternative type of trading bands, called data-transformed (DT) bands, which belong to the third group of bands. DT bands are more general case than the Bollinger bands:

---

[4] The author of this article uses underlined and bold fonts.

[5] Possibly should be "most" instead of "more"; probably Bollinger text error.

[6] "Richard Dennis, a famous commodity trader, taught a number of traders his proprietary techniques. Those traders are called Turtles" (Bollinger, 2002, p.190)

(1) they provide more rigorous theoretical interpretation (justification);
(2) the manner in which they are constructed, allows for a mathematical approach for bandwidth optimization that is based on data distributional properties.

Our goal is to produce a new theoretical concept and to show that DT bands can be used as a new technical indicator to create successful trading strategies. For this the method must be easy to apply[7] and outperform Bollinger bands at least in some cases. To illustrate its potential, we give four numerical examples on different types of the financial instruments (bonds, stocks, commodities and artificially constructed portfolio indexes), at different frequencies and time periods. In these examples we compare the DT bands with the classical Bollinger bands and show that in some cases DT bands give better results.

The empirical observations of the DT bands suggest an interesting hypothesis. We believe that DT bands can help to detect Elliot waves and Support/Resistance levels, two popular concepts in technical analysis. This hypothesis will not be analyzed in detail as it goes beyond the scope of this paper. However, the strategies based on these signals are introduced and they show positive (though not the best) outcomes in the majority of cases.

The structure of the part is the following. In the first chapter we present the classical Bollinger bands and the trading strategies built on these bands, and discuss the assumptions on which the theory of Bollinger bands is based. The second chapter introduces the new concept of the DT bands. The third chapter proposes the trading strategies and the framework for the comparisons of the DT and Bollinger bands. The fourth chapter presents four numerical examples. The conclusions are proposed at the end of the part.

1 Classical Bollinger bands concept

This part discusses the concept of Bollinger bands proposed in the early 1980s and still very popular among traders nowadays. The first subchapter defines the bands and explains how they are constructed (calculated). The second subchapter presents several trading strategies based on Bollinger bands. Finally, the third subchapter proposes some plausible assumptions as the basis of the theory and as such plays an important role in our paper because there do not seem to be any theoretical papers on these bands.

This is rather surprising because this approach clearly introduces the statistical concepts (volatility measurements) into the trading methods and provides a powerful instrument to construct profitable trading strategies. Unfortunately Bollinger bands are not presented as true theoretical concept: there is no framework, including assumptions; no cases are given to show when the method should work (at least theoretically). Bollinger mentioned these assumptions indirectly in his works (books, web-sites) explaining in a simple manner when and how the approach should be used. But we believe that these assumptions should be formalized in more scientific manner. In particular, as main critique of the technical analysis is its subjectivity and the absence of scientific justification. This is not merely a theoretical exercise. We will show that by questioning these assumptions we were able to produce a more general concept of DT bands.

[7] In contrast to the Bomar bands, DT bands are not computationally intensive.

1.1 Bollinger bands definition

Bollinger bands are particular examples of trading bands. Figure 2.2 shows an example of these trading bands for Bund quotes. We can see the four key components:

(1) price (quotes),
(2) the mid-line,
(3) the upper band, and
(4) the lower band.

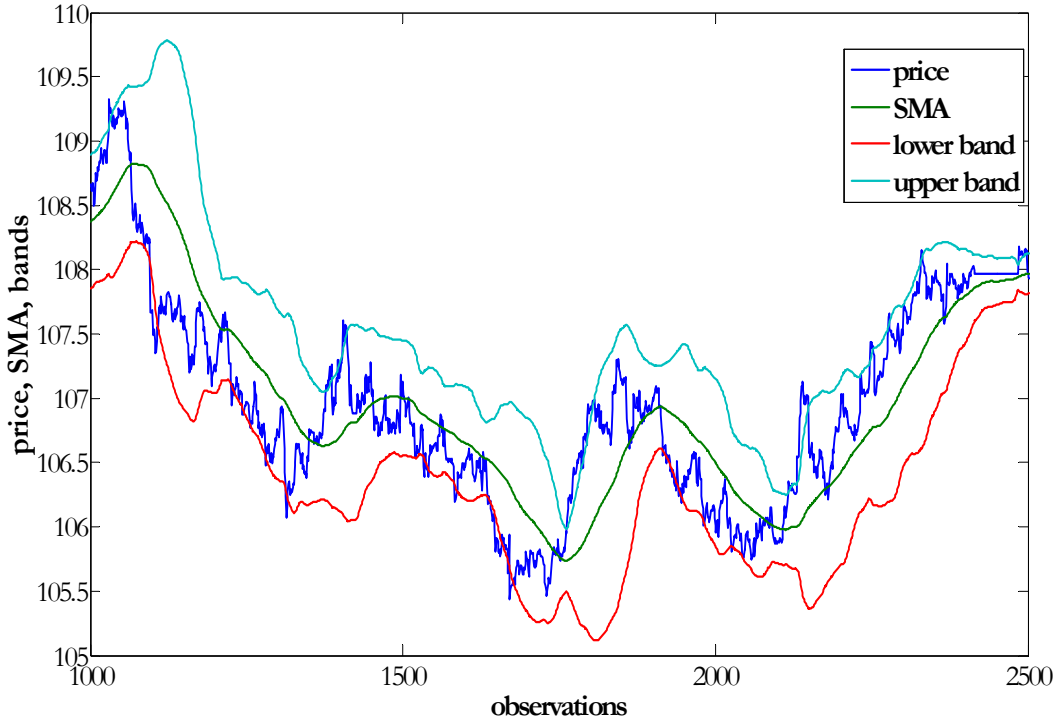


Figure 2.2. Bund quotes (30/7/2003-7/12/2006, frequency: 30 minutes), SMA (length n=115 observations) and Bollinger bands (Bollinger parameter k=2.1)

In most of the well-known strategies, based on the Bollinger bands, the moving average (MA) represents the mid-line. Traders use different types of moving averages such as a simple moving average (SMA) or an exponential moving average (EMA). The upper and lower bands are obtained by shifting the mid-line up and down by a value, which is a function of the price volatility. These components are represented mathematically by the following formulas:

(1) Price curves: $\{P_i\}_{0\le i\le t}, \forall t > 0$

(2) Mid-line:

$$m = \frac{\sum_{i=t-n+1}^{t} P_i}{n} \qquad \text{(II.1.1)}$$

It is an estimator of the mean on the interval $[t-n+1, t]$, which coincides with simple moving average (SMA) of length n at time moment t.

(3) Upper line:

$$b_{U_t} = m_t + k\sigma_t \qquad \text{(II.1.2)}$$

where m is the price mean estimator on the interval of length n ($[t-n+1,t]$), or SMA of length n estimated as in (II.1.1) and where σ^2 is an estimator of the price variance on the same interval of length n: $[t-n+1,t]$:

$$\sigma_t^2 = \frac{\sum_{i=t-n+1}^{t}(P_i - \mathrm{E}(P))^2}{n} \qquad \text{(II.1.3)}$$

k - Bollinger bands parameter, $\forall k = const > 0$.

For stationary processes the expectation $\mathrm{E}(P)$ in its pure statistical sense coincides with SMA_t or mid-line for the Bollinger bands. This is one reason why Bollinger insists on using this type of the MA as the mid-line. Note[8], however, that the estimator of the variance is not the weighted sum of the squared sum of the residuals $R_i = P_i - SMA_i$:

$$\sigma_t^2 = \frac{\sum_{i=t-n+1}^{t}(P_i - SMA_t)^2}{n}, \qquad \text{(II.1.3')}$$

not $\sigma_t^2 \neq \frac{\sum_{i=t-n+1}^{t} R_t^2}{n} = \frac{\sum_{i=t-n+1}^{t}(P_i - SMA_i)^2}{n}$.

(4) Lower line:

$$b_{Lt} = m_t - k\sigma_t \qquad \text{(II.1.4)}$$

where m, σ, k, n are as in (II.1.2).

From these comments we can draw the following conclusions:

- The Bollinger trading band strategy is adjusted for price volatility, as the bands definition involves standard deviation, as a volatility measurement.
- Upper and lower bands at the same moment of time are placed at equal distance with respect to MA, i.e. these bands are symmetrical. However, this distance can be different at different time moments, due to the changing nature of the price volatility[9].
- The distance between bands is narrowing with decrease in the price volatility and widen with increase in the price volatility.
- The usage of SMA as the middle line at the place of other moving average is justified by the fact that SMA is the statistical mean of the price sub-sample – the same value that is used for the calculations of the price standard deviation. In addition, some research shows that substitution of the SMA by more rapid MA does not produce higher strategy outcomes (Bollinger, 2002). We have also shown in part 1 that optimal kriged moving average (KMA) on average coincides in lag with the SMA; and KMA might almost coincide with the SMA in values at large lengths of the mowing window. The interest of introducing EMA on the place of SMA might be the recurrent method of its calculation (the current EMA can be calculated as a weighted sum of current price and previous EMA). However, at this respect SMA can be also programmed as recurrent formulae, but it needs to save more data at each time moment than the EMA.

[8] This observation will be important further when we will consider the data-transformed bands.
[9] It is well-known fact that historic price volatility is not a constant value.

The choice of the parameters k, n should depend on the data and optimization procedures. However, Bollinger (2002) proposed the following default values for these parameters:

a. $n = 20$ days (or 1 month)[10];
b. $k = 2$
c. If n decreases to 10, then k should decrease to 1.9, and if n increases to 50, then k should increase to 2.1 (see Table 2.1).

Table 2.1
Default Bollinger bands parameters values

Moving average length	*k*
10	1.9
20	2
50	2.1

Source: Bollinger, J. 2002. "Bollinger on Bollinger Bands"

These default values were based on empirical findings that at these values one can get "containment between 88 and 89 percent in most markets"[11]. In the case the data demands calculation of the Bollinger bands for the periods longer than 50 or shorter than 10, Bollinger suggested changing the data sampling frequency in order to stick to the default values 10-50 periods. The default parameters are also proposed as a starting point in optimization procedure of the Bollinger bands parameters search.

1.2 Bollinger bands and trading strategies

The classical trading strategy based on Bollinger bands can be defined as a contrarian, i.e. the entry positions are characterized by over- or under-pricing of the instruments. The bands themselves serve as the thresholds, which define whether instrument can be considered as miss-priced. The simplest Bollinger bands strategy can be formulated as following:

- the long position is taken when prices approach (or cross) the lower band (Figure 2.3, point 1). An instrument is under-priced and its price is expected to return to its average, i.e. to increase.
- the short position is taken when prices approach (or cross) the upper band (Figure 2.3, point 3). An instrument is overpriced and its price is expected to return to its average, i.e. to decrease.
- the position can be closed when prices return to the average level (Figure 2.3, points 2 and 4). The middle line or MA represents this average price level.

The probabilistic explanation of the strategy is presented in Figure 2.4. Let say the price has some density with mean m, given by solid blue curve. The Bollinger bands $m \pm b$ define the areas of the extreme price values that defines over- or under-priced instrument (the areas marked by blue lines). Suppose current price $P1$ has breached the upper Bollinger band $m + b$, indicating that the instrument is overpriced. The trader take short position expecting that the price will return to its mean value. Taking into account the form of the density curve, the future price will return to more probable value within the interval $m \pm b$. Even if the future price value will be only $P2 > m$ the net worth of the position is positive: $W1 = P1 - P2 > 0$.

[10] Bollinger (2002) has not precised the frequency of data, but we believe he talked about daily (or monthly) frequency.
[11] Bollinger, J. "Bollinger on Bollinger Bands", *McGraw Hill,* 2002, p. 54.

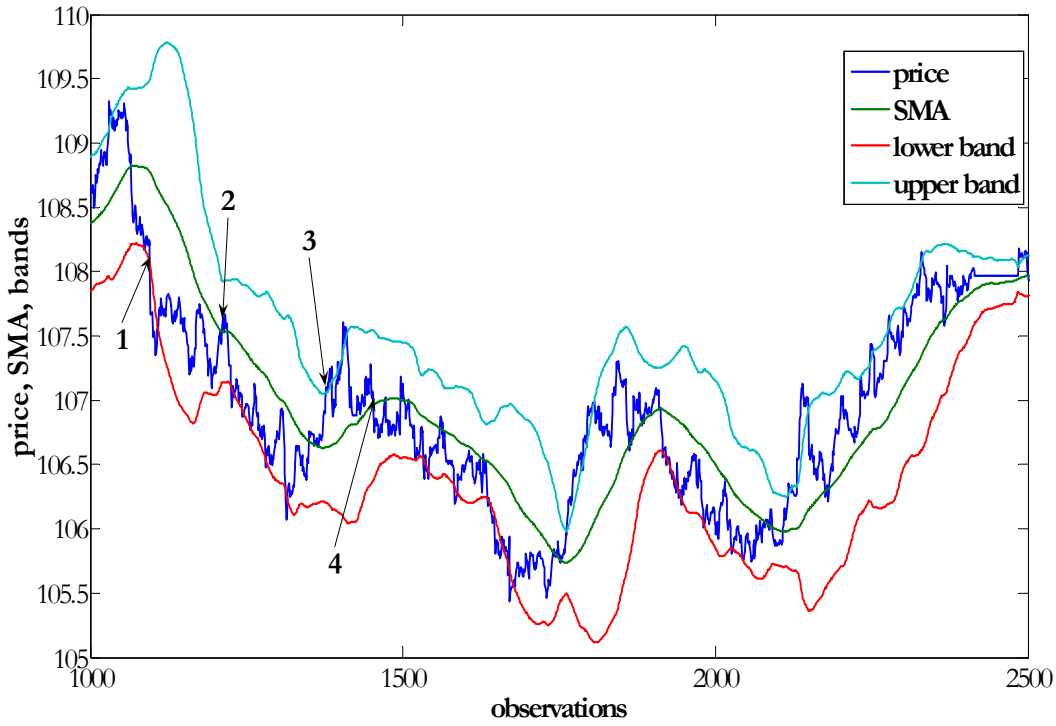


Figure 2.3. Bund quotes (30/7/2003-7/12/2006, frequency: 30 minutes), SMA (length n=115 observations) and Bollinger bands (k=2.1): Positions, taken within contrarian strategy: (1) entry-long; (2) exit-long; (3) entry-short; (4) exit short.

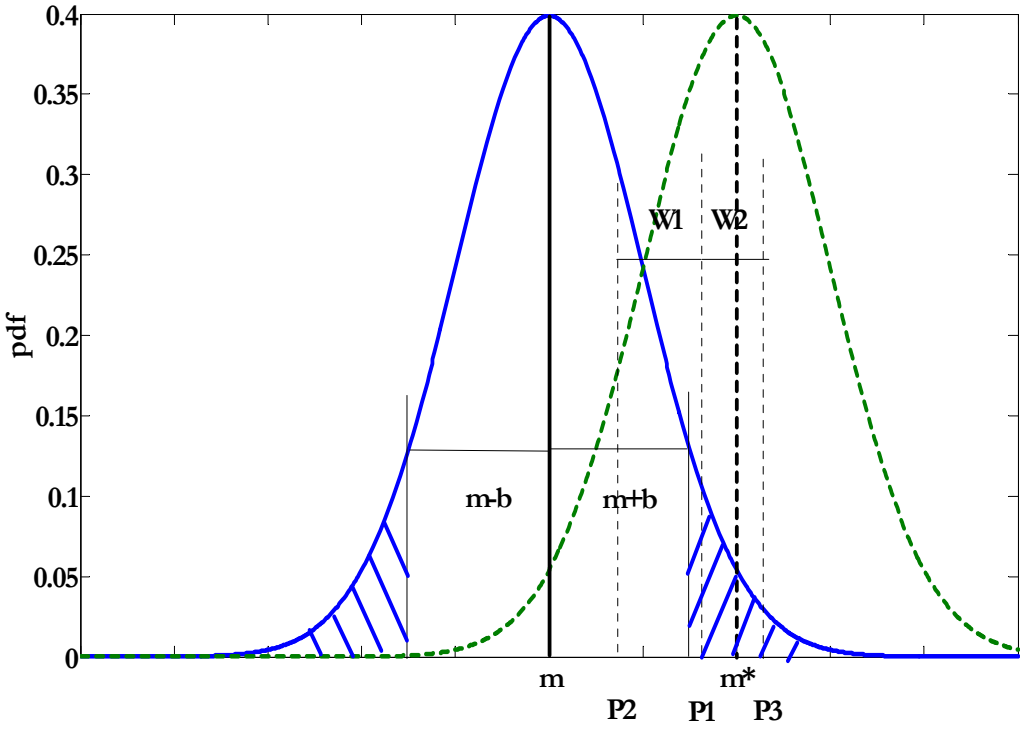


Figure 2.4. Statistical explanation of the Bollinger bands strategy

Figure 2.4 also explains what happens when the statistical characteristics of the variable has changed, in particular the mean has changed from $m \to m^*$ [12]. In such case price $P1$ will be mistakenly defined as extreme value, and trader will take the short position expecting that the price will fall to m. In reality it is more probable that the price will move up to its current mean or even higher $P3$. In such case the net worth of the position will be negative: $W2 = P1 - m^* < 0$ or $W3 = P1 - P3 < 0$.

Figure 2.4 also explains why the parameter k is tabulated for different MA lengths. The distribution of the residuals, in particular its dispersion, depends on the MA length used for their calculations. As the result, the bands, which contain 88-89% of data, depend on the MA length.

There are some specific features of applying Bollinger bands, which are widely discussed in the trading literature and can be seen from Figure 2.3. First, during well-pronounced positive (negative) trends the price fluctuates mainly between middle line and upper (lower) band. For example on Figure 2.3, prices vary between (approximately) 1100-1400 observations for negative trends and between 2100-2500 observations for positive trends. The probabilistic explanation is that current mean is changing, in particular increasing (decreasing), thus higher (lower) price are most probable.

Second, Bollinger band strategies, like all contrarian strategies, send false signals during the upward or downward trends due to the mistakes it makes in defining the "true" instrument value. For trending markets the "true" value increases or decreases, thus, "overvaluation" or "under-valuation" statements are wrong. For example, what we believe to be over-valued, in reality indicates the increase in the true value. Statistically it implies that our current bands parameters (estimates) do not reflect the true probability distribution, which is not constant. Generally speaking, applying contrarian strategies in the trending markets leads to the losses.

Due to these false signals traders normally do not rely solely on the Bollinger band signs, but consider them in the combination with other technical signals, which either confirm the overbought/oversold state or predict the future price movements, for example, trend-reversion. These indicators include momentum, volume, sentiment, open interest, inter-market data, etc. (Bollinger, 2002, www.bollinger.com). Consequently, Bollinger bands do not give the unconditional signal to entry the position; they give clues as to how the price might behave in future but these clues normally need confirmation.

Confirmation signals help to avoid these trending parts of the price curve. We choose the momentum indicator to confirm that the market is oversold or overbought. Momentum can be calculated as follows:

$$M(t,\Delta) = P(t) - P(t-\Delta) \qquad \text{(II.1.5)}$$

where Δ is the lag parameter.

As can be seen from Figure 2.5 momentum is an oscillator, which fluctuates around zero. (See Figures A1-A2 in Appendix A for real-life examples). Local momentum maximum/minimum indicate that the local upward/downward trend is slowing down (Figure 2.5, points 2 and 5), while the intersection with the 0 level points to the trend inversion (Figure 2.5, points 3 and 6). The overbought/oversold markets are confirmed when momentum reaches some threshold

[12] Note that the change in the true mean will not be reflected straight away in the SMA (mid-line) and bands, as their estimates depends on the past values.

levels while moving upward or downward[13]. These (thresholds) values depend on the instrument quotes and behavior. So they have to be calculated on a case-by-case basis.

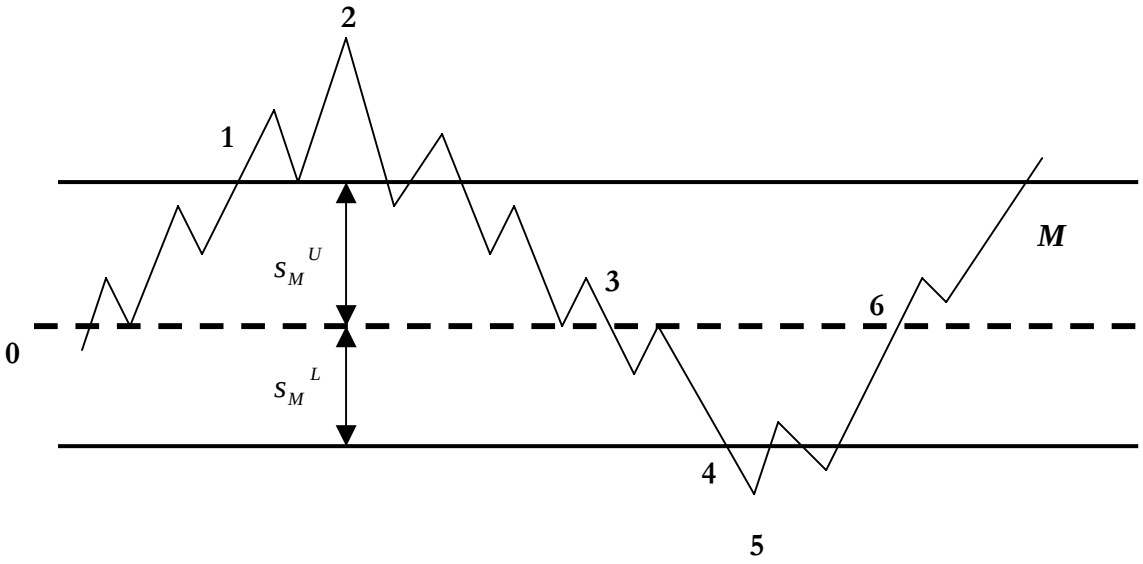


Figure 2.5. Schematic representation of a momentum indicator

We can formulate the strategy, based on momentum, as follows. Let $s_M{}^U$, $s_M{}^L$ be the momentum threshold values used to define the overbought and oversold signals. Then if momentum values are higher than the upper threshold level, the instrument is overbought (Figure 2.5, point 1); if momentum values are lower than the lower threshold level, it is oversold (Figure 2.5, point 4). Suppose for simplicity that the thresholds are symmetric $s_M{}^U = -s_M{}^L = s_M$. If $M(t,\Delta) > s_M$ then the overbought market is confirmed and if Bollinger bands strategy generates the same signal – take short position (*S*). By the symmetry if $M(t,\Delta) < -s_M$ then the oversold market is confirmed and if Bollinger bands strategy generates the same signal – take long position (*L*).

The usage of the momentum as confirmation for miss-priced instruments might be confusing at first sight as this indicator used mainly in the trend-following strategy formulation. There are several explanations, why the usage of the trend-following indicators in the contrarian strategies might be accepted. On one hand we can define the trend-less long-term period as the combination of the local short-term upward and downward trends. Than slowing down of the trend implies the reversion of the trend and thus, return to the mean value. On the other hand, the increasing/decreasing momentum within its thresholds interval indicates that currently the instrument is in trending periods, allowing to avoid taking false position by the contrarian strategy. Finally, the signal (sent by momentum) that the current trend is slowing down might indicate its reversion into trend-less periods; the contrarian strategies produce the best results for the price pattern without any trend.

Further we show that these confirmed strategies lead to higher outcomes than non-confirmed ones.

[13] See Murphy (1999) for more details.

1.3 Theoretical framework for Bollinger bands

The main problem with the Bollinger band approach is the absence of a precise theoretical framework, which could justify their usage in trading. Literature review indicates three groups of the research done in the field. Bollinger (2002) and many Internet resources explain how to calculate Bollinger bands and presents various trading strategies based on indicators or price patterns. The other two groups concentrates on the optimization of the bands parameters, such as scaling parameter k and MA length n; the difference between them lies in the chosen optimization approach. Bollinger (2002) provides the optimal couples of the parameters (k,n) that work for the "majority of markets". These beliefs are based on the analysis of the prices and verification that the "default" bands include a high proportion of prices (around 90%). Other group of research (for example, Williams, 2006) has defined parameters values (k, n) as being optimal, when they guarantee positive outcomes for the simulated trading strategies based on these bands.

Consequently, we can see that parameters optimization problem is based on at least two different approaches. One approach is based on trading profit maximization, i.e. find k so as to obtain the largest or stable positive profit when using the trading strategy. The main criticism of this approach is that the optimal values are data and strategy type dependent, because the optimization is based on past historic data and strategy algorithm.

According to the other approach, the optimal parameter values guarantee that the interval between the upper and lower bands contains some fixed percentage, *K%,* of price observations. Normally the *K* lies between 85-95%. This optimization approach is more appealing, as it explains why the strategy should generate profits, from a statistical point of view. For large values of *K*, the prices outside the bands are extreme values that have a low probability of happening (the instrument is overpriced/under-priced), thus, the future prices should probably move to the values close to the average one[14]. That is, the instrument will return to its average value.

Under such problem definition the strategy characteristics, such as profit and number of trades, depend on the *K%* value. The net profit per trade is the difference between the (attained) extreme and the mean values, while the expected profit is the weighted sum of all possible net profits. The number of trades under this strategy is an indirect function of *K%*: the larger *K%*, the less probable is the occurrence of the events, when price is outside the bands; and, thus, fewer trades are undertaken. This conclusion is important for the choice of *K%* value.

Both optimization problems should be based on the number of statistical assumptions about the data that can define the theoretical framework for the Bollinger bands application. In particular, the second optimization approach, as more statistically grounded, requires more assumptions about data, which, we define further.

Assumption 1: Data stationarity.
Stationarity imposes the invariance of the joint probability density function and, therefore, all its moments under temporal (or spatial) shift (translation), i.e. for any h probability distribution of the variables $P(t_1), P(t_2),..., P(t_n)$ is the same as of the variables $P(t_1+h), P(t_2+h),..., P(t_n+h)$. For the temporal data it means that all moments of the random variable does not depend on the time at which the variables are observed, but only on the distance between them. Simply speaking, "information about the process [is] the same no matter where it is obtained"[15].

[14] The conclusion holds for the stationary process.
[15] Anselin L. "Variogram analysis", presentation

However, even under stationarity assumptions joint probability distribution is often difficult to estimate; therefore, the statisticians has introduced second-order stationarity - the invariance of the first and second moments under translation (see Definition 2.1).

Definition 2.1
A random function $\{P(t)\}$ is stationary of order two if its mean and covariance do not depend on time, but only on the distant between the variable, i.e. $\forall t, h$:

$$\mathrm{E}[P(t)] = m \qquad \text{(II.1.6)}$$

$$Cov[P(t), P(t+h)] = \mathrm{E}[P(t)P(t+h)] - m^2 = C(h) \qquad \text{(II.1.7)}$$

For Bollinger bands this means the following:

$$m = \mathrm{E}(P) \approx \frac{\sum_{i=t-n+1}^{t} P_i}{n} = SMA_t = const$$

$$\sigma^2 = \mathrm{E}\left[(P - \mathrm{E}(P))^2\right] = \frac{\sum_{i=t-n+1}^{t} (P_i - m)^2}{n} = const \qquad \text{(II.1.8)}$$

$$P_t \sim L(m, \sigma^2)$$

In the case of the stationary data, SMA can be used as a true mean estimator. Then statistically the prices that breach the bands can be considered as the extreme values and application of the contrarian strategy will make sense.

Therefore, if the price data is *stationary of order two*, then:

1. Bollinger bands can be defined as (II.1.1)-(II.1.4) in section 1.1 and Bollinger strategy has statistical justification. Moreover, the bands will be at almost the same distance from MA through all the time.
2. The same strategy parameters (defined as optimal) can be applied throughout the time.

The stationarity hypothesis can be weakening if we assume that prices are only locally stationary. Local stationarity means that expression (II.1.8) holds true only within some interval of time and the values of statistical moments although constant within the interval on which they are estimated, can be different from one interval to another:

$m_l = \mathrm{E}(P) \approx \frac{\sum_{i=t-n+1}^{t} P_i}{n}$, $m_l = const$ on the interval $[t-u, t]$, where $u > n$

$\sigma_l^2 = \mathrm{E}\left[(P - \mathrm{E}(P))^2\right] = \frac{\sum_{i=t-n+1}^{t} (P_i - m_l)^2}{n}$, $\sigma_l^2 = const$ on the interval $[t-u, t]$, where $u > n$

$P_t \sim L(m_l, \sigma_l^2)$, where L - distribution law is the same on the interval $[t-u, t]$, where $u > n$

We can imagine such patterns as a combination of some trending periods with the trendless periods; or the presence of the jumps in the trendless price patterns.

Then:

1. Within stationary intervals the Bollinger bands can be defined as (II.1.1)-(II.1.4) in section 1.1 and Bollinger strategy has statistical justification.
2. The strategy parameters need to be recalculated locally, each time the stationarity pattern changes.

In some cases financial data can be considered trend stationary. A trend stationary variable is a variable that is formed by the deterministic trend and stationary process (Focardi, Fabozzi, 2004):

$$P_t = m_t + R_t, \qquad \text{(II.1.9)}$$

where m_t - is some deterministic function, which values we know or can calculate at each moment of time t; R_t - stationary residuals. A trend-stationary process can be transformed into stationary by subtracting known trend m_t. The main problem with the trend stationary data is that normally the trend values m_t are unknown; thus, needs to be estimated first.

The advantage of the Bollinger bands lies in their flexibility. If in the case of the stationary data SMA served as a mean estimator, in the case of trend-stationary data SMA can be considered as estimator of the trend m_t. Let say the length of the SMA n^* such as $\mathrm{E}(R_t) = 0$. In this case, the Bollinger bands can define the extreme values, as in this case:

$$\mathrm{E}(P_t) = m_t + \mathrm{E}(R_t) = m_t \approx SMA_t \qquad \text{(II.1.10)}$$

$$\sigma(P_t) = \sigma(R_t) \qquad \text{(II.1.11)}$$

Although the statistical meaning of the bands is still preserved, the application of the contrarian strategy should be under question as the mean of the price is trending. Therefore, even if the position is taken for the (locally) extreme price values, the position might be liquidated at the SMA_t, which is unfavorable from the point of view of the entry price. Therefore, under the trend-stationary hypothesis, the Bollinger bands have statistical meaning, but the application of the contrarian strategy is not profitable.

Finally, if data is non-stationary we can calculate the Bollinger bands, but the trading strategy cannot be justified statistically. For non-stationary data the formulas for the mid-line and the bands will no longer be the estimates of true statistical characteristics; thus, the Bollinger bands strategy loses its theoretical statistical sense and cannot be considered correct. We cannot apply the estimated (optimal) strategy parameters for other intervals, as these parameters vary over time and the estimates depend on past data.

Therefore, only stationary or local stationary patterns can be used in the Bollinger bands strategy applications.

Assumption 2: Symmetrical probability distribution
The form of the price probability distribution defines whether the bands are equally distanced from MA or not. In the case of equally distanced upper and lower bands, the distribution should be symmetrical around the price mean.

If the data is skew, the bands that satisfy *K%* are
$b_U = m + k_1\sigma, b_L = m - k_2\sigma$, with $k_1 \neq k_2 > 0$.

The examples of such non-symmetrical bands are the Bomar bands, or Marc Chaiken innovation (Bollinger, 2002).

So we see that the classical Bollinger bands should only be applied to data, which has a symmetrical probability distribution.

Assumption 3: Known data probability distribution function
Although data stationarity and a symmetrical probability distribution are sufficient assumptions to justify using of the Bollinger bands, knowing the probability distribution significantly facilitates the search for the optimal parameter *k*. Otherwise the estimation of the optimal value *k*, which guarantees that *K%* of data is inside the bands, becomes complicate and computer-intensive procedure. Mathematically defined probability density functions make it much easier to calculate the bands, as no preliminary probability distribution function estimation is needed.

To summarize
In order to apply Bollinger bands the data should be stationary, at least locally; it should have a symmetric distribution if we want equally distanced bands; otherwise we apply Bomar (not Bollinger) bands. Finally, knowing the data distribution facilitates the optimal band calculations, especially if the distribution function is well-defined.

2 Data transformed bands

As have been discussed earlier, the analysis of the price (residuals) distribution and interval around mean that contains *K*% of data is statistically more consistent method to search for the optimal *k*-parameter in the Bollinger bands than trading strategy simulation. An advantage of this approach is that by taking into account the distribution properties of the data we incorporate some "hidden" information valuable for all data samples. At the same time we know that in most cases historic price (or residuals) data have neither symmetrical nor well-defined known distribution function. Thus, classical Bollinger bands might not lead to the maximum trading profits.

Mentioned earlier the Bomar bands could be the important generalization of the Bollinger bands for the data with different distributional properties. Unfortunately, we could not find the bibliographical sources, which explain how this method is applied. We know, however, that this method is computationally intensive and, thus not popular (Bollinger, 2002). We might only guess that they either applied optimization procedure directly to empirical histogram or cumulative distribution function (cdf)[16] or first fit some density function to the empirical histogram and then derived the solution by the optimization of this function. Both solutions needs additional data analysis and cannot be held on the routine basis.

We propose different approach to the optimal band definition, which is simple and can be quickly implemented, i.e. it can be used on continuous time basis.

[16] This optimisation problem might not have the single solution and is complicate to implement.

2.1 Data transformed bands: algorithm description

The construction of the *data transformed bands* (further *DT bands*) involves the statistical method of data transformation (anamorphosis) with one distributional characteristic into the data with known (usually normal) distribution.

In the case of the known distribution of the non-normal variable $Z(x)$, the transformation of the $Z(x)$ into normal $Y(x)$ is performed by using the equation:

$$F(z) = G(y) \Leftrightarrow Y = G^{-1}(F(Z)), \qquad \text{(II.2.1)}$$

where $F(z)$ is the known CDF of the $Z(x)$; $G(y)$ is the standard normal CDF.

If the distribution $F(z)$ is not known, it can be estimated empirically and then the transform function (II.2.1) can be applied. Further in the applications we use the Kaplan-Meier estimate of the empirical cumulative distribution function. The reader can be referred to Cox, Oakes (1984).

The transformation procedure is presented in Figure 2.6. We can see the transformation of the observation of the lognormal random variable z^* into the respective value y^* of the standard normal variable. As we can see: $F(z^*) = G(y^*)$.

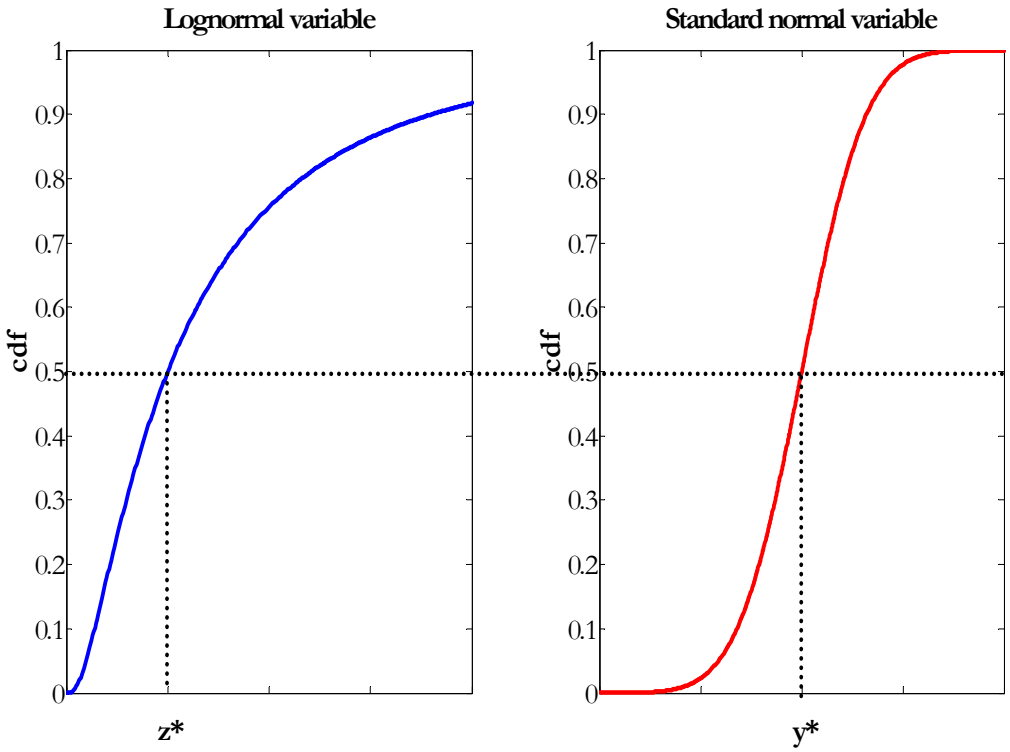


Figure 2.6. Transformation of the lognormal data (left) into standard normal variable (right)

The equation (II.2.1) implies that the following proposition 2.1 holds true.

Proposition 2.1

Let Z some random variable with CDF $F(z)$ and Y is the standard normal variable with CDF $G(y)$, such that $Y = G^{-1}(F(Z))$. Then for any z_1 and z_2, the observations of variable Z and their respective transformed values y_1 and y_2 the following equation hold true:

$$K = \Pr(z_1 \le Z \le z_2) = \Pr(y_1 \le Y \le y_2). \quad \text{(II.2.2)}$$

Proof: $K = \Pr(z_1 \le Z \le z_2) = F(z_2) - F(z_1) = G(y_2) - G(y_1) = \Pr(y_1 \le Y \le y_2)$.

The application of the method is straightforward. First, we transform the empiric data into normal standard variable according to some previously calibrated transform function. The transformed variable has symmetrical normal distribution, for which the optimal intervals around zero-mean that contains *K*% of data can be easily obtained (upper and lower bands are equally distanced from the moving average). Afterwards these (normal) bands are transformed back using the calibrated transformation function. The obtained bands, further called *data-transformed bands* (DT bands), are the bands that can be used with the original empiric data. Note that according to the proposition 2.1 the bands will contain the same percentage of data *K*%.

The choice of the variable to which the transformation should be applied is important. The bands should be built around the middle line, which in Bollinger case is represented by the SMA. We propose to keep the same middle line because it's an unbiased mean estimator, but also because we want to keep the same basis for the comparison with the Bollinger approach. In this case we can evaluate the effect of the substitution of the bands calculated on the basis of Bollinger method by the DT bands[17]. As the result, at each moment of time we have to construct the bands around SMA_t. In order to estimate the bands values around SMA_t, we need to analyze the distribution properties of the residuals $X_i = P_i - SMA_t$ and to calibrate the transform function according to this distribution. On one hand such definition of the residuals allows us to keep in line with the Bollinger estimation procedure: these residuals used for the estimation of the price volatility. On the other hand, we will see further that the usage of these residuals will make our bands less sensitive to the insignificant fluctuations in the mid-line curve.

The construction of the DT bands is made in the following steps:

1. **Calibration of the transformation function.** The empirical data $X_i = P_i - SMA_t$, $i \in [t-n+1; t]$ (n is the length of SMA) with CDF $F(x)$ is transformed into normal standard random variable by transform function φ:
$$X' = \varphi(X),\ X' \sim N(0,1), \quad \text{(II.2.3)}$$
such that $F(x) = G(\varphi(x))$, where $G(.)$ is the CDF of the normal standard variable.
2. **Search for optimal *k*-parameter.** The coefficient *k* defines the bands for the normal standard variable ($\sigma = 1$), which corresponds to the interval $[-k; k]$ that contains *K*% of transformed normal data.
3. **Backward bands transformation.** The bands for empirical data are obtained by the backward transformation of their values by the indirect function, estimated in (II.2.3):
$$\tilde{B}_U = \varphi^{-1}(k),\ \tilde{B}_L = \varphi^{-1}(-k) \quad \text{(II.2.4)}$$
4. The bands are positioned around the SMA as $SMA_t + \tilde{B}_{U,t}$ and $SMA_t + \tilde{B}_{L,t}$ [18] and together with price and SMA curves define the entry/exit points for the DT bands trading strategy.

[17] This is the reason why we do not substitute SMA with the kriged moving average analysed in part 1.

[18] The lower band is created also as a sum of SMA and estimated band, as lower band has a negative sign.

The important assumption under which the DT bands can be constructed on continuous basis and be statistically justified is data local stationarity.

2.2 Data transformed bands: empirical observations

The example of the DT bands for Bund instrument is presented in Figure 2.7. As we can see from the graph, the form of the DT bands differs from the classical Bollinger bands.

1. Contrary to the classical Bollinger bands that are almost strictly decreasing or increasing the DT bands show the presence of "stair"-type patterns (further "steps").
2. During local trends, prices approach and, in many cases, cross both Bollinger and DT bands, i.e. both bands send false signals during trending markets. Moreover, the DT bands may send even more false signals than the Bollinger bands.

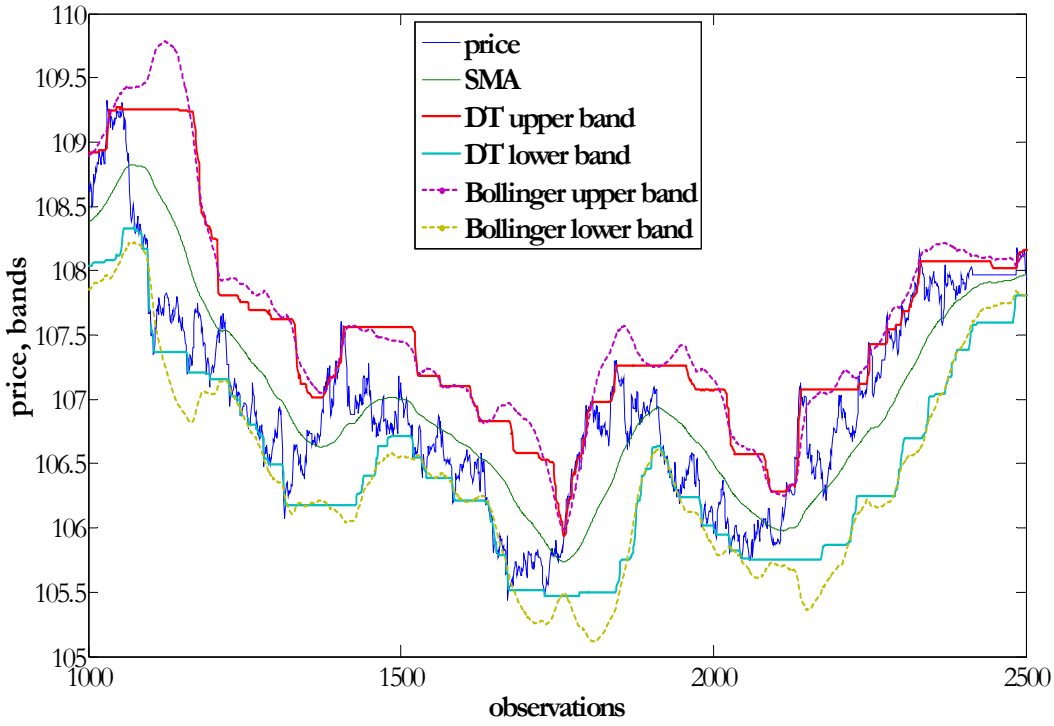


Figure 2.7. Bund quotes (30/7/2003-7/12/2006, frequency – 30 minutes, observations 1000-2500): DT bands (k=2.1) and Bollinger bands (k=2.1).

The specific "stair"-shape of the bands is worth of separate discussion. The bands are created as the sum of non-constant SMA and transformed (residuals) bands. In order to achieve the constant stair-patterns on the price band, the residuals bands should change proportionally to the change in SMA, but with the opposite sign. The explanation of the phenomena lies in the construction of the residuals $X_i = P_i - SMA_t$. In order to estimate the band at point $t+1$ we will need to calibrate the new transform function on the sub-sample $\{X_i\}_{t-n\le i\le t+1}$. Let say the SMA increased by $\delta: SMA_{t+1} = SMA_t + \delta$. Then although the sub-sample of prices $\{P_i\}_{t-n\le i\le t+1}$ does not change significantly comparing to the sub-sample $\{P_i\}_{t-n+1\le i\le t}$ used in the calculations of band at point t, the new sub-sample $\{X_i\}_{t-n\le i\le t+1}$ will differ significantly due to the change in the SMA used for their calculations. In particular, all "old" values X_i, $i \in [t-n;t]$ will be decreased

by δ. If the new price at moment $t+1$ does not bring significant innovation to the sample, we end up with approximately the same distribution function, but different mean: the CDF curve of the non-transformed residuals will be shifted to the left by δ. This means that the new band that will correspond to the same $\pm k$ argument for standard normal distribution will decrease by δ. Consequently, the bands will move from "stair" to "stair" only if the presence of the "new" prices in the sample is so important that it will change the form of the distributional function.

Due to the particular form of the DT bands several conclusions can be made with respect to the price movements around DT bands. The following observations were made from the large sample of Bund prices and were subsequently confirmed for other instruments:

1. Jumps in the DT bands from one step (stair) to another correspond to large price changes due to the explanation given above. Let us define trend "phases" as parts of the price curve, which correspond to the step in the DT band. We have observed that the trend normally slows down after at least three such "phases", i.e. the probability of trend reversing increases. The total length of these three "phases" is likely to be important. If the steps are very short more than 3 steps might needed to complete the trend.
2. If prices largely cross the steps while the bands show increasing/decreasing patterns, the trend is likely to continue. Obviously, several large price increments would change local distribution of the residuals and, thus, take the bands to the new (step) level.
3. We have observed also that in some cases the prices rebound from a step: exists some small $\varepsilon > 0$, such that:

   $$\left| P_t - \left(MA_t + \tilde{B}_U \right) \right| \le \varepsilon \quad \text{(the price rebound from upper band step)}$$

   $$\left| P_t - \left(MA_t + \tilde{B}_L \right) \right| \le \varepsilon \quad \text{(the price rebound from lower band step)}$$

   Then if the prices rebound from long (the same) horizontal part of the DT band (step) the trend is likely to reverse.

2.3 Data transformed bands: other technical rules definition

The empirical observations made in the previous section have potentially more applications than just definition of the DT bands strategies. These observations might help to define two popular but very subjective technical rules: Elliot Waves and Support/Resistance patterns. Short descriptions of these instruments are presented in Boxes 2.1 and 2.2 respectively.

Box 2.1
Elliot Waves

Elliot Waves are extremely popular in Technical Analysis, but, at the same time, are quite subjective and complicated to implement. The first reference to the Elliot waves appeared in 1938, in "The Wave Principle" by Charles J. Collins. The idea was finalized in 1946 by R.N. Elliot in "Nature's Law – The Secret of the Universe" and was republished by A. Hamilton Bolton in 1953 in "Elliot Wave Supplement" to the *Bank Credit Analyst*. The theory states that the market follows a repetitive rhythm of a particular complete cycle pattern, presented in Figure Box2.1.

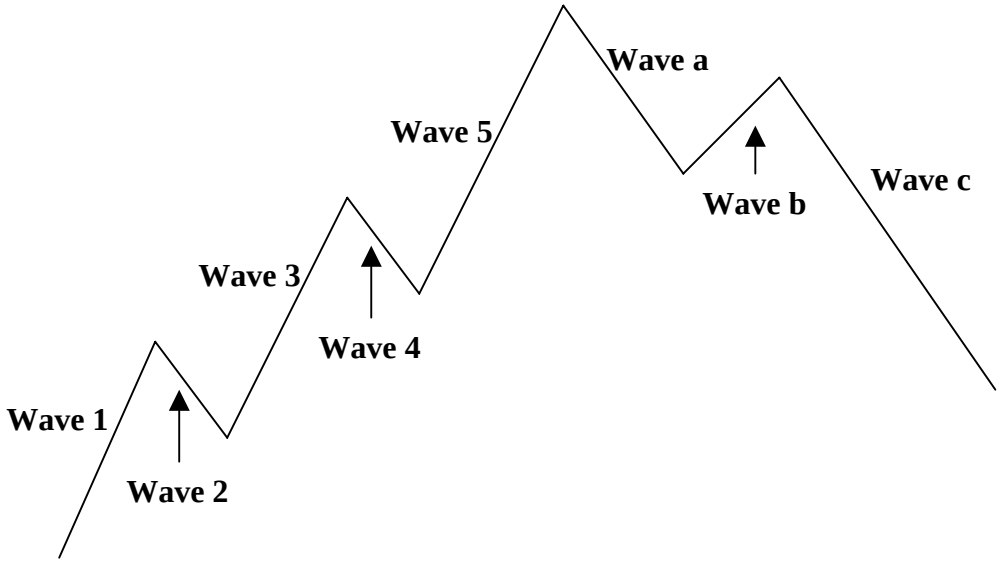


Figure Box2.1. The basic pattern of Elliot Waves

Each basic pattern consists of 3 upward waves and 2 downward waves, followed by the three-wave correction phase. The upward waves (Wave 1, Wave 3 and Wave 5) are the *impulse waves*, and the downward waves (Wave 2 and Wave 4) are *corrective waves*. The three-wave correction phase starts after the 5-waves pattern is completed; this phase is identified with the Waves a, b, c.

Elliot also identified the degrees of the trend, or the time-horizon of the trend, ranging from the very long-term trend of the two hundred years to the short-term trend of a few hours. However, no matter how long is the trend, it is represented by the same basic pattern given in Figure Box2.1. In fact, each long-term basic pattern is split into the basic patterns of shorter trends, i.e. for example each Wave 1-5 can be represented by the similar, but shorter 8-waves basic patterns, etc.

There are many strategies and patterns related to the Elliot Waves theory (for more details see Murphy, 1999). We will not present them here, as they are outside of scope of the paper. However, it is clear that if we manage to identify the 5 waves pattern, the trend reversion (or correction) of 3 waves can be predicted; and afterwards, the new reversion of the five-waves cycle can be expected. The biggest problem is to identify these patterns in real price data samples, as it is obvious that the schematic representation of the basic cycle from Figure Box2.1 does not exist, due to the data noise.

Source: Murphy, John J. "Technical Analysis of the Financial Markets", *New York Institute of Finance*, 1999.

Box 2.2

The Support and Resistance Technical Rules

Price patterns can be considered as a series of local peaks and troughs (local maximums and minimums). These peaks and troughs can be identified with resistance and support levels, two well-known concepts of the technical analysis. See Picture Box2.2.

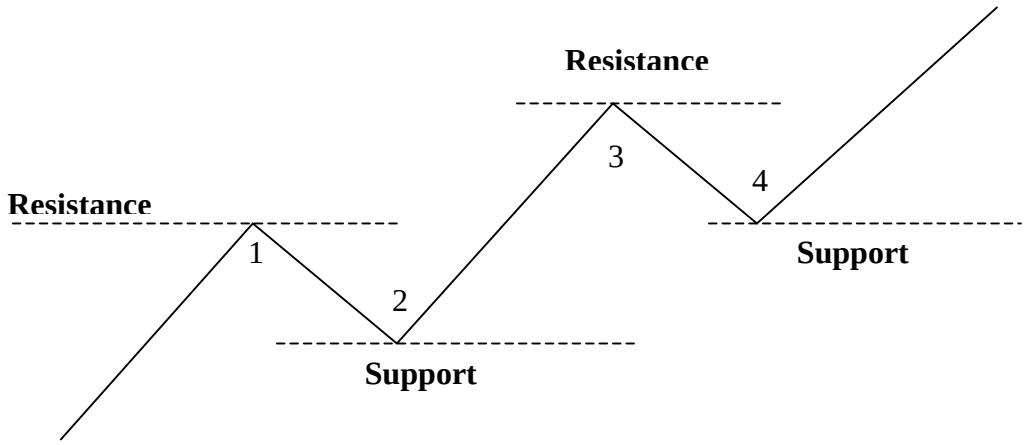


Figure Box2.2. Schematic representation of the support and resistance levels.

In the Figure, points 1 & 3 represent local maximum or resistance levels (R1 & R3), while points 2 & 4 represent local minimum or support levels (S2 & S4).

We now present some of the technical trading rules based on the Support/Resistance:

1. The upward trend is confirmed if each following support and resistance levels are strictly higher than the previous ones (R3>R1, S4>S1), and likewise, for downward trends: R3<R1, S4<S1.
2. The trend reversal could happen when a support (resistance) hits the same level as (at least) one of the previous support (resistance) levels: for example, S2=S4 for change from downward to upward trend, and R3=R1 for change from upward to downward trend.

As long as these Support/Resistance levels are identified the trading rules can be applied.

Source: Murphy, John J. "Technical Analysis of the Financial Markets", *New York Institute of Finance*, 1999.

Taking into account the framework of the Markov times presented in the General Introduction, neither Eliot waves nor Support-Resistance trading rules are Markov times, which complicates significantly their implementation, i.e. in order to identify the patterns we need to anticipate future. But we believe that DT bands can help to define these rules more easily and in a less subjective way. In particular, we believe that first five-wave part of the basic Elliot Waves pattern corresponds to our observation (1) in previous subchapter about the trend slow down or inversing after three steps of DT bands. In particular, three steps coincides with the waves 1, 3, and 5, while waves 2 and 4 are not reflected in the band patterns as the prices innovations, corresponding to these waves, are not significant to change the distribution properties of the residuals.

Observation (2) about the price bouncing off the DT bands step before trend reversion might correspond to the Support/ Resistance trading rule about the price bouncing from the same levels of the support/resistance (see rule #2, Box 1)[19]. In fact, the same step is serving the support/resistance level for different prices.

[19] In order to talk about the consistency of the rule definition, the price should bounce at least twice from the same band. Further, we consider only one bouncing as sufficient indicator of the trend reversal, due to very low number of

As the result, DT bands might give us the possibility to define the Elliot and "Support & Resistance" rules without anticipating the future, simply by using the statistical characteristics of the data. The Bollinger (2002) states that Bollinger bands can be used to define such technical analysis patterns as "W bottoms" or "M tops", head-and-shoulders. Though to our point of view these definitions are still complicate to program, as they are quite subjective. Instead, the DT bands give straight criteria to define Elliot waves or Support/Resistance levels.

However, the statements that particular DT bands patterns can be used to define these technical indicators need to be proven by some tests or experiments. For example, some pattern recognition techniques might be used to identify the Elliot or Support/Resistance patterns, which afterwards can be compared with the DT bands results to show that these bands can recognize the same patterns. Taking into account the scale of the analysis, this problem is worth of separate discussion. In addition, this problem is out of the scope of our paper as our objective is the amelioration of the trading bands not the definition of the Elliot or Support/Resistance instruments.

At the same time, without in-depth analysis, we cannot state that we are using precisely Elliot waves theory and Support/Resistance levels to confirm the trend reversion, while constructing confirmed DT bands strategies. Thus, in rest of the text we will refer to these conditions for price trend reversion as trading signals, generated by DT bands, further called "Elliot" and "Support/Resistance" signals:

1. **"Elliot" signals**: if DT bands form 3 consecutive steps of total length Sl_E, then take:
   - short position, if these 3 steps (formed by upper DT band) moved upward and respective residuals $R_t = P_t - SMA_t > 0$ (prices are above the SMA curve).
   - long position, if these 3 steps (formed by lower DT band) moved downward and respective residuals $R_t = P_t - SMA_t < 0$ (prices are below the SMA curve).
2. **"Support/Resistance" signals**: if at least one price observations bounce from the same step, then take:
   - short position, if respective residuals $R_t = P_t - SMA_t > 0$ (the step is formed by upper DT band)
   - long position, if respective residuals $R_t = P_t - SMA_t < 0$ (the step is formed by lower DT band)

Consequently, the DT bands has the following advantages over the Bollinger bands:

1. DT bands provide much clearer theoretical statistical justification of the bands usage.
2. DT bands can define the following technical trading rules that are not Markov times, and make it possible to implement these rules as algorithms:
   a. Elliot waves theory
   b. Support/Resistance concept

However, in order to justify their usage we need to show that DT bands improve existing classical trading strategies, based on Bollinger bands. The Chapters 3 and 4 present the framework and the results of such analysis.

---

trades, which is generated by higher restriction on the parameter. We consider this signal only together with "Eliot" signal as an additional signal confirmation.

3 Strategies descriptions

As we have discussed above, the application of the Bollinger bands strategy is statistically justified only in the cases of known symmetrical distribution. The DT bands in their turn can be applied to more general data case. The question is whether the usage of DT bands at the place of the Bollinger bands improves the trading results. For this purpose we should back-test (simulate) Bollinger and DT bands trading strategies on the historic data and compare their outcomes. The following chapter presents the strategies and their algorithms. It also discusses the choice of parameters values used for the trading simulations, as well as the choice of the indicators used for the analysis of the trading outcomes.

3.1 Strategies

The strategies used in the analysis of DT bands are based on the traditional Bollinger bands strategies (simple basic contrarian and confirmed contrarian), as well as confirmed strategies based on the signals generated solely by the DT bands. As the result the following groups of the strategies are considered in the paper:

1. **Basic strategy** (or **Benchmark strategy**), based on
    - Bollinger bands
    - DT bands
2. **Confirmed strategies** (basic strategy with confirmative signals):
    2.1. Confirmation by the momentum indicator of the oversold/overbought signals, generated by:
        - Bollinger bands
        - DT bands
    2.2. Confirmation by the "Eliot" signal of the oversold/overbought signals generated by:
        - DT bands
    2.3. Confirmation by the "Elliot" and "Support/Resistance" signals of the oversold/overbought signals generated by:
        - DT bands

Note that the last two confirmed strategies are constructed only for DT bands, as only this type of bands generates the "Elliot" and "Support/resistance" signals.

The more detail presentation of the strategies algorithm is presented further. Notation used for the strategies definition can be found in Table 2.2.

Table 2.2
Notation for the strategies parameters

P_t	price at moment t
n	interval length, on which the SMA and moving volatility indicator are calculated
MA_t $MA_t = \mathrm{E}(P) = \frac{\sum_{i=t-n+1}^{t} P_i}{n} = m$	SMA value at moment t; calculated on the interval $[t-n+1;t]$.
σ_t $\sigma_t^2 = \mathrm{E}\left[(P-\mathrm{E}(P))^2\right] = \frac{\sum_{i=t-n+1}^{t}(P_i-m)^2}{n}$	price standard deviation at moment t; calculated on the interval $[t-n+1;t]$.
k_B, k_{DT}	parameters of the Bollinger and DT bands respectively, which defines the distance between the middle line and bands.
l_{DT}	size of the sub-sample, used for the transformation of raw data into normal data
$\tilde{B}_U$, $\tilde{B}_L$ $\tilde{B}_U = \varphi^{-1}(k_{DT})$, $\tilde{B}_L = \varphi^{-1}(-k_{DT})$	upper and lower DT bands, defined for the residuals $R_i = P_i - m$, $m = \frac{1}{l_{DT}} \sum_{i=t-l_{DT}+1}^{t} P_i$, $i \in [t-l_{DT}+1;t]$
b_U, b_L $b_{U,t} = SMA_t + \tilde{B}_{U,t}$ $b_{L,t} = SMA_t + \tilde{B}_{L,t}$	upper and lower DT bands, defined for price
Δ	lag parameter used for momentum calculation
M_t^{Δ}, $M_t^{\Delta} = P_t - P_{t-\Delta}$	price momentum at moment t
s_M	momentum threshold value, used for definition of the trend reversal and confirmation of the overbought/oversold signals
St_E	number of the consecutive steps, defined by DT bands
Sl_E	total length of the consecutive steps, defined by DT bands
T_E	tolerance level to define the step, i.e. the level at which two values can be considered equal: band observations form a step if $\lvert b_i - b_{i-k} \rvert \le T_E$
T_{SR}	tolerance level to define "price touching the band": price touches the band if $\lvert P_i - b_i \rvert \le T_{SR}$
q	number of price observations that "touched" the bands step

1. Benchmark (basic) strategy

A basic strategy is a contrarian strategy, where the bands define the thresholds for overbought/oversold instruments. Breaching these thresholds (upper and lower bands) defines entry points for the strategy (short and long positions respectively). The position is closed when the price returns to its mean value, i.e. breaches MA curve.

1.1. Basic strategy, based on Bollinger bands

Entry position:

if $\left|P_t - MA_t\right| > k\sigma_t$

Long position (L): if $\left(P_t - MA_t\right) < 0$

Short position (S): if $\left(P_t - MA_t\right) > 0$

Exit position:

if $\left(P_t - MA_t\right)\left(P_{t-1} - MA_{t-1}\right) < 0$

1.2. Basic strategy, based on DT bands

Entry position:

if $\left(P_t - MA_t\right) > \tilde{B}_t^U$ or $\left(P_t - MA_t\right) < \tilde{B}_t^L$

Long position (L): if $\left(P_t - MA_t\right) < 0$

Short position (S): if $\left(P_t - MA_t\right) > 0$

Exit position:

if $\left(P_t - MA_t\right)\left(P_{t-1} - MA_{t-1}\right) < 0$

2. Confirmed strategy with the momentum confirmation signal

A confirmed strategy is a contrarian strategy, where the bands together with trend-reversal signals define the overbought/oversold instruments. Breaching these thresholds (upper and lower bands) together with the signal that the trend will reverse its direction defines entry points for the strategy (short and long positions respectively). The signal for trend reversal is the exceeding (in absolute terms[20]) by the price momentum M_t^{Δ} some threshold value s_M. The position is closed when the price returns to its mean value, i.e. breaches MA curve.

2.1. Confirmed strategy, based on Bollinger bands

Entry position:

if $\left|P_t - MA_t\right| > k\sigma_t$ and $\left|M_t^{\Delta}\right| > s_M$

Long position (L): if $\left(P_t - MA_t\right) < 0$

Short position (S): if $\left(P_t - MA_t\right) > 0$

Exit position:

if $\left(P_t - MA_t\right)\left(P_{t-1} - MA_{t-1}\right) < 0$

[20] In order to define whether a trend reverses its direction from increasing to decreasing or from decreasing to increasing it is important to know what threshold (upper or lower is breached). However, as long as we consider this signal in combination with the band breaching signals, it is sufficient to know the sign of the residuals $\left(P_t - MA_t\right)$. The breaching of the upper (lower) band from below (above) is possible only during increasing (decreasing) trend pattern; thus, only the fact of breaching the threshold by momentum is important to take the right position.

2.2. Confirmed strategy, based on DT bands

Entry position:

if $\left|M_t^{\Delta}\right| > s_M$ and $\left(P_t - MA_t\right) > \tilde{B}_t^U$ or $\left(P_t - MA_t\right) < \tilde{B}_t^L$

Long position (L): if $\left(P_t - MA_t\right) < 0$

Short position (S): if $\left(P_t - MA_t\right) > 0$

Exit position:

if $\left(P_t - MA_t\right)\left(P_{t-1} - MA_{t-1}\right) < 0$

3. Confirmed strategy with the "Elliot" confirmation signal

A confirmed strategy is a contrarian strategy, where the bands together with trend-reversal signals define the overbought/oversold instruments. Breaching these thresholds (upper and lower bands) together with the signal that the trend will reverse its direction defines entry points for the strategy (short and long positions respectively). The signal for trend reversal is the completion of the first five Elliot waves. The DT bands define these waves as a certain number of the consecutive steps St_E (in the same direction[21]) of the total length Sl_E. Therefore, before making any decision the strategy should estimate how many steps ($STEPS_t$) and of what length ($LGTH_t$) that precede the moment of decision-making. A step is formed if at least two consecutive price bands values coincide: $\left|b_i - b_{i-k}\right| \le T_E$. The position is closed when the price returns to its mean value, i.e. breaches MA curve.

3.1. Confirmed strategy, based on DT bands

Entry position:

if $[\left(P_t - MA_t\right) > \tilde{B}_t^U$ or $\left(P_t - MA_t\right) < \tilde{B}_t^L]$ and $[STEPS_t \ge St_E$ and $LGTH_t \ge Sl_E]$

Long position (L): if $\left(P_t - MA_t\right) < 0$

Short position (S): if $\left(P_t - MA_t\right) > 0$

Exit position:

if $\left(P_t - MA_t\right)\left(P_{t-1} - MA_{t-1}\right) < 0$

4. Confirmed strategy with the "Elliot" and "Support/Resistance" confirmation signals

A confirmed strategy is a contrarian strategy, where the bands together with trend-reversal signals define the overbought/oversold instruments. Breaching these thresholds (upper and lower bands) together with the signal that the trend will reverse its direction defines entry points for the strategy (short and long positions respectively). The signal for trend reversal is the completion of the first five Elliot waves and rebounding of the price from the same Support/Resistance level. The DT bands define these waves as a certain number of the consecutive steps St_E of the total length Sl_E. Therefore, before making any decision the strategy should define how many steps ($STEPS_t$) and of what length ($LGTH_t$) that precede the moment of decision-making. A step is formed if at least two price bands have the same value ($\left|b_i - b_{i-k}\right| \le T_E$). The Support/Resistance level is defined by the step from which q price observations rebounded. The signal is observed if

[21] Depending on whether we consider increasing (prices are above MA curve) or decreasing (prices are below MA curve) price patterns, each consecutive step should be higher or lower than the previous one. Otherwise, the Elliot wave pattern is not confirmed.

the number of price observations $TOUCH_t$ that touched the same step ($|P_i - b_i| \le T_{SR}$) is larger than q. The position is closed when the price returns to its mean value, i.e. breaches MA curve.

4.1. Confirmed strategy, based on DT bands

Entry position:

if

$[(P_t - MA_t) > \tilde{B}_t^U$ or $(P_t - MA_t) < \tilde{B}_t^L]$ and $[STEPS_t \ge St_E$ and $LGTH_t \ge Sl_E]$ and $TOUCH_t \ge q$

Long position (L): if $(P_t - MA_t) < 0$

Short position (S): if $(P_t - MA_t) > 0$

Exit position:

if $(P_t - MA_t)(P_{t-1} - MA_{t-1}) < 0$

3.2 Choice of the strategy parameters

As can be seen the strategies presented in Ch.3.1 have an important number of parameters. For example, the simplest basic strategies can be represented as following (see Table 2.2 for notation):

- for **Bollinger basic strategy**:

$$P\&L = f_1(n, k_B) \qquad \text{(II.3.1)}$$

- for **DT basic strategy**:

$$P\&L = g_1(n, l_{DT}, k_{DT}) \qquad \text{(II.3.2)}$$

For the confirmed strategies the number of parameters increases even more with additional parameters for each confirmation indicators. For example, the strategies confirmed by the momentum (see Table 2.2 for notation):

- for **confirmed by momentum strategy, based on Bollinger bands:**

$$P\&L = f_2(n, k_B, \Delta, s_M) \qquad \text{(II.3.3)}$$

- for **confirmed by momentum strategy, based on DT bands:**

$$P\&L = g_2(n, l_{DT}, k_{DT}, \Delta, s_M) \qquad \text{(II.3.4)}$$

For the **confirmed by "Elliot" and "Support/Resistance" signals strategy**, the parameters of the strategies based on DT bands are presented as following (see Table 2.2 for notation):

$$P\&L = g_3(n, l_{DT}, k_{DT}, \Delta, s_M, St_E, Sl_E, T_E, T_{SR}, q) \qquad \text{(II.3.5)}$$

The search of the optimal parameters can involve maximization of the P&L with respect to all parameters. However, as we can see for some more advanced strategy definitions the number of parameters goes to ten, which makes the optimization procedure extremely complicated. Taking into account the fact, that it is virtually impossible to produce the same parameters, which will be optimal for all instruments and markets, the procedure will be multiplied by the number of instruments, for which the analysis should be performed.

We propose less ambitious approach to simplify both analysis and results presentation, taking into account that our main objective is to define whether DT bands improve the trading outcomes comparing to the Bollinger bands strategies. Therefore, we propose to predefine some of the parameters that might have less importance. As the result, we want to show that for some instruments under other equal conditions the DT bands strategies work better than the Bollinger classical technical strategies, presented in finance literature. Our goal is to present the DT bands

as strategy building blocks and show that they are not a mere products of the theoretical amelioration of the existing bands, but can produce some superior trading results.

The parameters, which define the Bollinger/DT bands and are common for all tested strategies need a very careful definition. Ideally the strategies should be to analyzed for different lengths of the SMA and k parameters. However, the main problem with such approach is that these parameters are related to each other. Scaling parameter k for the Bollinger bands depend on the distribution of the instruments residuals; it is chosen at the level to guarantee that approximately 90% of the raw data is within the bands. At the same time the length of the SMA used to define the residuals has direct impact on the data distribution. As the result these parameters have to be chosen in couples.

Estimation of the MA, volatility and calibration of the data transform function is made on the same sub-samples: $l_{DT} = n$. The length of the MA is chosen at two levels (medium and long), in order to analyze the impact of different MA at the strategy outcomes. The values of each type of the MA lengths will depend on the data frequency: for frequency 30 minutes - $n_1 = 50, n_2 = 115$; for frequency 1 hour - $n_1 = 20, n_2 = 50$. We decided not to change the frequency of the data[22], taking into account that nowadays traders work with even higher frequency than 30 minutes, while the MAs of the lengths of 1-2 days (which coincide with our n values) are quite common for short-term traders. The chosen SMA length also coincides with the "default" lengths ($n = 20,50$ observations) according to the Bollinger bands concept. It should be noted however, that the results for the short MA length and interval used for data transformation ($n_1 = 20$), should be analyzed with cautious, as it is not statistically correct to derive some conclusions from such small samples.

The longer SMA $n = 115$ (although not recommended by the Bollinger theory) was chosen for two reasons. On one hand, comparison with longer SMA helps to define whether the "default" values are the optimal one, i.e. brings more profits than the other parameters. On the other hand, the longer SMA (and sample) allows more data variability to calibrate the transform function (the same data samples are used for both calculations). Low data variability will have impact on the choice of the value for parameter k: the larger k values are likely to be transformed into $\pm\infty$ values for DT bands. For example, this is the case for the DT bands with the parameters $n = 50$ and $k = 2.1$. The simulation of the trading activity is then impossible.

The choice of the Bollinger bands scaling parameters k is given in Table 2.3.

Table 2.3
Choice of the k parameter value

Length of SMA, observations	Bollinger bands parameter	DT bands parameter
20	2	1.65
50	2.1	1.65
115	2.1	2.1

As we stated already the k value implies that 90% of data lays within the bands $[-k;k]$ (Bollinger, 2002). For the transformed standard normal variable this value ($k^* = 1.65$) is the same independently of the MA lengths (the transformed variable has always the same distribution). For non-normal non-transformed data this value will depend on the variability of the residuals: for

[22] As recommended by Bollinger (2002).

example, the shorter MA leads to less variable residuals, and thus smaller k values. Thus, the values for the k parameter for Bollinger bands are chosen at the default levels for the SMA lengths $n_1 = 20, n_2 = 50$ (see Table 2.1), while we keep this parameter for the DT bands at the same level $k^* = 1.65$ for both SMA lengths. As the result, for the first two cases we will see whether the k value that guarantees the same share of data within the bands gives the same trading results. For longer SMA ($n_3 = 115$) we choose to test whether the same k for Bollinger and DT bands gives comparable trading results for different bands types. We keep the same value $k_3 = 2.1$ as no default values are defined for this SMA length.

For the strategies confirmed by the "Elliot" and "Support/Resistance" signals we should define whether bands values in some close neighborhood create a step. For this purpose we define the tolerance level at level $T_E = 0.0001$ that allows considering two bands values to be the same: $|b_i - b_{i-k}| \le T_E \Rightarrow b_i = b_{i-k}$. The benchmark for total steps number $St_E = 3$ is chosen from empirical observations at the level that defines the first five Elliot waves (see Ch.2.3 for explanation).

For the "Support/Resistance" signal two parameters should be predefined. We choose to observe at least one touch of the Support/Resistance level to accept the signal of trend reversal, i.e. the number of "touches" should be larger or equal to 1 ($q \ge 1$[23]). The choice of the tolerance level T_{SR} that defines whether the band was touched ($|P_i - b_i| \le T_{SR} \Rightarrow P_i = b_i$) depends on the instrument tick. The optimal T_{SR} is found from the trade simulation results as the one that maximize the trading profits.

As for the strategies confirmed by momentum, the optimal momentum parameters (lag and threshold) are defined by the maximum performance (surface), simulated for different lags (Δ) and thresholds (s_M). These parameters are optimized on case-by-case basis. We have introduced two vectors: (1) vector of momentum lags [50; 80; 115; 150; 180] and (2) vector of their threshold levels. The choice of the latest depends on the ranges of the instrument momentum values. Simulation of the trading activities for different combination of these parameters creates the surface (table) of the profit/losses (P&L). Taking into account that the surfaces for Bollinger and DT bands might not coincide, we report two outcomes for DT bands: (1) global maximum for DT bands P&L surface; and (2) the P&L that corresponds to momentum parameters, defined as optimal for the Bollinger P&L surface.

Finally, we assume 0-transaction costs and slippage for the simulation framework.

3.3 Analysis of the outcomes

For each strategy type the following trading outcomes are analyzed:
(1) profit/losses (P&L);
(2) total number of trades;
(3) profit/losses (P&L) per trade.

We also consider the form of the cumulative P&L path (as a function of time). Note that the strategy, which ends up in large profits at the end of some (long) period of time, but has shown the losses during the period itself is less interesting than the strategy, which ends up at some

[23] Our experience shows that more strict constraint might leads to the absence of trades.

lower profits, but has shown stable or increasing profitability during the whole period. This characteristic is especially important for the dynamic trading, such as the intra-day trading. For such trading it is more important to have the opportunity to end up strategy application with positive outcome at any moment of time rather than wait for some drastically larger profits during the long-term periods. Therefore, we also analyze the P&L paths together with the end-of-period performance values. The increasing P&L patterns have higher priority for us than the absolute end-of-period performance values (under condition that the strategy has positive outcome).

4 Trading outcomes for different instruments

The trading strategies were simulated for four different instruments that represent different markets: (1) Bund; (2) DAX; (3) Brent; (4) X instrument. The results are summarized further.

4.1 Bund

Bund data represents the price quotes for the period of 30/7/2003-7/12/2006 (see Figure 2.8). The quotes were sampled at the 30 minutes frequency.

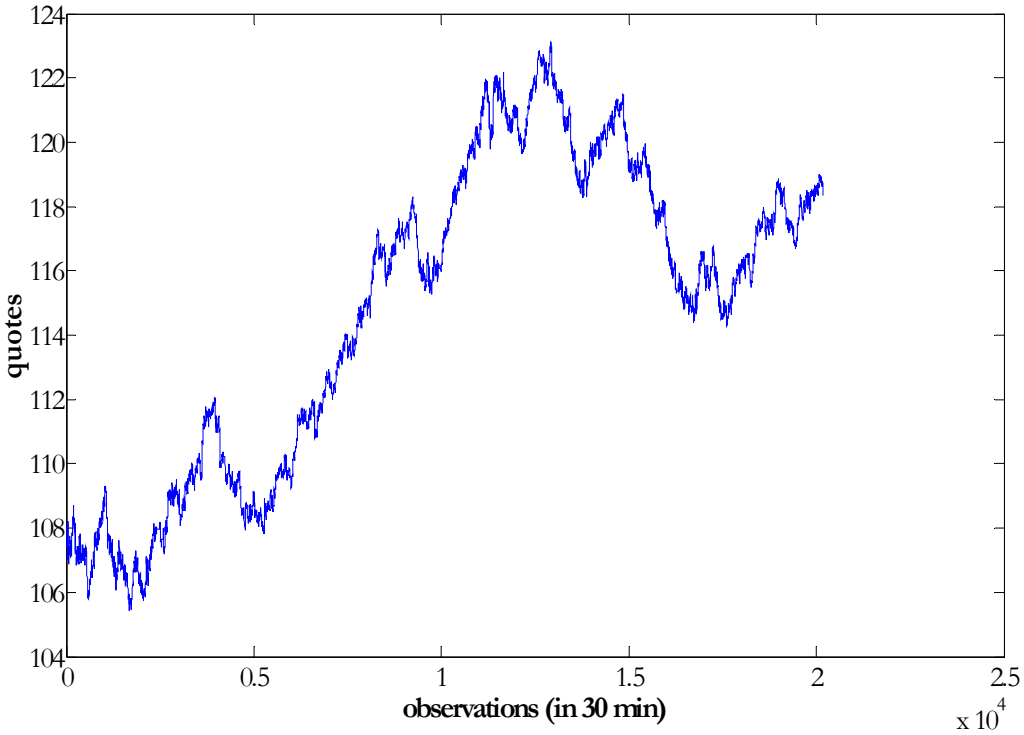


Figure 2.8. Bund quotes for period 30/7/2003-7/12/2006 (frequency – 30 minutes)

Table 2.4 and 2.5 present respectively the optimal (maximal) outcomes for different strategies application and their corresponding number of trades. Both tables present also the sub-optimal results for the confirmed by momentum strategy, based on the DT bands, which corresponds to the momentum parameters optimal for the Bollinger bands strategy. The comparison of these absolute results shows that the largest profits are produced by the strategy confirmed by momentum and based on the DT bands for the SMA length of 50 observations (9880€).

Moreover, this type of DT strategy shows superior results for both SMA lengths (50 and 115 observations).

The shorter SMA (50 observations) works better than the longer SMA (for the first 3 types of the strategies). Only the fourth DT bands strategy, confirmed by "Elliot" and "Support & Resistance" signals works worse for SMA length of 50 observations, and slightly improves for longer SMA.

Unconfirmed strategies (strategy 1), based on either Bollinger or DT bands, are non-profitable, confirming the fact that breaching any type of the bands by price curve does not produce unanimous trading signals. These signals need a confirmation by other technical overbought/oversold indicators. The unconfirmed basic strategy, based on the DT bands, brings higher losses than the unconfirmed basic strategy, based on the Bollinger bands. This means that the DT bands by there own send more false signals than the Bollinger bands (more losses for more trades number); it was expected in subchapter 2.2 from the visual observations of the graphs.

The analysis of the trades number (Table 2.5) leads to the similar conclusions. The confirmation of the overbought/oversold signals reduces significantly the number of trades for both Bollinger and DT bands (by rejecting the false signals). What is even more interesting is that the largest profit for DT bands is achieved for significantly lower number of trades than for Bollinger bands, meaning much higher efficiency of the DT strategy per trade in comparison with the confirmed Bollinger bands strategy (44€/trade versus 22€/trade). The efficiency of other confirmed DT bands is lower than the efficiency of the confirmed Bollinger bands strategy.

Table 2.4
Bund (30/7/2003-7/12/2006, frequency – 30 minutes): Profit & Losses

Strategies	Length of the SMA			
	50		115	
	Bollinger	DT bands	Bollinger	DT bands
Strategy 1: Basic	-240	-1530	-3960	-4360
Strategy 2: Confirmed-Momentum	8300	9880 (6870*)	2200	3020 (-600*)
Strategy 3: Confirmed-"Elliot"	-	5400	-	800
Strategy 4: Confirmed-"Elliot & Support/Resistance"	-	2620	-	3740

* values that correspond to the momentum parameters – optimal for "Confirmed-Momentum" Bollinger bands strategy

Table 2.5
Bund (30/7/2003-7/12/2006, frequency – 30 minutes): Number of trades

Strategies	Length of the SMA			
	50		115	
	Bollinger	DT bands	Bollinger	DT bands
Strategy 1: Basic	665	901	295	375
Strategy 2: Confirmed-Momentum	386	224(538*)	28	50 (42*)
Strategy 3: Confirmed-"Elliot"	-	364	-	20
Strategy 4: Confirmed-"Elliot & Support/Resistance"	-	188	-	96

* values that correspond to the momentum parameters – optimal for "Confirmed-Momentum" Bollinger bands strategy

As we have discussed already, the end-of-period P&L does not gives the absolute answer, which strategy is better. The more important is a P&L path. Both basic strategies produce negative outcomes with non-monotone P&L paths (see Figures 2.9 and 2.10). We can also notice that both strategies get big losses after strong trend pattern (compare the Figure 2.9, 2.10 and 2.8).

This just proves the fact that the unconditional (non-confirmed) Bollinger or DT bands strategies work well only on non-trending stationary markets (markets with constant mean).

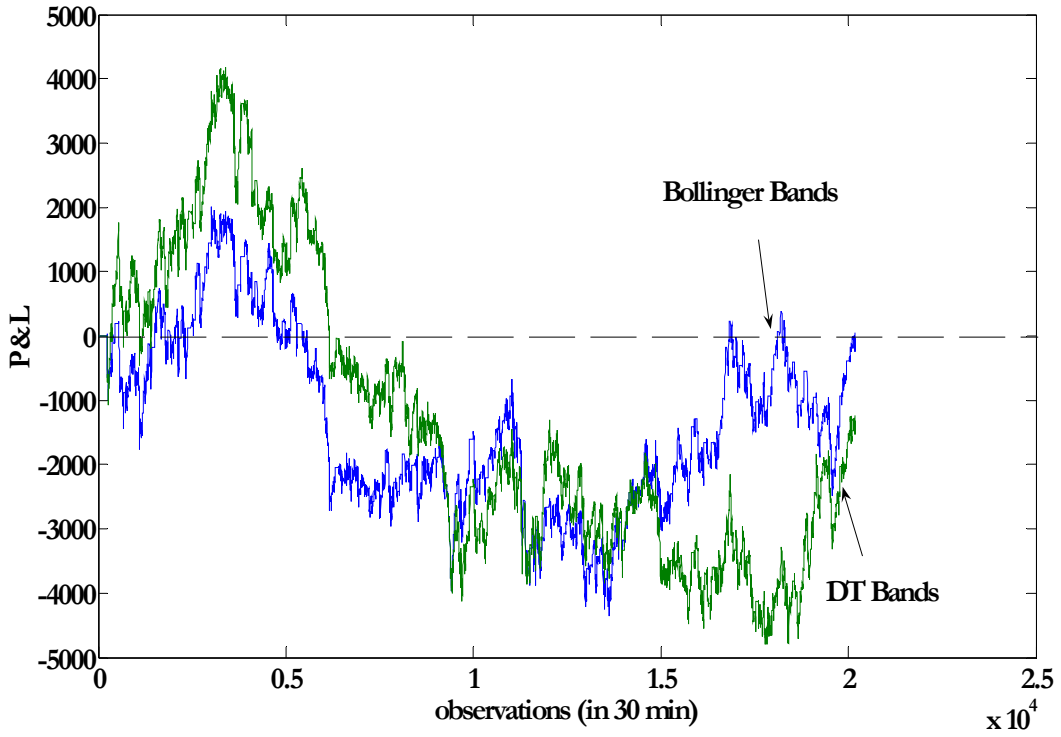


Figure 2.9. Bund (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the basic strategies, based on the Bollinger and DT bands, SMA length n=50 observations

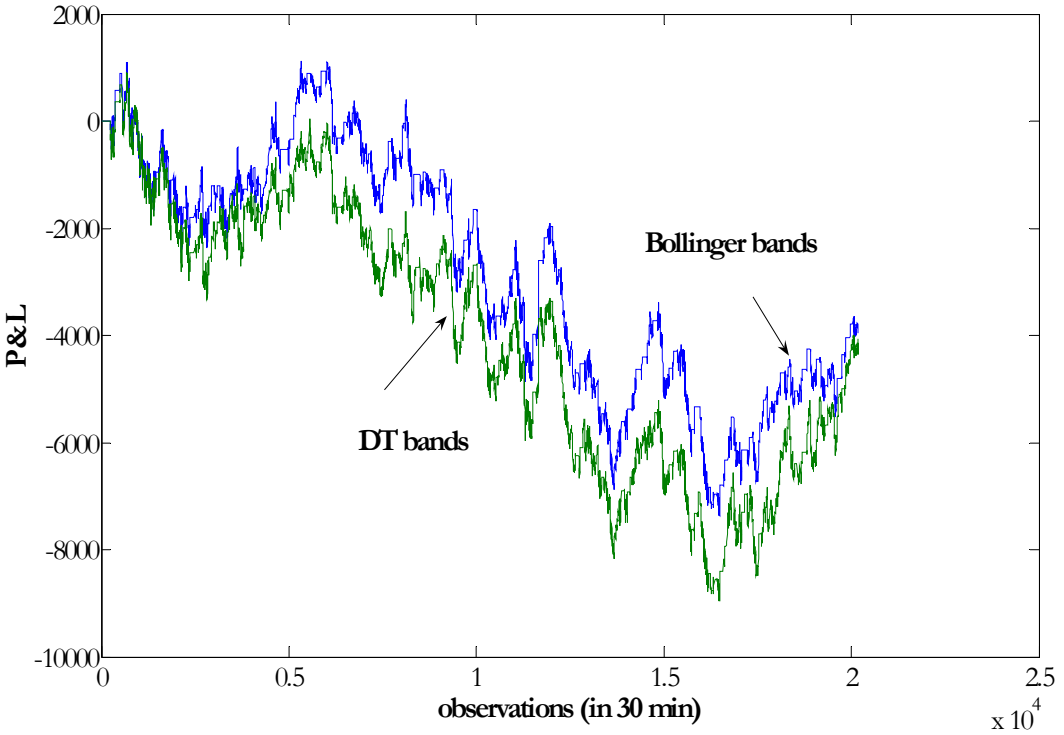


Figure 2.10. Bund quotes (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the basic strategies, based on the Bollinger and DT bands, SMA length n =115 observations

The paths for the confirmed strategies, based on the momentum indicator (see Figures 2.11-2.12) were simulated for the following optimal momentum parameters:

- ***n*(SMA)=50:**
  - (1) Δ=150 (Lag), s_M =0.6 (thresholds) (for both Bollinger and DT bands strategies);
  - (2) Δ=180, s_M =1.2 (for the DT bands strategy).
- ***n*(SMA)=115:**
  - (3) Δ=80, s_M =1.3 (for both Bollinger and DT bands strategies);
  - (4) Δ=180, s_M =1.7 (for the DT bands strategy).

These parameters were defined as optimal from the P&L surfaces, given in the Appendix A (Figures A1-A2 and Tables A1-A4). Note that the optimal momentum parameters for DT bands ((2) and (4)) are more stable over different SMA lengths than the optimal momentum parameters for Bollinger bands. For both SMAs, DT bands strategies, confirmed by momentum, produce higher in absolute values P&L with the path that exhibit steeper positive trend than the best possible confirmed Bollinger band strategy (see Figures 2.11 and 2.12). It should be noted that confirmed DT bands strategies, simulated for the same momentum parameters as Bollinger bands strategies produce worse outcome. However, we suppose that a rational trader will choose the best possible trading strategy. Thus, we can conclude that confirmed by momentum DT bands strategy is superior to confirmed Bollinger bands strategy.

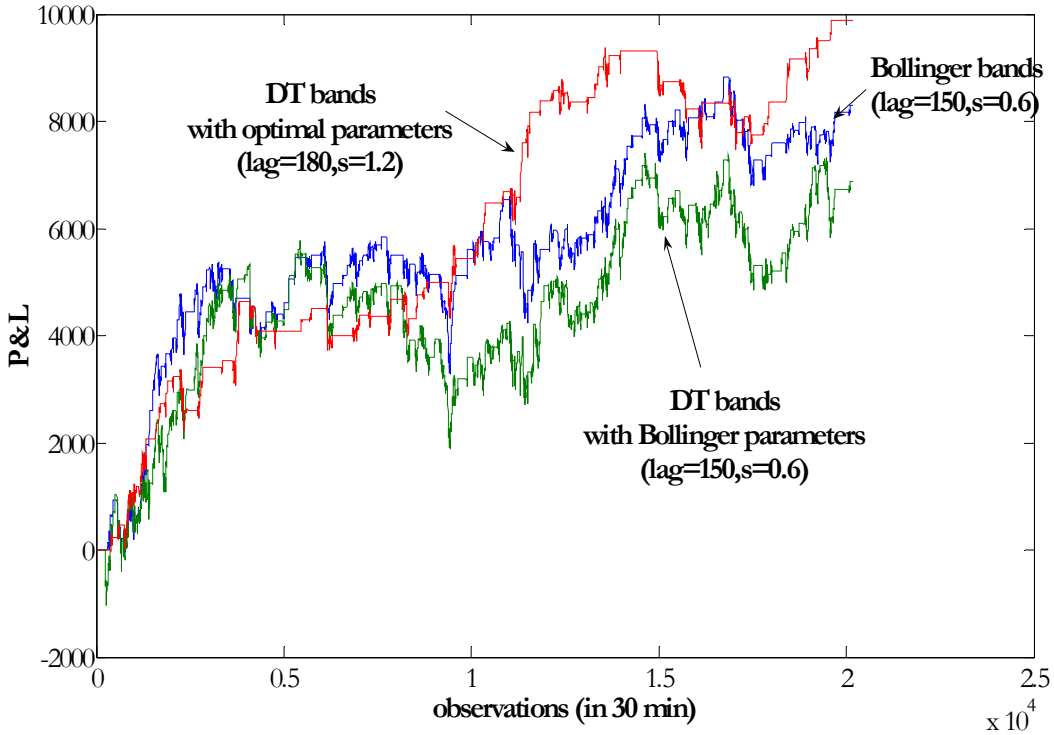


Figure 2.11. Bund (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the strategies, confirmed by momentum and based on the Bollinger and DT bands, SMA length *n* =50 observations

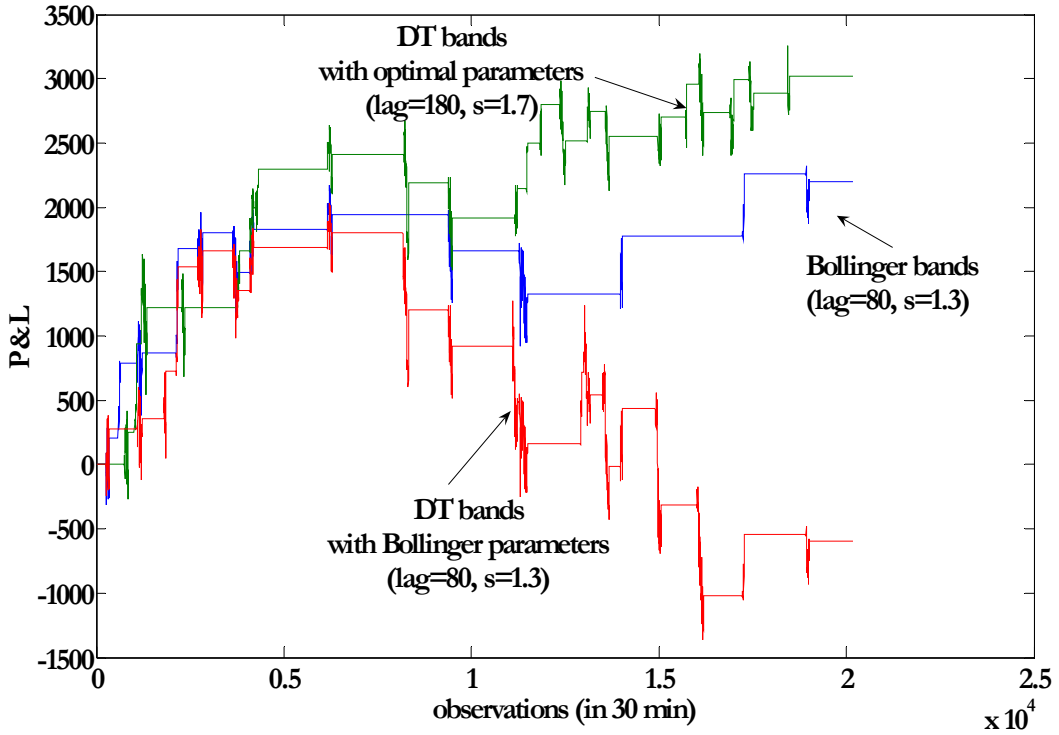


Figure 2.12. Bund (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the strategies, confirmed by momentum and based on the Bollinger and DT bands, SMA length n =115 observations

The third and fourth types of confirmed strategies are simulated only for DT bands strategies. For consistency their results will be compared with the best Bollinger bands strategy, confirmed by momentum (see Figures 2.13-2.14).

The trades were simulated for the following DT bands strategy parameters, derived from the tables A5-A7 in Appendix A:

- ***n*(SMA)=50**:
  (1) "Elliot" strategy: St_E =3 (# steps), Sl_E =30 (total length);
  (2) "Elliot" & "Support/Resistance": St_E =3, Sl_E =30, T_{SR} =0.002 (tolerance Support/Resistance).
- ***n*(SMA)=115:**
  (1) "Elliot" strategy: St_E =3, Sl_E =120;
  (2) "Elliot" & "Support/Resistance": St_E =3, Sl_E =100, T_{SR} =0.004.

As we can see the parameters choice is quite consistent. Shorter MA leads to less dispersed residuals (price-MA), thus, the smaller is the tolerance level to define whether the price touches the DT bands. The shorter MA implies the shorter local residuals trend; thus, the smaller is the total steps length.

For shorter SMA, confirmed by momentum Bollinger bands strategy works better than the DT bands strategy, confirmed by other technical signals ("Elliot" and "Support/Resistance"). However, for longer SMA the DT bands strategy confirmed by both the "Elliot" and "Support/Resistance" signals brings higher results than the Bollinger bands. The P&L paths for confirmed DT bands strategy exhibit positive trend for both SMA lengths, while the confirmed

Bollinger bands strategy produce the path with significantly flatter trend of longer SMA (see Figures 2.13-2.14).

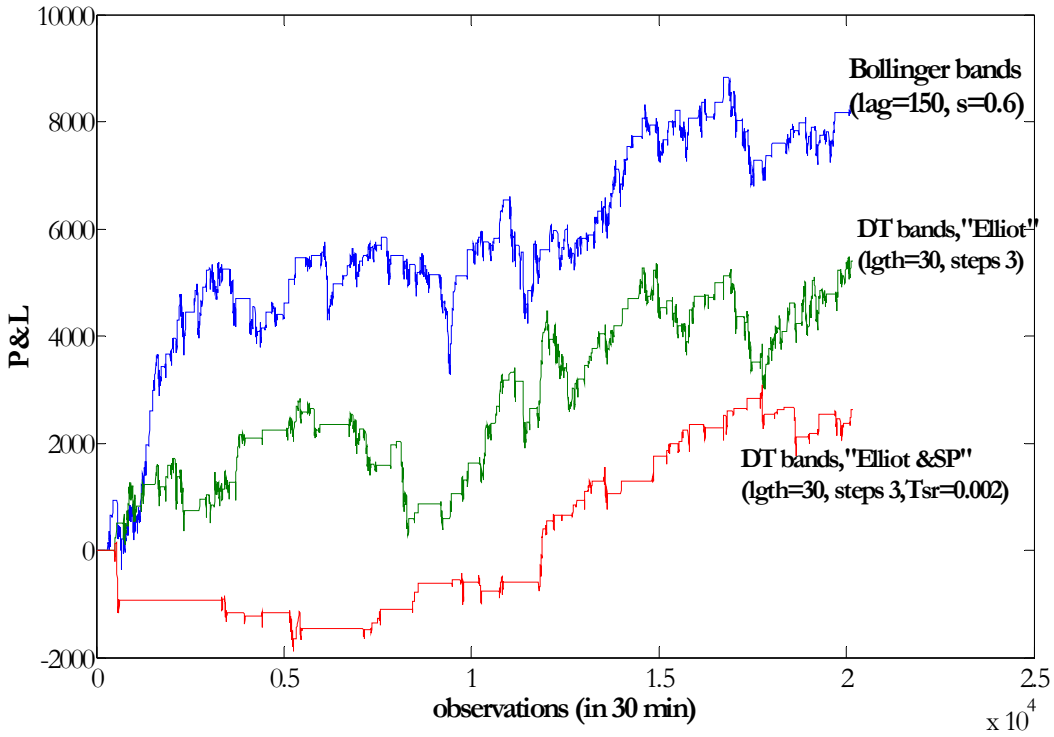


Figure 2.13. Bund (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the Bollinger strategy, confirmed by momentum, the DT strategy, confirmed by the "Elliot" signals and the DT strategy, confirmed by the "Elliot" and "Support/Resistance" signals, SMA length n=50 observations

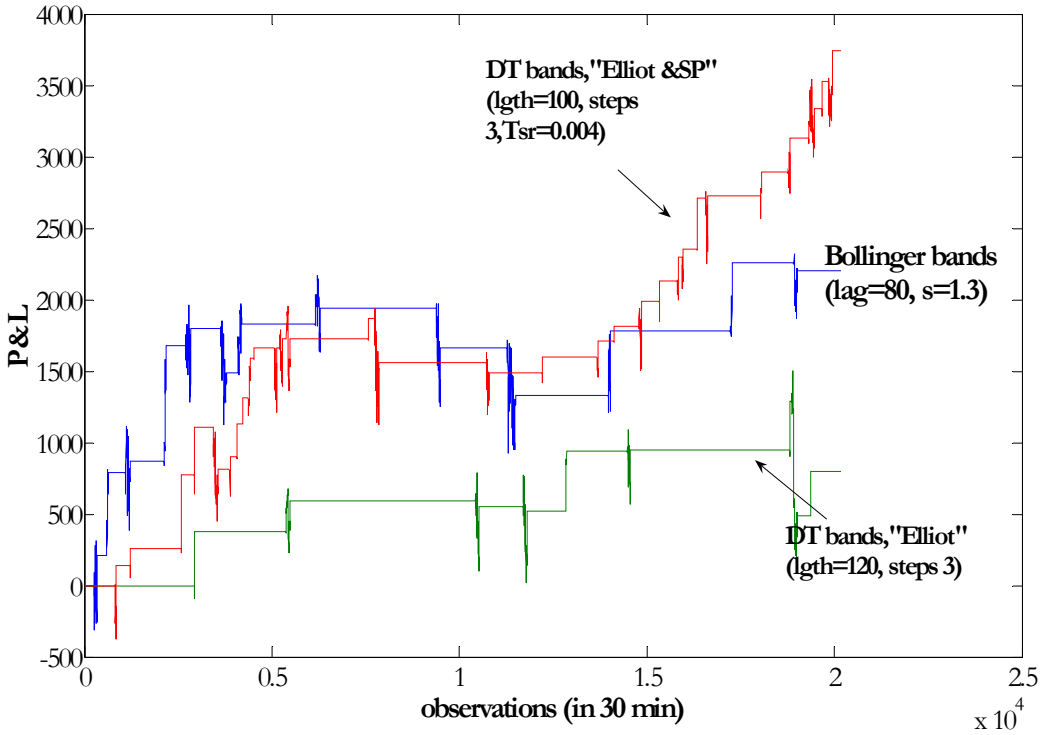


Figure 2.14. Bund (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the Bollinger strategy, confirmed by momentum, the DT strategy, confirmed by the "Elliot" signals and the DT strategy, confirmed by the "Elliot" and "Support/Resistance" signals, SMA length n=115 observations

From the Bund case we can derive the following conclusions:

1. Bollinger bands, with preliminary chosen default parameter k, do not produce profits.
2. The signals sent by Bollinger bands and confirmed by momentum indicator lead to the end-of-period profits and almost increasing path of P&L.
3. The Bollinger bands strategy, which incorporates shorter SMA, produce higher profits then the same strategy incorporating longer SMA. This shorter SMA coincides with the "default" value defined in Table 2.1.
4. The DT bands on their own produce more false signals and, thus, larger losses than the Bollinger bands.
5. Confirmed by momentum DT bands strategy produces the highest profits and the P&L path with positive (trend) slope. The optimal momentum parameters for the DT bands strategy are quite stable over both SMA lengths (the lag is the same, and thresholds are close in values); for the confirmed Bollinger bands strategy the momentum parameters for different SMA lengths differ significantly.
6. Confirmed by "Elliot" and "Support/Resistance" signals DT bands strategies produce positive profits and in some cases larger outcomes than confirmed Bollinger bands strategy. Their P&L paths exhibit increasing pattern.
7. The DT bands show more consistency in application results in the terms of profit stability over the time and strategies. Especially, this is evident for longer SMA, where Bollinger bands strategy works worse.

4.2 DAX

DAX index represents one more type of financial instruments – stocks (see Figure 2.15).

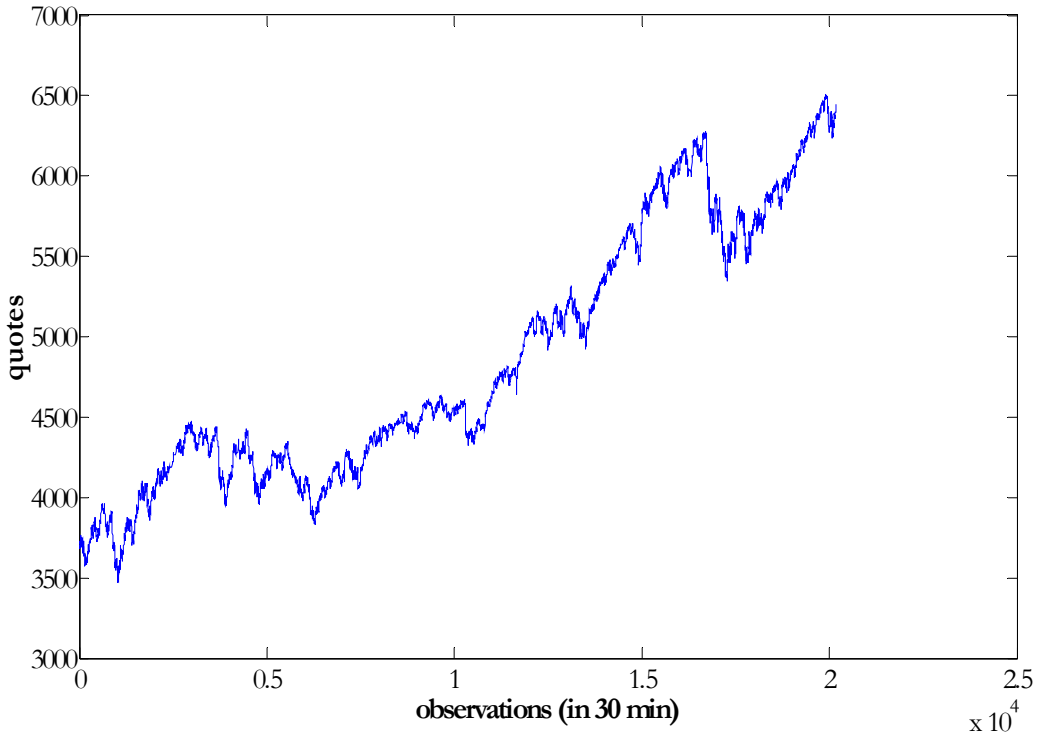


Figure 2.15. DAX quotes for period 30/7/2003-7/12/2006 (frequency – 30 minutes).

DAX data corresponds to the same interval of time (30/7/2003-7/12/2006) and frequency (30 minutes) as Bund data in previous example. As we can see from Figure 2.15, the DAX was positively trending during this period.

The end-of period outcomes for the analysed strategies are given in Tables 2.6-2.7. As we can see, confirmed by the momentum DT strategy produces the highest possible outcome for the SMA length n=50 observations. Confirmed Bollinger band strategy works better than DT bands strategy for longer SMA, although its profits are still lower than for the short SMA. Unconfirmed (basic) strategies, both Bollinger and DT bands, produce very low profits and losses respectively.

Table 2.6
DAX (30/7/2003-7/12/2006, frequency – 30 minutes): Profit & Losses

Strategies	Length of the SMA			
	50		115	
	Bollinger	**DT bands**	**Bollinger**	**DT bands**
Strategy 1: Basic	1287.5	-15862	-13250	-30038
Strategy 2: Confirmed-Momentum	49175	54275 (38125*)	15325	10700 (2275*)
Strategy 3: Confirmed-"Elliot"	-	-2200	-	10200
Strategy 4: Confirmed-"Elliot & Support/Resistance"	-	9550	-	14225

* values that correspond to the momentum parameters – optimal for "Confirmed-Momentum" Bollinger bands strategy

Table 2.7
DAX (30/7/2003-7/12/2006, frequency – 30 minutes): Number of trades

Strategies	Length of the SMA			
	50		115	
	Bollinger	**DT bands**	**Bollinger**	**DT bands**
Strategy 1: Basic	657	869	301	359
Strategy 2: Confirmed-Momentum	323	387(463*)	185	181(229*)
Strategy 3: Confirmed-"Elliot"	-	320	-	66
Strategy 4: Confirmed-"Elliot & Support/Resistance"	-	228	-	115

* values that correspond to the momentum parameters – optimal for "Confirmed-Momentum" Bollinger bands strategy

For the SMA length of 50 observations the DT bands strategy, confirmed by momentum, brings the highest profits, but with lower efficiency than for the Bollinger bands strategy, confirmed by momentum (140€/trade versus 152€/trade). This means that for DAX case the DT bands strategy confirmed by momentum leads to more false signals than similar Bollinger bands strategy, contrary to the case of Bund. At the same time, the highest efficiency (155€/trade) is achieved for the DT bands strategy, confirmed by "Elliot" signals (the SMA length is 115 observations).

Similar to the Bund case, we can see that the number of trades for confirmed strategies is significantly lower than for the non-confirmed strategies due to the elimination of the number of false signals (Table 2.7).

While basic (non-confirmed) strategies generate negative outcomes (see Figures 2.16 and 2.17), for the strategies, confirmed by momentum, the situation is quite different (see Figures 2.18 and 2.19).

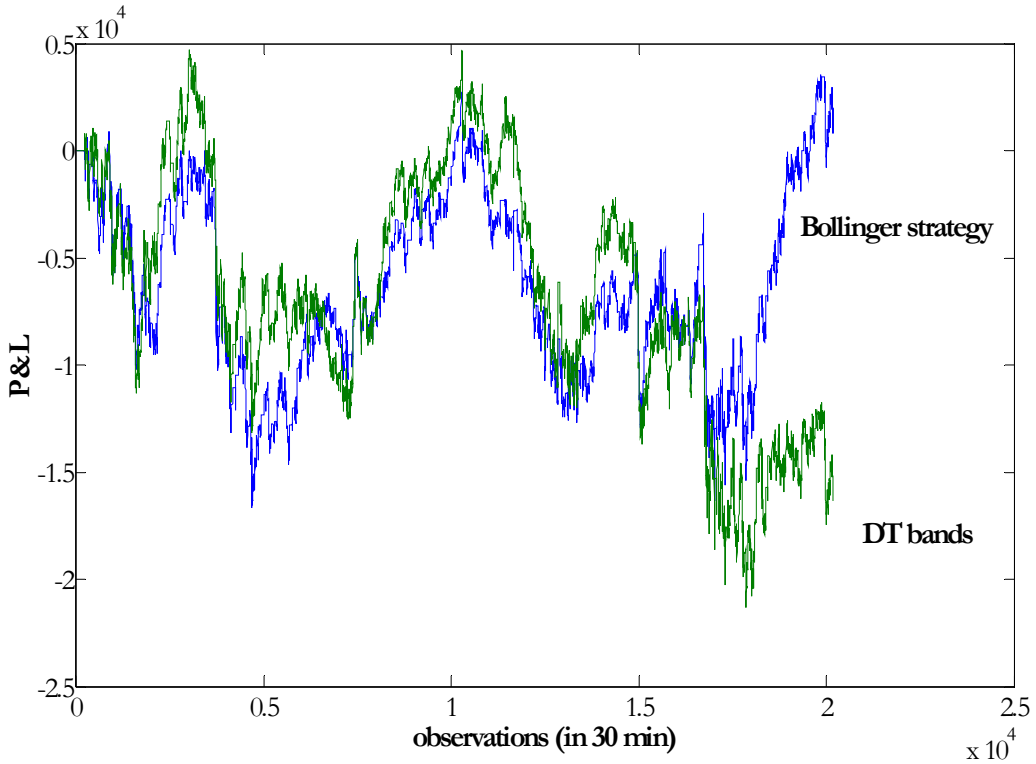


Figure 2.16. DAX (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the basic strategies, based on Bollinger and DT bands, SMA length *n*=50 observations

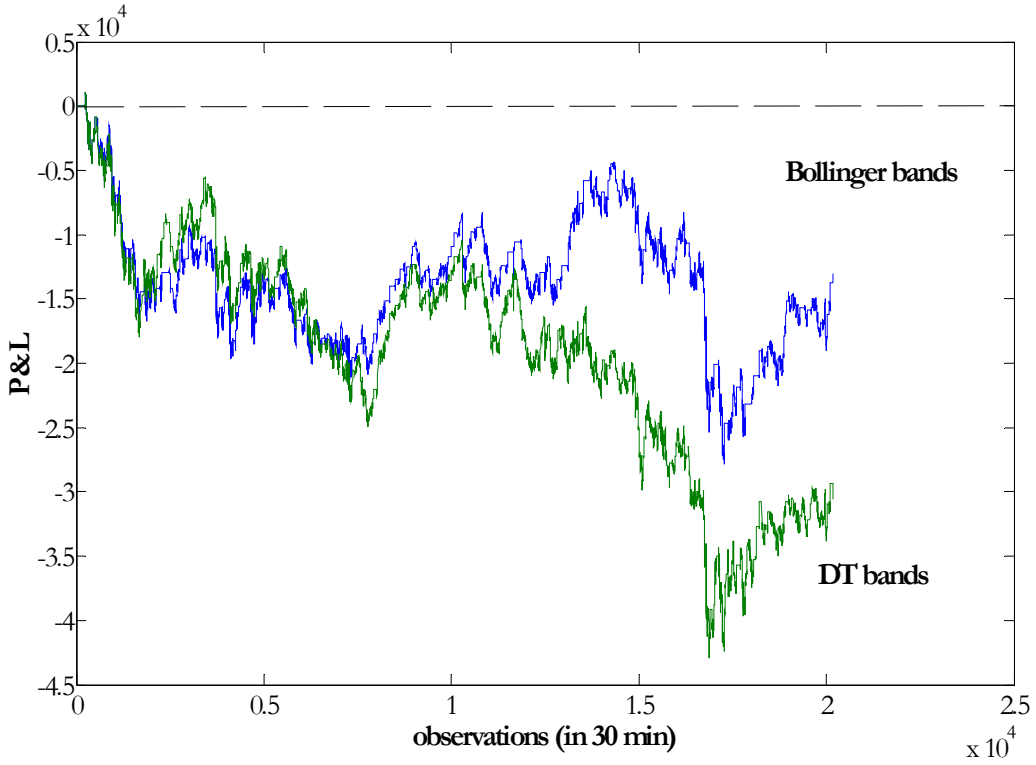


Figure 2.17. DAX (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the basic strategies, based on Bollinger and DT bands, SMA length *n*=115 observations

The paths for the strategies, confirmed by momentum, were simulated for the following parameters, obtained from the momentum values in Figures B1-B2 and optimisation Tables B1-B4 in the Appendix B:

- ***n*(SMA)=50:**
  (1) Δ=180 (lag), s_M=110 (threshold) (for both Bollinger and DT bands strategies);
  (2) Δ=180, s_M=130 (for the DT bands strategy).

- ***n*(SMA)=115:**
  (1) Δ=180, s_M=90 (for both Bollinger and DT bands strategies);
  (2) Δ =180, s_M=130 (for the DT bands strategy).

As for the Bund case the optimal momentum parameters are more stable over SMA lengths for the DT bands than for the Bollinger bands. As we can see, for the shorter SMA (50 observations) the confirmed DT and Bollinger bands strategies produce almost everywhere increasing P&L paths (Figure 2.18). For longer SMA (115 observations) the P&L paths show quite unstable behaviour over time even though they produce end-of-period profits (see Figure 2.19). Contrary to Bund case, the strategies, confirmed by momentum, produce increasing paths exclusively for the short SMA length (50 observations).

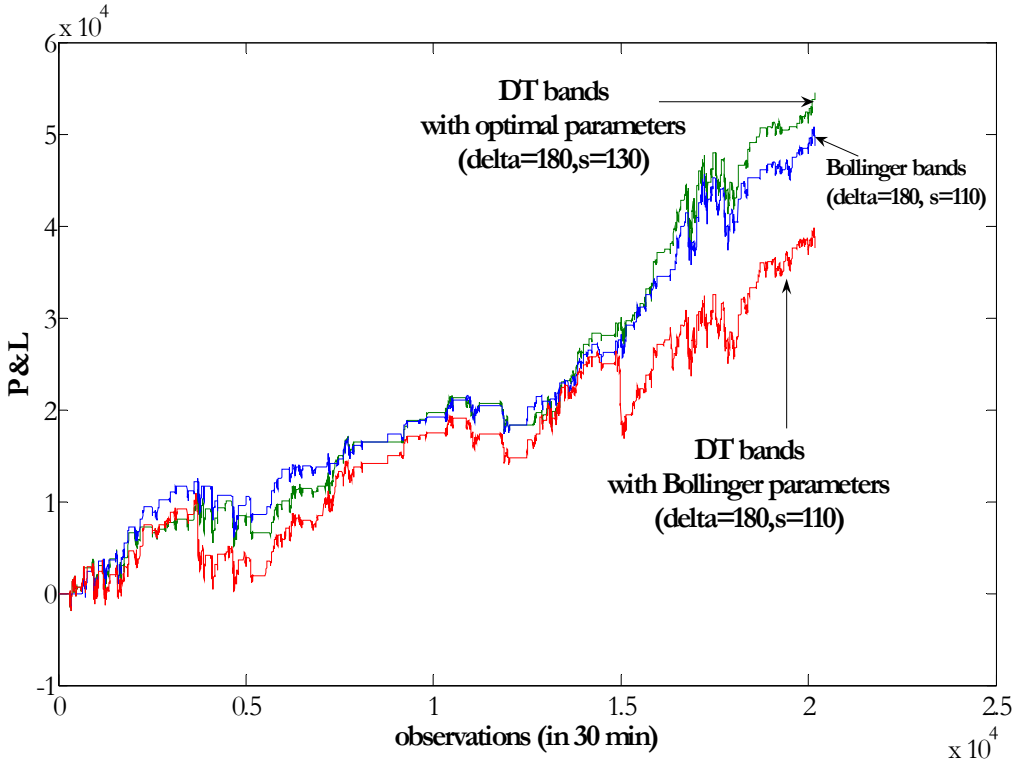


Figure 2.18. DAX (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the strategies, confirmed by momentum and based on the Bollinger and DT bands, SMA length *n*=50 observations

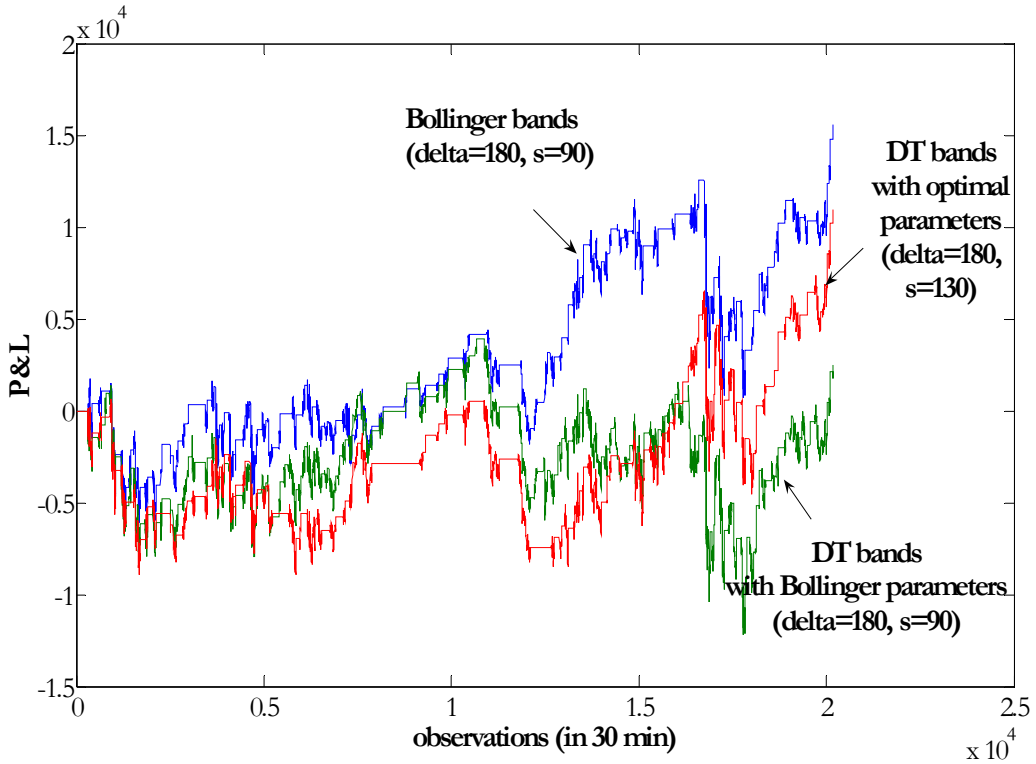


Figure 2.19. DAX (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the strategies, confirmed by momentum and based on the Bollinger and DT bands, SMA length n=115 observations

The paths for the strategies, confirmed by other technical indicators were simulated for the following parameters, derived as optimal from the tables B5-B7 in the Appendix B:

- ***n*(SMA)=50**:
  - (1) "Elliot" strategy: St_E =3 (#steps), Sl_E =30 (total length);
  - (2) "Elliot & "Support/Resistance": St_E =3, Sl_E =40, T_{SR} =0.45.
- ***n*(SMA)=115:**
  - (3) "Elliot" strategy: St_E =3, Sl_E =90;
  - (4) "Elliot & "Support/Resistance": St_E =3, Sl_E =90, T_{SR} =0.25.

All momentum strategies lead to comparable profits, except for one strategy, based on the "Elliot" signal (short SMA), which is non-profitable (the Figure is not presented). For longer SMA, the DT strategy, confirmed by the "Elliot" signal leads to profits only slightly below the results of the DT bands strategy, confirmed by momentum (see Figure 2.20). Moreover, the P&L path for this confirmed DT bands strategy seems to be less volatile than for the confirmed Bollinger bands strategy. Figures 2.21-2.22 present the DT bands strategy results, confirmed by both "Elliot" and "Support/Resistance" signals for different SMA lengths. For the shorter SMA the P&L is increasing, but end-of-period result is much lower than for the confirmed Bollinger bands strategy (Figure 2.21). However, for longer SMA the P&L is increasing and ends up at the level only slightly lower than the confirmed Bollinger bands strategy. In addition the P&L is less volatile and almost never goes below zero contrary to the confirmed Bollinger band strategy (see Figure 2.22).

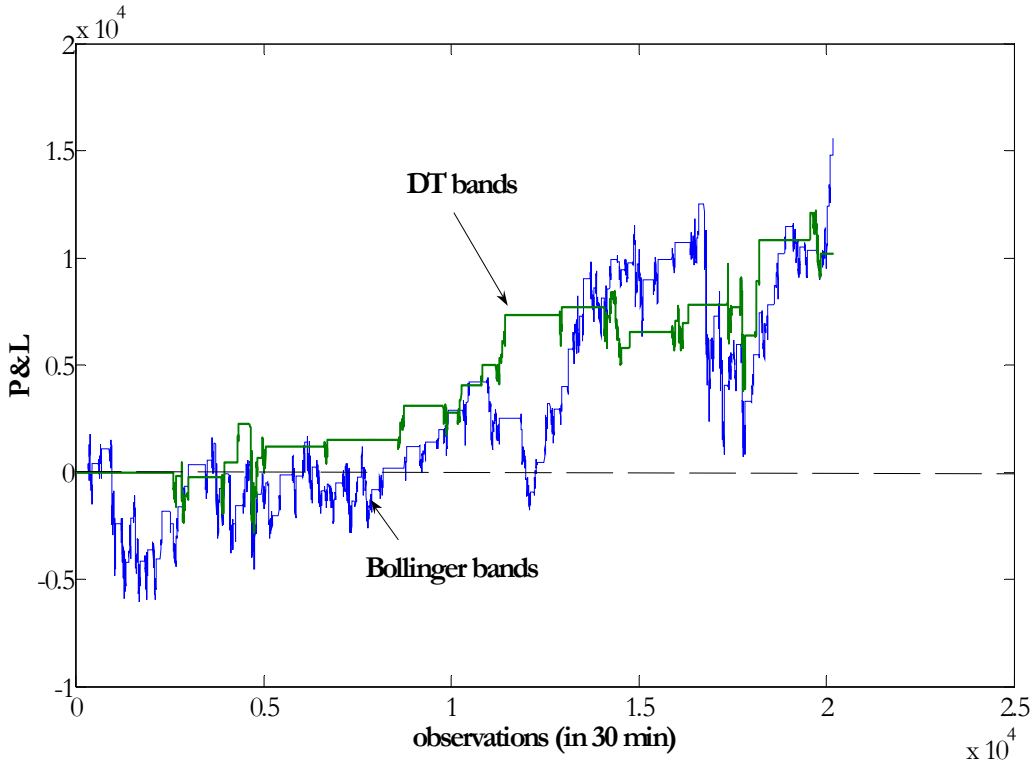


Figure 2.20. DAX (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the Bollinger strategy, confirmed by momentum and the DT strategy, confirmed by the "Eliot" signals, SMA length *n*=115 observations

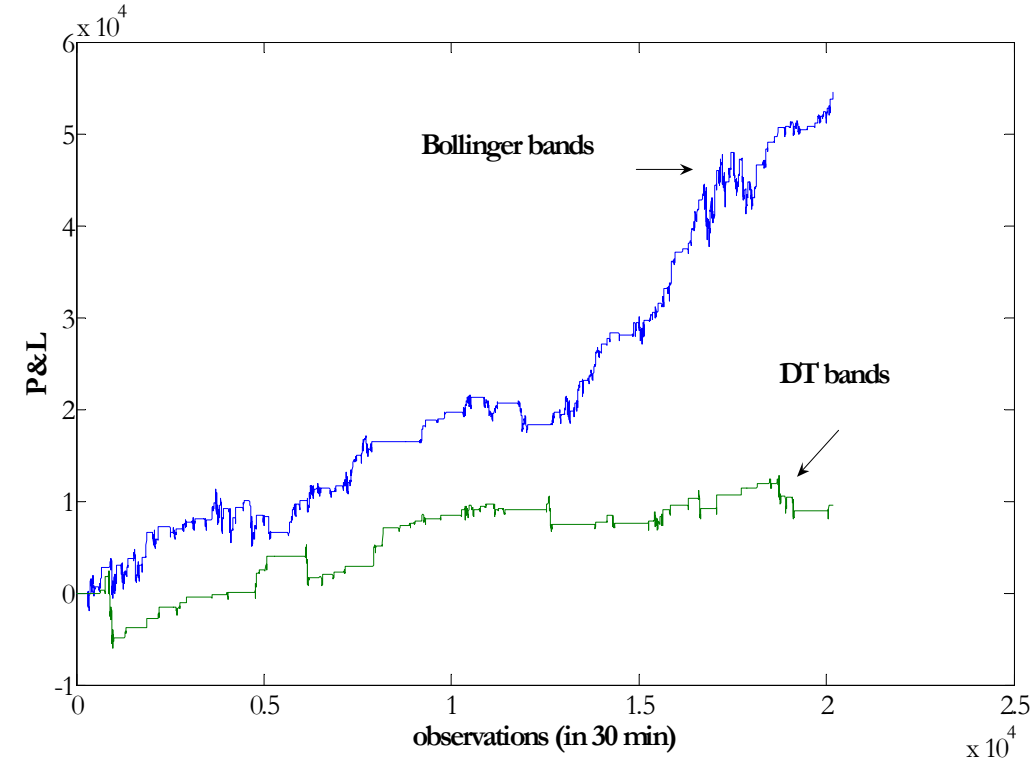


Figure 2.21. DAX (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the Bollinger strategy, confirmed by momentum and the DT strategy, confirmed by the "Eliot" and "Support/Resistance" signals, SMA length *n*=50 observations.

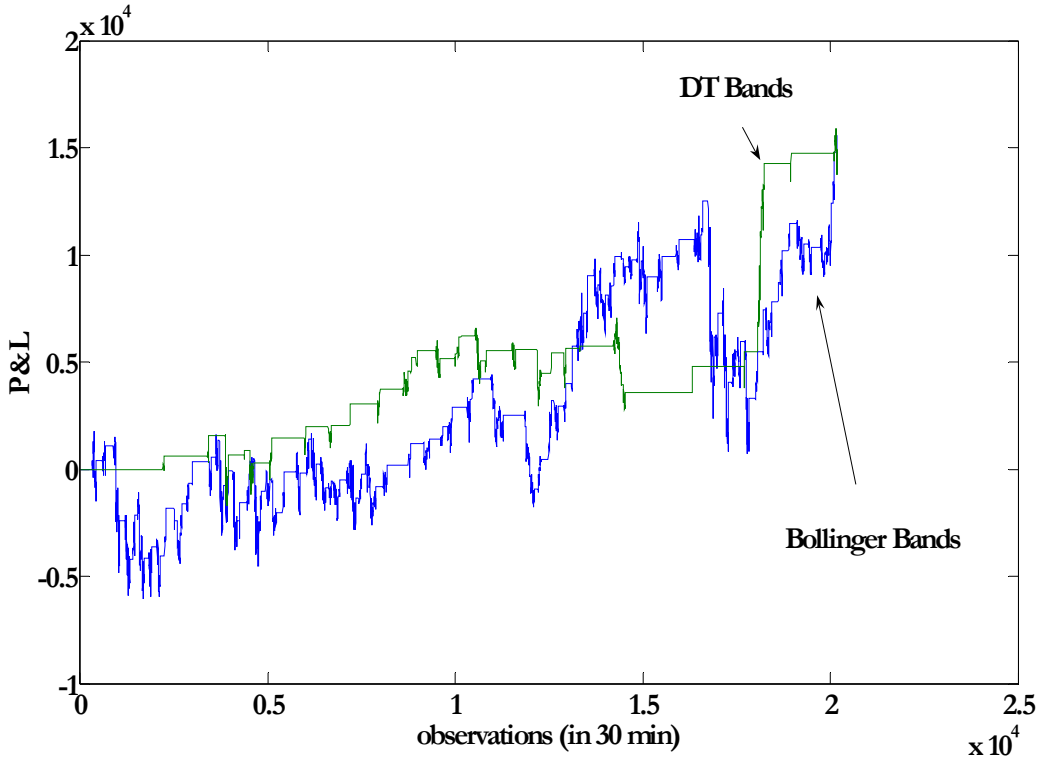


Figure 2.22. DAX (30/7/2003-7/12/2006, frequency – 30 minutes): P&L for the Bollinger strategy, confirmed by momentum and the DT strategy, confirmed by the "Eliot" and "Support/Resistance" signals, SMA length n=115 observations.

As we can see for the DAX case the conclusions are similar to the Bund outcomes:

1. Bollinger bands, with preliminary chosen default parameter k, do not produce or produce very low profits.
2. The signals sent by the Bollinger bands, confirmed by momentum indicator lead to the end-of-period profits, while the P&L path is almost everywhere increasing only for the short SMA (50 observations).
3. The Bollinger bands strategy that incorporates shorter SMA, produces higher profits then the same strategy incorporating longer SMA.
4. DT bands on their own produce more false signals and, thus, larger losses than the Bollinger bands.
5. Confirmed by momentum DT bands strategy produces the highest profits and almost everywhere increasing P&L path for the short SMA. The optimal momentum parameters for DT bands strategy are stable over both SMA lengths (the lag and thresholds are the same); for Bollinger bands only the lag parameter is the same.
6. Confirmed by "Elliot" and "Support/Resistance" signals the DT bands strategy produces positive profits (except for the case of short SMA length); for long SMA the results are comparable with the outcomes obtained for the confirmed Bollinger bands strategy. Their P&L paths are exhibit positive trend, contrary to the Bollinger bands strategies.
7. The DT bands show more consistency in application in the terms of profit stability over the time and the strategies.

4.3 Brent

Brent is a futures on crude oil. Therefore, we consider this instrument as a representative of commodity markets. The peculiarity of this instrument is that it is highly volatile. Our data represents the period of 17/12/04-27/01/06 with 30 minutes frequency (see Figure 2.23).

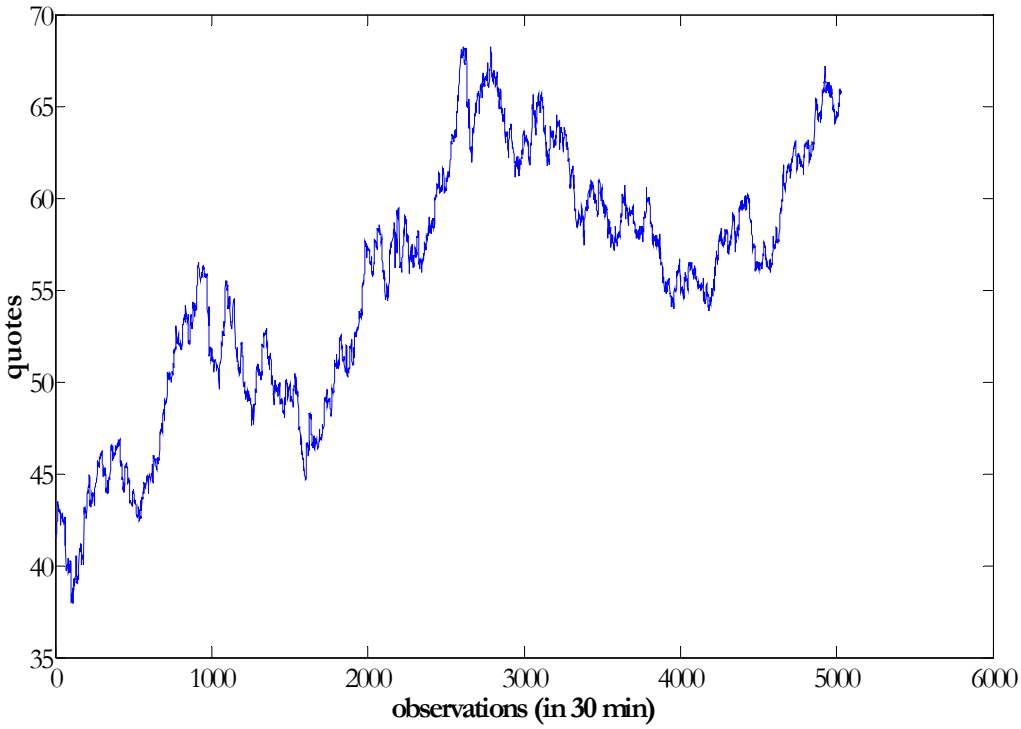


Figure 2.23. Brent quotes for period 17/12/04-27/01/06 (frequency 30 minutes).

Table 2.8
Brent (17/12/04-27/01/06, frequency 30 minutes): Profit & Losses

Strategies	Length of the SMA			
	50		115	
	Bollinger	DT bands	Bollinger	DT bands
Strategy 1: Basic	-16440	-15820	-7870	-1470
Strategy 2: Confirmed-Momentum	5520	9090 (-1400*)	12410	18720(18340*)
Strategy 3: Confirmed-"Elliot"	-	1130	-	6490
Strategy 4: Confirmed-"Elliot & Support/Resistance"	-	1220	-	1790

* values that correspond to the momentum parameters – optimal for "Confirmed-Momentum" Bollinger bands strategy

Table 2.9
Brent (17/12/04-27/01/06, frequency 30 minutes): Number of trades

Strategies	Length of the SMA			
	50		115	
	Bollinger	DT bands	Bollinger	DT bands
Strategy 1: Basic	153	200	61	87
Strategy 2: Confirmed-Momentum	64	38 (81*)	28	39(34*)
Strategy 3: Confirmed-"Elliot"	-	4	-	22
Strategy 4: Confirmed-"Elliot & Support/Resistance"	-	4	-	4

* values that correspond to the momentum parameters – optimal for "Confirmed-Momentum" Bollinger bands strategy

Tables 2.8 and 2.9 represent the absolute end-of-period outcomes for different strategies, based on Bollinger and DT bands. As we can see unconfirmed strategies brings losses for both length of SMA; moreover, contrary to previous cases (Bund and DAX) the shorter SMA of 50 observations produces even higher losses than the longer one. Neither higher profits are detected for shorter SMA for confirmed strategies.

As in the previous case-studies, the DT bands strategy (n(SMA)=115 observations), confirmed by momentum shows the highest results. The strategies confirmed by the "Elliot" and "Support/Resistance" signals showed positive profits for both SMAs, but these results are considerably lower than those, obtained for the strategies, confirmed by the momentum indicators.

The trades' efficiency, calculated for the Tables 2.8 and 2.9, is the highest for the strategies, based on the DT bands. For shorter SMA, the difference between the efficiency for the Bollinger and DT bands is very large ($239/trade for the DT bands strategy, confirmed by momentum, versus 86€/trade for the Bollinger bands strategy, confirmed by momentum) and the discrepancy decreases for the longer SMA ($480/trade for the respective DT bands strategy versus $443/trade for the respective Bollinger bands strategy). We should note that the DT bands strategies, confirmed by other (than momentum) technical indicators lead to very small number of trades (except for the DT bands strategy, confirmed by the "Elliot" signal, for n(SMA)=115 observations). That is why, further we will present the P&L paths only for three confirmed DT bands strategies that produce the number of trades higher than 10.

As for P&L paths, basic strategies show non-monotone behavior for both SMA lengths (see Figures 2.24-2.25).

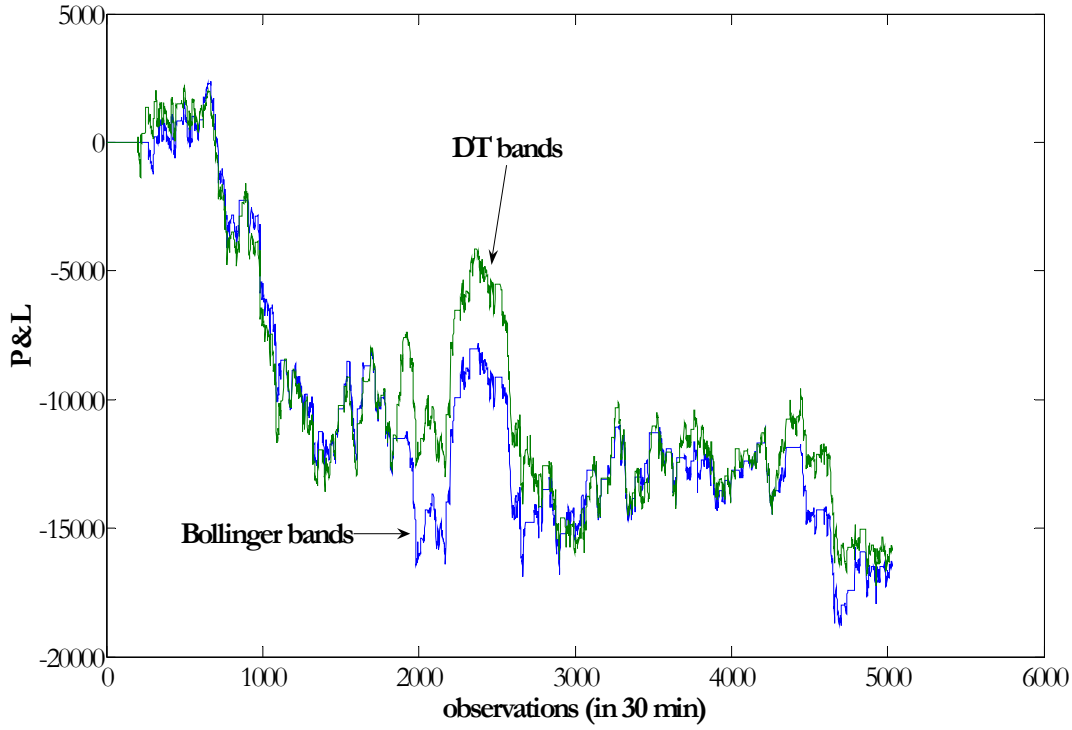


Figure 2.24. Brent (17/12/04-27/01/06, frequency - 30 minutes): P&L for the basic strategies, based on the Bollinger and DT bands, SMA length n=50 observations

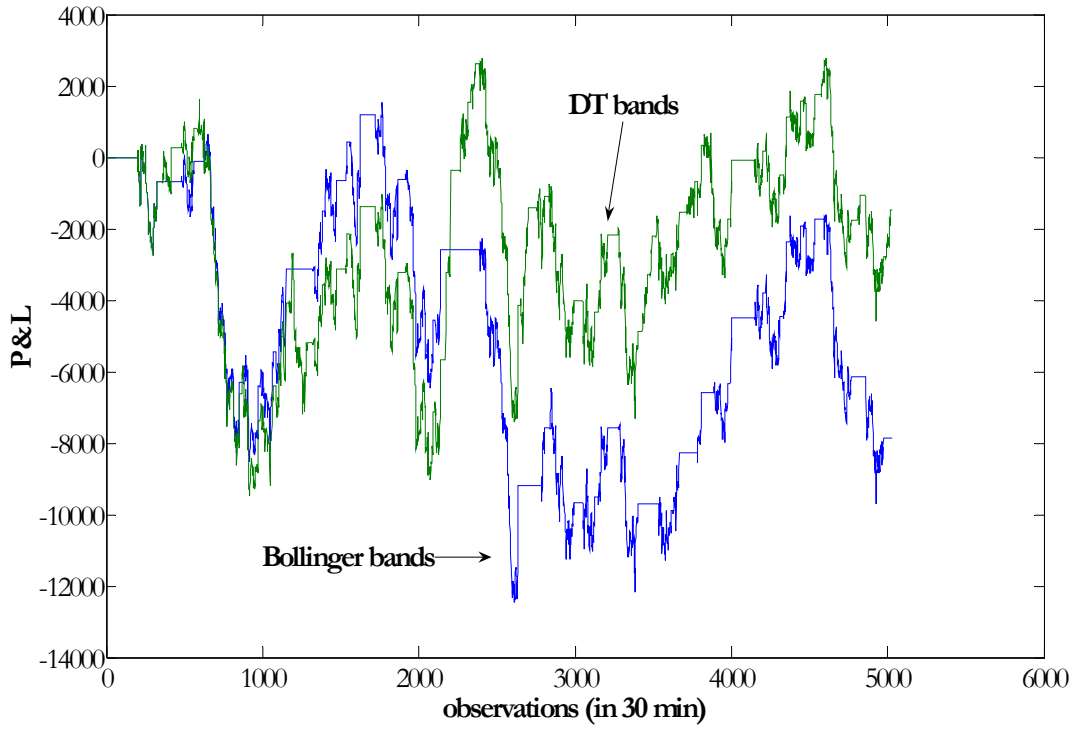


Figure 2.25. Brent (17/12/04-27/01/06, frequency - 30 minutes): P&L for the basic strategies, based on the Bollinger and DT bands, SMA length n=115 observations

The strategies, confirmed by momentum were simulated for the following set of parameters, derived from Figures C1-C2 and Tables C1-C4 in the Appendix C:

- ***n*(SMA)=50:**
  (1) Δ=90, s_M=3 (threshold) (for both Bollinger and DT bands strategies);
  (2) Δ=50, s_M=3.6 (for the DT bands strategy).

- ***n*(SMA)=115:**
  (1) Δ=50, s_M=3.6 (for both Bollinger and DT bands strategies);
  (2) Δ =50, s_M=3.4 (for the DT bands strategy).

Again the parameters are more stable over different SMA lengths for the DT bands strategy than for the Bollinger bands strategy. Confirmed by momentum optimal strategy, based on the DT bands, produces the P&L paths with steeper positive slope than for the confirmed strategies, based on the Bollinger bands (see Figures 2.26-2.27). For both SMAs, these confirmed DT bands strategies produce higher profits than the confirmed Bollinger bands strategies. In particular, for the shorter SMA, the confirmed strategy based on the DT bands takes less false position, because the P&L drops less than for the Bollinger bands strategy (see Figure 2.26). For the longer SMA, the paths are almost parallel, while the P&L for the DT bands strategy gets some boost in the middle of the observed period (see Figure 2.27).

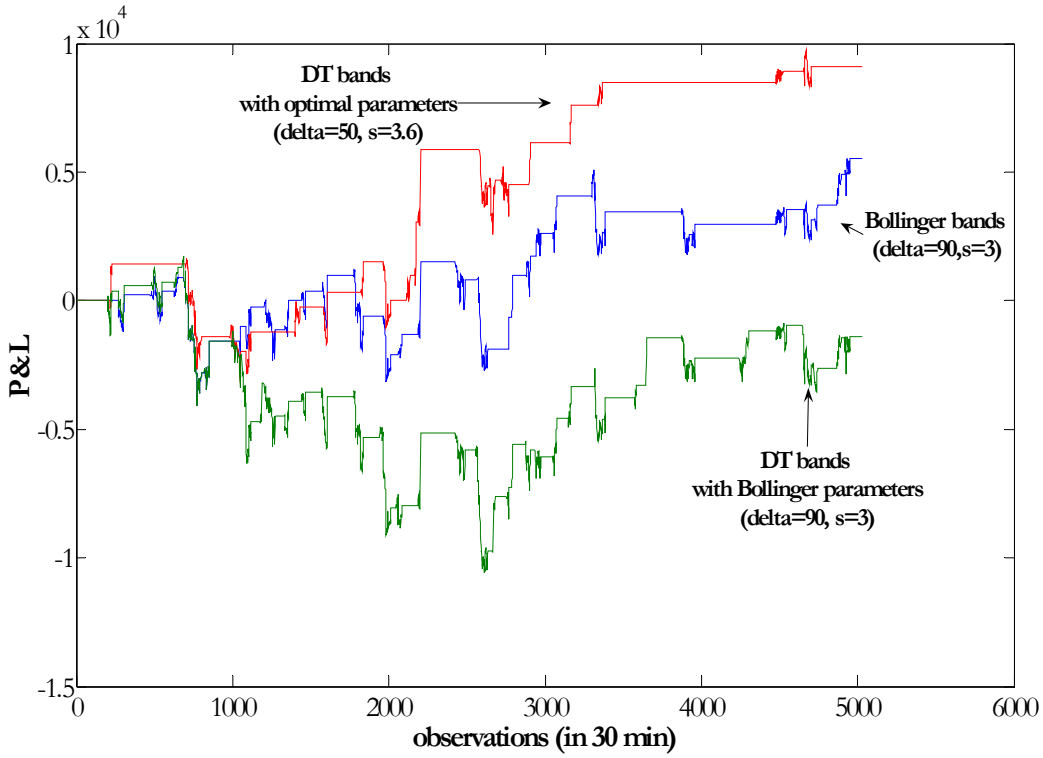


Figure 2.26. Brent (17/12/04-27/01/06, frequency - 30 minutes): P&L for the strategies, confirmed by momentum and based on the Bollinger and DT bands, SMA length n=50 observations

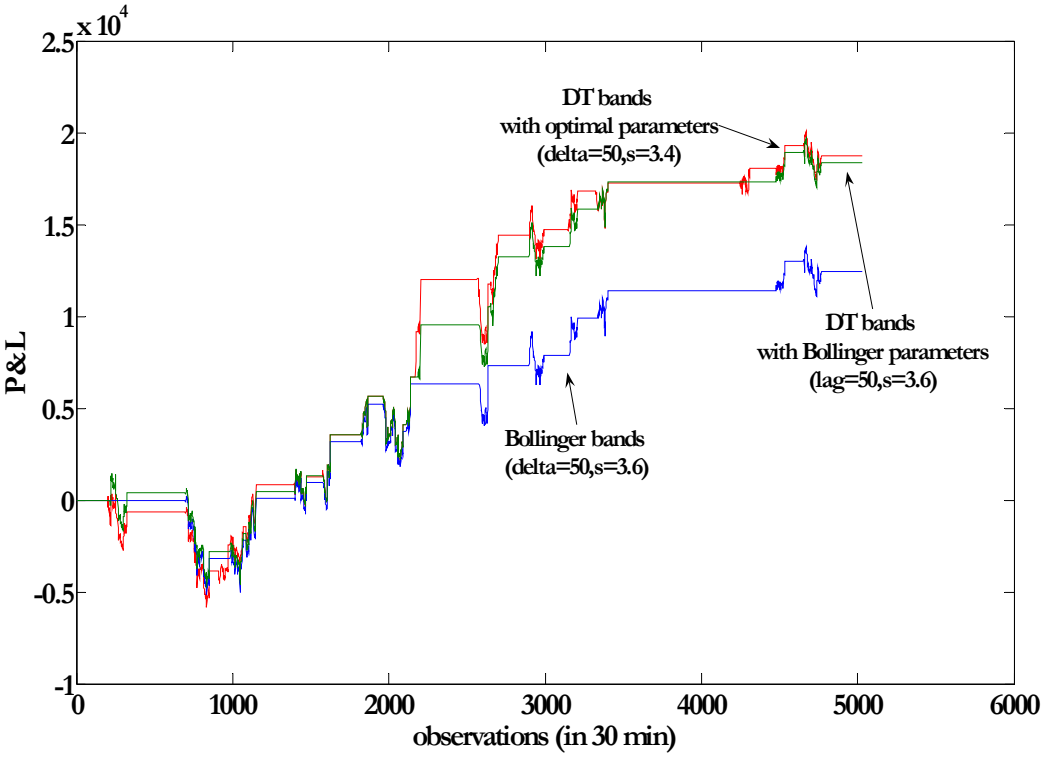


Figure 2.27. Brent (17/12/04-27/01/06, frequency - 30 minutes): P&L for the strategies, confirmed by momentum and based on the Bollinger and DT bands, SMA length n=115 observations

As has been mentioned already, only one DT bands strategy, confirmed by "Elliot" signals (n(SMA)=115 observations), produces sufficient amount of trades. The P&L path for this strategy is given in Figure 2.28. The path was simulated for the following parameters derived from Appendix C, Tables C5-C7: St_E =3 (steps #), Sl_E =60 observations (total length).

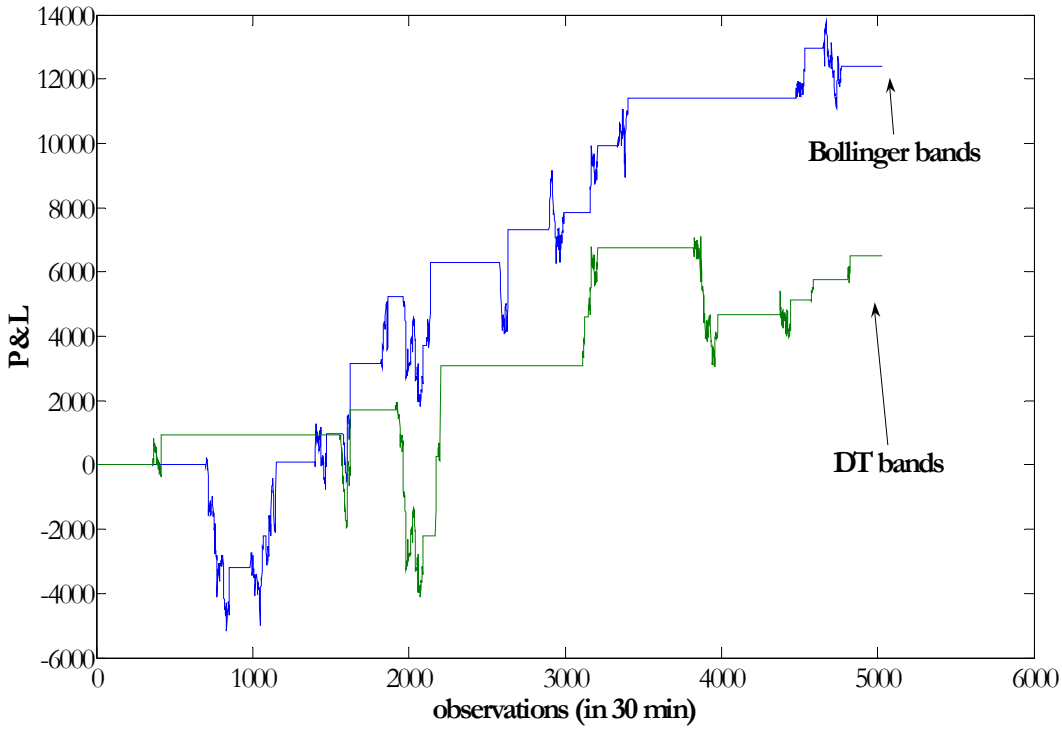


Figure 28. Brent (17/12/04-27/01/06, frequency - 30 minutes): P&L for the Bollinger strategy, confirmed by momentum and DT strategy, confirmed by the "Elliot" signals, SMA length n=115 observations

As the result, the Brent case leads to the similar conclusions as the previous Bund and DAX cases, with one exception (conclusion #3):

1. Bollinger bands, with preliminary chosen default parameter k, do not produce profits.
2. The signals sent by the Bollinger bands, confirmed by momentum indicator lead to the end-of-period profits, while the P&L path is almost everywhere increasing only for long SMA (50 observations).
3. Contrary to the previous cases, the SMA with short "default" length does not produce better results than the longer SMA.
4. DT bands on their own produce more false signals and, thus, larger losses than the Bollinger bands.
5. Confirmed by momentum the DT bands strategy produces the highest profits and almost everywhere increasing P&L path among all testing strategies, including the confirmed Bollinger bands strategy. The optimal momentum parameters for DT bands strategy are stable over both SMA lengths (the lag is the same and the thresholds are almost equal); for the Bollinger bands both momentum parameters are different for different SMA lengths.
6. Confirmed (by the "Elliot" and "Support/Resistance" signals) DT bands strategies produce positive profits, but smaller than confirmed Bollinger bands strategy outcomes; they also generate fewer trades than the other strategies.

4.4 X instrument

Instrument X represents an artificially created index, used by one bank for strategy constructions[24]. The data frequency is 1 hour.

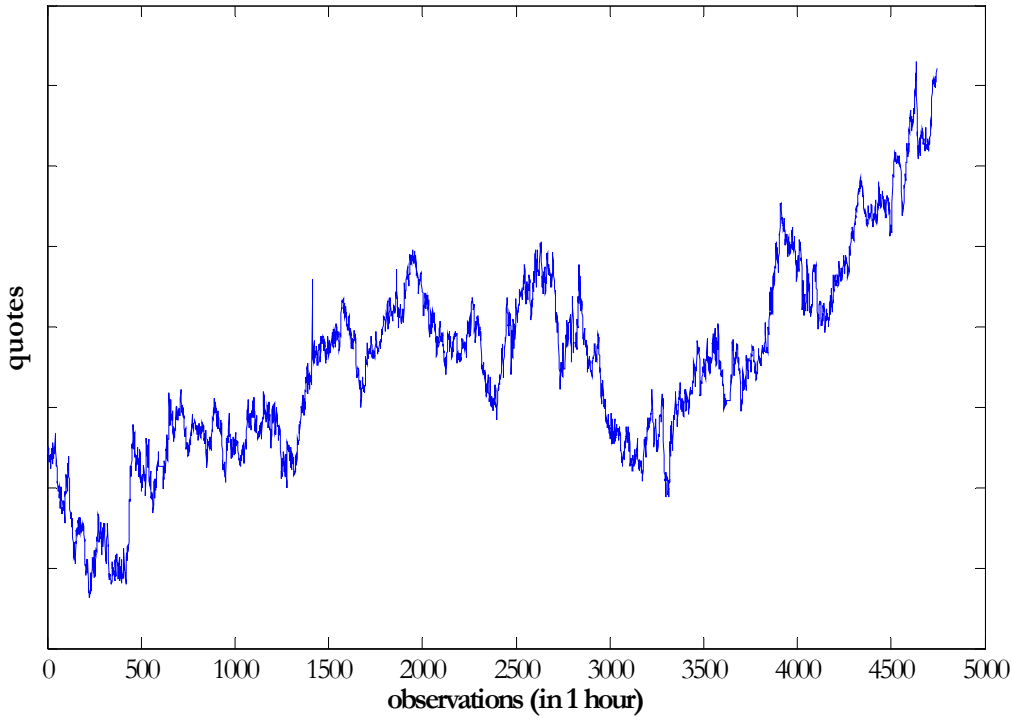


Figure 2.29. X instrument quotes (frequency – 1 hour)

The strategies end-of period outcomes are given in Tables 2.10-2.11. As in previous cases, unconfirmed (basic) strategies bring losses for short SMA (20 observations), while for long SMA the basic Bollinger bands strategy brings some small profits. The highest results are achieved for the Bollinger bands strategy, confirmed by momentum (n(SMA)=50). The DT bands strategy, confirmed by momentum brings larger profits for the shorter SMA (n(SMA)=20) and comparable profits in the case of the longer SMA. Contrary to other cases, the other confirmed strategies, based on the DT bands bring very poor results.

Table 2.10
Instrument X (frequency - 1 hour): Profit & Losses

Strategies	Length of the SMA			
	20		50	
	Bollinger	DT bands	Bollinger	DT bands
Strategy 1: Basic	-432	-1192	188	-12
Strategy 2: Confirmed-Momentum	258	290 (38*)	528	488(418*)
Strategy 3: Confirmed-"Elliot"	-	-100	-	10
Strategy 4: Confirmed-"Elliot & Support/Resistance"	-	-130	-	-70

* values that correspond to the momentum parameters – optimal for "Confirmed-Momentum" Bollinger bands strategy

[24] Due to the confidentiality reasons we cannot neither present the detail description, nor provide the information on the instrument real quotes. That is why on Figure 2.29 that presents the quotes path during some period of time, there are no ticks on the Y-coordinate.

Table 2.11
Instrument X (frequency 1 hour): Number of trades

Strategies	Length of the SMA			
	20		50	
	Bollinger	DT bands	Bollinger	DT bands
Strategy 1: Basic	345	505	181	251
Strategy 2: Confirmed-Momentum	101	90 (133*)	105	79 (127*)
Strategy 3: Confirmed-"Elliot"	-	172	-	2
Strategy 4: Confirmed-"Elliot & Support/Resistance"	-	16	-	2

* values that correspond to the momentum parameters – optimal for "Confirmed-Momentum" Bollinger bands strategy

Taking into account the number of trades in Table 2.11, we can see that trades efficiency is higher for the DT bands strategy confirmed by momentum, than for the similar Bollinger bands strategies.

As for the P&L paths, the Figure 2.30 shows decreasing path for all basic strategies constructed for SMA length of 20 observations. Non-monotone paths for all basic strategies for SMA length of 50 observations are presented in Figure 2.31.

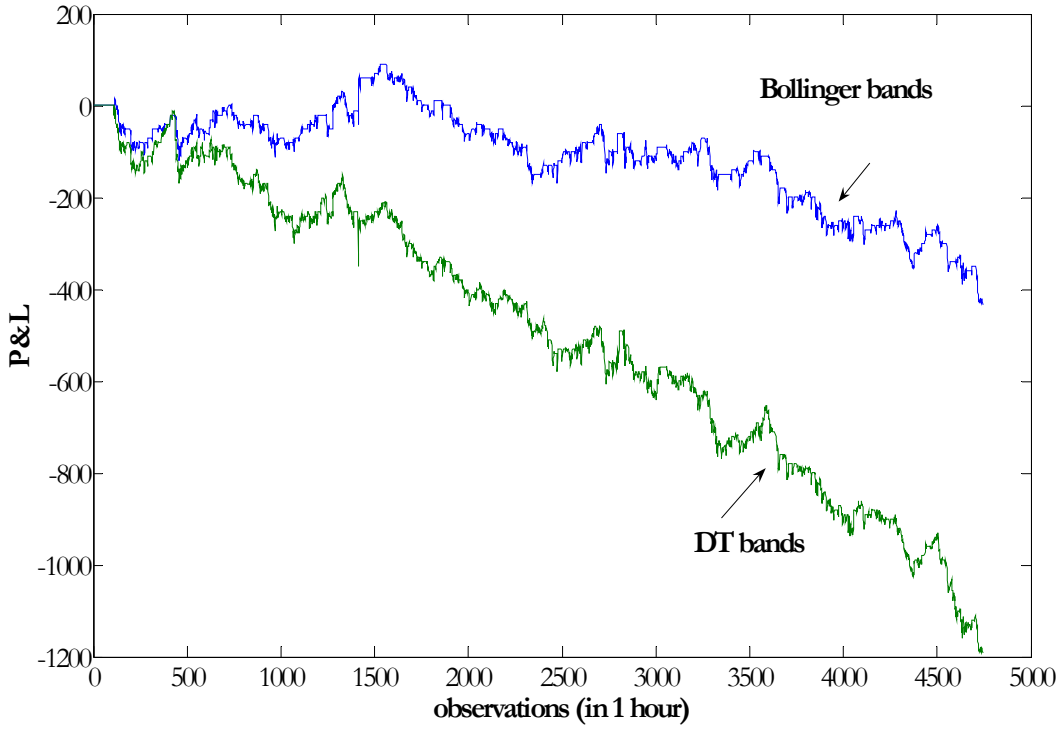


Figure 2.30. X instrument: P&L for the basic strategies, based on the Bollinger and DT bands, SMA length n=20 observations

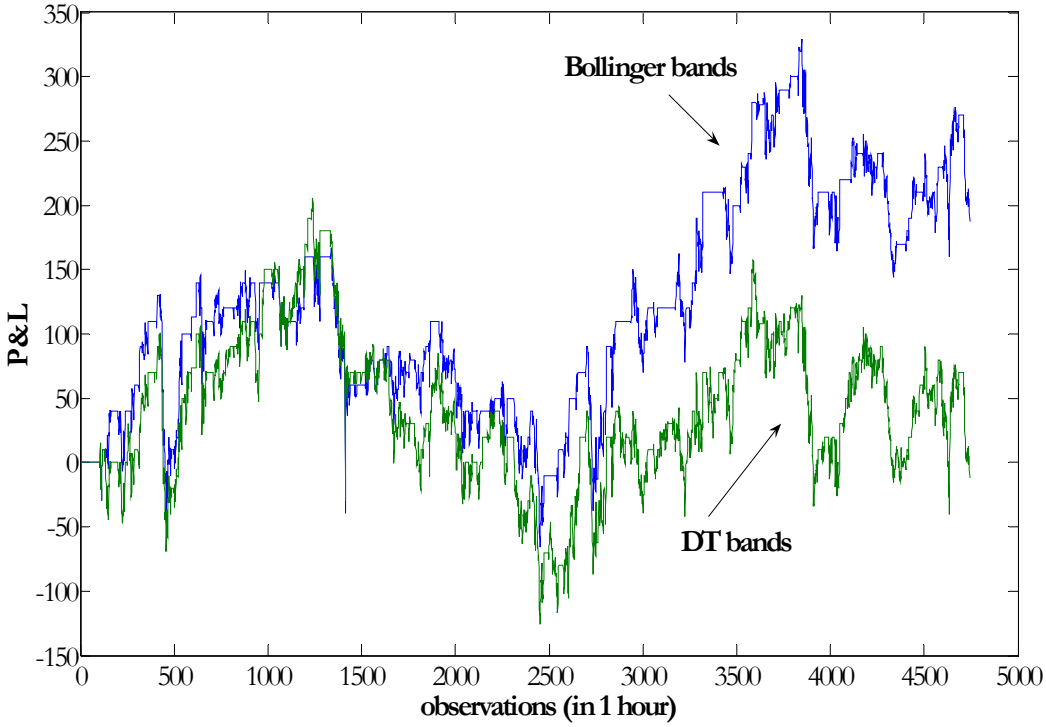


Figure 2.31. X instrument: P&L for the basic strategies, based on the Bollinger and DT bands, SMA length n=50 observations

The paths for the strategies, confirmed by momentum, were simulated for the following parameters, derived as optimal from the Appendix D, Figures D1-D2 and Tables D1-D4:

- ***n*(SMA)=20:**
  (1) $\Delta = 100$, $s_M = 8$ (threshold) (for both Bollinger and DT bands strategies)
  (2) $\Delta = 10$, $s_M = 5$ (for the DT bands strategy)

- ***n*(SMA)=50:**
  (1) $\Delta = 10$, $s_M = 4$ (for both Bollinger and DT bands strategies)
  (2) $\Delta = 10$, $s_M = 5$ (for the DT bands strategy)

The stability of the momentum parameters for the DT bands strategies also holds for the case of the X instrument. Confirmed by momentum strategies produce positive and almost everywhere increasing paths for all SMA lengths. The shorter SMA leads to higher profits for the DT bands strategy, confirmed by momentum; however the difference is not that large (see Figure 2.32). Similar situation is observed at Figure 2.33, the P&L for the DT bands strategy for longer SMA is slightly lower at the end of the observed period than for the Bollinger bands strategy.

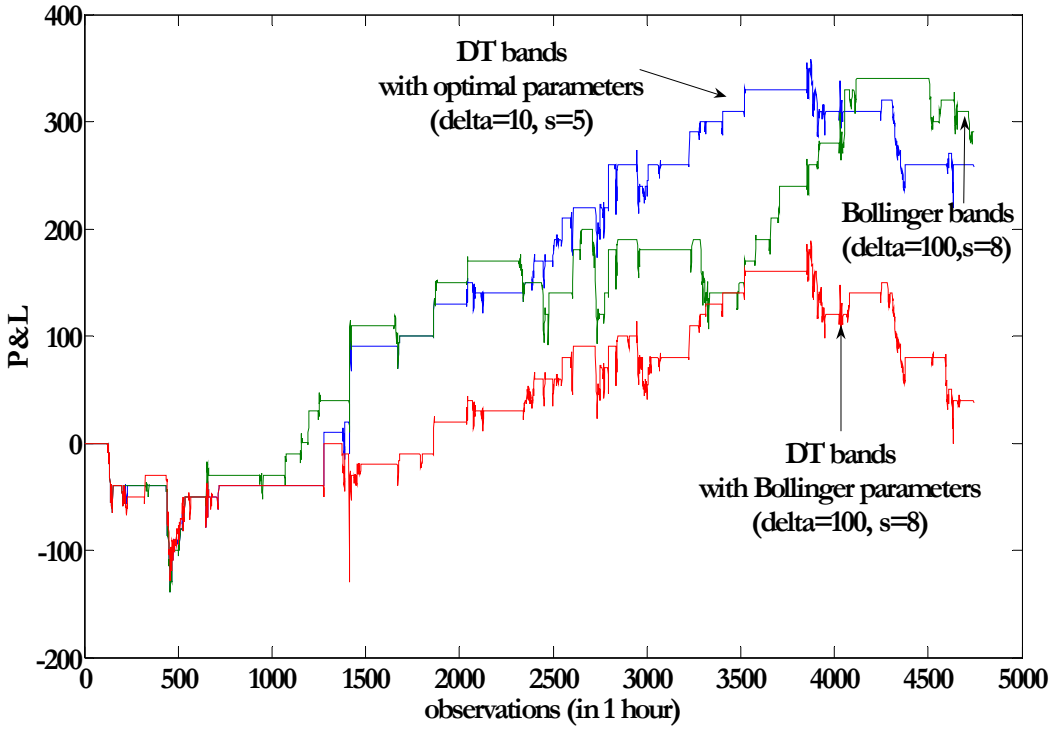


Figure 2.32. X instrument: P&L for the strategies, confirmed by momentum and based on the Bollinger and DT bands, SMA length n=20 observations

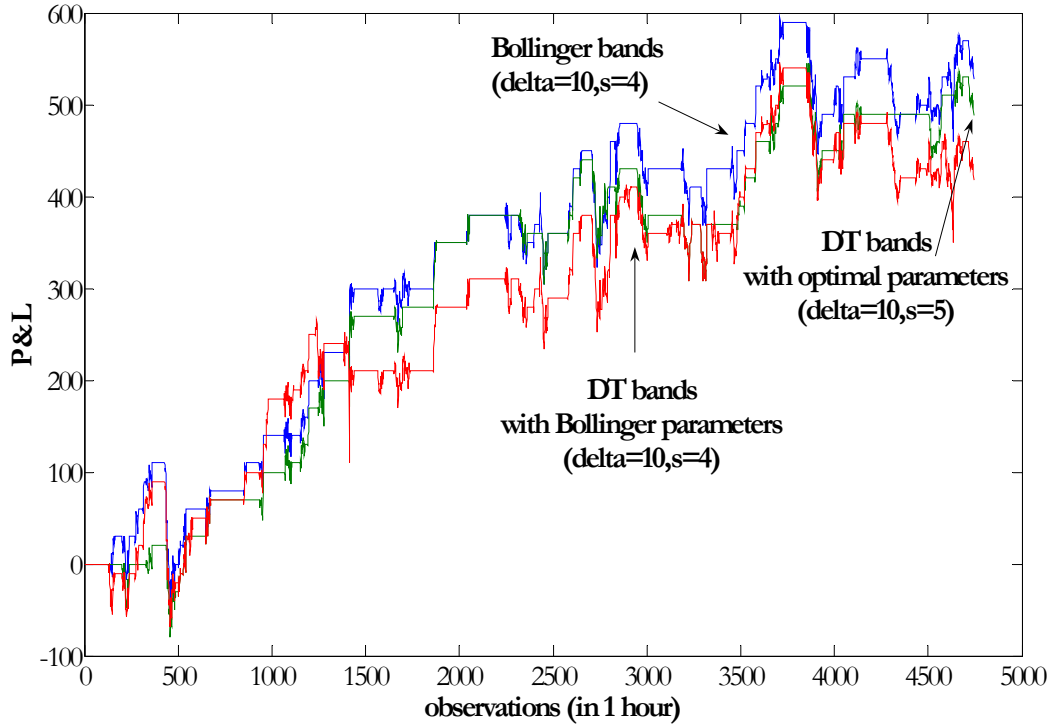


Figure 2.33. X instrument: P&L for the strategies, confirmed by momentum and based on the Bollinger and DT bands, SMA length n=50 observations

Contrary to other examples, the DT bands strategies, confirmed by the "Elliot" and "Support/Resistance" signals produce very poor results (see Table 2.10); therefore, we do not provide the graphical representation of the P&L paths.

As the result, the DT bands shows less interesting results for this particular instrument:

1. Bollinger bands, with preliminary chosen default parameter k, do not produce or produce very low profits.
2. The signals sent by the Bollinger bands and confirmed by momentum indicator lead to the highest end-of-period profits, while the P&L path is almost increasing for long SMA (50 observations).
3. The Bollinger bands produce end-of period profits for the SMAs of both lengths.
4. DT bands on their own produce more false signals and, thus, larger losses than the Bollinger bands.
5. Confirmed by momentum DT bands strategies produce comparable profits and increasing P&L path for longer SMA and higher profits with increasing P&L path for shorter SMA. At the same time, trade efficiency is the highest for the DT bands strategies. The optimal momentum parameters for the DT bands strategy are stable over both SMA lengths (the lag is the same and the thresholds are almost equal); for Bollinger bands both momentum parameters are different for different SMA lengths.
6. The DT strategies confirmed by the "Elliot" and "Support/Resistance" signals do not produce even comparable results.

5 Conclusions

This paper presents the data transformed (DT) bands as the alternative to the Bollinger bands. Although being the simple and comprehensive technical analysis instrument that can be incorporated into a successful strategy, Bollinger bands lacking the theoretical justification or explanation, why the bands should work and how to optimize its parameters, the length of the simple moving average and scaling coefficient. By providing the conditions under which the Bollinger bands can be statistically justified, we showed that usage of the traditional Bollinger method is limited to the particular case of the stationary (at least locally), symmetric and known distributions. The easing of these assumptions complicates significantly the applications and calculations of the optimal Bollinger bands. The DT bands under assumption of local stationarity provide simple, but powerful transformation of the Bollinger approach into better theoretical framework, which in addition is easier to optimize. The DT bands, obtained for the transform function calibrated for the residuals $R_i = P_i - SMA_t, SMA_t = \frac{1}{n}\sum_i P_i$, $i \in [t-n+1; t]$, has a specific stair-type form. Such DT bands might be helpful in the definition of other technical rules – the Elliot waves and Support/Resistance levels.

The numerical examples show that the DT bands are not only the instrument that is better justified theoretically, but they can also be a successful strategy component. With the exception of the artificially constructed instrument X, all other instruments shows more or less the same result: confirmed DT bands strategy usually produces higher profits, trade efficiency and increasing P&L paths than confirmed (by momentum) strategies, based on the Bollinger bands. Table 2.12 summarizes the obtained results for the optimal strategies. The hypothesis about 0-transaction costs, under which the strategies were simulated, is certainly unrealistic. However, we can get some estimates of what can happen in the real-life cases. For example, if we assume that

the estimates of the transaction costs (c) for the Bund and Brent instruments respectively are 0.35€/trade and \$5/trade, the outcomes for Bund and Brent will remain positive.

Besides of the transaction costs there is a notion of "slippage", meaning that frequently the position cannot be realized at the current market prices due to the difference in the demand and supply prices. Slippage is the difference between the expected execution and actual price. While the transaction fees are relatively fixed, the slippage varies significantly. As we can see from Table 12 if the "slippages" estimates are below the P/L per trade values the strategies are profitable.In fact for certain realistic estimates of the "slippage" we can state that optimal strategies are profitable for Bund, DAX and Brent for both SMA lengths and profitable for X instrument for longer SMA.

The P&L for the strategies based on the DT bands are almost everywhere increasing, which guarantees positive outcome at any moment of time, i.e. the profitability of our position does not depend on the moment that we have decided to end up trading; moreover, the longer we stay in game – the higher could be the outcomes.

The strategy parameters, optimized for the DT bands (momentum parameters, total steps length and tolerance level) seems to be more stable across different SMA lengths than momentum parameters for the confirmed Bollinger bands strategies; their values are less dependent on the choice of the SMA length.

The strategy that incorporates the "Elliot" and "Support/Resistance" signals does not brings the highest result. However, in many cases these strategies lead to the comparable outcomes.

Several directions of the future research can be figured out. Some research should be devoted to the analysis of the relationship between the strategy profitability and the value of K% section or k parameter of the DT bands.

The data transformation approach can be also applied to other technical strategies that are based on the extreme values. For example, in this work we have considered the symmetrical threshold for the momentum, used for the confirmation of the signals sent by the bands. However, these strategies might be ameliorated if the choice of the (optimal) momentum thresholds will be made on the basis of the data transformation approach. For this purpose the momentum data should be transformed into normal, the *K%* section should be applied and the thresholds should be back-transformed to the real data. This might lead to the asymmetric momentum thresholds.

Finally another direction should address in more details the problem of the DT bands application to the definition of other technical indicators, such as Elliot waves and Support/Resistance trading rules.

Table 2.12
Summary of the best trading outcomes for different strategies

	Short SMA				Long SMA			
	Bund	**DAX**	**Brent**	**X**	**Bund**	**DAX**	**Brent**	**X**
SMA length	50	50	50	20	115	115	115	50
Winning strategy	Strategy #2: **DT bands**, confirmed by Momentum	Strategy #2: **DT bands**, confirmed by Momentum	Strategy #2: **DT bands**, confirmed by Momentum	Strategy #2: **DT bands**, confirmed by Momentum	Strategy #4: **DT bands**, confirmed by "Elliot" and "Support/ Resistance" signals	Strategy #2: **Bollinger bands**, confirmed by Momentum	Strategy #2: **DT bands**, confirmed by Momentum	Strategy #2: **Bollinger bands**, confirmed by Momentum
P&L	9880	54275	9090	290	3740	15325	18720	528
# of trades	224	387	38	90	96	185	39	105
P&L/trade	44	140	239	3	39	83	480	5
Monotonicity	Monotone	Monotone	Monotone	Monotone	Monotone	Non-monotone	Monotone	Non-monotone
Tick	0.01	0.5	0.01	1	0.01	0.5	0.01	1
Tick value	10	12.5	10	10	10	12.5	10	10
P&L/trade (in number of ticks	4.4	11.2	23.9	0.3	3.9	6.6	48	0.5
Strategy profitable?	yes	yes	yes	No	yes	yes	yes	yes

6 Appendices II

Appendix A

Example 1: Bund

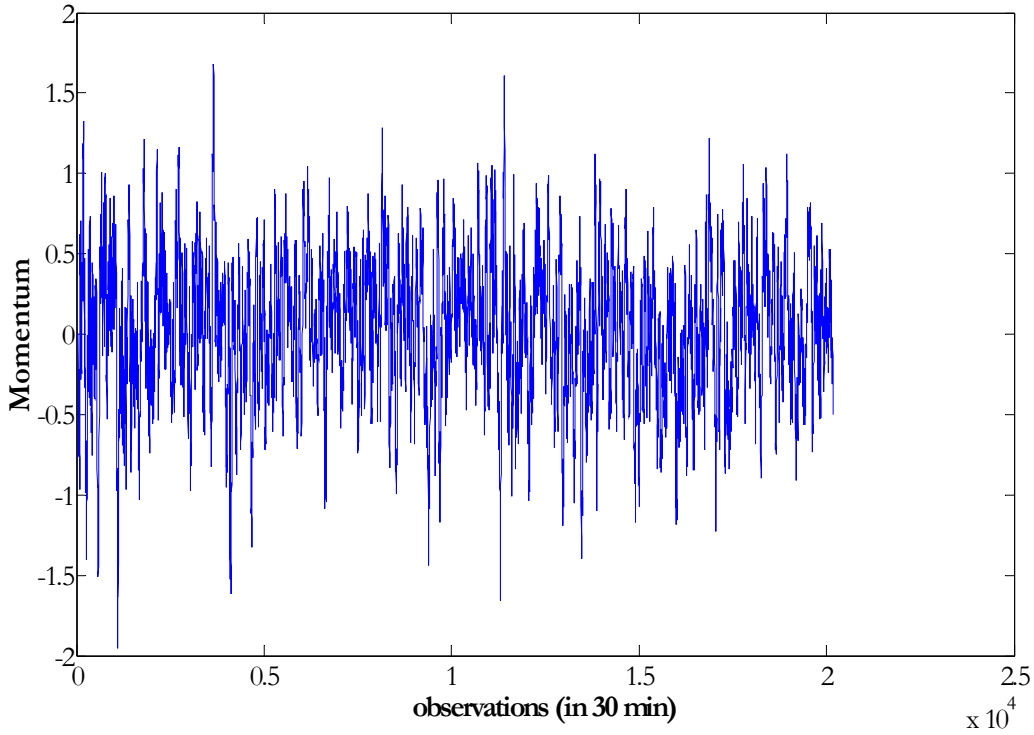


Figure A1. Bund (30/7/2003-7/12/2006, frequency 30 minutes): Momentum indicator (lag Δ=50 observations).

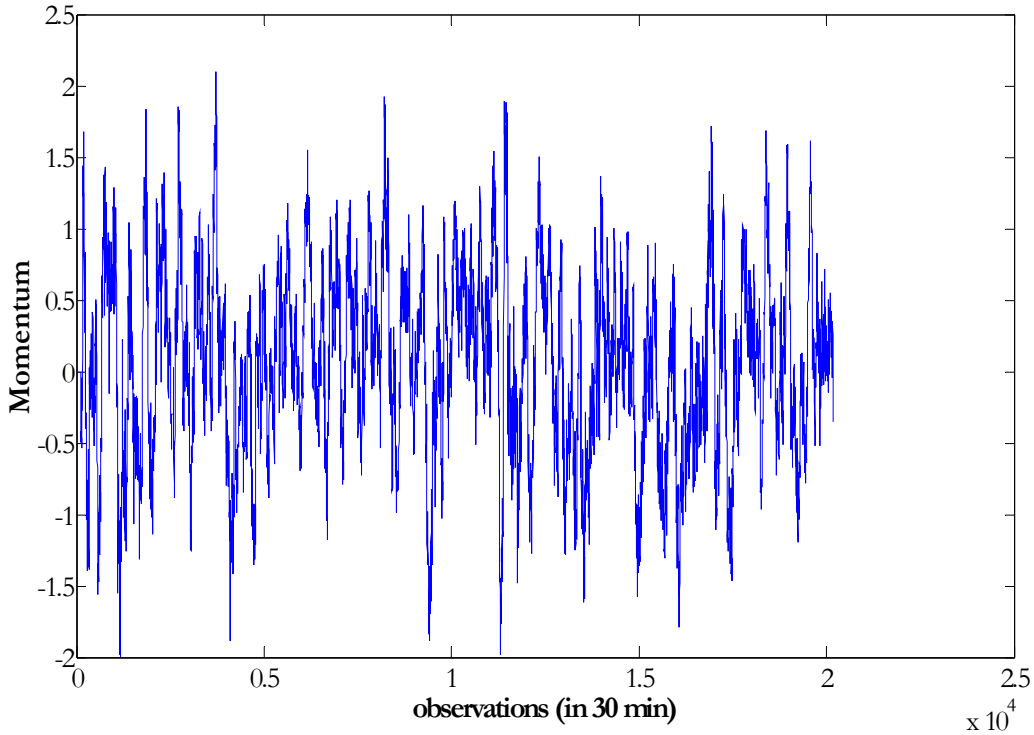


Figure A2. Bund (30/7/2003-7/12/2006, frequency 30 minutes): Momentum indicator (lag Δ=115 observations)

Table A1
Bund (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L values for confirmed Bollinger bands strategy for different momentum parameters, *n*(SMA)=50, *k*=2.1

Momentum parameter (delta)	momentum parameter (threshold)																				
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**	**1.1**	**1.2**	**1.3**	**1.4**	**1.5**	**1.6**	**1.7**	**1.8**	**1.9**	**2**
50	-280	580	850	3190	1030	2170	200	1930	1960	1080	1590	-580	-250	470	730	960	980	0	0	0	0
80	30	2060	3510	1360	2870	3380	3930	2410	3250	2970	2900	3850	4510	2550	750	240	210	720	850	700	0
115	-120	190	2840	2260	5670	5060	6520	7950	8150	4920	4470	3100	2930	4270	2550	1850	1690	930	1790	1510	520
150	-280	-340	2590	3010	5030	6680	8300	6760	6640	4780	3480	5900	4610	2020	1570	2850	2240	2350	2630	2040	1110
180	-240	370	3670	4070	5960	5590	5200	6930	5960	7730	7330	6570	7910	7040	5450	5180	2270	3040	2590	2040	1600

Table A2
Bund (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L values for confirmed DT bands strategy for different momentum parameters, *n*(SMA)=50, *k*=1.65

Momentum parameter (delta)	momentum parameter (threshold)																				
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**	**1.1**	**1.2**	**1.3**	**1.4**	**1.5**	**1.6**	**1.7**	**1.8**	**1.9**	**2**
50	-1880	1200	4880	5780	3040	2460	-110	-50	-1120	770	20	-470	-320	1020	680	1010	1270	250	420	420	0
80	-1510	660	590	4250	4700	1900	4320	2650	2680	920	-50	3150	3600	3070	1660	1140	820	1610	1010	700	0
115	-1340	200	2740	6430	4380	3900	5340	4610	8750	7920	6560	3160	1450	2410	2720	2710	3450	1880	2770	1830	520
150	-1550	520	5240	6440	8810	7850	6870	5910	5750	4810	5450	6640	4420	3040	1600	2960	3100	2890	2780	2130	980
180	-1480	-370	-320	800	3620	5160	6300	7190	7140	8110	7050	8840	9880	7510	7830	7890	4060	4360	3250	2680	2600

Table A3
Bund (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L values for confirmed Bollinger bands strategy for different momentum parameters, *n*(SMA)=115, *k*=2.1

Momentum parameter (delta)	momentum parameter (threshold)																				
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**	**1.1**	**1.2**	**1.3**	**1.4**	**1.5**	**1.6**	**1.7**	**1.8**	**1.9**	**2**
50	-3960	-3980	-2570	-1690	-1750	-3480	-4630	-5870	-7080	-3020	-1360	-380	220	150	1170	1230	1010	50	220	220	0
80	-3960	-3480	-3100	-1120	-490	-1600	-2590	-4800	-4070	-2800	-2200	950	1600	2200	360	620	660	1250	960	730	0
115	-3960	-3580	-3040	-2470	-2460	-5380	-4050	-3480	-770	-3930	-1970	-760	-540	430	990	1140	390	140	830	650	0
150	-4370	-4140	-4890	-4700	-2610	-2470	-2560	-460	-2350	-4100	-3870	-2510	-4240	-3320	-2530	-930	930	700	920	1010	490
180	-4040	-2720	-3640	-1260	-980	-190	-3570	-4680	-4500	-4970	-4110	-1320	-210	-1540	-990	1020	700	1850	1790	880	780

Table A4
Bund (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L values for confirmed DT bands strategy for different momentum parameters, *n*(SMA)=115, *k*=2.1

Momentum parameter (delta)	momentum parameter (threshold)																				
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**	**1.1**	**1.2**	**1.3**	**1.4**	**1.5**	**1.6**	**1.7**	**1.8**	**1.9**	**2**
50	-4360	-4310	-2570	-3100	-2790	-3080	-3800	-6780	-7780	-3640	-3660	-3030	-1360	250	270	750	1010	50	220	220	0
80	-4360	-4530	-2930	-3130	-2000	-4450	-3760	-6300	-6940	-6710	-5400	-840	-730	-600	-1160	-420	170	960	630	730	0
115	-4320	-3070	-1820	-2440	-600	-1540	-1620	-1530	-770	-2690	-2570	-3880	-3510	-2600	-50	700	1560	1290	2240	1220	360
150	-4330	-4390	-2670	-2470	-1460	-890	-2250	-2690	-5880	-7210	-6960	-3770	-3650	-2930	-1960	1140	2890	2380	1980	1430	1140
180	-4360	-4210	-5040	-2580	-1070	-260	-2170	-2180	-3610	-3100	-2790	10	1230	1300	1170	120	2050	3020	1900	1770	1130

Table A5
Bund (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L for confirmed by "Elliot" signal strategy, based on DT bands (Steps # = 3)

n(SMA)	total steps length parameter																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
50	140	410	-1030	5400	3980	3220	460	-150	-150	0	0	0	0	0	0	0	0	0	0	0	0
115	-3730	-3190	-4020	-2950	-2800	-2770	-5750	-2770	-1380	-3200	-810	500	800	420	-200	-200	-200	-200	-170	250	0

Table A6
Bund (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L for confirmed by "Elliot" and "Support/Resistance" signal strategy, based on DT bands (*n*(SMA)=50, Steps #=3)

tolerance SR	total steps length parameter																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
0.001	1380	1240	1000	270	-90	-110	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.002	370	510	1960	2620	1550	360	110	60	40	0	0	0	0	0	0	0	0	0	0	0	0
0.003	-6120	-6890	-4560	-1240	-1980	400	-110	-560	-190	0	0	0	0	0	0	0	0	0	0	0	0
0.004	-5850	-6840	-3860	-1120	-830	1120	470	-410	-80	60	0	0	0	0	0	0	0	0	0	0	0
0.005	-12950	-11050	-6810	-1950	-1610	370	200	-210	150	160	0	0	0	0	0	0	0	0	0	0	0
0.006	-14180	-14260	-8530	-4280	-2930	580	210	-350	10	160	0	0	0	0	0	0	0	0	0	0	0
0.007	-14960	-15200	-10540	-6580	-4130	750	540	-320	10	160	0	0	0	0	0	0	0	0	0	0	0
0.008	-13520	-14340	-11160	-7740	-5820	-460	810	-150	20	160	0	0	0	0	0	0	0	0	0	0	0
0.009	-13950	-14190	-12070	-8860	-5950	-460	810	-150	20	160	0	0	0	0	0	0	0	0	0	0	0
0.01	-15420	-14630	-12690	-6940	-4830	-180	940	-120	20	160	0	0	0	0	0	0	0	0	0	0	0

Table A7
Bund (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L for confirmed by "Elliot" and "Support/Resistance" signal strategy, based on DT bands (*n*(SMA)=115, Steps #=3)

	total steps length parameter																				
tolerance SR	**0**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
0.001	-8370	-8500	-7040	-2360	-860	-1140	-1350	970	-800	590	2940	1000	920	510	450	390	390	290	290	50	0
0.002	-8010	-7550	-4630	-90	700	700	-90	2080	2000	1330	3160	1300	1250	510	450	390	390	290	290	50	0
0.003	-9490	-9490	-5930	-2420	1020	850	-40	2260	2220	1570	3460	1430	1270	530	450	390	390	290	290	50	0
0.004	-10570	-9920	-7610	-4620	-1310	380	-840	1940	1950	1820	3740	1430	1270	530	450	390	390	290	290	50	0
0.005	-10050	-9420	-6840	-4090	-660	280	-1420	1340	1500	1300	3230	900	1290	530	450	390	390	290	290	50	0
0.006	-10020	-9400	-6780	-4030	-610	280	-1270	1340	1500	1300	3230	900	1290	530	450	390	390	290	290	50	0
0.007	-11440	-10870	-8100	-5090	-1990	-1290	-2830	-600	-710	-930	1570	-290	600	-330	450	390	390	290	290	50	0
0.008	-11150	-10640	-7290	-3970	-1670	-980	-2920	-420	-270	-520	1990	-290	600	-330	450	390	390	290	290	50	0
0.009	-12130	-11410	-7810	-4530	-2850	-1910	-3990	-1470	-1320	-1590	1930	-310	580	-350	430	370	370	370	370	50	0
0.01	-15460	-13900	-13530	-10350	-7160	-8260	-10940	-7070	-6220	-5930	-1360	-980	-1000	-1680	-1760	-1600	-1600	-1600	-470	50	0

Appendix B

Example 2: DAX

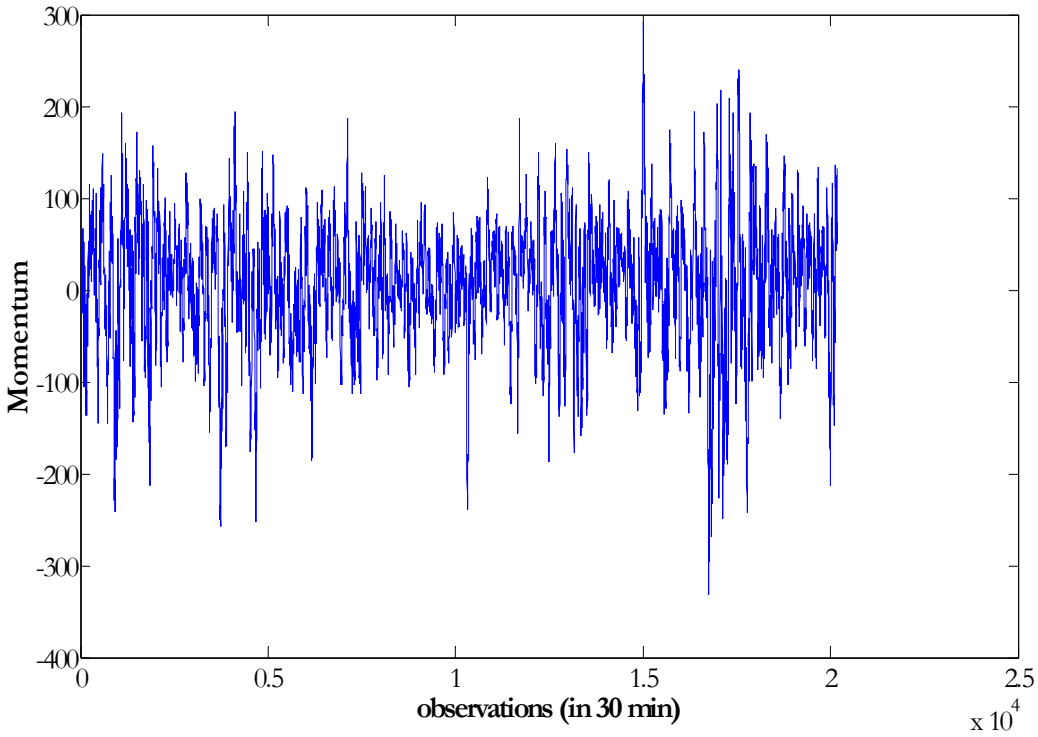


Figure B1. DAX (30/7/2003-7/12/2006, frequency 30 minutes): Momentum value (lag Δ=50 observations)

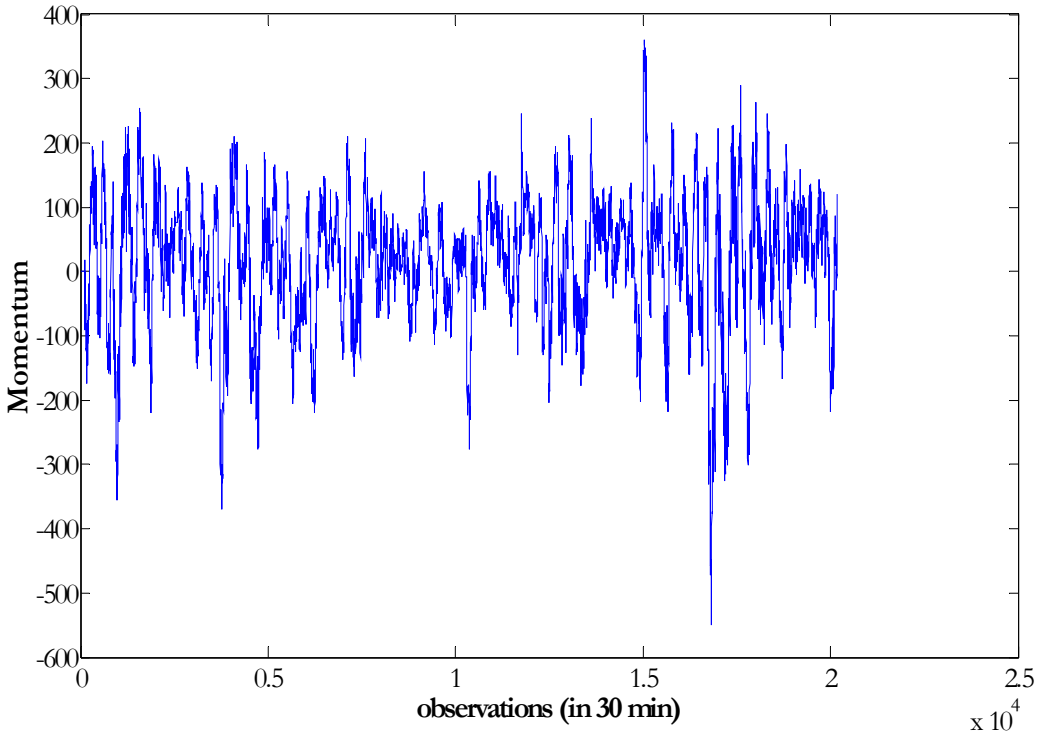


Figure B2. DAX (30/7/2003-7/12/2006, frequency 30 minutes): Momentum value (lag Δ=115 observations)

Table B1
DAX (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L values for confirmed Bollinger bands strategy for different momentum parameters, *n*(SMA)=50, *k*=2.1

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
50	1288	6100	8062	6288	6475	12125	13850	6325	8538	11475	-1925	-5838	-1238	4375	5150	9962	8450	11650	13437	14675	9500
80	1288	2175	6838	7513	10300	7737	4163	9250	2350	188	10400	15350	16175	20662	18338	9925	5675	-3513	-3913	-1950	-5562
115	1500	8875	8125	6575	9725	17137	19975	16275	18212	26350	29113	19625	19863	24075	25575	20725	20563	17138	9388	8087	5800
150	1288	4313	4413	7700	23775	17875	21550	12212	17475	27350	34438	27838	16713	17687	18613	16975	15362	10087	9150	10650	9612
180	1575	1913	4763	7413	14212	15150	21100	21987	26600	32600	40837	49175	47250	42812	35350	30838	32050	22338	19400	17550	18987

Table B2
DAX (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L values for confirmed DT bands strategy for different momentum parameters, *n*(SMA)=50, *k*=1.65

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
50	-15625	-6412	2563	6713	3413	12475	10213	2625	9187	5487	6300	312	-4413	1688	612	4225	-588	7387	8988	8050	5550
80	-15712	-12375	2425	7012	3800	-2575	-4350	4300	187	-2638	-1325	10912	12500	14850	7538	4375	2488	-2688	-3113	-3475	-5813
115	-15837	-5800	-5513	-10450	-2375	12125	15312	17537	24050	31487	28825	19375	21325	19750	19287	16950	19375	18075	12125	12162	7775
150	-15862	-14637	-14338	-13800	2575	8912	8462	7662	13137	23962	23737	31375	17975	18425	10063	8663	10388	6462	3450	4925	6638
180	-15575	-16638	-18075	-11475	-7938	50	11125	16037	22012	22575	28937	38125	43237	54275	46450	44012	43775	27850	26613	23225	23625

Table B3
DAX (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L values for confirmed Bollinger bands strategy for different momentum parameters, *n*(SMA)=115, *k*=2.1

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
50	-13250	-12887	-12000	-10550	-13412	-16062	-11475	-17912	-11625	-8675	-7962	-10262	-12762	-587	-8562	-11163	-14988	-8637	-8612	-12138	-9950
80	-13250	-12837	-12850	-14575	-14675	-13588	-12888	-6400	-6875	-5275	-4675	-4312	-3925	-1487	-5700	-3237	-4600	-4200	-6513	-12100	-10837
115	-13250	-15050	-14175	-21462	-17713	-9200	-9475	-2987	-6187	-9037	-9512	-10062	-5850	-4025	-5062	-8850	-3875	-4325	-5350	-4275	-1900
150	-13250	-12013	-13038	-7687	-2975	-3325	-12313	-10412	-4087	-4050	-11387	-15800	-15775	-9325	-11025	-10300	-8600	-10250	-12800	-7850	-9688
180	-13250	-13325	-15350	-16713	-12050	-12525	12	6613	9113	10375	15325	9775	3050	3725	2162	3888	4663	937	3162	3025	2763

Table B4
DAX (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L values for confirmed DT bands strategy for different momentum parameters, *n*(SMA)=115, *k*=2.1

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
50	-29813	-26938	-25350	-26613	-32088	-24425	-24738	-25813	-18825	-18588	-24338	-24413	-19050	-7750	-11525	-8663	-11638	-4575	-5388	-2913	-3700
80	-29813	-29138	-28713	-27425	-25363	-22650	-24313	-27950	-22650	-18125	-16387	-14912	-14112	-8600	-13713	-7738	-6300	-10363	-5000	-4238	-5425
115	-30038	-28950	-29850	-36013	-33363	-18088	-19775	-17675	-14913	-16963	-21913	-21187	-10075	-8850	-13500	-14725	-6813	-7537	-11938	-7388	-3087
150	-31300	-30713	-32575	-25588	-19388	-14413	-17013	-24400	-13913	-11663	-12000	-16875	-22600	-16062	-15225	-15725	-14238	-16225	-16537	-12987	-10362
180	-30038	-33813	-36100	-30400	-21588	-19938	-12313	-10475	-3850	-1588	2275	4800	3388	10700	2813	3138	3650	-1738	750	1200	2263

Table B5
DAX (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L for confirmed by "Elliot" signal strategy, based on DT bands (Steps # = 3)

n(SMA)	total steps length parameter																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
50	-23013	-22138	-26788	-2200	-3612	-9400	-4575	-5663	0	0	0	0	0	0	0	0	0	0	0	0	0
115	-35888	-33863	-23388	700	-1488	-3638	-5350	6412	6500	10200	2775	4850	688	738	-3838	-1775	-1775	0	0	0	0

Table B6
DAX (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L for confirmed by "Elliot" and "Support/Resistance" signal strategy, based on DT bands (*n*(SMA)=50, Steps #=3)

tolerance SR	total steps length parameter																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
0.05	763	400	-163	-2763	-800	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.10	-6738	-8763	275	-1300	688	175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.15	-15675	-19450	-9962	1950	-175	1175	363	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.20	-20337	-21813	-16250	-3600	-3763	-1238	-2663	-3138	0	0	0	0	0	0	0	0	0	0	0	0	0
0.25	-16613	-16900	-11638	-5738	913	113	-2450	-3138	0	0	0	0	0	0	0	0	0	0	0	0	0
0.30	-16587	-22912	-12200	450	4038	-7188	-4463	-6025	0	0	0	0	0	0	0	0	0	0	0	0	0
0.35	-14363	-21075	-12150	1750	5963	-7000	-4463	-6025	0	0	0	0	0	0	0	0	0	0	0	0	0
0.40	-14650	-22550	-11850	-1200	6513	-6450	-4463	-6025	0	0	0	0	0	0	0	0	0	0	0	0	0
0.45	-9500	-18788	-11300	-113	9550	-3500	-3088	-6025	0	0	0	0	0	0	0	0	0	0	0	0	0
0.50	-6375	-14838	-6012	2812	7937	-6287	-3088	-6025	0	0	0	0	0	0	0	0	0	0	0	0	0

Table B7
DAX (30/7/2003-7/12/2006, frequency 30 minutes): End of period P&L for confirmed by "Elliot" and "Support/Resistance" signal strategy, based on DT bands (*n*(SMA)=115, Steps #=3)

	Total steps length parameter																				
tolerance SR	**0**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
0.05	-36362	-39425	-26463	-5650	-8175	-2850	2475	4550	-800	4125	13	-150	-100	-388	650	725	713	0	0	0	0
0.10	-51088	-54913	-43775	-22500	-19400	-13850	-8450	1925	-1075	5925	463	438	-1325	-1500	-1300	-2488	-2700	0	0	0	0
0.15	-52438	-56288	-43812	-20275	-13575	-9537	-5138	3025	363	7113	1650	1363	88	-1500	-1300	-2488	-2700	0	0	0	0
0.20	-48138	-51938	-40412	-17350	-10800	-7112	475	9863	6825	13912	7700	5650	2188	600	338	-863	-1013	863	0	0	0
0.25	-45675	-49775	-38425	-16275	-9900	-5437	2150	11425	7675	14225	8888	6213	2713	613	350	-850	-1013	863	0	0	0
0.30	-48900	-53000	-41650	-19725	-13787	-8512	-125	9338	4025	12350	5513	2813	-688	-2788	-2400	-688	-1013	863	0	0	0
0.35	-51950	-55975	-42200	-21575	-15050	-8100	-987	8688	2500	11550	8000	5300	1800	-300	88	1800	1200	2225	1363	700	1575
0.40	-50175	-53750	-43400	-23425	-16463	-9512	-2400	7275	1425	11550	8000	5300	1800	-300	88	1800	1200	2225	1363	700	1575
0.45	-51962	-55275	-43487	-23188	-16463	-9512	-2400	7275	1425	11550	8000	5300	1800	-300	88	1800	1200	2225	1363	700	1575
0.50	-55150	-58150	-44550	-23263	-16200	-7700	-2087	6688	-238	9487	7687	5300	1800	-300	88	1800	1200	2225	1363	700	1575

Appendix C

Example 3: Brent

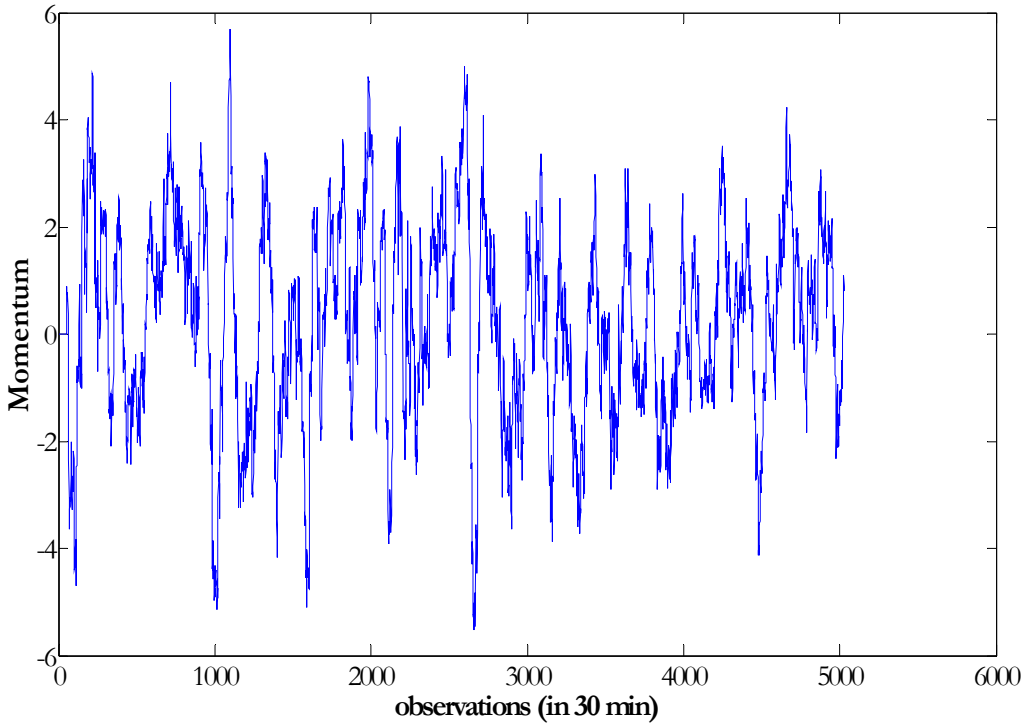


Figure C1. Brent (17/12/04-27/01/06, frequency - 30 minutes): Momentum values (lag Δ=50 obs)

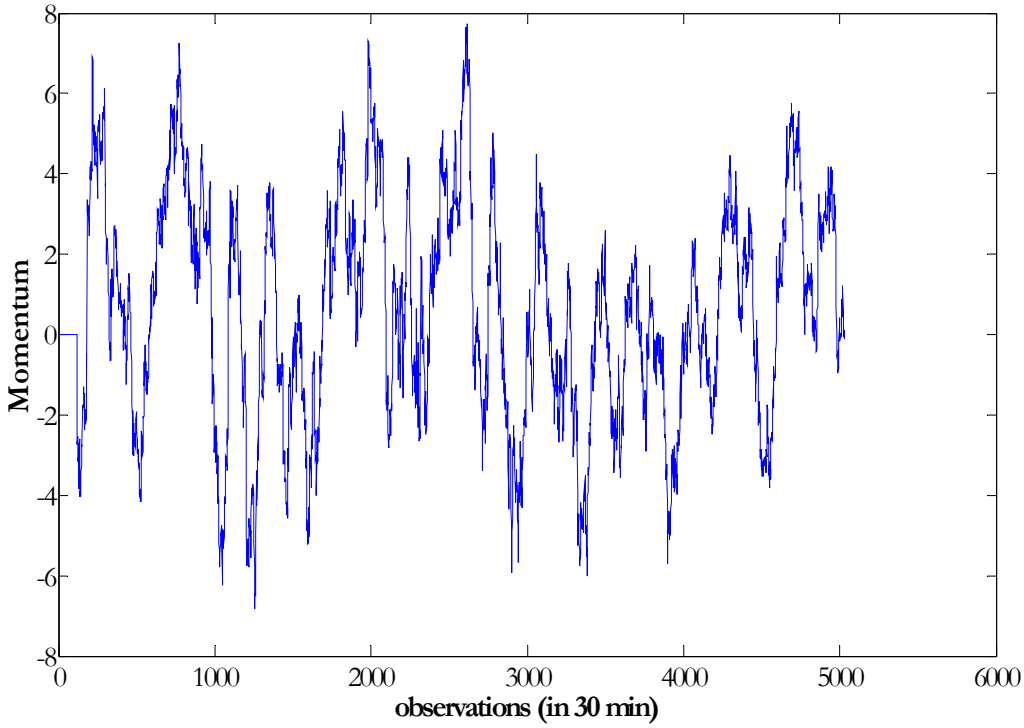


Figure C2. Brent (17/12/04-27/01/06, frequency - 30 minutes): Momentum values (lag Δ=115 obs)

Table C1

Brent (17/12/04-27/01/06, frequency - 30 minutes): End of period P&L values for confirmed Bollinger bands strategy for different momentum parameters, *n*(SMA)=50, *k*=2.1

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0
20	-16440	-17630	-18310	-14690	-10590	-5650	-7300	-4990	-6680	-8660	-7000	-5110	-280	-10	590	290	1780	2660	2700	600	-330
50	-16440	-16830	-15190	-14440	-11710	-11610	-6860	-6010	-10270	-7150	-3530	-6050	-5220	-6620	-3490	770	2820	3500	4350	2740	1050
90	-16440	-14390	-11520	-16170	-17480	-12480	-10360	-6190	-7670	-8640	-4490	-5380	-6590	-3840	2260	5520	4130	4310	450	-630	-2890
115	-16440	-15000	-13990	-13430	-11330	-9850	-8680	-13550	-15120	-14810	-11460	-9500	-5650	-7000	-6750	-5010	-2800	520	-960	4000	5290
150	-16440	-16140	-15710	-17670	-13400	-11930	-7430	-4910	-6220	-6540	-7120	-4400	-3830	-4190	-7640	-7270	-7080	-5890	-10520	-8320	-7010

Table C2

Brent (17/12/04-27/01/06, frequency - 30 minutes): End of period P&L values for confirmed DT bands strategy for different momentum parameters, *n*(SMA)=50, *k*=1.65

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0
20	-15820	-18060	-18430	-16650	-14820	-11070	-8870	-5840	-9430	-11880	-6820	-2690	1020	2800	3380	3080	3880	4120	2050	600	-330
50	-15970	-14690	-13400	-15970	-12790	-9070	-7260	-7750	-7850	-3280	1050	-3400	-880	-8380	-4940	-1830	2850	4310	9090	7220	4270
90	-15820	-13740	-13720	-15480	-14010	-9800	-10750	-8840	-11160	-10350	-5370	-7500	-3870	-3060	-930	-1400	-1070	3670	-2290	-680	-6200
115	-15820	-15330	-14070	-10340	-9400	-7130	-3480	-12130	-14320	-11260	-12660	-9410	-8650	-7090	-8500	-5030	-3840	-870	-1210	640	2080
150	-15820	-16540	-16130	-14680	-10750	-12410	-4490	-4990	-830	-2560	410	-60	-1470	-1780	-5230	-4900	-5300	-5730	-11030	-9920	-9240

Table C3

Brent (17/12/04-27/01/06, frequency - 30 minutes): End of period P&L values for confirmed Bollinger bands strategy for different momentum parameters, *n*(SMA)=115, *k*=2.1

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0	3.2	3.4	3.6	3.8	4.0
20	-7870	-7870	-7870	-6720	-5990	-4990	-4860	-390	-2550	-120	1140	-1580	5000	9170	8310	4470	3530	2490	3200	1880	970
50	-7870	-7640	-7310	-7250	-6880	-7450	-5180	-6270	-5620	-5190	-3610	-3060	60	-2820	600	4750	7640	11150	12410	12150	9840
90	-7870	-7870	-6630	-6240	-5900	-5390	-6530	-7280	-7950	-7880	-5850	-6570	-4960	-5630	-4310	-2000	-2200	-430	-2040	-2480	-1710
115	-7870	-7740	-9140	-9140	-9680	-10420	-11800	-11860	-9270	-9300	-6360	-3740	-4070	-3250	-500	-2180	-1070	1420	-510	1220	3060
150	-7870	-8340	-7970	-6510	-7010	-11680	-12670	-13860	-12760	-13490	-10850	-11650	-11970	-12260	-11960	-10300	-12680	-11860	-10590	-9080	-6760

Table C4
Brent (17/12/04-27/01/06, frequency - 30 minutes): End of period P&L values for confirmed DT bands strategy for different momentum parameters, *n*(SMA)=115, *k*=2.1

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**	**1.8**	**2.0**	**2.2**	**2.4**	**2.6**	**2.8**	**3.0**	**3.2**	**3.4**	**3.6**	**3.8**	**4.0**
20	-1470	-1330	-1330	-900	180	1170	3160	3530	-680	3250	5310	2590	7440	10670	11020	7180	6240	5200	5910	4590	3680
50	-1470	-3070	-3290	-3310	-2570	-1940	-770	-2710	-2420	-800	1480	-2540	2320	4670	8300	11940	14570	18720	18340	18080	12940
90	-1470	-1420	-1780	-330	-2810	-1070	-3180	-910	-540	-1190	-390	950	2120	1420	-200	1160	2420	3920	1460	1720	-1540
115	-1470	-680	-1520	-40	1210	570	-2730	-8020	-11060	-14300	-10610	-6480	-9130	-7080	-7690	-3940	-350	2400	4830	5730	7410
150	-1470	-770	-320	990	2320	850	-150	-2590	360	-2930	-1030	320	-2440	-1640	-900	200	-170	-220	-4640	-2690	-2290

Table C5
Brent (17/12/04-27/01/06, frequency - 30 minutes): End of period P&L for confirmed by "Elliot" signal strategy, based on DT bands (Steps # = 3)

n(SMA)	total steps length parameter																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
50	-11430	-14080	-8930	-9090	-8930	-3440	1130	0	0	0	0	0	0	0	0	0	0	0	0	0	0
115	-2850	-2620	-3260	-1210	1630	3690	6490	4960	3790	2850	4230	0	0	0	0	0	0	0	0	0	0

Table C6
Brent (17/12/04-27/01/06, frequency - 30 minutes): End of period P&L for confirmed by "Elliot" and "Support/Resistance" signal strategy, based on DT bands (*n*(SMA)=50, Steps #=3)

tolerance SR	total steps length parameter																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
0.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.002	-2120	-2500	-210	-1680	-1690	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.003	-18480	-20540	-15540	-9620	-9670	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.004	-17340	-19400	-15540	-9620	-9670	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.005	-12450	-17560	-14780	-8560	-8360	730	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.006	-13270	-16150	-15660	-11110	-11720	-1420	760	1220	1220	0	0	0	0	0	0	0	0	0	0	0	0
0.007	-13270	-16150	-15660	-11110	-11720	-1420	760	1220	1220	0	0	0	0	0	0	0	0	0	0	0	0
0.008	-13270	-16150	-15660	-11110	-11720	-1420	760	1220	1220	0	0	0	0	0	0	0	0	0	0	0	0
0.009	-11080	-13290	-15860	-10690	-11720	-1420	760	1220	1220	0	0	0	0	0	0	0	0	0	0	0	0
0.01	-13550	-12440	-13850	-6630	-8960	-760	760	1220	1220	0	0	0	0	0	0	0	0	0	0	0	0

Table C7
Brent (17/12/04-27/01/06, frequency - 30 minutes): End of period P&L for confirmed by "Elliot" and "Support/Resistance" signal strategy, based on DT bands (*n*(SMA)=115, Steps #=3)

	total steps length parameter																				
tolerance SR	**0**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
0.001	-12290	-12410	-15000	-14400	-17710	-17210	-17740	-17070	-2580	-3800	1150	950	0	0	0	0	0	0	0	0	0
0.002	-23430	-23310	-22920	-19650	-22300	-18240	-18720	-19630	-5570	-6250	1150	950	0	0	0	0	0	0	0	0	0
0.003	-22200	-21750	-19640	-16370	-19520	-15460	-17220	-18420	-4150	-5870	1530	950	0	0	0	0	0	0	0	0	0
0.004	-22200	-21750	-20470	-17190	-20230	-15460	-17220	-18420	-4150	-5870	-760	950	0	0	0	0	0	0	0	0	0
0.005	-20820	-20730	-17440	-15040	-20230	-15460	-17220	-18420	-4150	-5870	-760	950	0	0	0	0	0	0	0	0	0
0.006	-20550	-20460	-15350	-12540	-17900	-14320	-16650	-17850	-3580	-5300	-190	1790	0	0	0	0	0	0	0	0	0
0.007	-19630	-19540	-14430	-11680	-16770	-12300	-15360	-16560	-2800	-4880	520	1790	0	0	0	0	0	0	0	0	0
0.008	-19630	-19540	-14430	-11680	-16770	-12300	-15360	-16560	-2800	-4880	520	1790	0	0	0	0	0	0	0	0	0
0.009	-19720	-19630	-14520	-11680	-16770	-12300	-15360	-16560	-2800	-4880	520	1790	0	0	0	0	0	0	0	0	0
0.01	-19280	-19090	-13650	-10910	-16810	-12340	-15360	-16560	-2800	-4880	520	1790	0	0	0	0	0	0	0	0	0

Appendix D

Example 3: Instrument X

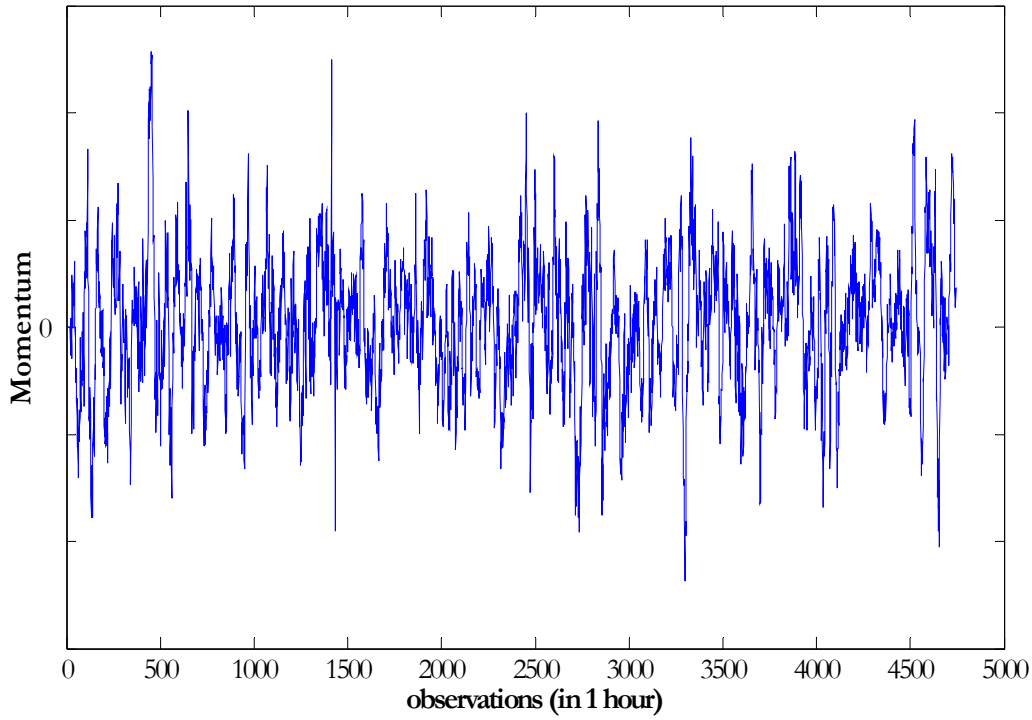


Figure D1. Instrument X (frequency=1 hour): Momentum values (lag Δ=20 observations)

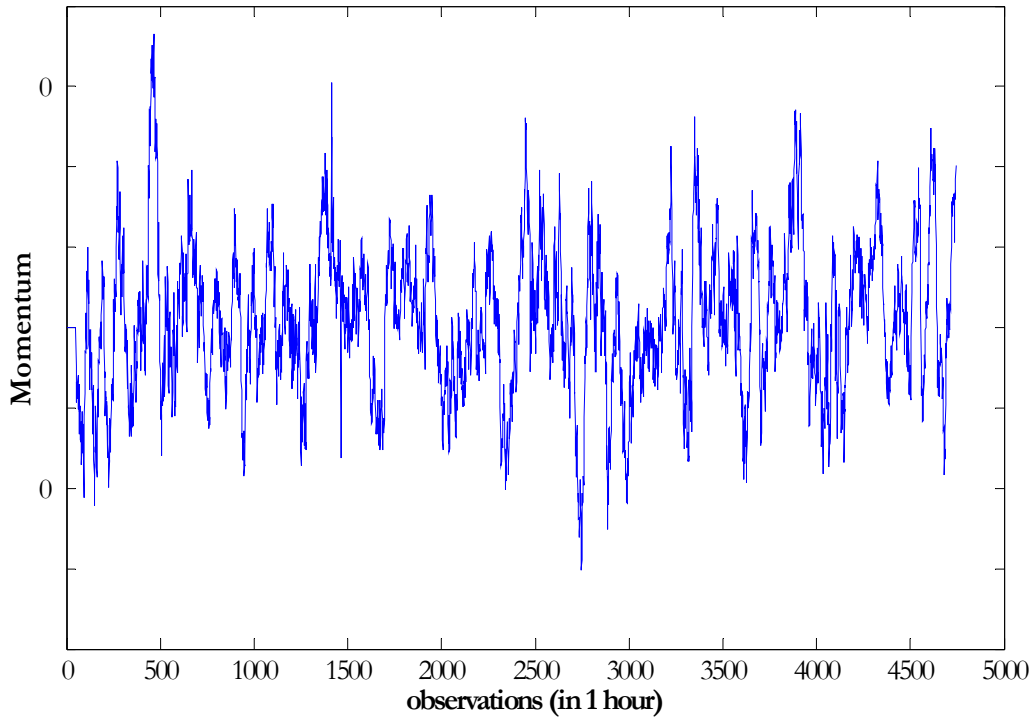


Figure D2. Instrument X (frequency=1 hour): Momentum values (lag Δ=50 observations)

Table D1

Instrument X (frequency 1 hour): End of period P&L values for confirmed Bollinger bands strategy for different momentum parameters, *n*(SMA)=20, *k*=2

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
10	-432	-312	-150	130	220	170	80	60	70	20	70	90	90	0	0	0	0	0	0	0	0
20	-432	-352	-180	-200	30	60	210	240	100	90	90	70	90	0	0	0	0	0	0	0	0
50	-432	-302	-372	-282	-32	18	118	128	48	8	58	150	140	100	90	90	0	0	0	0	0
80	-432	-322	-292	-262	-222	-192	-72	38	48	70	170	200	140	110	150	150	90	70	70	70	90
100	-432	-352	-322	-262	-92	68	98	238	258	258	158	180	180	130	100	90	90	110	110	90	90

Table D2

Instrument X (frequency 1 hour): End of period P&L values for confirmed DT bands strategy for different momentum parameters, *n*(SMA)=20, *k*=1.65

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
10	-1182	-912	-510	-210	230	290	60	80	70	40	90	110	110	0	0	0	0	0	0	0	0
20	-1192	-892	-640	-340	-80	-70	130	220	120	100	40	60	100	0	0	0	0	0	0	0	0
50	-1192	-992	-1032	-762	-322	-22	-52	108	88	118	68	120	140	100	80	100	10	10	0	0	0
80	-1192	-1152	-902	-802	-662	-472	-352	-152	-172	-40	70	190	150	90	160	160	110	100	100	100	90
100	-1192	-1132	-1032	-822	-642	-542	-372	-152	38	8	-42	60	40	80	80	100	90	140	140	120	90

Table D3

Instrument X (frequency 1 hour): End of period P&L values for confirmed Bollinger bands strategy for different momentum parameters, *n*(SMA)=50, *k*=2.1

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
10	188	218	288	328	528	518	160	100	220	110	110	120	120	0	0	0	0	0	0	0	0
20	188	138	88	128	368	228	378	368	348	290	140	140	110	0	0	0	0	0	0	0	0
50	188	218	288	458	248	308	308	288	198	150	190	220	150	150	90	110	20	0	0	0	0
80	188	238	178	-2	-42	88	198	308	268	198	200	200	190	220	150	150	110	110	110	70	90
100	188	188	148	168	350	360	310	400	380	290	230	250	310	280	180	160	100	160	160	110	90

Table D4

Instrument X (frequency 1 hour): End of period P&L values for confirmed DT bands strategy for different momentum parameters, *n*(SMA)=50, *k*=2.1

Lag (Momentum parameter)	Threshold (momentum parameter)																				
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
10	-12	78	108	148	418	488	180	100	220	110	140	120	120	0	0	0	0	0	0	0	0
20	-12	-42	-42	48	188	88	378	288	368	270	120	120	110	0	0	0	0	0	0	0	0
50	-12	8	78	228	218	328	258	288	198	188	68	190	180	160	90	110	20	20	0	0	0
80	-12	-122	-72	-272	-212	-22	138	238	128	98	210	250	210	220	170	170	140	150	150	110	90
100	-12	-42	-52	-62	68	68	178	278	398	268	138	120	90	200	150	160	120	200	200	150	90

Table D5

Instrument X (frequency 1 hour): End of period P&L for confirmed by "Elliot" signal strategy, based on DT bands (Steps # = 3)

n(SMA)	total steps length parameter																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
20	-842	-842	-842	-842	-842	-842	-842	-842	-872	-712	-642	-632	-452	-422	-362	-100	-140	-170	-200	-190	-160
50	-2	-162	-42	-150	-90	-100	-10	-20	10	0	0	0	0	0	0	0	0	0	0	0	0

Table D6

Instrument X (frequency 1 hour): End of period P&L for confirmed by "Elliot" and "Support/Resistance" signal strategy, based on DT bands (*n*(SMA)=20, Steps #=3)

tolerance SR	total steps length parameter																				
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
0.10	-2300	-2070	-700	-130	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.20	-2982	-2522	-840	-150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.30	-3192	-2722	-1060	-170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.40	-3352	-2832	-1150	-170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.50	-3502	-3012	-1140	-170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.60	-3532	-3062	-1140	-170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.70	-3602	-3132	-1150	-170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.80	-3682	-3212	-1170	-170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.90	-3732	-3272	-1210	-170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.00	-3762	-3292	-1210	-170	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabl e D7

Instrument X (frequency 1 hour): End of period P&L for confirmed by "Elliot" and "Support/Resistance" signal strategy, based on DT bands (*n*(SMA)=50, Steps #=3)

	total steps length parameter																			
tolerance SR	**0**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**	**160**	**170**	**180**	**190**	**200**
0.10	-1382	-1412	-822	-550	-260	-200	-80	-70	-70	0	0	0	0	0	0	0	0	0	0	0	0
0.20	-1842	-1762	-1462	-1120	-770	-380	-100	-110	-70	0	0	0	0	0	0	0	0	0	0	0	0
0.30	-1982	-1922	-1532	-1190	-840	-380	-110	-120	-80	0	0	0	0	0	0	0	0	0	0	0	0
0.40	-2082	-2002	-1612	-1190	-860	-430	-130	-130	-80	0	0	0	0	0	0	0	0	0	0	0	0
0.50	-2122	-2042	-1632	-1190	-880	-450	-160	-130	-80	0	0	0	0	0	0	0	0	0	0	0	0
0.60	-2122	-2042	-1632	-1190	-880	-450	-160	-130	-80	0	0	0	0	0	0	0	0	0	0	0	0
0.70	-2122	-2022	-1622	-1160	-880	-450	-160	-130	-80	0	0	0	0	0	0	0	0	0	0	0	0
0.80	-2142	-2012	-1592	-1140	-860	-430	-160	-130	-80	0	0	0	0	0	0	0	0	0	0	0	0
0.90	-2142	-2012	-1592	-1140	-860	-430	-160	-130	-80	0	0	0	0	0	0	0	0	0	0	0	0
1.00	-2132	-2002	-1582	-1130	-850	-430	-160	-130	-80	0	0	0	0	0	0	0	0	0	0	0	0

Part III. Disjunctive kriging in finance: a new approach to construction and evaluation of trading strategies

Introduction

The definition of technical trading strategies frequently involves signals, trigged by breaching predefined thresholds by price or other indicators. The examples of such strategies were given in part II. In fact, almost all trading bands strategies can be reformulated in the format of "breach the threshold" definition.

Let $\{X_t\}_{t>0}$ be some price process. $\delta(\theta(t))$ defines the threshold level that depends on some parameter (or vector of parameters) θ. For example, for Bollinger bands this parameter is defined by price volatility. Moving average can be represented as some function g of past prices: $MA_t = g(X_t, X_{t-1},...)$. Then trading bands are defined by $MA_t = g(X_t, X_{t-1},...)$ as the middle line and $MA_t \pm \delta(\theta(t))$ as the upper/lower bands. The bands trigger first trading signal at moment τ, such that:

$$\tau = \inf_t \{t \in [0,+\infty) : |X_t - MA_t| \geq \delta(\theta_t)\}.$$

Let's consider a simple trend-following strategy that states to buy ($u = 1$) the stock when its price reaches the upper band and sell ($u = -1$) if price reaches the lower band. The position is liquidated at some future time moment T. The thresholds (bands) in this strategy serve as the confirmation of the established upward or downward trends. The trading outcome under zero-transaction cost assumption, is then defined as:

$$\pi = u \cdot (X(T) - X(\tau)).$$

If the bands are breached the trader would be always better of taking the position earlier at $\tilde{t} < \tau$:

$$\tilde{\pi} = \pi + u \cdot (X(\tau) - X(\tilde{t})).$$

In the case of "correct" prediction of the trend the trader would generate larger profits: $\tilde{\pi} > \pi > 0$. In the case when the trend, was predicted incorrectly the trader would generate smaller losses: $\pi < \tilde{\pi} < 0$.

We can see that better strategy should not be solely based on the optimal (in the terms of parameters) trading bands, but also involves the prediction that the bands will be breached. This forecasting problem can be solved by different approaches.

The first approach is to estimate (predict) the value of price $X^*(t+\Delta)$ or residuals (price-MA) $R^*(t+\Delta)$ and compare the forecast with the respective threshold $MA^*(t+\Delta) \pm \theta^*(t+\Delta)$ or $\theta^*(t+\Delta)$. Note that as definition of the threshold sometimes involves the information about price, the threshold estimates can also depend on the price estimates. Different prediction method exists in financial literature. Among them are ARIMA (Box-Jenkins analysis), ARCH,

GARCH and state space models. The references to the models can be found in Greene (2007), Box, Jenkins, Reinsel (2008), Harvey, Koopman, Shephard (2004), Harvey (1991).

Another approach is to solve the optimal stopping problem for τ. Then if the optimal stopping time is below our investment time horizon $\tau < T^*$ the respective trading position can be taken.

Finally, the third approach is to estimate the indicator function[1] that the threshold is breached. Prediction of the indicator function coincides with the estimation of the probability that the threshold is breached at some point of time in future: $P_X\left(\Delta, \theta \middle| X_t, X_{t-1}, ...\right)$ or $P_R\left(\Delta, \theta \middle| R_t, R_{t-1}, ...\right)$.

In the context of trading band strategy, frequently we do not need to know the absolute future price values or precise moment of time when the threshold is reached. In order to construct such strategy, it is sufficiently to know that the threshold will be broken during some future intervals of time. Thus, the estimation of the probability of breaching the threshold can be applied directly to the strategy formulation.

A conditional probability distribution function is frequently used in financial applications for the estimation of stochastic model parameters by the maximum-likelihood or for the evaluation of derivatives prices. Popular financial models such as Black and Scholes (1973), Vasicek (1977) rely on the closed-form expression for the distribution function. However, for many other financial models the closed-form expressions do not exist. Among others methods that are used for the estimation of the parameters of stochastic processes are generalized method of moment (Hansen and Sheinkman, 1995), non-parametric density matching (Air-Sahalia, 1996). One of the approaches to the estimation of the conditional probability is to find a numeric solution to the Kolmogorov partial differential equation (Lo, 1988). Another approach is based on the Monte-Carlo simulations of models paths for more finely re-sampled data set (Honoré, 1997; Santa-Clara, 1995). The problem with these approaches is that we need to know or predefine a stochastic model of a process in advance. They also considered the Markov processes in their applications. These approaches also do not provide the closed-form expression for probability calculations.

Ait-Sahalia (1999, 2002) proposes a method to produce an accurate approximation of the unknown transition function for the Markov processes to calculate the maximum-likelihood function. This approach is based on the normalization of the diffusion X_t with an unknown distribution by a variance; the obtained diffusion Y_t with unit variance has the transition density that can be represented as an expansion in the closed-form expression. Then the transition density P_Y can be transform into the transition density P_X. Ait-Sahalia (1999, 2002) uses the Hermite expansion of the P_Y density around the Gauss density function. He shows that this approach works well for the estimation of the parameters of the classical interest rate models.

In the context of the construction of trading algorithms, the absence of the closed-form solutions complicates significantly the usage of the approach on the routine basis, as well as increases the time-intensity of its implementation. In this part III we propose a method, based on the disjunctive kriging (DK) approach to estimate the conditional probabilities

[1] The indicator function $I(\cdot)$ is defined as following: $I\left(V(x) \le i\right) = \begin{cases} 1, \forall V(x) \le i \\ 0, otherwise \end{cases}$

$P\left(Z_{t+\Delta} < z_c \middle| Z_t, Z_{t-1}, ..., Z_{t-m}\right)$. We do not limit our research to only Markov processes. The disjunctive kriging (DK) approach was developed by Matheron in 1970s and has been widely used by the geostaticians since.

The DK method, first published in 1976 in Matheron's paper "A simple substitute for conditional expectation: the disjunctive kriging", allows producing an optimal estimator for non-linear functions, such as a probability or an indicator function. It can be applied to both normal and non-normal data; and it can produce the estimator for both Markov and non-Markov processes. The DK method allows estimation the probability of breaching the threshold not only at some particular point of time in future $t+\Delta$, but also during some future interval of time $\left[t+\Delta_1; t+\Delta_2\right]$. The latest has even more appeal for strategy construction, as frequently for traders it is sufficiently to know only that a threshold will be breached (sooner or later) in near future.

One of the first applications of the DK method, performed by Orfeuil (1977), was the evaluation of the probability that certain thresholds of the air pollution, defined as critical, would be breached during following hours. Further the method was also applied to the mining industry, geology, meteorology, environmental studies, etc. (see for example, Dai, Wei, Wang, 2007; Emery, 2006; Triantafilis, Odeh, Warr, Ahmed, 2004; von Steiger, Webster, Schulin, Lehmann, 1996). It should be noted that probability expansion by Hermite polynomials used by Ait-Sahalia (1998, 2002) for the calculation of the transition densities is a particular case of the DK method when the transformed variable follows isofactorial Gaussian model. At the same time, to our best knowledge nobody tries to apply this method to the construction of trading strategies.

The objective of this part III is to show that this approach can be used in finance to improve trading strategies that incorporate different thresholds. First, we present the theory behind the DK method in Chapter 1. Chapter 2 discusses peculiarities of financial data and their impact on the DK procedure; it also presents the way to adjust for these peculiarities. Chapter 3 presents the examples of DK probabilities. Chapter 4 provides the deep analysis of the interval DK probabilities, i.e. the probabilities that the threshold will be breached during some future interval of time. Chapter 5 presents the outcomes of the trading strategies that incorporate the DK probabilities. Finally, the results of the paper are summarized in the Conclusions.

1 Theory

Suppose that we want to predict some random variable (for example, price) $Z(T)$ or the value of linear function $l(Z(T))$ at some future point of time from the available data Z_α on some interval of time $\alpha : t_\alpha \in [t_0; t_n]$. As have been shown in Part I, we can use linear methods, such as the kriging, to obtain a linear predictor for $Z(t)$. However, if we want to get a prediction for some non-linear function $\eta(Z(T))$ of $Z(T)$, which is the case of the probability estimation, it is not optimal to apply linear prediction methods to the $\eta(Z(t_1)), \eta(Z(t_2)), ..., \eta(Z(t_n))$. Instead, a non-linear prediction method, such as DK, should be used.

This chapter briefly summarizes the DK method. For more detailed description interested reader can be referred to the works of Matheron (1976), Cressie (1991), Rivoirard (1994), Chilès and Delfiner (1999). Chapter 1 is split in four sub-chapters. First, we present the general case of a predictor for a non-linear function $\eta(Z(t))$. Chapter 1.2 considers the particular case of a non-

linear predictor $\eta(Y(t))$, with the normally distributed argument $Y(t)$. Chapter 1.3 shows how to apply DK method to a non-normal random variable. Chapter 1.4 presents the DK estimator of the regularized variable, such as the proportion of observations below a benchmark on some future interval of time. Finally, we consider the particular case of the application of the DK method to the variable with an exponential covariance model in Chapter 1.5.

1.1 Disjunctive kriging

Let's assume that a process Z_t is stationary with known mean and finite variance. From the statistical point of view, the best predictor for $\eta(Z(T))$ is a conditional expectation:

$$\eta^*(Z(T)) = E(\eta(Z(T))|Z(t_\alpha)) = \int \eta(z) dF_T(z|Z_\alpha), \qquad \text{(III.1.1)}$$

where F - is the conditional distribution of the Z at the moment of time T on past values Z_α from the interval $\alpha : t_\alpha \in [t_0; t_n]$.

This conditional distribution can be estimated as following:

$$F_T(z|Z_\alpha) \equiv \Pr(Z(T) \le z|Z_\alpha) = \mathrm{E}[I(Z(T) \le z)|Z_\alpha] \qquad \text{(III.1.2)}$$

where $I(\cdot)$ is an indicator function, such that:

$$I(V(x) \le i) = \begin{cases} 1, \forall V(x) \le i \\ 0, otherwise \end{cases}$$

It is not appropriate to substitute $\Pr(Z(T) \le z|Z_\alpha)$ by $\Pr(Z(T) \le z)$, as marginal distributions of $Z(T)$ have more variability than the conditional distribution $Z(T)|Z_\alpha$. The problem with the optimal predictor (III.1.1) lays in the necessity to know the *(n+1)*-dimensional distribution of $Z(T)|Z_\alpha$; it is complicate to estimate this distribution from empirical data.

The DK method allows avoiding the estimation of the distribution of $Z(T)|Z_\alpha$. According to Matheron (1976) the optimal predictor for $\eta(Z(T))$ can be approximated as a linear combinations of measurable square-integrable functions $\{f_i : i = 1,...,n\}$:

$$\eta^*(Z(T)) = \sum_{i=1}^{n} f_i(Z(t_i)) \qquad \text{(III.1.3)}$$

According to the Hilbert-space theory, optimal predictors satisfy the orthogonality property for all measurable functions $\{h_j : j = 1,...,n\}$:

$$\mathrm{E}[\{\eta(Z(T)) - \eta^*(Z(T))\} \cdot h_j(Z(t_j))] = 0$$

As the result, for the optimal functions $\{f_i : i = 1,...,n\}$ that guarantee the minimum mean-squared prediction error, the following equation holds:

$$\mathrm{E}\left[\eta(Z(T))\middle|Z(t_j)\right]=\sum_{i=1}^{n}\mathrm{E}\left[f_i(Z(t_i))\middle|Z(t_j)\right] \qquad \text{(III.1.4)}$$

Contrary to (III.1.1) the equation (III.1.4) implies that only knowledge of the bivariate distributions $\{Z(t_i),Z(t_j)\},0\le i<j\le n$ and $\{Z(T),Z(t_j)\},0\le i<j\le n$ is a necessary condition for the calculations of the predictor for a non-linear function. This is a less strict assumption than the knowledge of the *(n+1)*-dimensional distribution.

The solution for the equation (III.1.4) can be obtained if the process $Z(t)$ can be represented as a particular isofactorial model. For such model the cumulative distribution function (CDF) of the pair $(Z(s),Z(u))$ satisfies:

$$F_{1,2}(dz_1,dz_2)=\sum_{k=0}^{\infty}\nu_k(s-u)\chi_k(z_1)\chi_k(z_2)F(dz_1)F(dz_2), \qquad \text{(III.1.5)}$$

where $\{\chi_k : k=0,1,...\}$ are complete and orthonormal functions with respect to F, i.e.

$$\begin{gathered}\chi_0(z)=1\\ \int\chi_k(z)F(dz)=0,k=1,...,\infty\\ \int\chi_k^2(z)F(dz)=1,k=1,...,\infty\\ \int\chi_k(z)\chi_l(z)F(dz)=0,k\neq l,k=0,1,...,l=0,1,...\end{gathered} \qquad \text{(III.1.6)}$$

The properties listed above implies the following equations:

$$\begin{gathered}\mathrm{E}[\chi_k(Z(s))\chi_l(Z(u))]=0,k\neq l\\ \nu_0(|s-u|)=1\\ \nu_k(|s-u|)=\mathrm{cov}[\chi_k(Z(s))\chi_k(Z(u))]=corr[\chi_k(Z(s))\chi_k(Z(u))],k=1,2,...\end{gathered}$$

Let's define for $i=1,...,n$:

$$\begin{aligned}a_{ik}&=\int f_i(Z(t_i))\chi_k(Z(t_i))F(dz_i),\\ b_k&=\int\eta(Z(T))\chi_k(Z(T))F(dz_T).\end{aligned}$$

Completeness and orthonormality of $\{\chi_k : k=0,1,...\}$ imply that

$$\begin{aligned}f_i(Z(t_i))&=\sum_{k=0}^{\infty}a_{ik}\chi_k(Z(t_i)), \qquad i=1,...,n\\ \eta(Z(T))&=\sum_{k=0}^{\infty}b_k\chi_k(Z(T))\end{aligned}$$

As the result, the disjunctive-kriging predictor of the function $\eta(Z(T))$ is:

$$\eta^*(Z(T))=\sum_{i=1}^{n}\sum_{k=0}^{\infty}a_{ik}\chi_k(Z(t_i)), \qquad \text{(III.1.7)}$$

where $\{a_{ik}\}$ are the solution of the following system:

$$\nu_k(T, t_j) b_k = \sum_{i=1}^{n} \nu_k(t_i, t_j) a_{ik} \ , \ k = 0,1,...; \ j = 1,...,n \qquad \text{(III.1.8)}$$

For $k = 0$ the expression (III.1.8) is reduced to $b_0 = \sum_{i=1}^{n} a_{i0}$.

The system (III.1.8) is truncated at some value $k = K$ such that $\text{var}[Z(T)] = \sum_{k=0}^{K} b_k^2$. As the result, the system (III.1.8) has *nK* – equations with *nK*- unknowns.

As have been shown in (III.1.5), the advantage of the isofactorial models is that some of them have polynomials factors, i.e. χ_k are polynomials. The choice of the polynomials $\{\chi_k : k = 0,1,...\}$ depends on the form of the isofactorial model $F_{u,s}(dz_u, dz_s)$. Chilès, Delfiner (1999) presents the following models and their polynomials:

I. Continuous marginal distributions:
- The Gaussian model with Hermite polynomials;
- The gamma model with Laguerre polynomials;
- The beta model with Jacobi polynomials.

II. Discrete marginal distributions:
- The binomial model with Krawtchouk polynomials;
- The negative binomial model with Meixner polynomials;
- The Poisson model with Charlier polynomials;
- The discrete Jacobi type model with discrete Jacobi polynomials.

Further we consider the case of the Gaussian isofactorial model $F_{u,s}(dz_u, dz_s)$ with the Hermite polynomials for a non-linear prediction of a normal variable.

1.2. Disjunctive kriging: normal random process

The gaussian isofactorial model is the most frequently used model in non-linear geostatistical applications. Let's consider variable $Y(t)$ that is normally distributed with 0-mean and unit variance. Its probability density function g and cumulative density function G are respectively:

$$g(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

$$G(u) = \int_{-\infty}^{u} g(t) dt$$

Matheron (1976) shows that if the functions $\{\chi_k : k = 0,1,...\}$ are represented by the Hermite polynomials, the $F_{1,2}$ in (III.1.5) is defined as bivariate normal distribution with standard normal

marginal distributions and correlation coefficient ρ; i.e. the couples $\{Y(t), Y(t+h)\}$ are assumed bivariate normal with correlation $\rho = \rho(h)$ and the probability density function:

$$g_\rho(t,u) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{t^2 + u - tu\rho}{2(1-\rho^2)} \right).$$

Hermite polynomials $H_k(Y(t))$ of order *k* are defined by following formulae ($k \geq 0$):

$$H_k(y) = \frac{1}{\sqrt{k!} \cdot g(y)} \frac{d^k g(y)}{dy^k}. \qquad \text{(III.1.9)}$$

Hermite polynomials can be defined recursively by the following recurrent relationship:

$$\begin{aligned} H_0(y) &= 1 \\ H_1(y) &= -y \\ H_2(y) &= \frac{y^2 - 1}{\sqrt{2}} \\ H_{k+1}(y) &= -\frac{1}{\sqrt{k+1}} y H_k(y) - \sqrt{\frac{k}{k+1}} H_{k-1}(y) \end{aligned} \qquad \text{(III.1.10)}$$

It can be shown Hermite polynomials of order $k \geq 1$ has the properties, defined by (III.1.6):

$$\begin{gathered} \mathrm{E}[H_k(Y(t))] = \int H_k(y) g(y) dy = 0 \\ \mathrm{var}[H_k(Y(t))] = 1 \\ \mathrm{cov}[H_k(Y(t)), H_p(Y(t))] = \mathrm{E}[H_k(Y(t)) H_p(Y(t))] = 0, \forall p \neq k \geq 0 \\ \mathrm{cov}[H_k(Y(t)), H_k(Y(t+h))] = [\rho(h)]^k = \nu_k(|s-u|) = \nu_k(h) \end{gathered} \qquad \text{(III.1.11)}$$

Almost any function of $Y(t)$ can be represented in the terms of Hermite polynomials:

$$f[Y(t)] = f_0 + f_1 H_1[Y(t)] + f_2 H_2[Y(t)] + \ldots = \sum_{k=0}^{\infty} f_k H_k[Y(t)], \qquad \text{(III.1.12)}$$

with the coefficients of the expansion:

$$\begin{aligned} f_0 &= \mathrm{E}[f(Y(t))], \\ f_k &= \mathrm{E}[f(Y(t)) H_k(Y(t))] = \int f(y) H_k(y) g(y) dy. \end{aligned}$$

Let's consider the example of the indicator function $I(Y(t) < y_c)$. The expansion coefficients of this function are:

$$f_k = \int I_{y<y_c} H_k(y) g(y) dy = \int_{-\infty}^{y_c} H_k(y) g(y) dy = \begin{cases} G(y_c), k = 0 \\ \frac{1}{\sqrt{k}} H_{k-1}(y_c) g(y_c), k \geq 1 \end{cases}$$

As the result,

$$I_{Y(t)<y_c} = G(y_c) + \sum_{k \geq 1} \frac{1}{\sqrt{k}} H_{k-1}(y_c) g(y_c) H_k[Y(t)] \qquad \text{(III.1.13)}$$

or

$$I_{Y(t) \geq y_c} = 1 - I_{Y(t)<y_c} = 1 - G(y_c) - \sum_{k \geq 1} \frac{1}{\sqrt{k}} H_{k-1}(y_c) g(y_c) H_k[Y(t)]$$

The disjunctive kriging of the function of $Y(t)$ as in (III.1.12):

$$[f[Y(t)]]^{DK} = f_0 + f_1[H_1[Y(t)]]^{CK} + f_2[H_2[Y(t)]]^{CK} + ... =$$
$$= f_0 + f_1[H_1[Y(t)]]^{K} + f_2[H_2[Y(t)]]^{K} + ... = \sum_{k=0}^{\infty} f_k[H_k[Y(t)]]^{K}$$

The cokriging[2] (CK) of Hermite polynomials becomes a simple kriging (K) due to the orthogonality (independence) of the polynomials. As the result the disjunctive kriging of the non-linear function of the normally distributed variable becomes the simple kriging of the Hermite polynomials constructed for this normal variable.

The kriging of the Hermite polynomial of the order k can be defined as following:

$$[H_k(Y(t))]^{K} = \sum_{\alpha} \lambda_{k\alpha} H_k(Y_\alpha), \qquad \text{(III.1.14)}$$

where t_α are the experimental points: $t_\alpha = \{t_i\}$, $i = 1,...,n$; $H_k(Y_\alpha) = H_k(Y(t_\alpha))$; $\lambda_{k\alpha}$ are the kriged coefficients.

The kriged weights $\lambda_{k\alpha}$ are obtained as the solutions of the following system:

$$\sum_{\beta} \lambda_{k\beta} \operatorname{cov}[H_k(Y_\alpha), H_k(Y_\beta)] = \operatorname{cov}[H_k(Y_\alpha), H_k(Y(t))],$$

i.e.

$$\sum_{\beta} \lambda_{k\beta} [\rho_{\alpha\beta}]^k = [\rho_{\alpha t}]^k, \qquad \text{(III.1.15)}$$

where $\rho_{\alpha\beta} = \rho(t_\alpha - t_\beta) = \operatorname{cov}(Y(t_\alpha), Y(t_\beta))$.

---

[2] Cokriging is a multivariate generalization of the kriging method. Cokriging is applied when for the prediction of some random process $Z_0(t)$ we can use not only past historic data on Z_0, but also data on other random processes $Z_i(t), 0 < i \leq p$ correlated with Z_0. As the result the system of cokriging equations involves not only the autocovariance matrix of the process $Z_0(t)$, but also the cross-covariance matrix. More information about the cokriging, as well as the cokriging equations can be found in Chilès, Delfiner (1999).

With increase in n $[\rho(h)]^k \to 1$ and kriged estimator tends to its mean that is 0. Therefore, only some limited number of polynomials should be kriged and the number of Hermite polynomials can be truncated at some level *m*. Taking into account that $\text{var}\, f[Y(t)] = \sum_1^\infty (f_k)^2$, the truncated number *m* should satisfy the following expression:

$$\sum_1^m (f_k)^2 \approx \text{var}\, f[Y(t)].$$

The most common choice of the covariance structure for the Hermite polynomials is:

$$T_k(h) = \text{cov}[H_k(Y(t)), H_k(Y(t+h))] = [\rho(h)]^k$$

However, others models for covariance structure are used. In practice, four models are the most popular (Chilès, Delfiner, 1999):

1. The pure diffusive model:
$$T_k(h) = \rho^k(h),\ k \geq 0$$
2. The mosaic model:
$$T_k(h) = \rho(h),\ k \geq 0$$
3. The barycentric model: (III.1.16)
$$T_k(h) = \beta\rho^k(h) + (1-\beta)\rho(h),\ k \geq 0$$
4. The beta model:
$$T_k(h) = \frac{\Gamma(\beta)}{\Gamma(\beta+k)} \frac{\Gamma(\beta\rho(h)+k)}{\Gamma(\beta\rho(h))},\ k \geq 0$$

Note that Gaussian model is a particular case of the pure diffusive model.

There are two approaches to choose between the models in (III.1.16) (Chilès, Delfiner, 1999). One approach involves the analysis of the regression of $\chi_k(Y(t+h))$ on $Y(t)$ or $\chi_k(Y(t))$. However, the analysis should be performed for all values of *k* and *h*, which is complicated to implement in practice. According to the second approach, the "right" model can be chosen by the inspection of the variogram of order 1, estimated for the process $Y(t)$. Variogram of order 1 is defined as following:

$$\gamma_1(h) = \frac{1}{2}\text{E}[|Y(t+h) - Y(t)|].$$

Let's consider the variograms of first and second[3] order, normalized by their respective sills C_1, C:

$$\tilde{\gamma}_1 = \frac{\gamma_1}{C_1},\ \tilde{\gamma} = \frac{\gamma}{C} = 1 - \rho(h).$$

[3] A standard semi-variogram: $\gamma(h) = \frac{1}{2}\text{E}[(Y(t+h) - Y(t))^2]$

The following relationship between these variograms holds for the respective models in (III.1.16):

$$\tilde{\gamma}_1(h) = \begin{cases} \sqrt{\tilde{\gamma}(h)} \\ \tilde{\gamma}(h) \\ \beta\sqrt{\tilde{\gamma}(h)} + (1-\beta)\tilde{\gamma}(h) \\ \dfrac{\Gamma(\beta)}{\Gamma\left(\beta + 1/2\right)} \dfrac{\Gamma\left(\beta\tilde{\gamma}(h) + 1/2\right)}{\Gamma\left(\beta\tilde{\gamma}(h)\right)} \end{cases}$$

Plotting $\tilde{\gamma}_1$ versus $\tilde{\gamma}$ helps to define what covariance structure for the Hermite polynomials should be chosen.

1.3 Disjunctive kriging: non-normal random process

We have presented the application of the disjunctive kriging method to the bivariate standard normal variable. However, it is rare that in real-life applications the analyzed variables satisfy this hypothesis. Matheron (1976) proposed the solution to this problem: a non-normal random variable and its thresholds are transformed into normal variable and normal thresholds. Then the DK method is applied to this normal variable as described in Chapter 1.2 of part III.

An anamorphosis (transformation) is the function $\Phi(\cdot)$ that relates a non-normal variable $Z(x)$ to a normal one $Y(x)$:

$$\begin{matrix} Z(t) = \Phi(Y(t)) \\ z_c = \Phi(y_c) \end{matrix}. \qquad \text{(III.1.17)}$$

This transformation is defined by the cumulative probability distributions $F(z) = P[Z(t) < z]$ and $G(y) = P[Y(t) < y]$ of the variables $Z(t)$ and $Y(t)$ respectively. The transformation associates with each value z such value y that $F(z) = G(y)$. The graphical example is given in Part II, Figure 2.6. The transformation functions are discussed in more details further in Chapter 2.

The obtained DK results for the normal variable will be valid for the respective non-normal variable:

$$\Pr(Z_t < z_c) = \Pr(Y_t < y_c).$$

1.4 Disjunctive kriging: case of a regularized variable

In Chapters 1.1-1.3 we have shown how to estimate probability of breaching threshold at some point of time in future. However, DK method also allows estimating the proportion of points, when the price is below (above) some threshold during some interval of time in future. This proportion is called regularized indicator and is defined as:

$$\frac{1}{L}\int_L I_{Z(t)<z}dt = \frac{1}{b-a}\int_a^b I_{Z(t)<z}dt\,,$$

where L is the interval $[a;b]$, on which the proportion is estimated.

The DK procedure should be applied to the normal regularized indicator. Note that:

$$\frac{1}{L}\int_L I_{Z(t)<z}dt = \frac{1}{L}\int_L I_{Y(t)<y}dt\,,$$

where $z=\Phi(y)$ - the transformation function, defined in (III.1.17).

Rivoirard (1994) shows that

$$\frac{1}{L}\int_L I_{Y(t)<y}dt = G(y)+\sum_1 \frac{H_{k-1}(y)g(y)}{\sqrt{k}}\left(\frac{1}{L}\int_L H_k(Y(t))dt\right). \qquad \text{(III.1.18)}$$

The kriged estimator of $\frac{1}{b-a}\int_a^b H_k(Y(t))dt$ can be obtained from the system similar to (III.1.15):

$$\sum_\beta \lambda_{k\beta}\,\mathrm{cov}\left[H_k(Y_\alpha),H_k(Y_\beta)\right] = \frac{1}{b-a}\int_a^b \mathrm{cov}\left[H_k(Y_\alpha),H_k(Y(t))\right]dx$$

or

$$\sum_\beta \lambda_{k\beta}\left[\rho_{\alpha\beta}\right]^k = \frac{1}{b-a}\int_a^b \left[\rho_{\alpha t}\right]^k dt \qquad \text{(III.1.19)}$$

Then the DK estimator is:

$$\left(\frac{1}{b-a}\int_a^b H_k(Y(t))dt\right)^{DK} = G(y)+\sum_1 \frac{1}{\sqrt{k}}H_{k-1}(y)g(y)\left(\sum_\alpha \lambda_{k\alpha}H_k(Y_\alpha)\right) \qquad \text{(III.1.20)}$$

Further for the strategy construction we will use the DK estimator of the regularized variable rather than the estimator of the indicator at one point of time.

1.5 Disjunctive kriging: particular case of the variable with exponential covariance model

The financial random functions that can be defined as Markov processes, have the covariance that fit an exponential model.

Suppose that some standard normal variable Y has the covariance that follows the following exponential model:

$$\rho(h) = e^{-\left|\frac{h}{a}\right|}.$$

The empirical observations Y_α are available at the points $\{x_i\}_{1\le i\le m}$. Suppose we need an estimate of the probability $\Pr(Y(x_0) < y_c)$ at prediction horizon x_0.

Let's consider two cases of the prediction horizon x_0:

- Case 1: x_0 is at unit distance from the estimation window $\{x_i\}_{1\le i\le m}$ (see Figure 3.1);
- Case 2: x_0 is at some larger distance from the estimation window $\{x_i\}_{1\le i\le m}$ (see Figure 3.2).

As have been shown in Chapter 1.2, the probability estimate (III.1.13) involves the value of the kriging estimate of each Hermite polynomials of the order k. Therefore, the estimation problem coincides with the search of the optimal kriged weights $\lambda_{k\alpha}$ as the solution of the systems (III.1.15).

Suppose the following covariance structure for the Hermite polynomials of order n:

$$\rho^n(h) = e^{-n\left|\frac{h}{a}\right|} = e^{-\left|\frac{h}{A}\right|}, \text{ where } A = \frac{a}{n}$$

Let's introduce the following notations:

$$C = \begin{pmatrix} 1 & \rho & \rho^2 & \bullet & \rho^{m-1} \\ \rho & 1 & \bullet & \bullet & \bullet \\ \rho^2 & \bullet & 1 & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet \\ \rho^{m-1} & \bullet & \bullet & \bullet & 1 \end{pmatrix} \quad B = \begin{pmatrix} \rho \\ \rho^2 \\ \bullet \\ \rho^{m-1} \\ \rho^m \end{pmatrix} \quad b = \begin{pmatrix} 1 \\ \rho \\ \bullet \\ \rho^{m-2} \\ \rho^{m-1} \end{pmatrix} \quad \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \bullet \\ \lambda_{m-1} \\ \lambda_m \end{pmatrix} \quad \tilde{\lambda} = \frac{\lambda}{\rho}$$

Case 1. Unit prediction horizon (Figure 3.1)

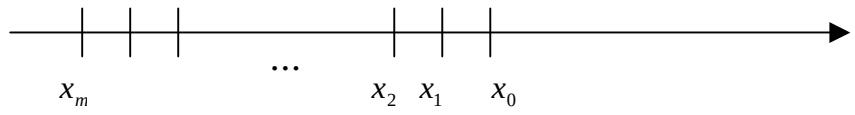


Figure 3.1. Estimation of the point probability at equally distanced point

The system of the DK equations (III.1.15) takes the following form for the case presented in Figure 3.1:

$$\sum_{\alpha=1}^{n} \lambda_{n\alpha} \rho^n\left(\left|x_\alpha - x_\beta\right|\right) = \rho^n\left(\left|x_0 - x_\beta\right|\right) \qquad \text{(III.1.21)}$$

Taking into account the matrix notation presented above, the system of equations (III.1.21) can be represented as:

$$C\lambda = B = \rho b$$

or

$$C\tilde{\lambda} = b \qquad \text{(III.1.22)}$$

The system (III.1.21) has a unique solution:

$$\tilde{\lambda} = \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix} \quad \text{or} \quad \lambda = \begin{pmatrix} \rho \\ 0 \\ \dots \\ 0 \end{pmatrix} \qquad \text{(III.1.23)}$$

Case 2. Distant prediction horizon (Figure 3.2)

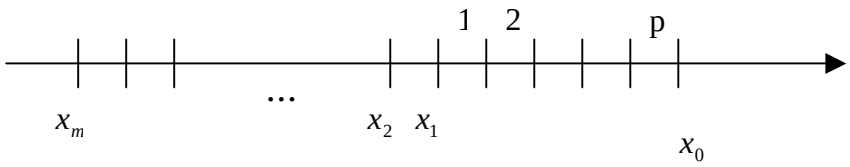


Figure 3.2. Estimation of the point probability at point distant from the estimated window

In the case, when point x_0 is situated at some larger distance from the estimation window, as in Figure 3.2, the distance to the time horizon can be defined as following: $|x_0 - x_1| = p\Delta x$, with Δx a unit distance. Then the optimal weights from the system (III.1.21) are:

$$\lambda = \begin{pmatrix} \rho^p \\ 0 \\ \dots \\ 0 \end{pmatrix} \qquad \text{(III.1.24)}$$

As we can see from the optimal solutions for the kriging weights (III.1.23) and (III.1.24) only last (most recent) observation matters for the DK estimation procedure in the case of the process with an exponential model.

2 Peculiarities of the disjunctive kriging applications to financial data

According to Chapter 1 two types of the DK probabilities can be defined and estimated. The first type is the DK probability, estimated for the particular future moments of time h. We call them further *point DK probabilities*:

$$P_{DK}^{p}(t, h, y_c) = \Pr(Y(t+h) < y_c). \qquad \text{(III.2.1)}$$

The second type is the regionalized DK probabilities, called further *interval DK probabilities*:

$$P_{DK}^{I}(t,l,y_c) = \Pr(Y_i < y_c, i \in [t+1; t+l]). \qquad \text{(III.2.2)}$$

Further the DK results should be analyzed and validated. Two approaches to the method validation can be considered in the financial context:

1. The obtained DK probabilities should be consistent; they should be compared with some alternative probability estimates.
2. The method, incorporated into a trading strategy, should improve the trading results.

The most rigorous evaluation approach would be the comparison of the DK probabilities with respective theoretically calculated "true" probabilities of being below the designed threshold; however, this is difficult to implement for the processes with unknown statistical characteristics. The best possible probability estimates that we can obtain empirically are the frequencies or the share of cases when price is below some threshold. Certainly, this approach is applicable only for the evaluation of the "interval" DK probabilities (III.2.2), estimated for regularized variables[4]. For these "interval" probabilities we can calculate the proportion of an instrument price (residuals) values smaller than particular threshold on the studied interval. Further in Chapter 3, 4 and 5 we will consider the examples of the point probabilities and will perform the analysis of the interval probabilities.

In this Chapter 2 we will concentrate on the peculiarities of the financial data such as non-normality and non-stationarity that have direct impact on the DK applications. Data non-normality implies the necessity of its transformation into a standard normal variable. Data non-stationarity implies that neither transform function nor the estimated covariance model can be assumed the same throughout the time. As the result two important questions should be addressed:

1. What transformation function to use?
2. What data sub-sample should be used to estimate the transformation function and how often the transformation should be adjusted to the new data?

2.1 Calibration of the transform function: kernel approach

When historical data $Z(x)$ has non-normal distribution the following cases are possible:

1. The non-normal distribution of the variable $Z(x)$ is known.
2. The non-normal distribution of the variable $Z(x)$ is unknown.

If distribution function is known and well defined, then the transformation into a normal variable $Y(x)$ can be done through equation:

$$F(z) = G(y) \Leftrightarrow Y = G^{-1}(F(Z)),$$

where $F(z)$ is the known CDF of the $Z(x)$; $G(y)$ is the standard normal CDF.

[4] It is impossible to calculate the same frequencies that correspond to the point probabilities at the presence of sole price path.

In some cases the existing well-defined statistical transforms functions $\varphi(\cdot)$ can be used to transform data into normal. The most popular existing function that can be applied to the skewed distributions is the Box-Cox transform:

$$\varphi(z,\lambda)=\begin{cases}\ln z, \lambda=0\\ \dfrac{z^{\lambda}-1}{\lambda}, \lambda\neq 0\end{cases}$$

The other well-known data transforms, developed to target different data types (skewed non-symmetric distributions; symmetric distributions, etc.), were proposed by Manly, John and Draper, Bickel and Doksum, Yeo and Johnson, etc.[5]. However, the main drawback of these well-defined transforms is that they provide good results for unimodal distributions. Therefore, when the empirical distributions are multimodal, the application of the discussed transformations does not make data normal.

In the case when the distribution function is unknown we have several approaches to its estimation. The first method is to fit a theoretical distribution to the empirical histogram (or CDF). Such methods as maximum-likelihood (ML) are used to obtain the estimates of the parameters of the predefined distributions. The main drawback of the method is that a distribution model should be chosen ad hoc, before the estimation and optimization procedure.

According to the second approach numerical estimators of the empirical histogram (or CDF) at observed points can be used for the interpolation[6] of the points, where the numerical estimates are not available. Different methods exist to estimate/interpolate numerically the empirical distribution function. One of the methods is to use the Kaplan-Meier estimator of the empirical CDF, used in the previous part II to transform data into normal variable. The drawback of the method is that the CDF is estimated very closely to the available empiric histogram, making the tails of the density function converge drastically toward 0 for the values outside of the observed data interval. For the DK procedure it means that if data is transformed locally, the thresholds that were not reached during the analyzed period are transformed into infinite values, even if the values are quite close to the observed price/residuals[7]. The infinite thresholds values imply 0 or 1 value for the DK probabilities no matter how "far" are the thresholds from the observed data; such probability values are useless in the trading applications.

The non-parametric methods, such as kernel approach, allow better (or managed[8]) approximation of the probability density for non-observed extreme values. Statisticians apply the kernel smoothing method to the estimation of the regressions or density probability functions. As a nonparametric method, it does not demand any predefined functional representation and it allows uncovering some structural characteristics that parametric methods cannot reveal. As the result, it can be fit to any empirical distribution. For more information on kernel smoothing method the interested reader can be referred to Wand (1995), Martinez, Martinez (2001), Scott (1992).

---

[5] See for more details: Li, P. 2005. "Box-Cox Transformations: An Overview", presentation, http://www.stat.uconn.edu/~studentjournal/index_files/pengfi_s05.pdf. See also Carroll, Ruppert (1988) for more information on data transforms.

[6] Statistical/Mathematical softwares frequently include the interpolation functions in their packages, making the procedure very easy.

[7] For previous DT bands transformations this factor had less implications, as the bands were defined by the K% cuts of the standard normal distribution, not by some prefixed value that corresponded to the historic (raw) data..

[8] The kernel fit can be "managed" by controlling for the bandwidth parameter that defines the closeness of the fit.

Let the random variable $\{Z_i\}_{1\le i\le n}$ have a continuous univariate density function ρ. Then a kernel estimator of the density function is:

$$\hat{\rho}(z,h)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{z-Z_i}{h}\right), \quad \text{(III.2.3)}$$

where $K(u)$ - is a kernel function, h - a bandwidth parameter.

Expression (III.2.3) can be considered as a weighted average of all available data points with the weights defined by the Kernel function and bandwidth parameter. We can notice that only distance between one point and all other available data points (but not their relative positioning) has impact on the value of the weight assigned to each of the points.

Kernel function $K(u)$ can itself be considered as probability density function that is unimodal and symmetric with respect to 0. The types of kernel include Gauss, uniform, triangular, Epanechnikov functions.

The bandwidth is a scaling parameter; it controls the smoothness of the estimated density curve. Small bandwidth parameter gives much closer fit for the estimated density function, but at the same time probability outside estimated interval quickly converge toward 0 (close fit). Larger bandwidth on the other hand givers poorer fit. Therefore the choice of the bandwidth is combined with the risk of over or under-smoothing. The methods of the search of the optimal bandwidth parameter can be split into two groups (Wand, 1995): (1) quick and simple approaches that does not guarantee the optimal bandwidth, but which are less time-consuming in introduction; (2) high tech approaches that are based on the optimization procedure of minimizing AMISE (asymptotic mean integrated squared error). We will not present the methods in more details; the interested reader can be referred to Wand (1995).

Taking into account the universality of the method with respect to its application to different instruments with different statistical characteristics, we choose the kernel method to estimate the empirical CDF, and as the result, to calibrate the transform function. The kernel weights are defined by Gauss function that assigns the highest weight to the data point where the density function is estimated, with the decreasing weights for more distant points. The bandwidth parameter is defined by the default value proposed by the MATLAB procedure. In addition, for DAX and Bund instruments, we consider one more bandwidth value, larger than the default value that allows more smooth density estimate in order to analyze the impact of this parameter on the DK results and trading outcomes.

Let's consider the results of the kernel application to the density function estimation for the particular case of the DAX instrument. DAX index represents the stocks market; its historic quotes are given in Figure 3.3. DAX data corresponds to the time interval of 30/7/2003-7/12/2006 and the frequency of 30 minutes.

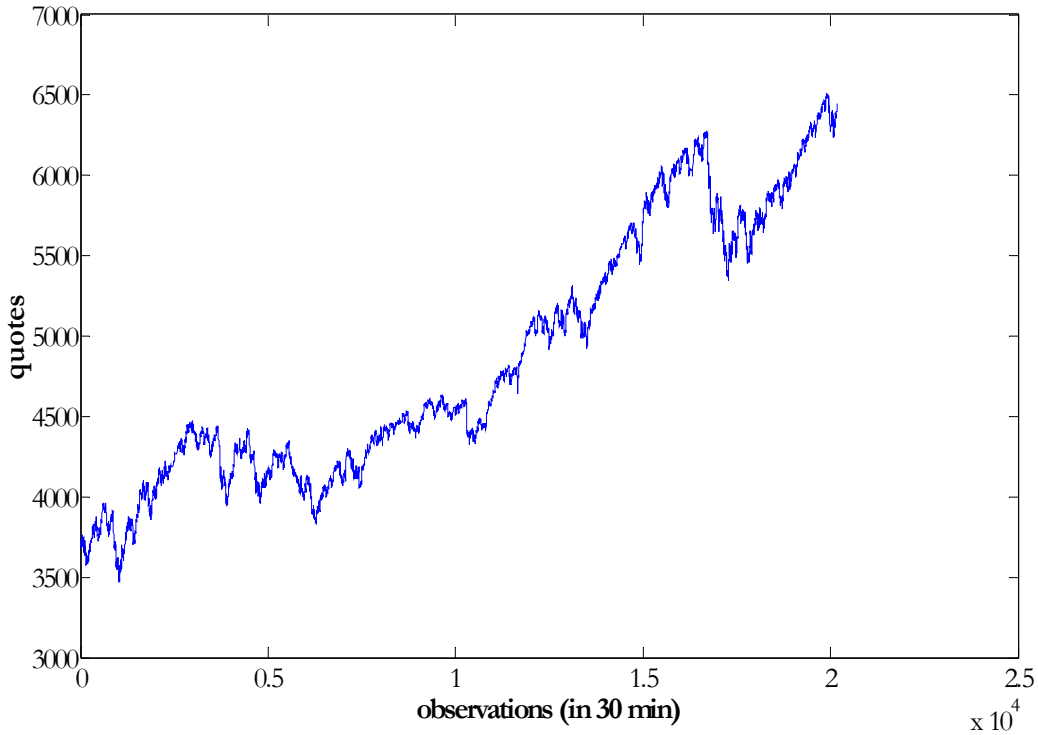


Figure 3.3. DAX quotes for period 30/7/2003-7/12/2006 (frequency – 30 minutes).

The strategies, based on trading bands, can be transformed into the strategies, based on the residuals, constructed as the difference between a price and middle line. For example, signal triggered by breaching the upper/lower bands by price, can be substituted by the signal triggered by breaching the threshold, defined by the distance between middle line and bands, by residuals series.

Let's consider the trading bands strategies with the EMA as the estimate of the middle line. The residuals, obtained by the subtraction of the EMA (EMA length=115 observations) from the DAX price data are given in Figure 3.4. The residuals data is not normal: the hypothesis that the residuals represent a standard normal variable can be rejected under Kolmogorov-Smirnov test with 5% significance level (KS(statistics)=0.58 with cut-off value of 0.01). Thus, calibration of the transformation function for residuals transformation is needed before application of the DK method.

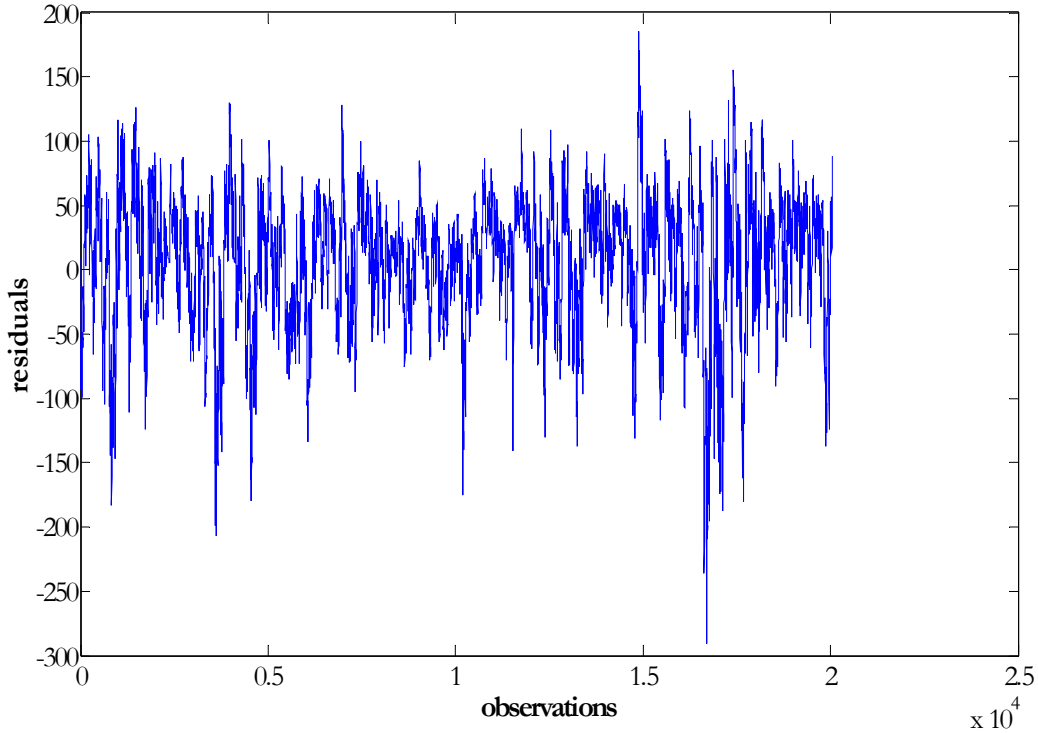


Figure 3.4. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 observations)

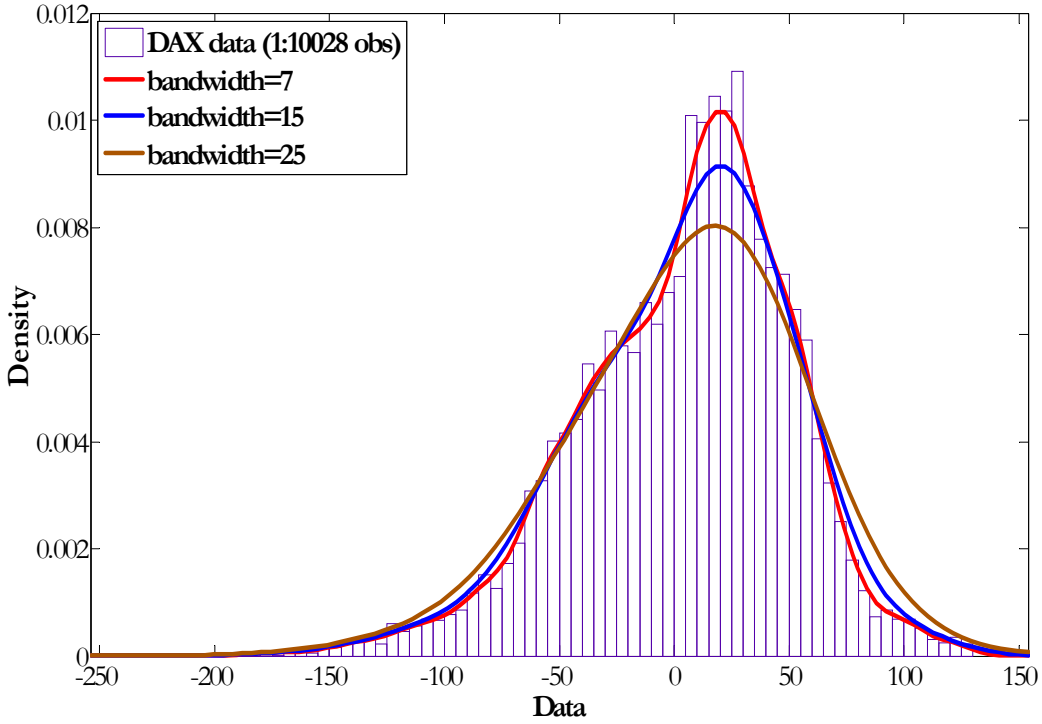


Figure 3.5. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 observations): empirical histogram and CDF estimates according to the kernel method for different bandwidth parameters (sub-sample [1;10028]).

The kernel estimates of the CDF for different bandwidth parameters, estimated on the first half of the sample are demonstrated in Figure 3.5 (default bandwidth=7). As we can see the smaller

the bandwidth parameter - the closer the estimated curve follows the empirical histogram and the thinner are the tails of the estimated distribution. The larger is the bandwidth parameter the thicker are the tails; at the same time the estimated mode of the distribution is much lower than the empirical mode.

Figure 3.6 shows the variograms of the normal variables, obtained for different transform functions that are calibrated for three bandwidth values. As we can see the bandwidth value has impact on the variogram parameters, in particular on its sill. The explanation for this phenomenon lays in the fact that large bandwidth parameter thickens the tails of the density function; this results in the higher probability values, attached to the (historically) observed extreme values. As the result, the discrepancy (variance) of the normal data diminishes.

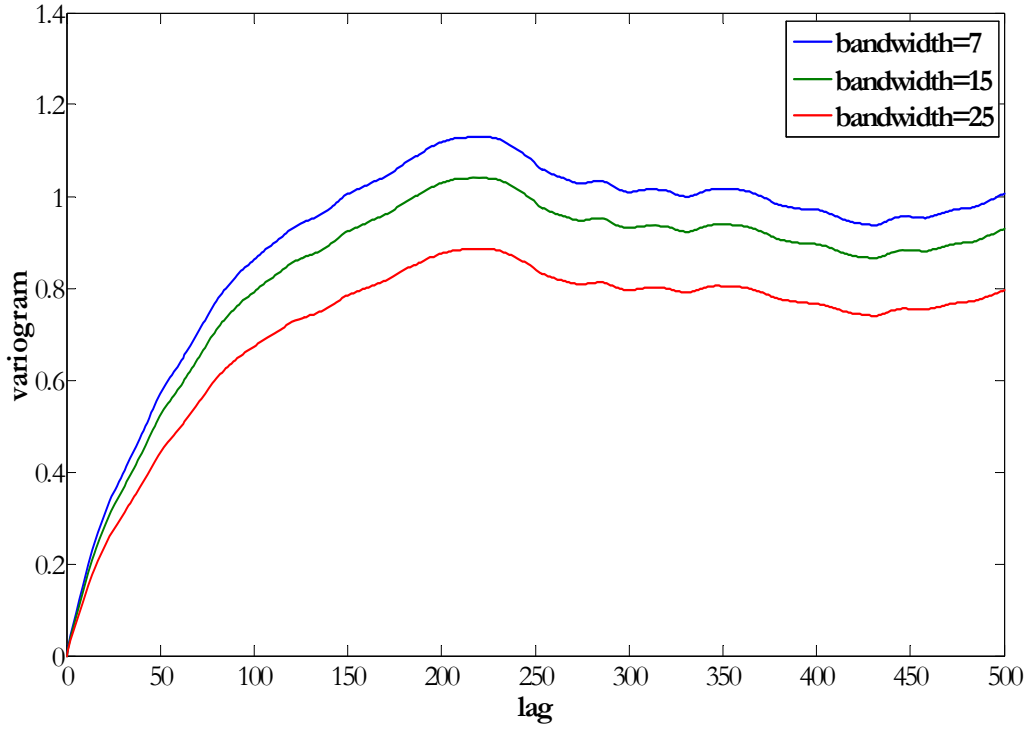


Figure 3.6. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 observations, observations 1-10029): variograms of the normal variable, transformed according to the kernel estimator with different bandwidth values

Further in the DK applications we use two bandwidths values to calibrate the transform function for DAX instrument: $b_1 = 7$ (default value) and $b_2 = 15$. The following models (the sum of exponential and damped-cosines model $\gamma(h) = \sigma_1\left(1 - e^{-\frac{h}{a_1}}\right) + \sigma_2\left(1 - e^{-\frac{h}{a_2\alpha_2}} \cdot \cos\left(\frac{2\pi h}{a_2}\right)\right)$) are fit to these two variograms:

- $b_1 = 7$: $\gamma(h) = 0.75\left(1 - e^{-\frac{h}{50}}\right) + 0.25\left(1 - e^{-\frac{h}{450 \cdot 0.7}} \cdot \cos\left(\frac{2\pi h}{450}\right)\right)$
- $b_2 = 15$: $\gamma(h) = 0.65\left(1 - e^{-\frac{h}{45}}\right) + 0.25\left(1 - e^{-\frac{h}{450 \cdot 0.8}} \cdot \cos\left(\frac{2\pi h}{450}\right)\right)$ (see Figure 3.7 for the variogram example)

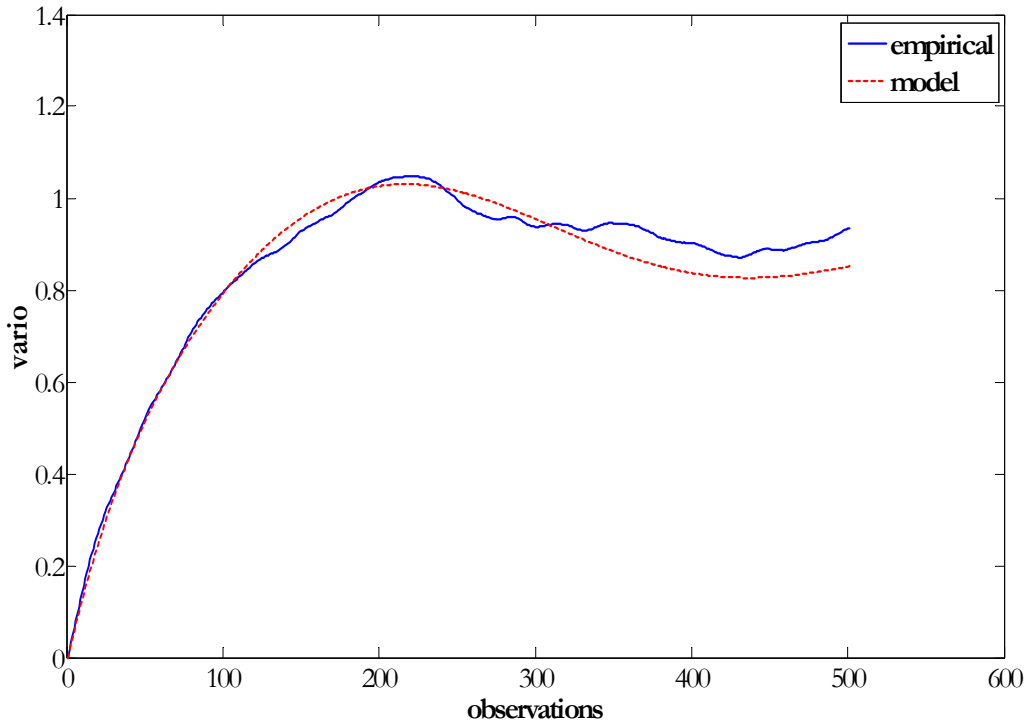


Figure 3.7. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 observations, observations 1-10028): variograms of the normal variable, transformed according to the kernel estimator with bandwidth $b = 15$ and its theoretical fit

2.2 Data used for the estimation of the CDF: locally adjusted transform function

The other peculiarity of the financial data is its non-stationarity that implies that part of the parameters used in the DK application procedure are not valid everywhere. Further in Chapter 3 we provide the examples of the DK probability estimates, obtained for the same CDF, that was estimated on the whole available sample, under the following assumptions: (1) data is stationary; (2) the estimated CDF represents "true" distribution function. We will show that this framework implies poor probability estimates.

Under the assumption of local stationarity, the local estimation of the density functions is more appropriate. At the same time the local estimation of the CDF is frequently based on significantly reduced sample size, making the estimation results less consistent (especially if the estimations are made for the thresholds that were not reached during the most recent period of time). Therefore, in order to avoid bias and errors due to the small number of observations in the estimated interval, we propose to estimate the CDF function on a longer sample and then locally adjust its statistical moments, in particular its variance. We assume that the CDF (further called basic CDF) and covariance types of models are the same throughout some long period of time, while some of its parameters, such as variance are only locally stationary[9].

[9] Taking into account that further the DK method is applied to the price residuals, we can assume that its mean is constant and zero.

Suppose $\tilde{Z} \sim L(0,1)$ is a random variable that follows some arbitrary distribution law, with 0 mean and variance 1. Let's define another random variables $Z^{(1)}$, $Z^{(2)}$ on the basis of $\tilde{Z}$ that follows the same distribution law with 0-mean and different variance:

$$Z^{(1)} = \sigma_1 \tilde{Z},\ Z^{(1)} \sim L(0, \sigma_1^2)$$
$$Z^{(2)} = \sigma_2 \tilde{Z},\ Z^{(2)} \sim L(0, \sigma_2^2)$$

Then

$$Z^{(2)} = \frac{\sigma_2}{\sigma_1} Z^{(1)} \qquad \text{(III.2.4)}$$

Let's consider some transformation function φ, such that:

$$Z^{(1)} = \varphi(Y),$$

where Y is a random standard normal variable: $Y \sim N(0,1)$.

Let's define the following CDFs for the variables $Z^{(1)}$, $Z^{(2)}$ and Y:

$$\Phi^{(1)}(z) = \Pr(Z^{(1)} < z) \qquad \text{(III.2.5)}$$
$$\Phi^{(2)}(z) = \Pr(Z^{(2)} < z) \qquad \text{(III.2.6)}$$
$$F(y) = \Pr(Y < y) \qquad \text{(III.2.7)}$$

Taking into account (III.2.4), (III.2.5) and (III.2.6), we can represent the CDF $\Phi^{(2)}$ as following:

$$\Phi^{(2)}(u) = \Pr(Z^{(2)} < u) = \Pr\left(\frac{\sigma_2}{\sigma_1} Z^{(1)} < u\right) = \Pr\left(Z^{(1)} < \frac{\sigma_1}{\sigma_2} u\right) = \Phi^{(1)}\left(\frac{\sigma_1}{\sigma_2} u\right)$$

The following expression holds true for the transformation functions:

$$Z^{(1)} = \varphi(Y) \rightarrow \Phi^{(1)}(t) = F(\varphi^{-1}(t))$$

Suppose that the residuals Z are stationary in the terms of the distribution law (i.e. the distribution function is the same and does not depend on the moment of time when it was estimated) and has constant mean (equal to zero). At the same time the variable Z is globally non-stationary in the terms of the variance (i.e. variance is time dependent), but can be considered constant within some (short) intervals of time (locally stationary).

Let's define a *basic* CDF $\Phi^{(b)}(z)$ as the CDF that is estimated on the sub-sample long enough to capture the "true" distribution function. By analogy, let's define a *basic* volatility σ_b as the standard deviation of the same sub-sample. Then a *local* CDF $\Phi^{(i)}(u)$ of the residuals on some *i*-interval with constant *local* variance σ_i can be defined as the following:

$$\Phi^{(i)}(u) = \Phi^{(b)}\left(\frac{\sigma_b}{\sigma_i} u\right) \qquad \text{(III.2.8)}$$

We can derive the following conclusions from the expression (III.2.8):

1. If $\sigma_i = \sigma_b$, then $\Phi^{(i)}(u) = \Phi^{(b)}(u)$, i.e. there is no need for local adjustment.

2. If $\sigma_i > \sigma_b$ (residuals become more volatile), then $\Phi^{(i)}(u) = \Phi^{(b)}(u^*)$, where $u^* = \frac{\sigma_b}{\sigma_i} u < u$.

3. If $\sigma_i < \sigma_b$ (residuals become less volatile), then $\Phi^{(i)}(u) = \Phi^{(b)}(u^*)$, where $u^* = \frac{\sigma_b}{\sigma_i} u > u$.

The expression (III.2.8) is applied further to obtain the local estimates of the CDF that is used to calibrate the transform function. According to this expression, the local CDF estimators are obtained as following:

1. Each historic observation in the estimated window is adjusted by the ratio $\frac{\sigma_b}{\sigma_i}$.
2. This adjusted historic data is transformed into normal variable by the means of the basic CDF function.

Figure 3.8 presents the examples of the volatility ratio $r_i = \frac{\sigma_i}{\sigma_b}$ for the DAX instrument. Basic volatility σ_b is estimated as a standard deviation of the data on the interval $[1; N/2]$, where N is the total length of the sample. Local volatility σ_i is represented by the standard deviation of the *i*-th moving window of the length $m: [i-m+1; i], i \geq N/2+1$.

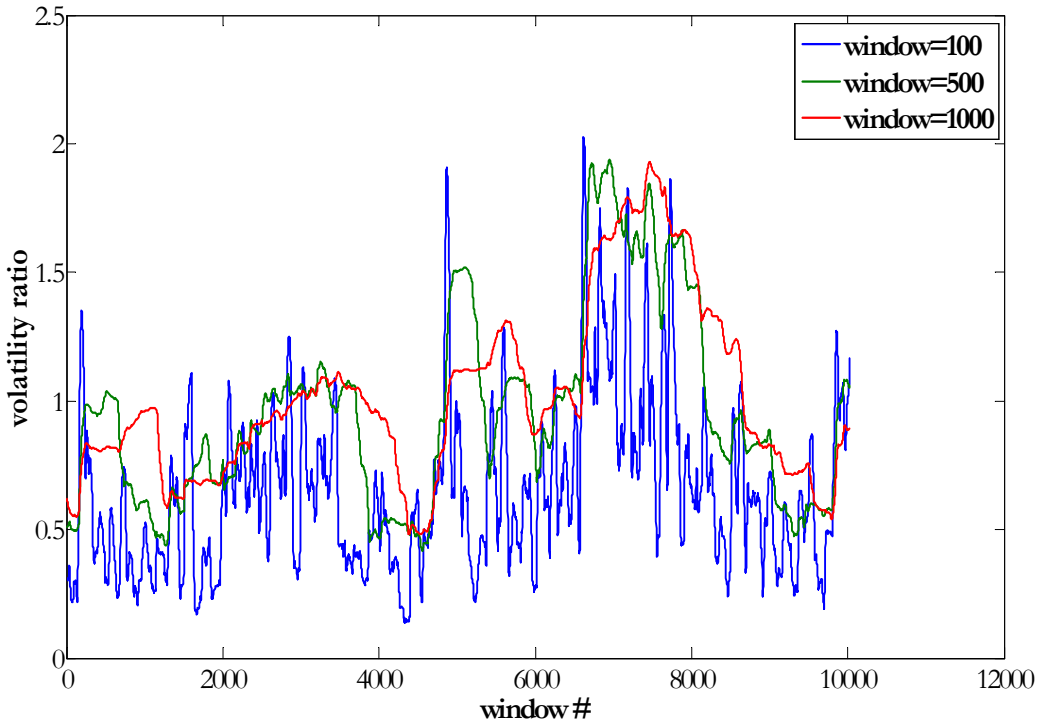


Figure 3.8. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 observations, observations 10029-20058): Volatility ratio (rolling standard deviation (ith window)/ rolling standard deviation (basic)) for historic residuals for different windows lengths

As we can see the volatility ratio fluctuates around 1: at some points local volatility is lower than the basic volatility, while at others - local volatility is higher than the basic volatility. Therefore, the adjustment of the CDF for volatility is justified.

Further in Chapter 4 we will show that contrary to the stationarity assumption, local adjustment of the CDF for the volatility improves the DK probability estimates.

Note also that we do not adjust the variogram parameters for local variance, as variogram is estimated for already transformed standard normal variable that always has unit variance by definition.

In the following applications we will estimate the basic CDF and variogram model only once. Though in real life routine application, the models should be tested for newly arrived data and in the case of significant discrepancy the models should be re-estimated.

3 Examples of the DK probabilities under stationarity assumption

In this chapter we consider the examples of the DK probabilities, obtained under assumption of data stationarity. This implies the same data transform function and covariance model for the transformed standard normal variable. Chapter 3.1 presents the examples of the point probabilities for the DAX residuals, while Chapter 3.2 analyses the interval probabilities for the Bund residuals. The Bund/DAX residuals are obtained by the extraction from price of the EMA of the one-week length (115 observations). The transform function is calibrated on the whole data sample and covariance function is estimated for the whole sample of the transformed normal variable. Bund and DAX residuals are still non-normal, but can be considered stationary as their semi-variograms converge toward residuals' variance. Taking into account that the examples of the DK probabilities in Chapter 3 are estimated for the same sample on which the calibration of the transformation function is done, this approach creates the data-snooping bias, when the obtained results are validated on the same sample, on which the parameters, used in the estimation procedure, are estimated. However, this allows us to analyze the results under the assumption that the obtained variogram and CDF estimates are the "true" estimates of the covariance and probability distribution. It should be noted that the variogram of the transformed normal variable for Bund (Appendix C, Figure C3) fit the exponential model $\gamma(h)=1\left(1-e^{-\left|\frac{h}{55}\right|}\right)$, while for the DAX residuals the variogram is $\gamma(h)=0.75\left(1-e^{-\frac{h}{50}}\right)+0.25\left(1-e^{-\frac{h}{450\cdot 0.7}}\cdot\cos\left(\frac{2\pi h}{450}\right)\right)$ (see Chapter 2.1).

3.1 Point DK probabilities: DAX case

The examples of the point probabilities are provided for the sub-sample of the DAX quotes (see price path in Figure 3.9). Appendix A explains the algorithm behind point probabilities calculations. Our objective is to show how the values of the point DK probability depend on the choice of the length of the window, time horizon and threshold level. Note that the threshold is

fixed at the level of the last observation in the window; this will allows us more intuitive interpretation of the results.

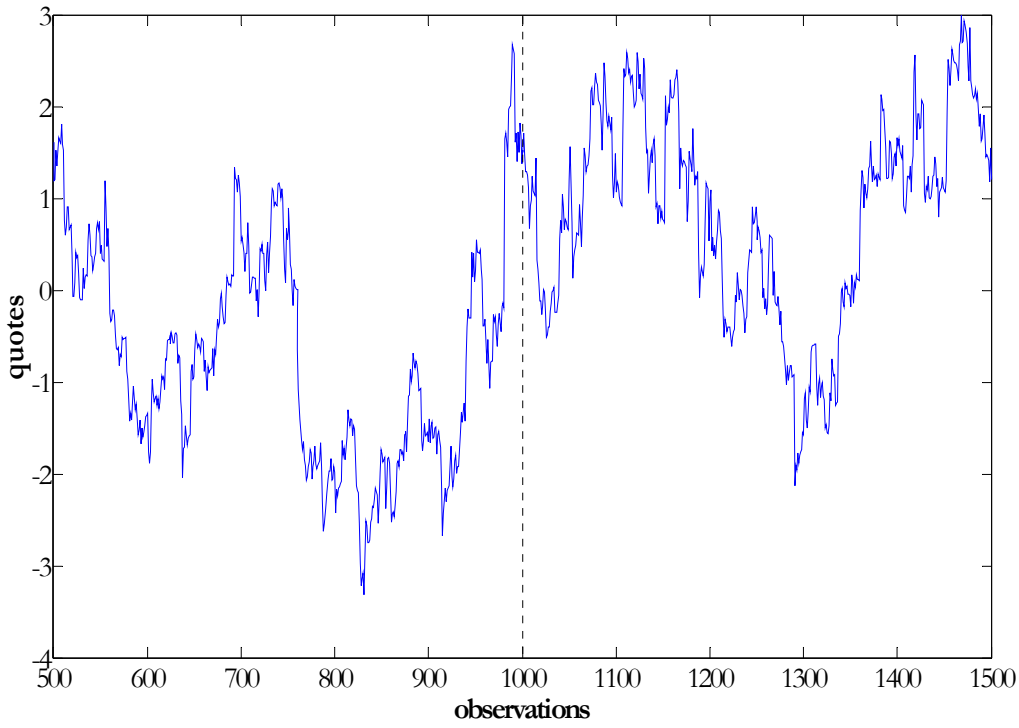


Figure 3.9. Transformed DAX residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): observations 500-1500

Figure 3.10 presents the estimated point probabilities for different threshold level and time horizon, estimated on the different windows, but of the same length 100 observations: the DK probability for the threshold $y_c = -1.36$ is estimated on the interval [500; 599]; for the threshold $y_c = 0.55$ - on the interval [650; 749]; and for the threshold $y_{c,150} = 0.09$ - on the interval [850; 949]. We can notice that at long time horizon the DK probability values converge towards unconditional normal probability (CDF): $F(y_c) = P(Y(h) < y_c)$. Moreover, this convergence achieved at the time horizon that is close to visual variogram range (where variogram convergence toward variance), confirming that existent autocorrelation is taken into account for DK probability estimation.

Increasing or decreasing probability at short time horizon depends on the price path that the instrument follows in the estimated window (see Picture 3.10). For the interval [500; 599], the negative trend is well pronounced; thus, the probability of being below the threshold (last observed price) is higher than unconditional normal CDF (i.e., we expect the price to decrease further and breach the "last price" level). For the interval [650; 749] at the presence of the positive trend the probability of staying below the designed threshold (last available observation) is less likely, therefore the probability is lower than the unconditional probability for this threshold. Finally, no trend is observed on the interval [850; 949] making the probability of being below the last observations very close to the unconditional value.

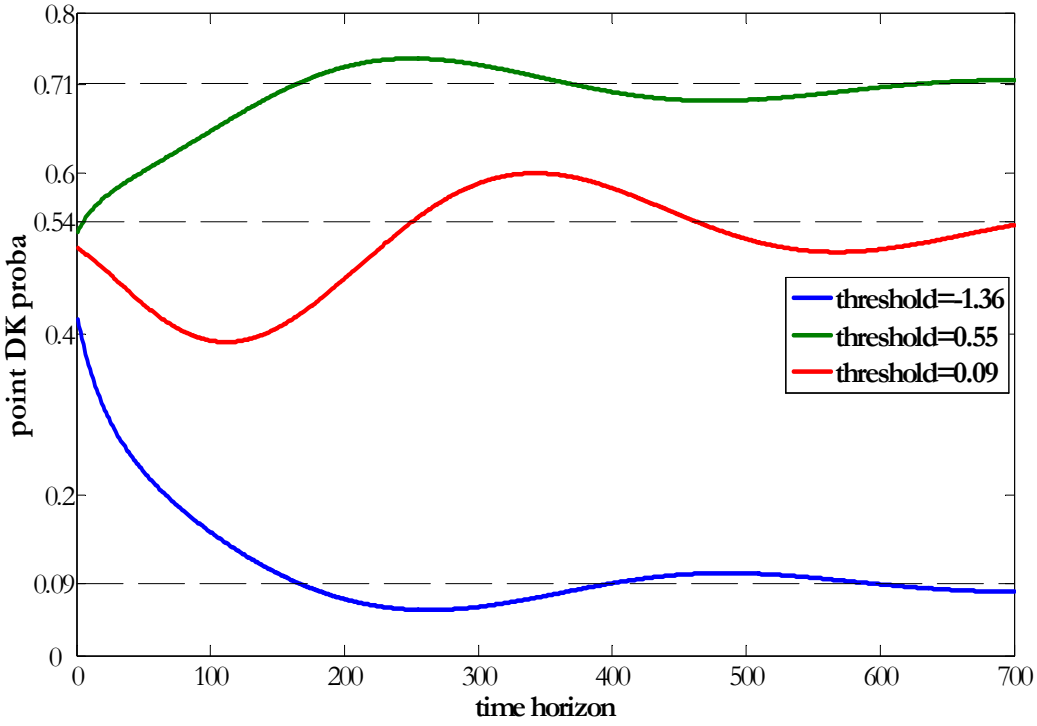


Figure 3.10. Transformed DAX residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): Probability $P(Y(t+h) < y_c)$ for different time horizon (lags) *h*, for different estimation window of the length 100 observations and different threshold levels: $y_c = -1.36$, $y_c = 0.55$, $y_{c,150} = 0.09$.

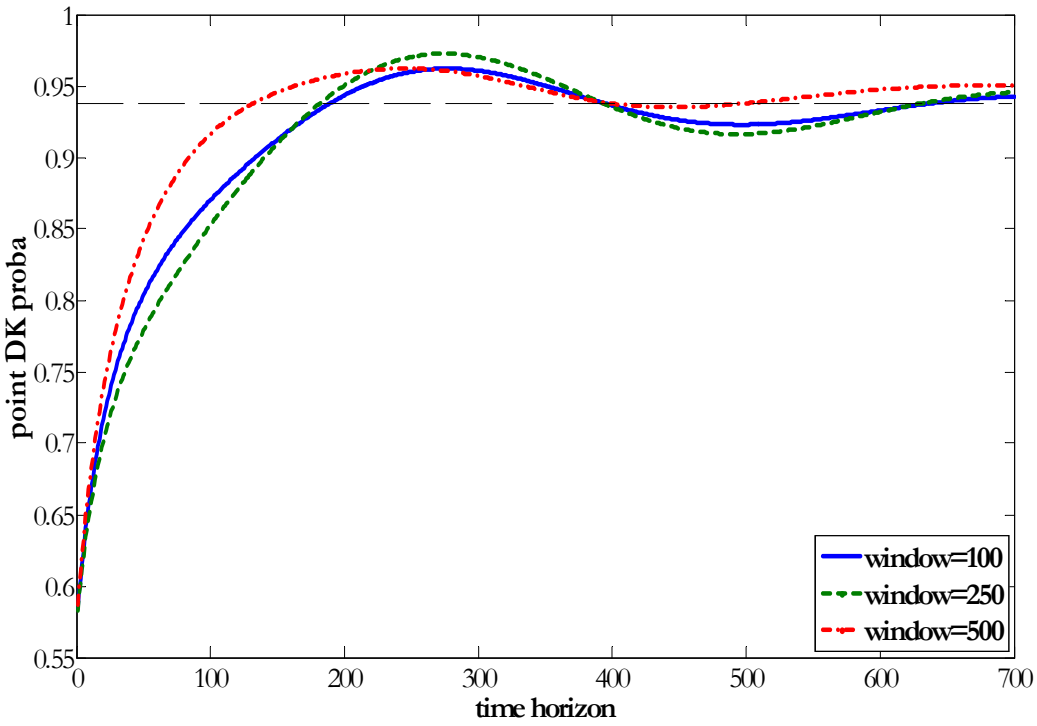


Figure 3.11. Transformed DAX residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): Probability $P(Y(t+h) < y_c)$ for different time horizon (lags) *h*, for different length of the estimation window and the same threshold level $y_c = 1.53$.

The length of the window has no impact at very short time horizon, but implies discrepancy at longer time horizon. Figure 3.11 demonstrates the DK probabilities estimated for the same threshold $y_c = 1.53$ that is defined by the same 1000th observation of the transformed residuals (the end of all windows is at 1000th observation). Therefore, the longer window (500 observations) contains the shorter windows (100 and 250 observations) and data overlaps. For the time horizon close to 1-20 observations three curves coincide, while for the longer time horizon the speed of the convergence versus unconditional value depends on the window lengths (see Figure 3.11). The shorter (100 observations) and medium (250 observations) windows contain a negative trend; therefore, convergence versus unconditional level is less rapid than for longer window, where no global trend can be defined.

Although the point DK probabilities cannot be back-tested, the results of this particular test follow common sense: (1) at distant time horizon the probability of being below some threshold converges versus the unconditional normal probability; (2) the point DK probability value at low time horizon as well as the speed of its convergence versus unconditional level depend on the instrument path in the estimated window; (3) the impact of the past observations (estimated window) disappears at the time horizon longer than the variogram range.

3.2 Interval DK probabilities for global transformation function: Bund case

Contrary to the point probabilities, the analysis of the interval DK probabilities can be enriched by the back testing results: the DK probabilities can be compared with the empirical frequencies $P_{emp}\left(Y_{[t+1;t+l]} < y_c\right)$, estimated on the same intervals. The transformation function as well as the covariance function is estimated on the whole available sample of the Bund residuals; they assumed to be the same for the whole time.

In order to analyse the impact of the different factors, the interval DK probabilities are calculated for the following parameters values:

- Length of the window: 54 observations (the window is rolled by the step 10 observations)[10];
- Length of the intervals: 30, 100 and 500 observations.
- Thresholds values: $y_c \in [-3:3]$.

Taking into account the large number of test parameters (interval length, threshold value) the volume of the analysis results is considerable. Many of these results are very similar. Therefore, we present here only some of the results that allow us to derive main conclusions about DK applications outcomes[11] under stationarity assumption.

Figure 3.12 and B1-B3 presents the examples of the interval DK probabilities and respective empirical frequencies, estimated for the following parameters: window=54 observations, interval=100 observations and threshold=1.9. The X-coordinate represents the number of the 54-observations window, on which the DK probability was calculated for the 100-observations

[10] We do not analyse the impact of the window length on the interval DK results. The covariance model used in the DK procedure is exponential; thus, only last observation in the window defines the DK probability. As the result the window length does not have impact on the DK probability values.
[11] Other results can be provided on the request.

interval that follows the window, i.e. window #1 corresponds to the window [1;54] and the interval [55;154], window #2 corresponds to the window [11;64] and the interval [65;164], etc. As we can see from Figure 3.12 the DK probability and empirical frequency resembles some random variable: they oscillate around 0.97-levels with different amplitudes.

The empirical frequencies curves deserve separate discussion as their forms might raise several questions. As we can see from Figure 3.12, the empirical frequencies often take the value of 1 (or 0 for other thresholds), increasing the dispersion of this variable. The fact that 100% of all observations in the estimated interval is found above or below the threshold value points at some data peculiarity, i.e. the residuals are moving around some local mean values different from 0 and oscillate around 0 in jumps. It means that we can always find some short intervals of data during which the data is below or above some chosen threshold level. The intervals with such behavior can be found in the Figures B1-B3 in Appendix B.

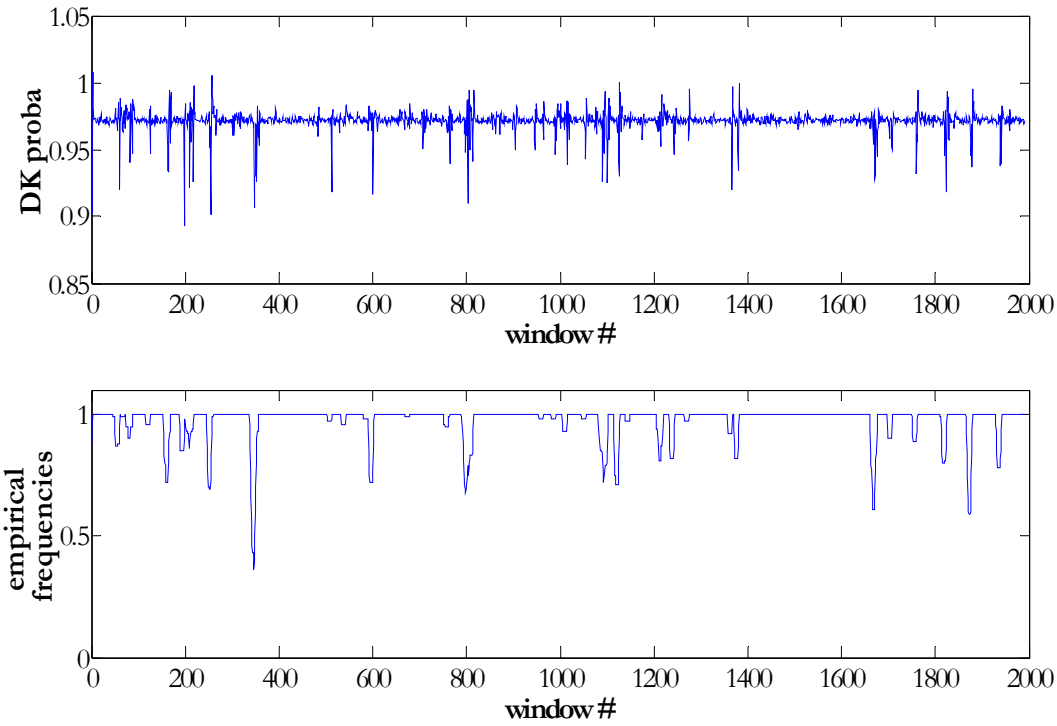


Figure 3.12. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 obs): DK probability $P_{DK}\left(Y_{[t+1;t+l]} < y_c\right)$ and respective empirical frequencies, threshold value $y_c = 1.9$, window length=54 obs, interval=100 obs.

It is clear that the direct comparison of the DK probabilities with the frequencies curves in Figure 3.12 is very difficult. Therefore, we propose to consider DK curves as some random variable and use its first two moments (mean and standard deviation) to summarize the results over all moving windows. Figures 3.13 and B4-B5 in the Appendix B represent the analysis and comparison of the DK probabilities over mean and standard deviation.

Let $F(y_c) = P(Y(h) < y_c)$ represents the CDF of the random variable. Figure 3.13 represents the average probability $P_a(Y(h) < y_c)$, or an average CDF of the variable $Y(h)$. The average CDF in Figure 3.13 resembles the normal CDF; this hypothesis is confirmed by the scatter point graph B7 in Appendix B.

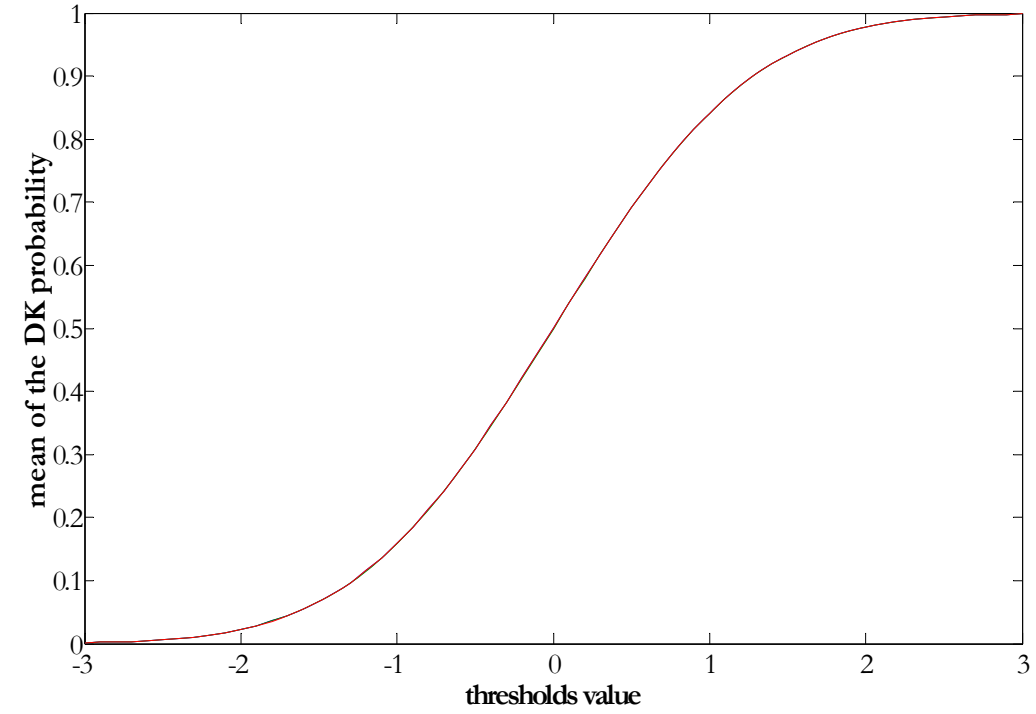


Figure 3.13. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 obs): Mean of the DK probabilities $P_{DK}\left(Y_{[t+1;t+l]} < y_c\right)$ for different thresholds (y_c); window length= 54 observations, interval =100 observations.

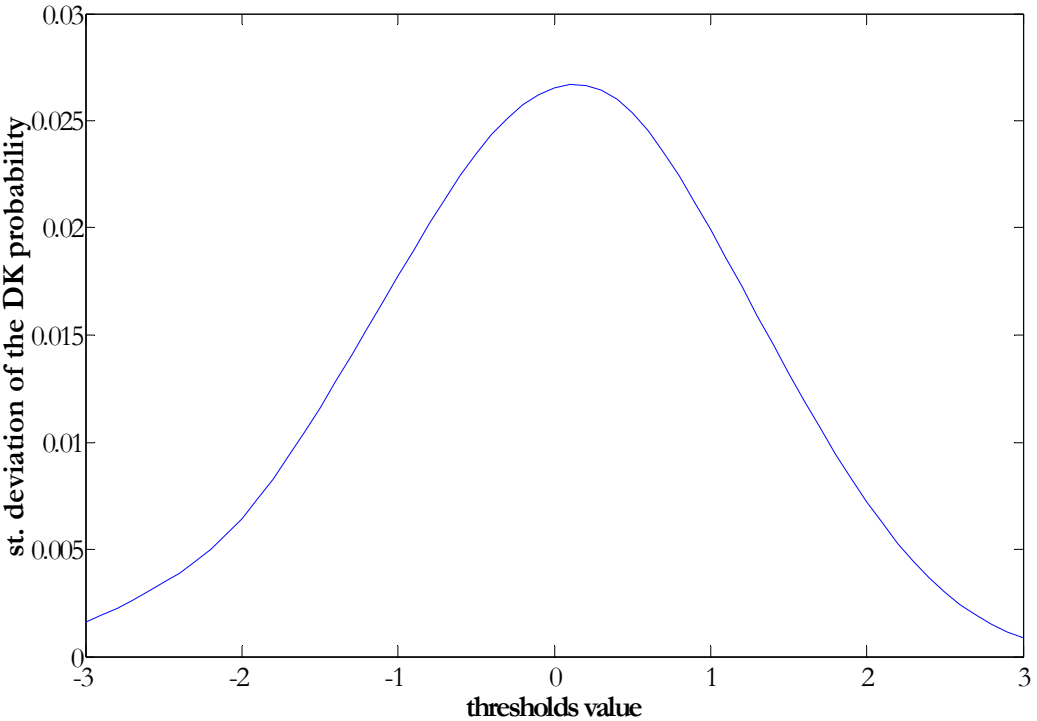


Figure 3.14. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): Standard deviation of the DK probabilities $P_{DK}\left(Y_{[t+1;t+l]} < y_c\right)$ for different thresholds (y_c); window length=54 observations, interval length=100 observations.

The measure of the probability dispersion for different threshold values is represented in Figure 3.14.

The following conclusions can be derived about DK probability dispersion:

1. The dispersion (standard deviation) is not very large, meaning that DK probabilities stay quite close to its mean value- unconditional normal probability.
2. The value of dispersion depends on the threshold value, achieving its peak at the 0-threshold value and diminishes toward 0 with the increasing of the threshold value in absolute terms. It means that oscillations of the DK probability are the highest for 0-threshold. This anomaly can be easily explained: as zero is the mean of the transformed variable, this value is crossed more often than the other threshold levels.

Thus, we can conclude that DK probabilities oscillate around unconditional normal probability. Figures B1-B2 (Appendix B) analyze whether the lengths of the intervals have impact on the interval DK probabilities. The following conclusions can be derived:

1. The mean value of the interval DK probabilities does not depend on the length of the interval for which the estimations are performed (see Figures B4 in the Appendix B).
2. The standard deviation of the interval DK probabilities depends on the interval length for which the estimation is conducted: the standard deviation of the DK probability decreases with increase of the interval length (see Figure B5 in the Appendix B), i.e. increasing of the interval for which the DK probabilities are estimated increases the error of the estimations.

The comparison of the DK probability with the empirical frequencies, estimated for the same intervals, is one of the most important parts of the analysis. If both entities coincide, the DK probability can be accepted as the estimator of the "true" probability and used further for successful strategy constructions. Figures 3.15-3.16 compare the main statistical characteristics of DK probabilities and empirical frequencies. Figure 3.15 presents mean value of the DK probability versus mean value of the empirical frequencies. We can conclude that on average the values of DK probability and frequencies coincide.

At the same time dispersion of the DK probability and frequencies differs significantly (see figure 3.16): the standard deviation of the empirical frequencies exceeds the same characteristics of the DK probability almost 10 times. It means that if we assume that the empirical frequency is the estimate of the "true" probability, the usage of the DK probabilities as these estimators will lead to the significant errors.

Increasing of the interval length decreases the standard deviation of the empirical frequencies, as the more data we use to estimate the "true" probabilities the better the estimates are (see Figure B6 in Appendix B). However, it is still larger than the DK probability dispersion. This result was predictable from figure 3.12: the empirical frequencies often take the value of 0 or 1 that increases the dispersion of the variable.

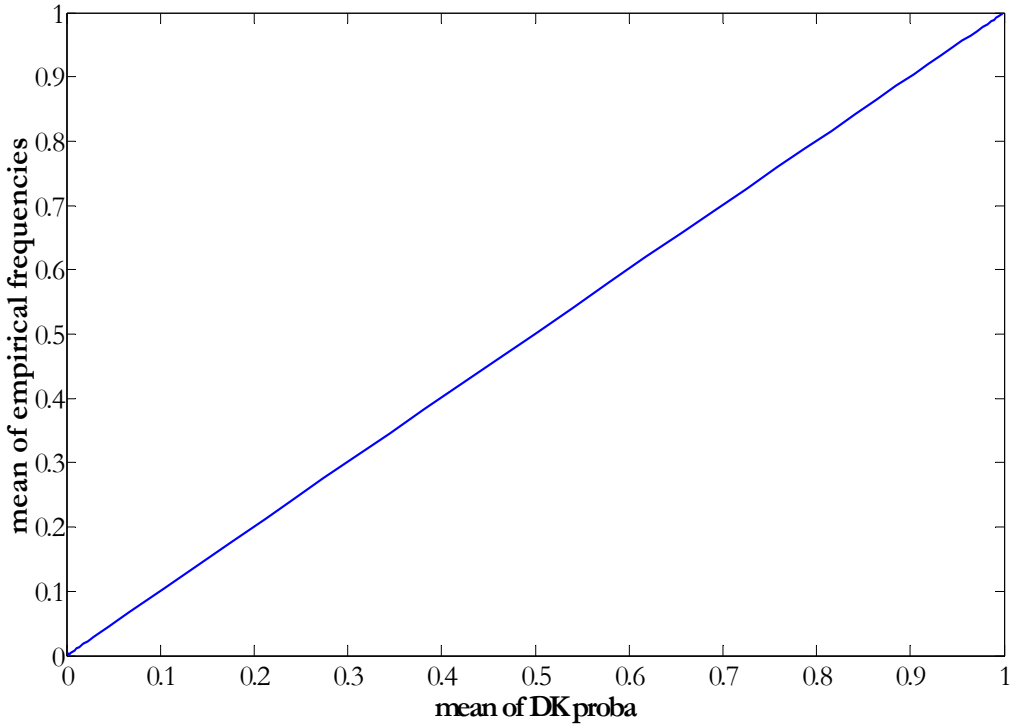


Figure 3.15. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): Mean of the DK probability versus mean of the empirical frequencies for the same thresholds: window length=54 observations, interval length=100 observations.

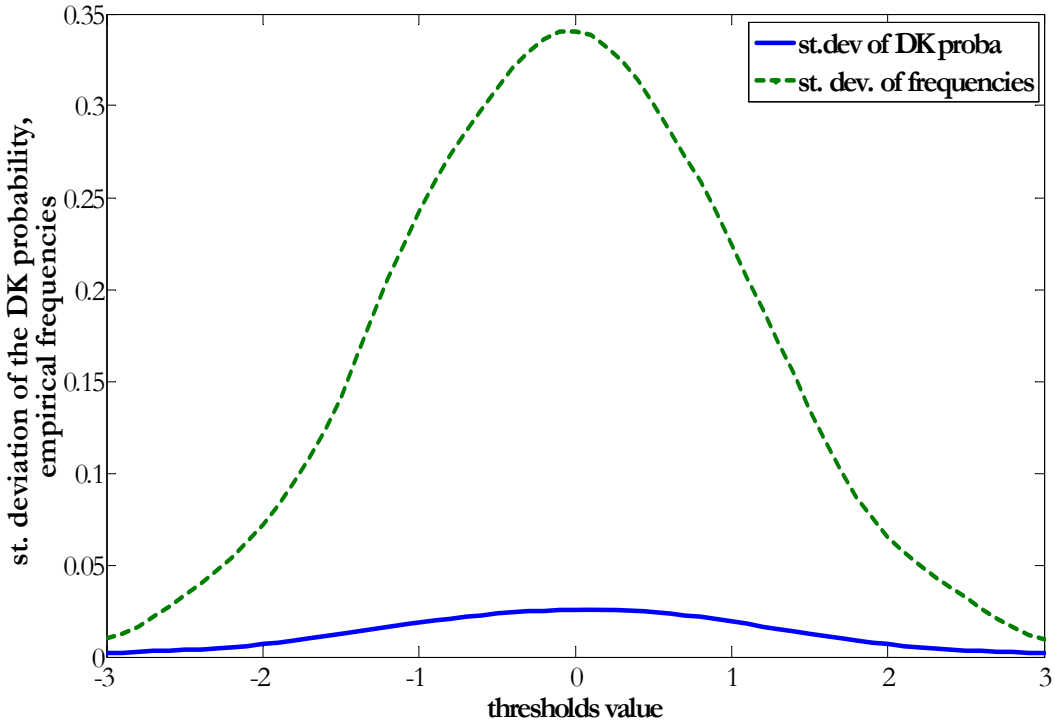


Figure 3.16. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): Standard deviation of the DK probabilities and empirical frequencies for different thresholds (y_c): window length=54 observations, interval length=100 observations

The conclusions about average values of the DK probabilities and their dispersion are interesting from the theoretical point of view. However, they are not very useful for strategy constructions. The fact that there is a large discrepancy between DK probabilities and empirical frequencies, makes it impossible to use the DK estimates as the "true" probability estimators. In addition it is also obvious that there is not only the discrepancy in values that is important, but also the discrepancy in the movements of the DK probabilities and frequencies, i.e. peaks or jumps in the empirical frequencies do not coincide (or even follow) with the peaks in DK probabilities (see Figures B1-B3 in the Appendix B). That is why the obtained results need the improvement in both senses: the value of DK probability and its relative movements with respect to the previous values.

The following approaches might improve the discrepancy between DK probability and empirical frequencies:

1. **The length of EMA, used to estimate the residuals, should be shortening.** The shortening of EMA can help to change the behavior of the residuals. The shorter the EMA, the closer it is to the price curve, therefore the residuals should oscillate more around 0. This might help to avoid the presence of the residuals clusters around some local mean different from zero (the peculiarity observed above).
2. **The data transformation into normal should be made locally.** On one hand, the usage of the large data samples gives more information about random variable. On the other, it smoothes the data too much and eliminates the local trends.

After the DK analysis[12] we found out that the first approach does not provide satisfying results: shortening of the EMA for the residuals extractions does not improve the discrepancy between DK probabilities and the empirical frequencies. Although the method diminishes the variability of the empirical frequencies, it is also diminishes the standard deviation of the DK probabilities; thus, the discrepancy between two values does not improve.

The local transformation of both data and thresholds allows taking into account local trends or their absence. The justification of such approach is based on the hypothesis that in a close neighborhood the data behaves according to the same rules (local stationarity), however, these rules differs for the distant intervals (non-stationarity). Thus, taking into account only local information can improve the short-term data forecast.

We can see that the local transformation will have impact on the DK probabilities though the analysis of the thresholds, transformed once on the basis of the whole data sample and the thresholds, transformed locally on the basis of the continuously rolling sub-samples of the length m[13]. The Figures 3.17-3.18 show the historic threshold $z_C = 0.5$ transformed into normal variable on the rolling sample of the lengths $m_1 = 500$ and $m_2 = 2000$ respectively. Each point of the blue curve on the figures corresponds to the number of the sub-sample, used for threshold transformation: #1 corresponds to data $\{Z_i\}_{1 \le i \le m}$, #k – to data $\{Z_i\}_{k \le i \le k+m-1}$, etc. The red dot line represents the value of the threshold $y_C = 1.56$, obtained for the transform function, calibrated on the whole available data set (20171 observations).

[12] Due to the large volume of the analysis the results are not presented in the paper. These results can be obtained from the author at request.

[13] For the local transformation of a threshold, the empirical CDF, used to calibrate the threshold transformation function is estimated on the rolling sub-sample.

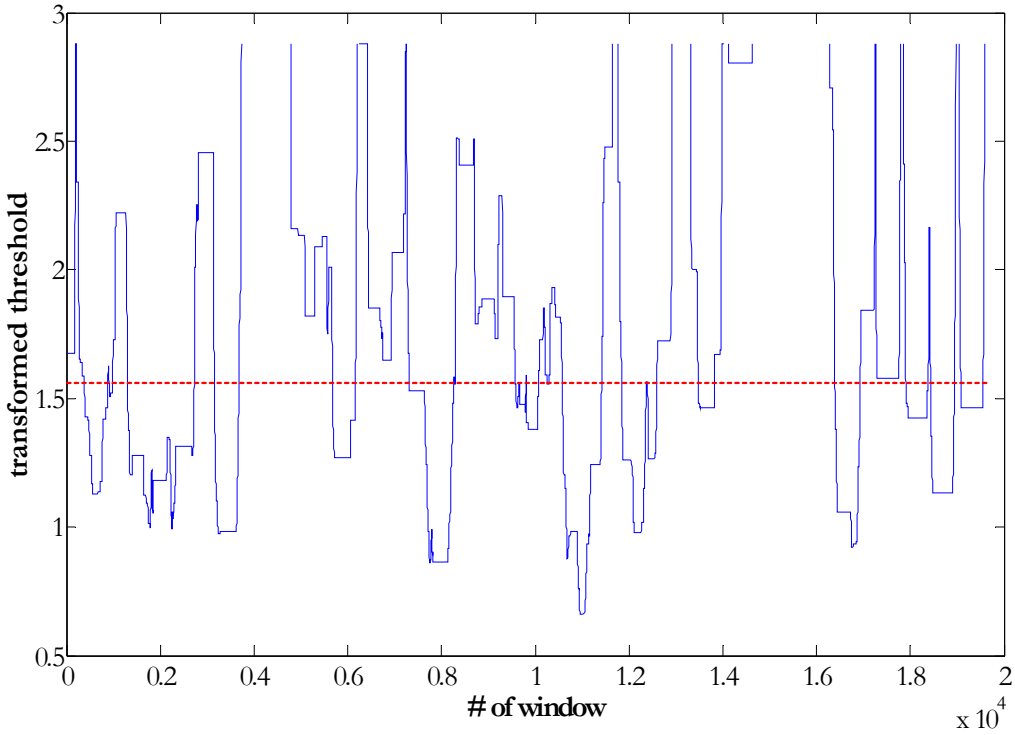


Figure 3.17. Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): The threshold $z_C = 0.5$ transformed on the rolling DCI(threshold) of length $m = 500$ observations (blue line) and whole sample DCI(threshold) (red line)

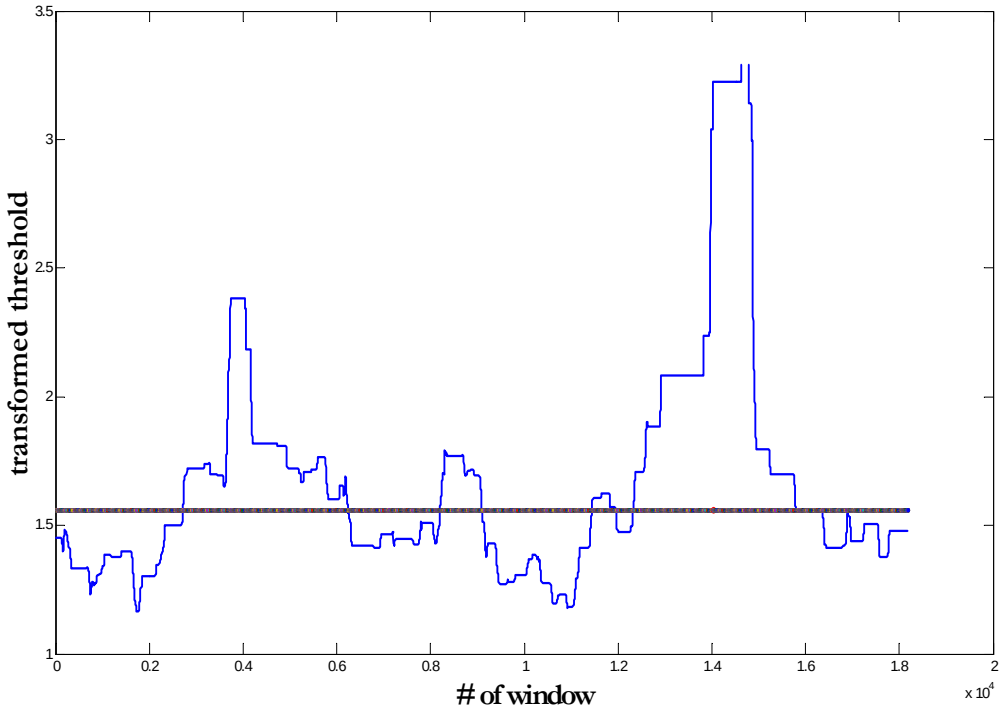


Figure 3.18. Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): The threshold $z_C = 0.5$ transformed on the rolling DCI of length $m = 2000$ observations (blue line) and whole sample DCI(threshold) (black line)

We can see from figures how the values of the transformed thresholds are data-sensitive, i.e. they depend on the sub-sample, on which the transform function is calibrated. Note that the gaps in the blue curve imply infinite values of the transformed thresholds. This happens if a sub-sample, used for the calibration of the transform function does not contain any observations of value $z_C = 0.5$ [14]. As we can see the transformed values varies a lot. This variability depends on the length of the sample used for data transformation: longer samples correspond to less variable thresholds values (compare Figures 3.17 and 3.18). The following Chapter 4 provides the analysis of the interval DK probabilities, estimated under assumption about local stationarity.

4 Examples of the DK probabilities under local stationarity assumption

As have been shown in Chapter 3, the estimates of the DK interval probabilities (for the same transform function) coincide on average with the empirical frequencies. However, these results cannot be used in the strategy construction. We have also shown that local transformation of the threshold brings the results significantly different than the threshold value that is obtained for the transformation function, calibrated on the whole sample. Therefore, in this chapter we propose to analyze the interval DK probability under the assumption of local stationarity. The DK estimates in this Chapter 4 are obtained for the local transform function, adjusted to the local volatility as discussed in Chapter 2.2. Chapter 4.1 explains the framework for the estimation of the interval DK probabilities. Chapter 4.2 compares the interval DK probabilities with the empirical frequencies, as well as analyzes the impact of the parameters choice (interval, window lengths, etc.) on the DK estimates.

4.1 Framework for the estimation and analysis of the interval DK probabilities

Let's define the general framework for the analysis. Let S_i is available sample of the historic data for the *i*-th instrument. Let $S^{(1)}_{1/2,i}$ and $S^{(2)}_{1/2,i}$ represent the first and second halves of the sample S_i. Sub-sample $S^{(1)}_{1/2,i}$ (first half of the original sample) is used to estimate the models further used for DK application procedure: (1) basic CDF and (2) variogram model for the normal variable, transformed from the historic data $S^{(1)}_{1/2,i}$ according to the basic CDF. The *interval* DK probabilities are estimated for the second sub-sample $S^{(2)}_{1/2,i}$. Such approach allows avoiding the mentioned data-snooping bias. This approach also enables us to apply the method on the routine basis, as it does not anticipate the future.

The interval DK probability $P^I_{DK}(t,l,y_c) = \Pr(Y_i < y_c, i \in [t+1; t+l])$ is estimated for the interval of length *l* and threshold y_c. Threshold y_c is obtained by the transformation of the historic threshold z_c. We choose some vector of constant values (fixed for the whole time period) for the threshold z_c: $z^c_t = const$. Such threshold choice coincides with the formulation of the trading bands strategies with the MA as the middle line and parallel shift of the MA up/down for the bands values (see Part II for more details). The algorithm for these strategies

[14] In this case the empirical probability of observing this threshold value is equal to 0; thus, the transformed variable takes infinite values.

can be reduced to the comparison of the residuals (difference between the price and MA) with a constant threshold.

The analysis of the prediction quality of the interval DK probabilities is made though the comparison of the DK probabilities with the respective empirical frequencies, as well as though the analysis of the impact the parameters choice, such as length of the estimated window (*n*) and interval length (*l*), has on the DK values.

Due to the large number of different parameters and sub-samples, used in the DK applications we propose the following vocabulary of terms for the results description:

1. *Distribution calibration interval (DCI)* is a sub-sample of historic data, used for the estimation (calibration) of an empirical distribution function or adjustment of a basic CDF (as discussed in chapter 2.2).
2. *Window* is a sub-sample of normal (transformed) data, used in the DK procedure for kriging of the Hermite polynomials.
3. *Interval* is a sub-sample of historic data, used for empirical probability (frequency) calculations.
4. *DK probabilities* are probabilities $P_{DK}^{I}(t,l,y_c) = \Pr(Y_i < y_c, i \in [t+1;t+l])$, estimated according to the DK procedure.
5. *Z-frequencies* are empirical frequencies $P_{emp}^{Z} = \dfrac{\sum_{i=t+1}^{t+l} I(Z_i < z_c)}{l}$ calculated on the basis of the historic data (residuals).

Let's use the schematic representation of these terms in Figure 3.19. It explains how the sub-samples are situated at the time coordinate with respect to some current moment of time t.

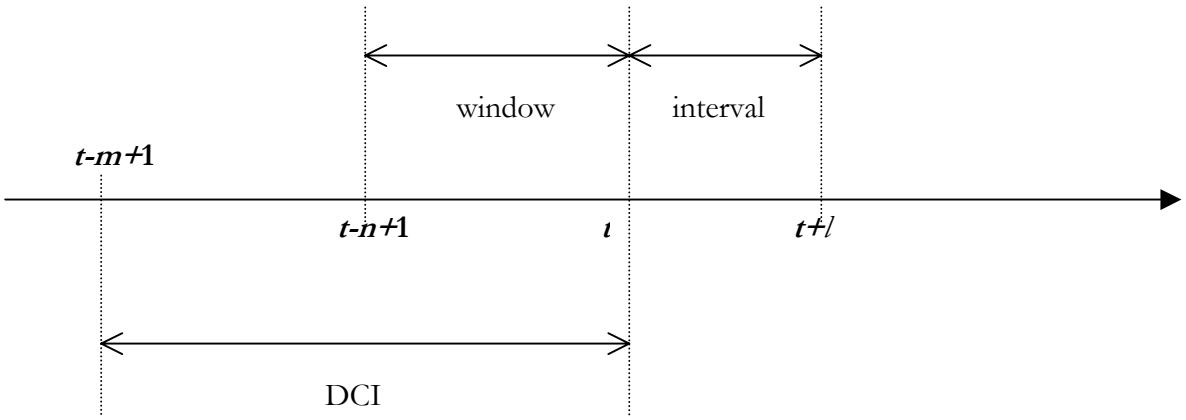


Figure 3.19. Schematic representation of the sub-samples in the time space

This figure 3.19 also represents the way the DK calculations are performed. Suppose we want to estimate DK interval probabilities at moment of time t and compare them with real-life frequencies using available historic data sample $\{Z_i\}_{0<i<N}$, where $t << N$. Then we use $\{Z_i\}_{t-m+1\le i\le t}$, the sub-sample *DCI* of the length m, to adjust the basic transformation function φ_b for volatility (see Chapter 2.2 for more details) and transform historic data Z and threshold

z_c into normal variable $\{Y_i\}_{t-m+1\le i\le t}$: $Y=\varphi_t(Z)$ and $y_c=\varphi_t(z_c)$. For the DK calculations we use the threshold value y_c and part of the normal data $Y_{[t-n+1;t]}\in Y_{[t-m+1;t]}$ as the *window* of the length n. We estimate the *DK interval probabilities* $P^I_{DK}(t,l,y_c)=\Pr(Y_i<y_c, i\in[t+1;t+l])$ for the *interval* $[t+1;t+l]$ of the length l. We calculate the *Z-frequencies* for the data $\{Z_i\}_{t+1\le i\le t+l}$ to compare with the DK probabilities $P^I_{DK}(t,l,y_c)$.

The DK analysis is performed according to the following algorithm:

1. Basic (non-adjusted) transformation function φ_b for the transformation of the historic data Z and its thresholds z_c into normal variable Y and y_c is calibrated on the first half of the sample $S^{(1)}_{1/2}=\{Z_i\}_{1\le i\le N/2}$, where N is the length of the sample.

2. Variogram parameters are estimated once, on the basis of the sample of $\{Y_i\}_{1\le i\le N/2}$:

$$Y_i=\varphi_b(Z_i), i\in[1;N/2].$$

3. The local transformation function φ_t is the basic transformation function φ_b, adjusted for the local volatility σ_t according to the following procedure (see Chapter 2.2). Let $\Phi^{(t)}(Z)$ be the local CDF function that we want to obtain (adjust) at the moment t for the historic data/threshold transformation:

$$Z=\varphi_t(Y)\to\Phi^{(t)}(Z)=F\left(\varphi_t^{-1}(Y)\right)$$

   We have estimated already the basic CDF $\Phi^{(b)}(Z)$ at the step #1. The adjustment of the basic CDF coincides with the adjustment of the historic data (threshold) that we want to transform:

$$\Phi^{(t)}(u)=\Phi^{(b)}\left(\frac{\sigma_b}{\sigma_t}u\right),\ \Phi^{(b)}(Z)=F\left(\varphi_b^{-1}(Y)\right).$$

   The volatility ratio is defined by the following values:

$$\sigma_b=\sqrt{\mathrm{E}\left[\left(Z_i-\overline{Z}_b\right)^2\right]}, i\in[1;N/2]$$
$$\sigma_t=\sqrt{\mathrm{E}\left[\left(Z_i-\overline{Z}_l\right)^2\right]}, i\in[t-m+1;t]$$

   where m is DCI length, N - sample length, $\overline{Z}_b=\mathrm{E}[Z_i], i\in[1;N/2]$ - mean of the first half of the sample $S^{(1)}_{1/2}$, $\overline{Z}_l=\mathrm{E}[Z_i], i\in[t-m+1;t]$ - local mean of the sub-sample.

4. Sub-sample Z_{DCI} and threshold z_c are transformed into normal Y_{DCI} and y_c according to the local transformation function φ_t.

5. For the estimation of the DK probabilities $P_{DK,t}\left(Y_{[t+1;t+l]} < y_c\right)$ at time $t > N/2$ we use normal variable sub-sample $Y_\alpha \subset Y_{DCI}$: $Y_\alpha = \{Y_i\}, i \in [m-n+1; m]$, where n - length of the window, used for DK procedure. Note that $0 < n \le m$.

6. DK probabilities at moment i $P_{DK,i}\left(Y_{[i+1;i+l]} < y_c\right)$ are calculated for the sample $i \in [N/2+1; N-l]$, where l is the length of the interval, for which the estimation procedure is performed.

7. Respective empirical probabilities at moment i $P^*_{emp,i}\left(Z_{[i+1;i+l]} < z_c\right)$ are calculated for historic data $\{Z\}_{i\in[N/2+1,N]}$:

$$P^Z_{emp,t} = \frac{\sum_{i=t+1}^{t+l} I\left(Z_i < z_c\right)}{l}.$$

8. The end of the estimation window *t* is moved on the time line with the step of 10 observations[15]: $t_k = t_{k-1} + 10$.

Several parameters intervene directly or indirectly in the DK calculations. The following parameters used directly in the DK calculations: (1) the length of the moving window of normal variable, used for Hermite kriging; (2) the length of the interval, for which the probabilities are estimated. The other parameters can intervene indirectly through estimation of the CDF, used in transformation: (1) the length of the DCI interval on which the adjustment of the basic CDF is made; (2) the bandwidth value used in the kernel procedure for CDF estimation. In order to analyze the impact of these parameters on the DK results we choose the length of the DCI, window and intervals at the levels that represent the "short", "medium" and "long" windows and intervals. As this classification is quite general and subjective, we put the following characteristics in the base of the differentiation between "short", "medium" and "long". The choice of the DCI covers sub-samples from 50 to 500 observations. For the windows the length is defined by the range parameter of the variogram, as it represents how long the past observations has impact on the future prediction: "long" length is represented by the distance, at which the variogram (graphically) converges toward the variance level; "medium" and "short" lengths represent approximately half and forth part of the "long" window. For the interval length, the "short" interval should still contain sufficiently observations to calculate the empirical frequencies (i.e. at least 100 observations).

Further, Chapter 4.2 presents the examples of the DK probabilities for the DAX residuals after extraction of the EMA of the one-week length (115 observations).

4.2 Interval DK probabilities: DAX instrument

The statistical comparison of the DK probabilities with the empirical frequencies involves estimation of some indexes or statistics tests that allow judging about the predictive properties of the DK method. However, as we will see further due to the specific nature of the estimates the classical indexes, such as mean squared error, are meaningless or complicate to calculate.

[15] The tabulation step of 10 observations is used in order to reduce the calculations time due to the length of the samples.

Therefore, we use graphical representations of the DK probabilities and Z-frequencies to compare them. The numerical evaluation of the DK prediction power is substituted by the analysis of the trading outcomes of the strategies that incorporate the DK probabilities (see Chapter 5 for the analysis).

Note that some of the transformed thresholds values presented further take the infinite values that graphically are reflected in the gaps in the DK probabilities curves. Such situation is observed for the historic thresholds that after adjustment for the volatility ratio were never attained in the first half of the sample. For example, let's consider the non-volatile periods when the local volatility of rolling sub-samples is twice as small as basic volatility. Then the threshold of 70 is adjusted to the 140 level that was never attained on the first half of the sub-sample (see Figure 3.4). Such situation is more likely for the thresholds more distant from 0. We will observe these gaps further in all figures, as they represent the particular examples of the DK probabilities for the historic threshold $z_c = 70$.

The following subchapters 4.2.1-4.2.4 analyse the impact of the window, interval, DCI lengths and bandwidth parameter on the interval DK probabilities.

4.2.1 Impact of the window length

The experiment is organized as following:

1. Variable parameter – window length:
   - $n_1 = 30$ observations;
   - $n_2 = 60$ observations;
   - $n_3 = 150$ observations.
2. Fixed parameters:
   1. DCI: $m = 100$ observations ($m' = 150$ for the $n_3 = 150$)
   2. Interval: $l = 100$ observations
   3. Historic thresholds (non-transformed): $z_c \in [-250 : 20 : 160]$

The examples of the DK probabilities and respective empirical probabilities for the threshold $z_c = 70$ are given in Figures 3.20-3.22. The *X*-coordinate represents the number of the interval, for which the estimations was done, for example, #1 - $[t+1; t+l]$; #2 - $[t+11; t+11+l]$; etc.[16]

The following peculiarities are observed: (1) some discrepancy in the absolute terms between DK probabilities and frequencies; (2) the presence of the gaps in the DK curves (explained above).

The DK probabilities in some cases lag the empirical frequencies curves, but this lag is very small. Moreover, the periods when both curves are different from one often coincides leading us to the conclusion that the prediction power (at least timing) of the DK method is quite good. As we can see, the window size does not have significant impact on the DK probabilities. Large window produces slightly different results, but the difference is not very large.

[16] Note that the moment of time *t* is tabulated with step 10 (see Ch.4.1)

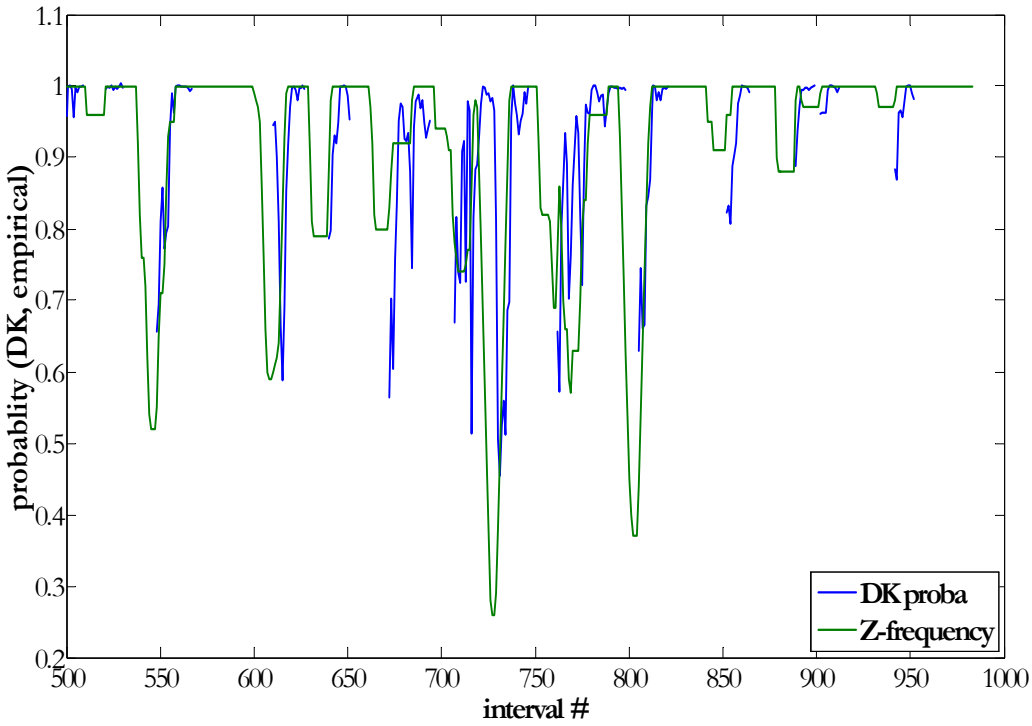


Figure 3.20. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 obs): DK probabilities and empirical Z-frequencies: DCI (calibration)=100 obs, interval=100 obs, window=30 obs, threshold(z)=70

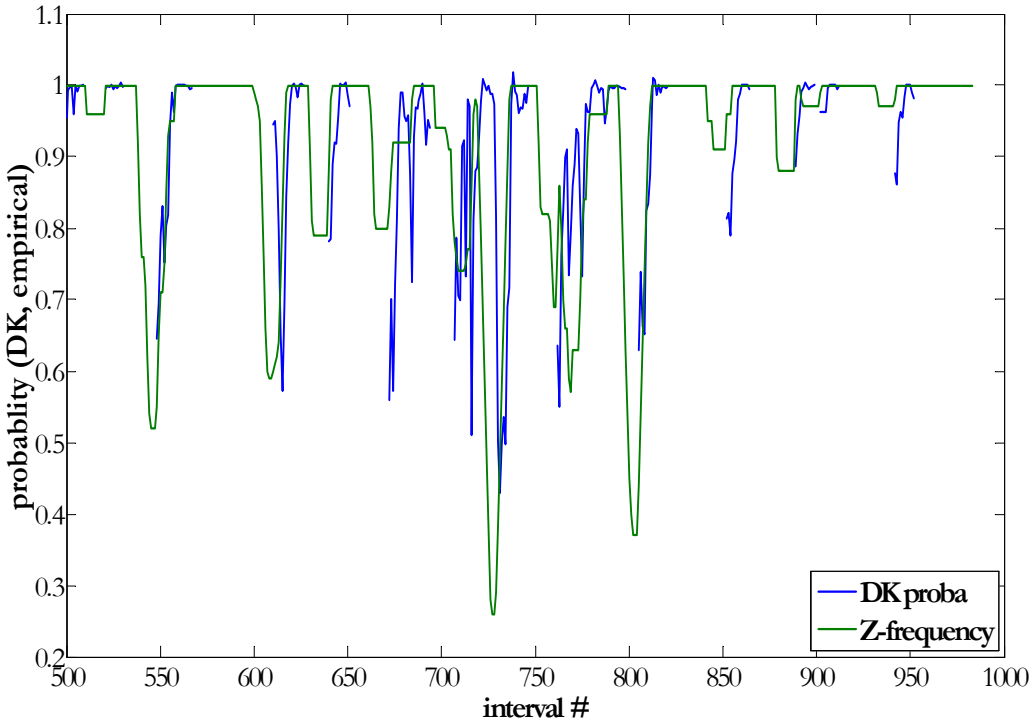


Figure 3.21. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 obs): DK probabilities and empirical Z-frequencies: DCI (calibration)=100 obs, interval=100 obs, window=60 obs, threshold (z)=70

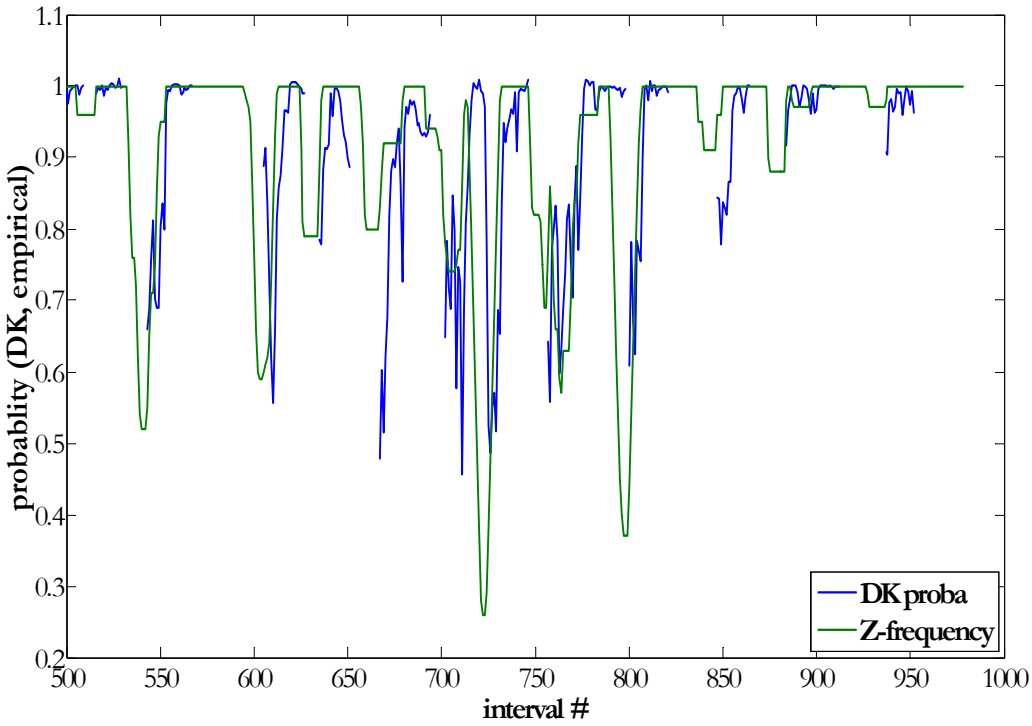


Figure 3.22. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 obs): DK probabilities and empirical Z-frequencies: DCI (calibration)=100 obs, interval=100 obs, window=150 obs, threshold (z)=70

4.2.2 Impact of the interval length

The experiment is organized as following:

1. Variable parameter – interval length:
   - $l_1 = 100$ observations;
   - $l_2 = 250$ observations;
   - $l_3 = 500$ observations.
2. Fixed parameters:
   - DCI: $m = 100$ observations
   - Window: $n = 60$ observations
   - Thresholds (non-transformed): $z_c \in [-250 : 20 : 160]$

As in the previous case the DK probabilities show some good prediction power in the term of timing. If we compare figures 3.21, 3.23 and 3.24 we can see that increase of the interval improve the prediction power of the DK method in both timing and absolute terms, although to certain point: increase of the interval to 500 observations decreases the prediction power due to the significant smoothing of the empirical frequencies curve.

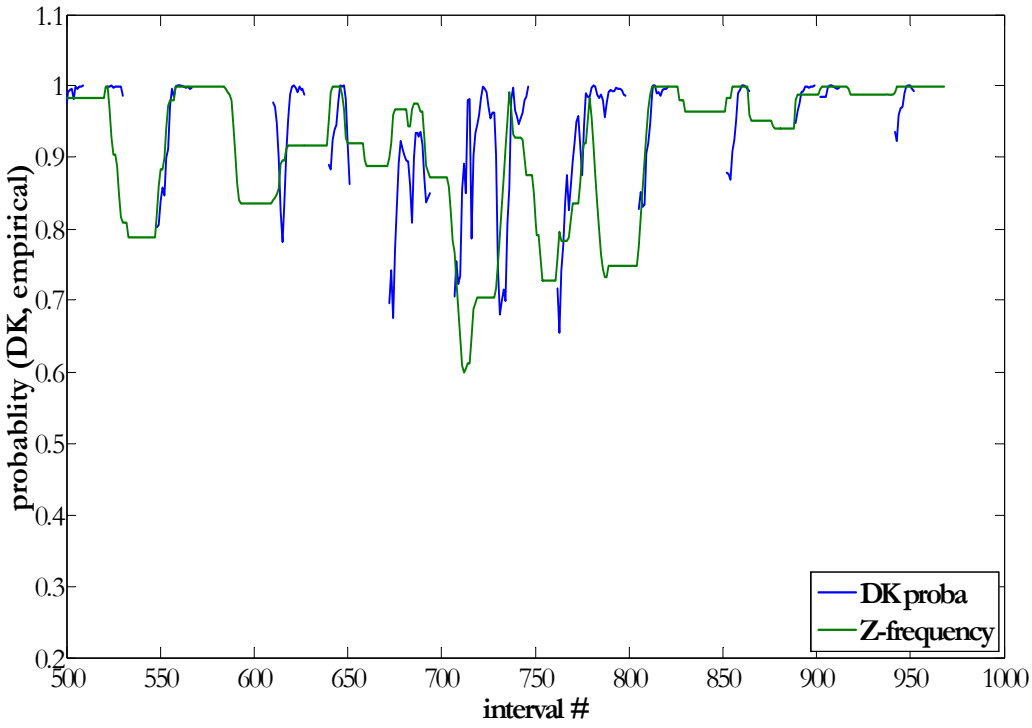


Figure 3.23. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 obs): DK probabilities and empirical Z-frequencies: DCI (calibration)=100 obs, interval=250 obs, window=60 obs, threshold (z)=70

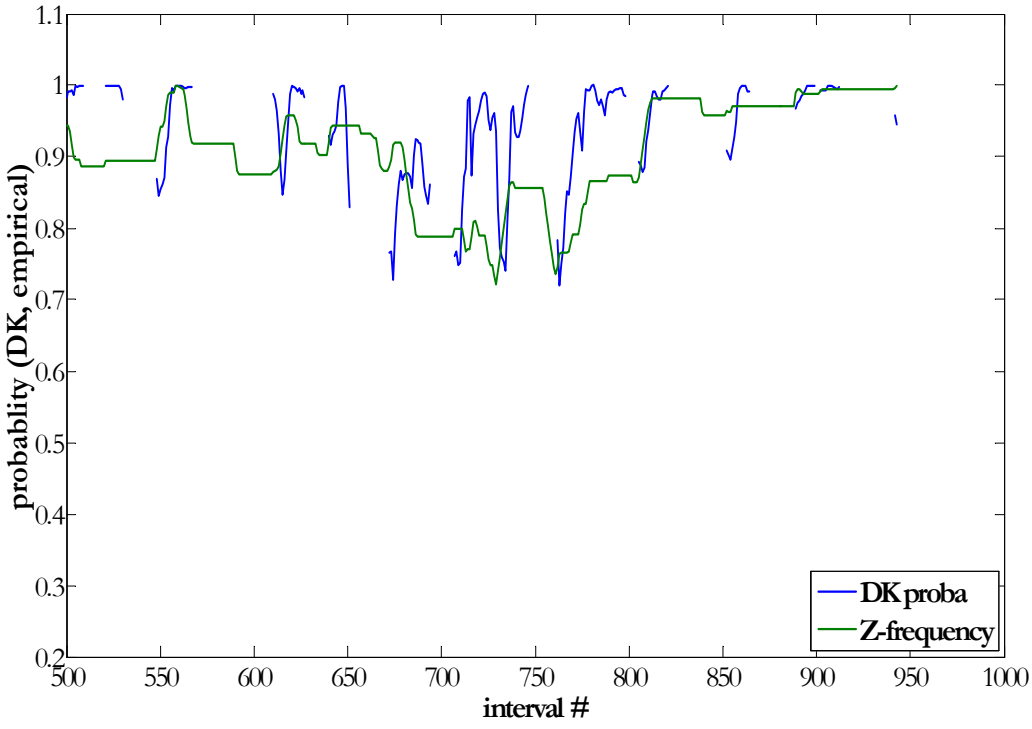


Figure 3.24. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 obs): DK probabilities and empirical Z-frequencies: DCI (calibration)=100 obs, interval=500 obs, window=60 obs, threshold (z)=70

4.2.3 Impact of the DCI length

The experiment is organized as following:

1. Variable parameter – DCI interval length:
   - $m_1 = 60$ observations;
   - $m_2 = 100$ observations;
   - $m_3 = 250$ observations ;
   - $m_4 = 500$ observations.
2. Fixed parameters:
   - Interval: $l = 100$ observations;
   - Window: $n = 60$ observations;
   - Thresholds (non-transformed): $z_c \in [-250 : 20 : 160]$

Contrary to the previous two cases, we can notice that the prediction power of DK probabilities depends on the DCI length. Figures 3.21 and 3.25-3.27 show that the short DCI intervals (60 and 100 observations) allow obtaining the best DK estimates in terms of timing and values. For the longer DCI of 250 and 500 observations the lag and discrepancy (in absolute terms) between two curves increase. Therefore, further for trading simulations we are considering only short DCI intervals of 60, 100 and 150 observations.

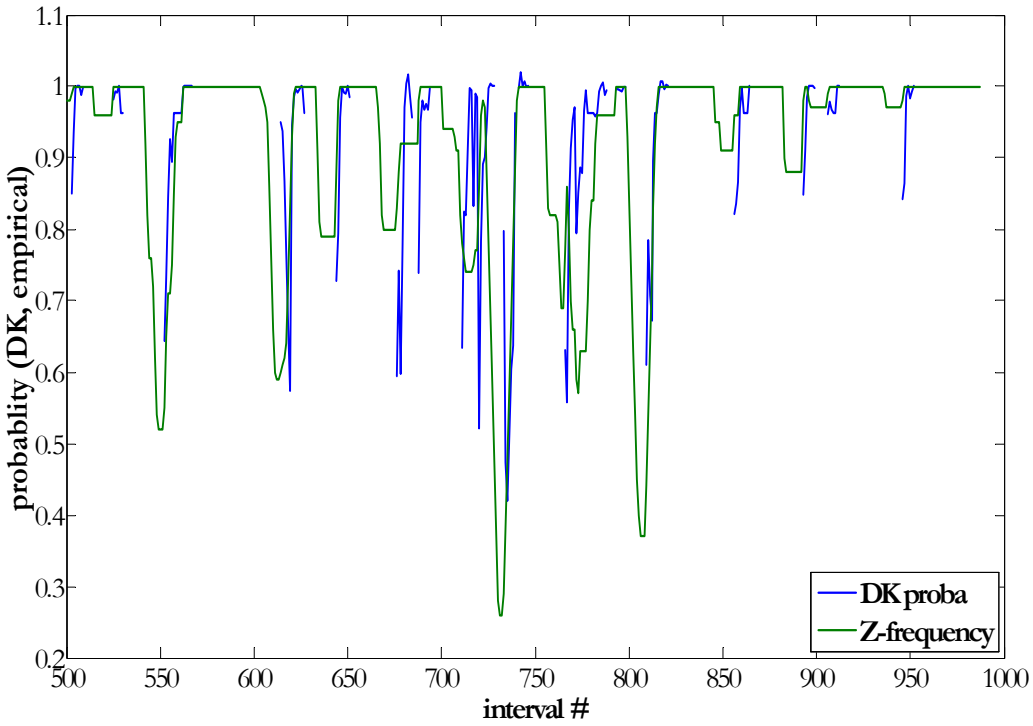


Figure 3.25. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 obs): DK probabilities and empirical Z-frequencies: DCI (calibration)=60 obs, interval=100 obs, window=60 obs, threshold (z)=70

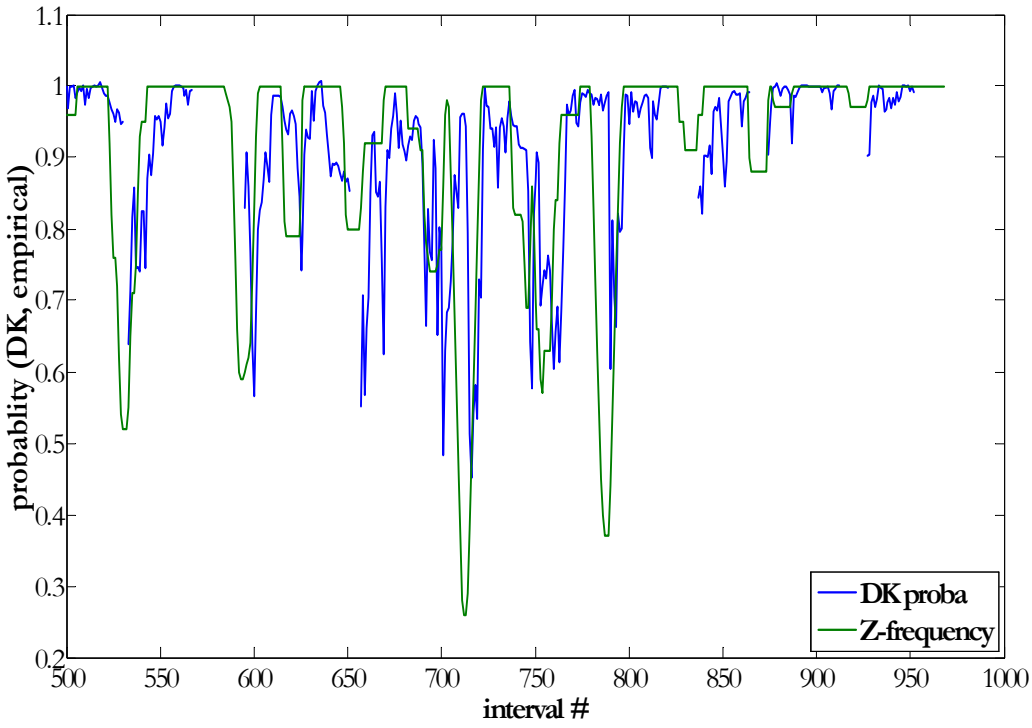


Figure 3.26. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 obs): DK probabilities and empirical Z-frequencies: DCI (calibration)=250 obs, interval=100 obs, window=60 obs, threshold (z)=70

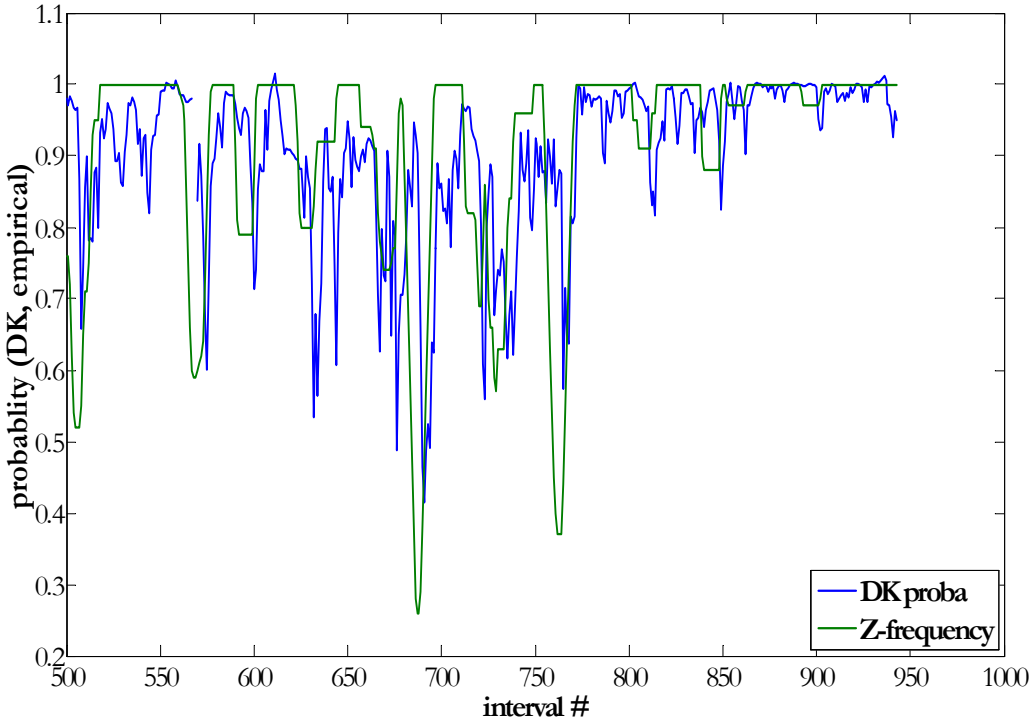


Figure 3.27. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 obs): DK probabilities and empirical Z-frequencies: DCI (calibration)=500 obs, interval=100 obs, window=60 obs, threshold (z)=70

4.2.4 Impact of the bandwidth parameter

The experiment is organized as following:

1. Variable parameters – bandwidth parameter:
   - $b_1 = 7$
   - $b_2 = 15$
2. Fixed parameters:
   - DCI interval length: $m_2 = 100$ observations;
   - Interval: $l = 100$ observations;
   - Window: $n = 60$ observations;
   - Thresholds (non-transformed): $z_c \in [-250 : 20 : 160]$

We propose to consider the scatter point between the DK probabilities, calculated for the transform functions that are based on the estimates of the empirical CDF with different bandwidth parameters: $b_1 = 7$ and $b_2 = 15$. As we can see from Figure 3.28 the bandwidth parameter does not have a very big impact on the DK probability values: the scatter point graph form a line. We will see the confirmation of this hypothesis in Chapter 5; the outcomes of the trading strategies, based on DK probabilities for DAX (Ch. 5.3.1) and Bund (Ch. 5.3.2) do not differ significantly for different bandwidths values. This is why for other instruments we will report only the trading outcomes for bandwidth value, chosen automatically by MATLAB procedure to fit the distribution curve.

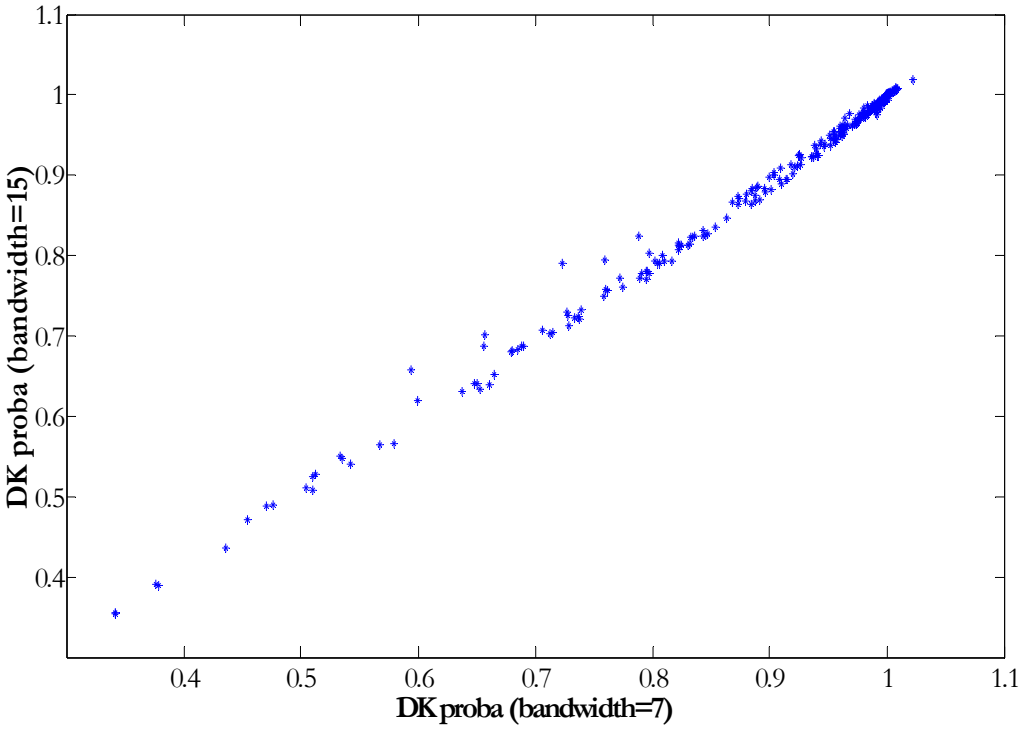


Figure 3.28. DAX residuals (30/7/2003-7/12/2006, frequency 30 minutes, n(EMA)=115 obs): DK probabilities for bandwidth=7 versus DK probabilities for bandwidth=15: DCI (calibration)=100 obs, interval=100 obs, window=60 obs, threshold (*z*)=70

The following conclusions summarize the results of the analysis:

1. The prediction power of the DK method is quite good. The DK probabilities in some cases lag the empirical frequencies curves, but this lag is very small. Moreover, the periods when both curves are different from 1 often coincides. At the same time some discrepancy exists in the absolute values of the DK probabilities and empirical frequencies. These discrepancies as well as the gaps in the DK probabilities curve make the classical numerical measure of the prediction power of the DK method impossible.
2. Neither window length nor bandwidth parameter has impact on the values of DK probabilities.
3. Interval length has an impact on the prediction power of the DK probabilities. At low values increase of the interval length improves the prediction power, while at high levels increase in the interval length diminishes the prediction power.
4. DCI length has an impact on the prediction power of the DK probabilities. The shorter is the interval the better is the prediction properties of the DK method. It is likely that the long DCI samples smooth too much the local volatility estimates.

Someone might argue that the discrepancies in the values of the DK probabilities and frequencies should lead to the conclusion about low prediction power of the method. However, for the construction of the trading strategies the most importance has the timing of the prediction. At the same time bad approximation of the probability value can be avoided by its comparison with the (probability) threshold. For example, the probability can be considered significant (that leads to taking a position) if it is larger than some threshold value. Chapter 5 proposes the example of a trading strategy that incorporates the DK probabilities, as well as the benchmark strategy to evaluate whether the introduction of the DK method improves the trading results.

5 Application of the DK method to strategies construction

We have shown in chapter 4 that the DK method produces probabilities that are comparable with empirical frequencies. In this chapter we present the example of the strategies that incorporate DK probabilities, as well as analyse their outcomes. We compare their results with some benchmark strategy, constructed on the basis of the well-known hypothesis in the finance – random walk. The sub-chapter 5.1 define a *"DK trading strategy"* – the strategy constructed on the basis of the DK probabilities; while sub-chapter 5.2 presents a "*Benchmark strategy"* – the strategy, the results of which will be used as the comparison benchmark. The trading results for four different instruments are analysed in sub-chapter 5.3.

Further we use the following notation to define the strategies:

$\{PRICE_i\}_{i>0}$ - price series;

$\{Z_i\}_{i>0}$ - price residuals: $Z_i = PRICE_i - EMA_i$;

$Pos(t)$ - position taken at moment *t*

$PRICE_{entry}$ - price at which the position is taken;

$PRICE_{exit}$ - price at which the position is liquidated;

$\{\Pi_j\}_{j>0}$ - profit per transaction: $\Pi_j = Pos_j \cdot (PRICE_{exit} - PRICE_{entry})$

5.1 DK trading strategy

Consider trading band strategy[17] with EMA as the middle line, and upper/lower bands represented by the shift of EMA up/down by *r* units. Let's consider trend-following strategy, i.e. the objective is to buy at low price and sell at high. Then if the price at current moment *t* is within the bands (1) take long position *L* at moment *t* if probability of breaching upper band during the following interval of length *l* is higher than probability of breaching lower band; or (2) take short position *S*, if the probability of breaching lower band is higher than the probability of breaching upper band; otherwise (3) take no position, if probabilities of breaching any of the bands (upper or lower) is 0. Exit (non-zero) position (*L* or *S*): (1) when the price breaches upper/lower bands or (2) at the end of period *t+l*. We suppose that entry and exit prices are equal to the prices (quotes) at the moment of entry/exit (no slippage or transaction costs).

DK trading strategy (algorithm):

If $|Z_t| \le r$

Entry: if $P_{DK}\left(Z_{[t+1;t+l]} < -r\right) < 1 - P_{DK}\left(Z_{[t+1;t+l]} < r\right)$

Then $Pos(t) = 1;\ PRICE_{entry} = PRICE(t)$

Otherwise

if $P_{DK}\left(Z_{[t+1;t+l]} < -r\right) > 0$

Then $Pos(t) = -1;\ PRICE_{entry} = PRICE(t)$

Otherwise $Pos(t) = 0$

Exit: if $|Z_v| \ge r, v \in [t+1; t+l]$

Then $Pos(v) = 0;\ PRICE_{exit} = PRICE(v)$

Otherwise $Pos(t+l) = 0;\ PRICE_{exit} = PRICE(t+l)$

5.2 Random-walk (or benchmark) strategy

Consider trading band strategy as in Ch.5.1. The trend-following strategy is formulated in the following way: take long/short position *L* at moment *t* with probability 0.5; exit position (*L* or *S*) when the price breaches upper/lower bands or at the end of period *t+l*.

RW trading strategy (algorithm) :

If $|Z_t| \le r$

Entry: Draw uniformly distributed variable U on the interval $[0;1]$

If $U > 0.5$

Then $Pos(t) = 1;\ PRICE_{entry} = PRICE(t)$

Otherwise $Pos(t) = -1;\ PRICE_{entry} = PRICE(t)$

Exit: if $|Z_v| \ge r, v \in [t+1; t+l]$

[17] See General Introduction and Part II for trading bands definitions.

Then $Pos(v) = 0;\ PRICE_{exit} = PRICE(v)$

Otherwise $Pos(t+l) = 0;\ PRICE_{exit} = PRICE(t+l)$

5.3 Strategies outcomes

Both strategies as in 5.1 and 5.2 are applied to four different instruments: (1) DAX; (2) Bund; (3) Brent; (4) X instrument[18]. Description of the DAX instrument is given previously in Chapters. 2-4, while short review of other instruments can be found in the Appendices C, D and E respectively. Their main statistical characteristics are presented in Table 3.1.

Table 3.1
Statistics for the instruments residuals

Statistics for residuals	DAX	Bund	Brent	X instrument
Mean	7.61	0.03	0.27	0.23
Standard deviation	49.39	0.32	1.32	2.30
St.deviation / Mean	6.5	10.7	4.9	10
Max	185	1.12	4.27	12.77
Min	-291	-1.32	-3.24	-8.79

As in chapter 4 we divide each sample in two sub-samples. First part is used for the evaluation of the basic CDF and volatility, as well as the variogram model for the transformed into normal historic data. The second part is used for the DK probability estimations and trade simulations. All trade simulations for each instrument start at the same point, making the outcomes for different parameters values, used in the simulations, comparable.

As in previous applications, we calculate DK probabilities and make trading decision each 10th observation $(t, t+10, t+20, ...)$; at the same time we use all available data before $(t, t-1, ..., t-DCI)$ and after $(t, t+1, ..., t+l)$ time moment t for CDF calibration and DK probabilities calculations, as well as transaction simulations. For each point $t, t+10, t+20, ...$ the profit series $\{\Pi_j\}_{j>0}$ are generated for both strategies. In order to compare the outcomes for both strategies we define the cumulative value of strategy (III.5.1) and average profit per transaction as in (III.5.2):

$$V_j = \sum_{u=0}^{j} \Pi_u \qquad \text{(III.5.1)}$$

$$V_a = \frac{\sum_{u=0}^{J} \Pi_u}{J} \qquad \text{(III.5.2)}$$

where J is a number of trades.

Note that although these values represent P&L prototypes, they cannot be treated as such. Our "prices" are defined in quotes/ticks (not in monetary units); therefore, our transactions values are calculated in these quote units. Zero-transaction costs assumption is applied though all simulations of trading strategies. This hypothesis will not invalidate our conclusions, as the principal goal of the analysis is to see if DK strategy gives better results than random-walk strategy.

[18] Due to confidentiality reason we cannot present the instrument in details.

5.3.1 DAX

Figure 3.29 presents the (end-of period) sum of the payoffs simulated for different DCI lengths. As we can see the DK strategy brings positive outcomes for the thresholds interval of medium values $r \in [50;80)$. The random walk strategy generates low and volatile outcomes that fluctuate around 0[19] (see Figure 3.29). Global maximum for all outcomes is reached for the DK strategy at threshold level of $r = 60$. The value of the DCI has an impact on the total outcome: the highest outcome is generated for the DCI length of 150 observations.

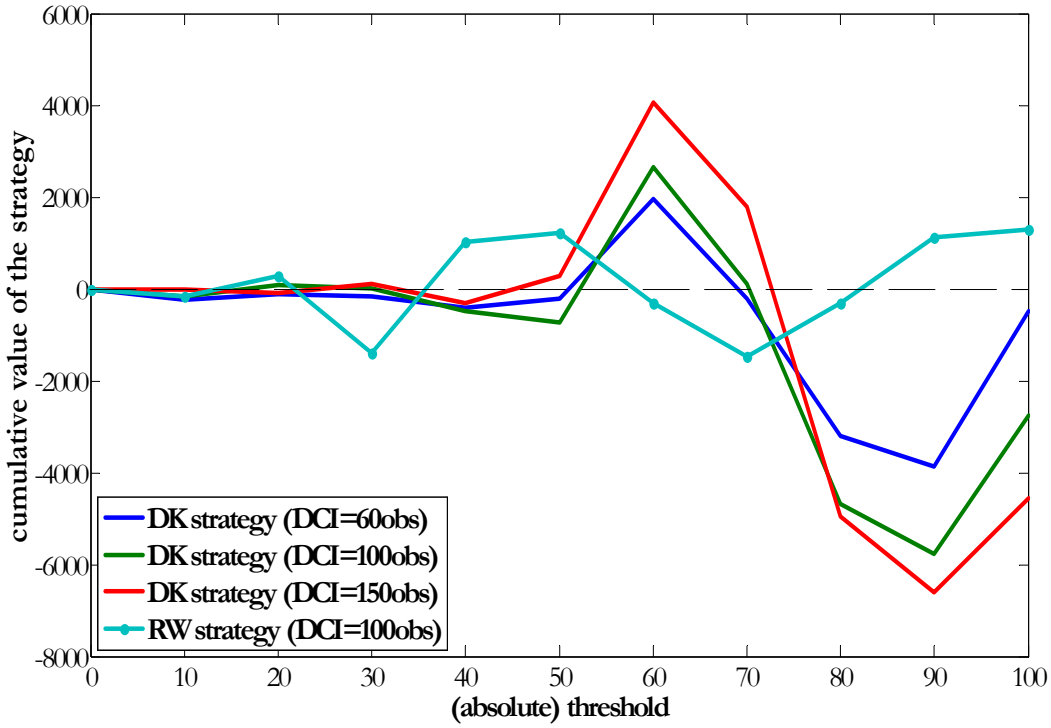


Figure 3.29. DAX (30/7/2003 - 7/12/2006, frequency 30 min, observations 10029-20056). Cumulative value (at the end of period) of the trading band strategy for different thresholds, based on DK and random walk (RW) approaches: *n*(EMA)=115 observations, DK parameters - window length=60 observations, interval length=100 observations, DCI length=60, 100 and 150 observations

Figure 3.30 presents the paths of the DK and RW strategies for different DCI parameter values with the highest cumulative outcomes ($r = 60$). DK path is more volatile than for the RW strategy, however it still exhibits positive general trend. The chosen path for the random walk strategy accumulates losses. We can notice the drop in the strategy value (around 600 observations) after some relatively positive trend; it is possible that recalculation of some of the parameters (for example, variogram model, can be used to improve the prediction power of DK method.

[19] Figure 3.29 presents only one case for the RW strategy for DCI=60 observations in order to not overload the graph. However, for other DCI lengths the form of the RW outcomes curve is similar: it is volatile around 0 and generates lower profits than the DK strategies.

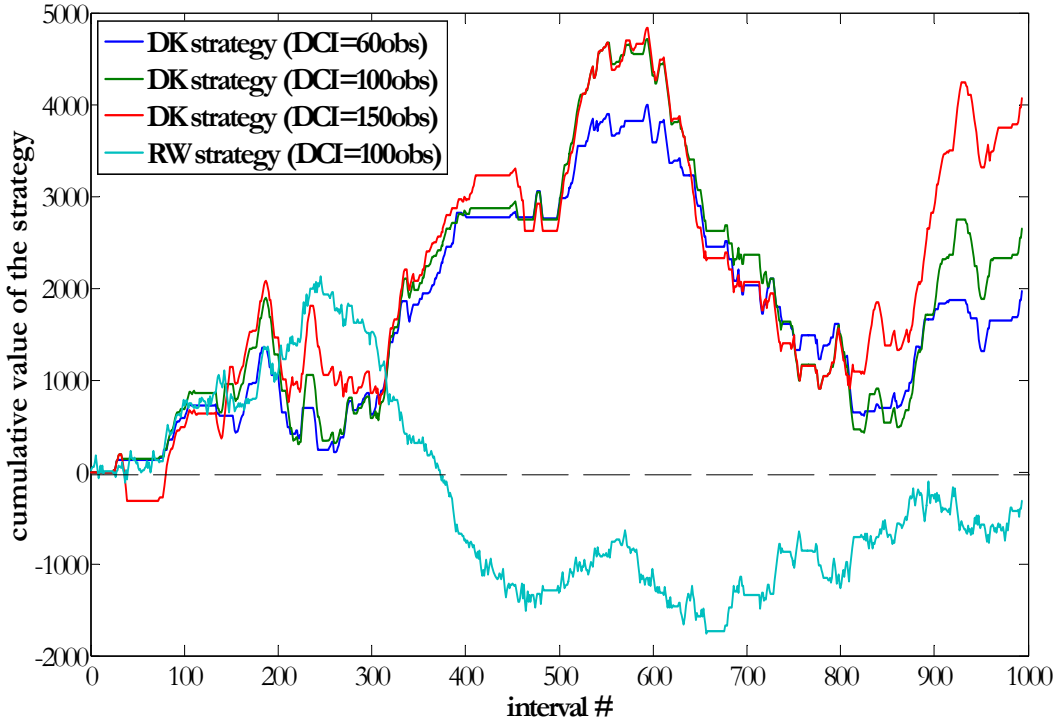


Figure 3.30. DAX (30/7/2003 - 7/12/2006, frequency 30 min, observations 10029-20056). Cumulative value of the DK trading band strategy for the threshold r=60 and different DCI parameter value (DCI=60, 100 and 150 observations): DK parameters - window=60 obs, interval=100 obs

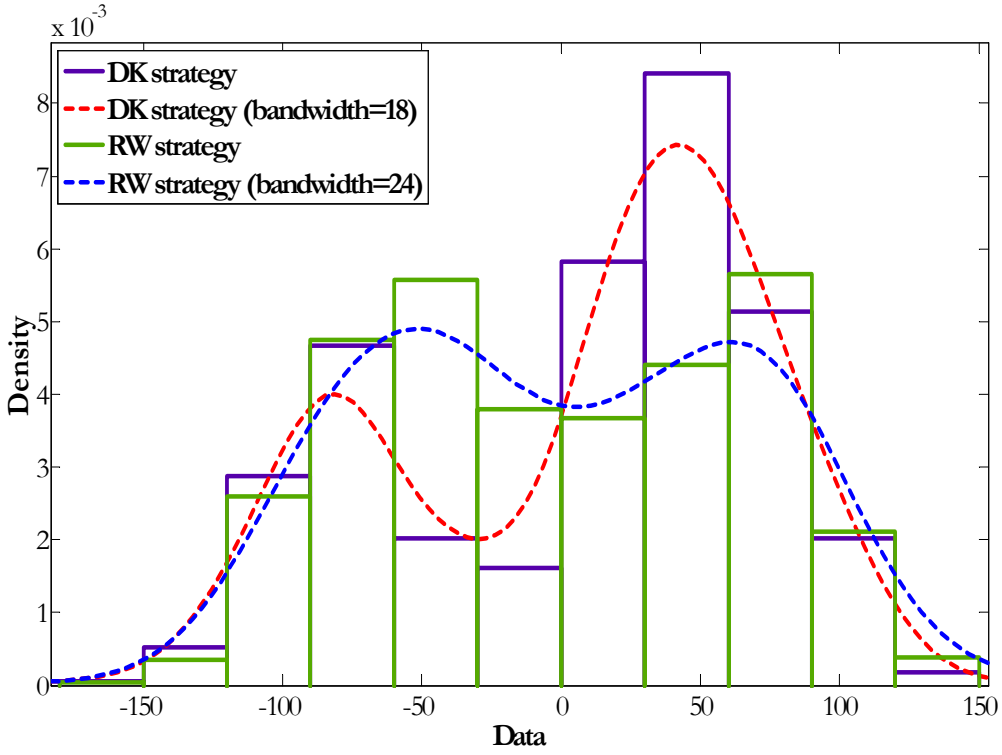


Figure 3.31. DAX (30/7/2003 - 7/12/2006, frequency 30 min, observations 10029-20056). Distribution of the payoffs for taken positions within DK and RW trading band strategies (kernel fit): n(EMA)=115 observations; threshold r=60; DK parameters - window=60 obs, interval=100 obs, DCI length= 150 observations; RW parameters - window=60 obs, interval=100 obs, DCI length= 100 observations

Figure 3.31 presents the distribution of the profits per trade for DK and RW strategies (threshold $r = 60$). As we can see the distribution of the payoffs for RW strategy is almost symmetrical around 0. The DK strategy in its turn allows diminish the number of the trades with medium losses and increase the number of trades with the medium profits. The mean of the profit per trade for the DK strategy is positive and larger than for the RW strategy (7 units versus -0.4 units); the median for the DK strategy is positive (26 units). More statistics for the strategies outcomes are given in Table 3.2, in Ch.6.

Figure 3.32 presents the trading outcomes for the DK strategy for different transform functions, calibrated for different bandwidth parameters. It shows that the choice of the bandwidth parameter does not really matter for the outcome of the trading strategy: bandwidth=15 that corresponds to thicker tails of the distribution produces only slightly higher outcomes.

As the result we see that the DK strategy provides better trading outcomes than the corresponding RW strategy.

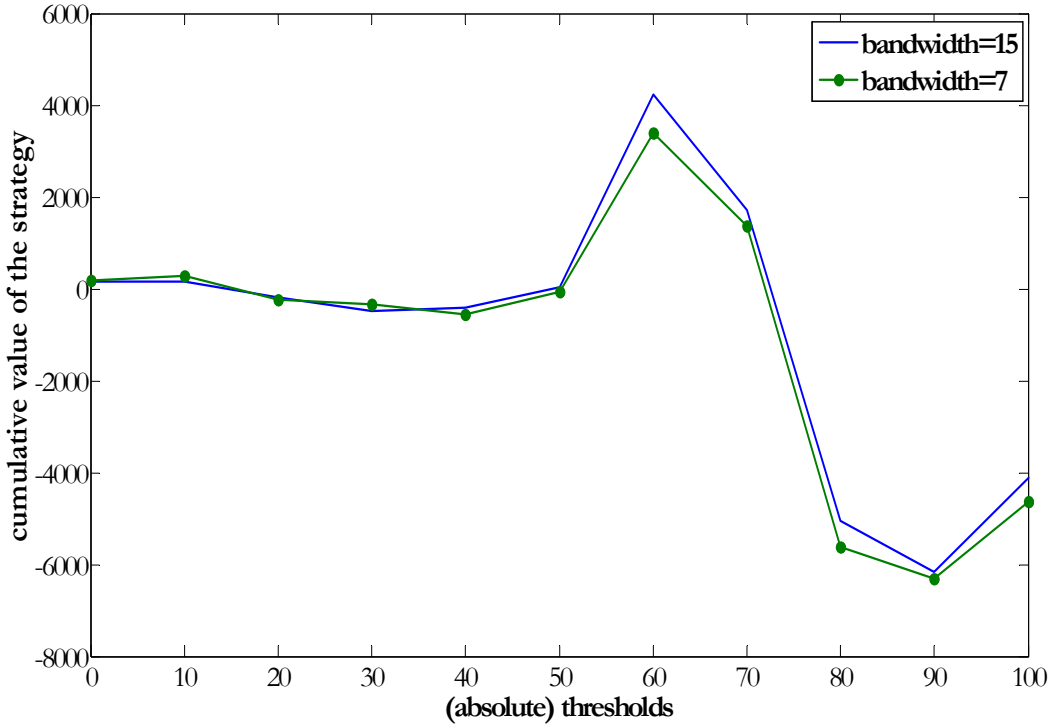


Figure 3.32. DAX (30/7/2003 - 7/12/2006, frequency 30 min, observations 10029-20056). Cumulative value of the DK trading band strategy for different bandwidth parameter: *n*(EMA)=115 observations, DK parameters - window=60 obs, interval=100 obs, DCI length= 150 observations, kernel bandwidth=15 and 7.

5.3.2 Bund

Short description of the Bund instrument, as well as the variogram model, used in the DK procedure is presented in Appendix C.

Figure 3.33 presents the end-of-period value of each strategy for different threshold values. The DK strategy produces positive outcomes for the thresholds $r \in [0.1;0.5)$. For random-walk

strategy its end-of-period outcomes have random nature – they fluctuate around 0[20]. Instead the DK strategy guarantees stable profits for the whole interval of low thresholds values. We can see that the choice of DCI lengths has impact on the final strategy outcomes, particularly in the terms of the global maximum: as in DAX case, the DCI length of 150 observations provides the highest cumulative profits for $r = 0.4$.

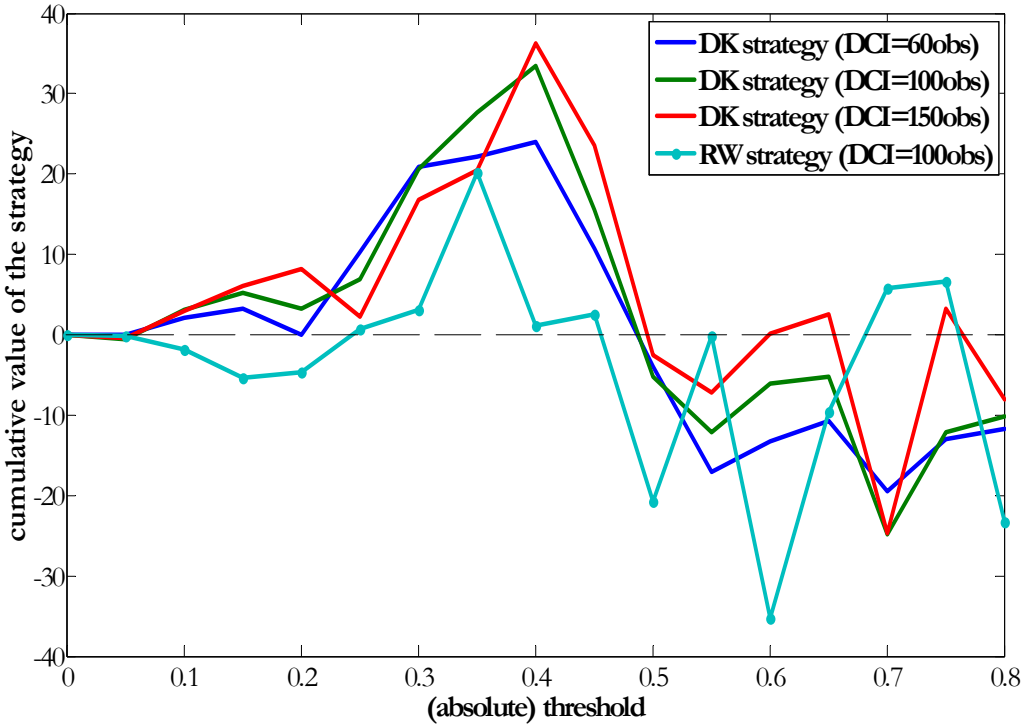


Figure 3.33. Bund (30/7/2003 - 7/12/2006, frequency 30 min, observations 10029-20056). Cumulative value (at the end of period) of the trading band strategy for different thresholds, based on DK and random walk (RW) approaches: *n*(EMA)=115 observations, DK parameters - window length=55 observations, interval length=100 observations, DCI lengths=60, 100 and 150 observations, kernel bandwidth=0.05

Figure 3.34 presents the paths of the DK strategy with the highest cumulative value for threshold $r = 0.4$. As we can see all DK paths are quite monotone and positive after interval #300.

Figure 3.35 presents the distribution of the profits per trade for DK and RW strategies (threshold $r = 0.4$). As we can see the distributions of the payoffs for the RW strategy is almost symmetrical around 0, while the DK strategy allows diminishing the number of trades with medium losses and increasing the number of trades with medium profits. The mean of the profit per trade for the DK strategy is positive and larger than for the RW strategy (0.05 units versus 0 units); the median for the DK strategy is positive (0.16 units). More statistics for the strategies outcomes are given in Table 3.2, in Ch.6.

[20] Figure 3.33 presents only one path for the RW strategy for DCI=60 observations in order to avoid overloading of the graph. Other RW paths for other DCI parameters have the same random outcomes that are lower than the respective global maximums of profit for the DK strategies.

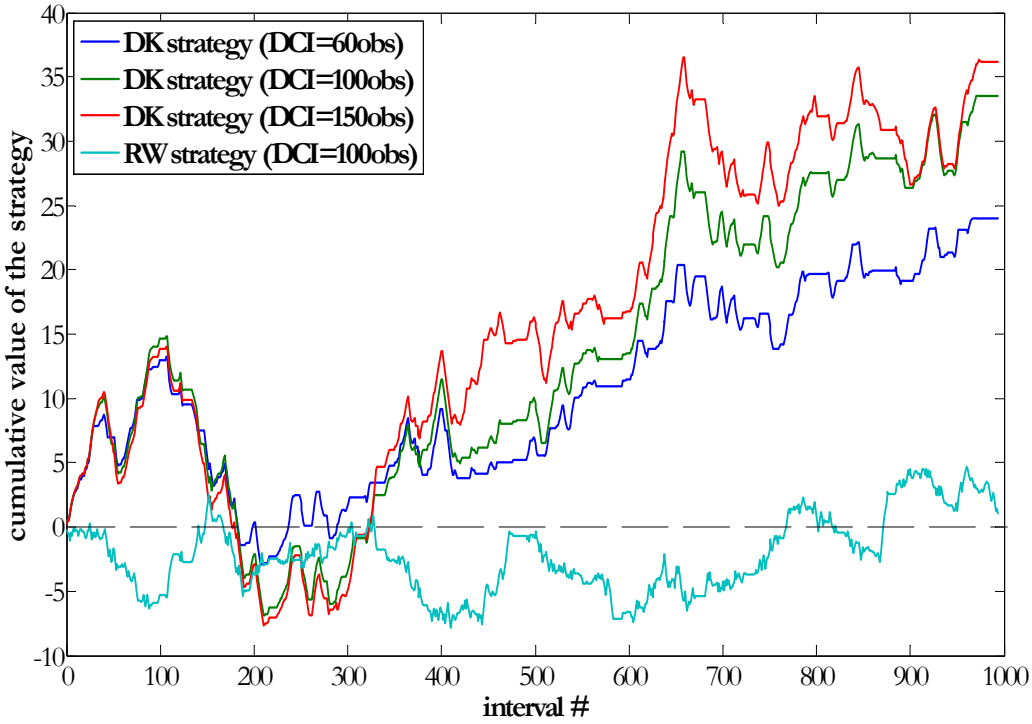


Figure 3.34. Bund (30/7/2003 - 7/12/2006, frequency 30 min, observations 10029-20056). Cumulative value of the DK trading band strategy: *n*(EMA)=115 observations, threshold *r*=0.4, DK parameters - window length=55 observations, interval length=100 observations, DCI length=60, 100 and 150 observations, kernel bandwidth=0.05

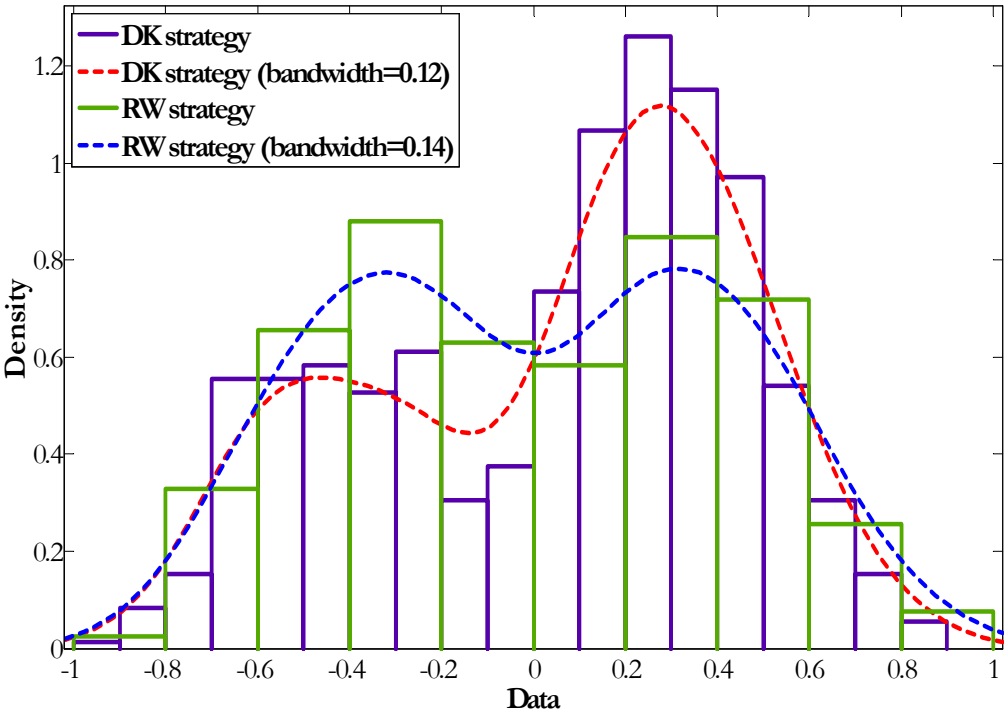


Figure 3.35. Bund (30/7/2003 - 7/12/2006, frequency 30 min, observations 10029-20056). Distribution of the payoffs for taken positions within DK and RW trading band strategies (kernel fit): *n*(EMA)=115 observations; threshold *r*=0.4; DK parameters - window=55 obs, interval=100 obs, DCI length= 150 observations; RW parameters - window=55 obs, interval=100 obs, DCI length= 100 observations.

Figure 3.36 presents the cumulative outcomes for the DK strategies for different bandwidth parameters, used to define the transformation function. The figure supports the conclusion that the value of the bandwidth parameter does not have significant impact on the outcome.

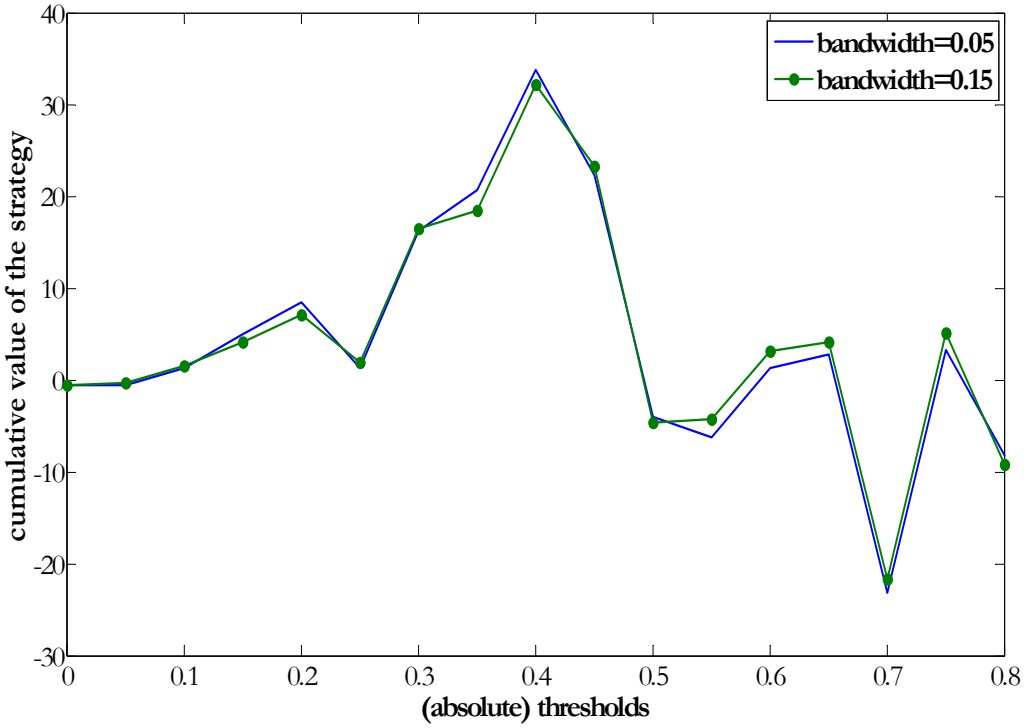


Figure 3.36. Bund (30/7/2003 - 7/12/2006, frequency 30 min, observations 10029-20056). Cumulative value of the DK trading band strategy for different bandwidth parameter: *n*(EMA)=115 observations, DK parameters - window length=55 observations, interval length=100 observations, DCI length= 150 observations, kernel bandwidth=0.05 and 0.15.

The results of the trading strategies support the conclusion that the DK probability can improve the trading results for Bund comparing to the RW strategy.

5.3.3 Brent

Short description of the Brent instrument, as well as variogram model, used in the DK procedure is presented in Appendix D.

Figure 3.37 presents the end-of-period values of each strategy for different threshold values. Contrary to Bund and DAX cases, the positive outcomes of the DK strategy have less stable character: they correspond to several points and one interval of the thresholds. The highest outcomes are reached for the thresholds interval $r \in [2.4;3.5]$. Contrary to the Bund and DAX cases with the relatively low levels of the optimal thresholds, for Brent the highest profits are reached for the thresholds of much higher values. One of the possible explanation is the instrument volatility: Brent shows much higher volatility than Bund and DAX, therefore the bands that are to close to the EMA (small thresholds values) produce more false signals and cannot be optimal (see Table 3.1).

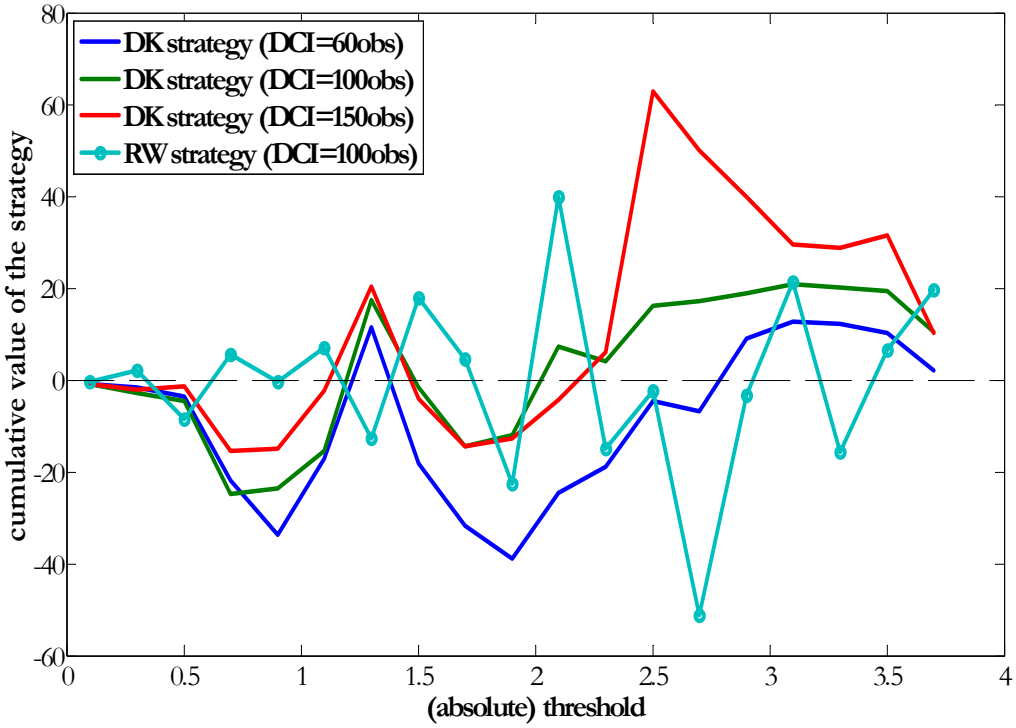


Figure 3.37. Brent (17/12/04-27/01/06, frequency 30 minutes, observations 2459-4919). Cumulative value (at the end of period) of the trading band strategy for different thresholds, based on DK and random walk (RW) approaches: *n*(EMA)=110 observations, DK parameters - window length=50 observations, interval length=100 observations, DCI length=100 observations

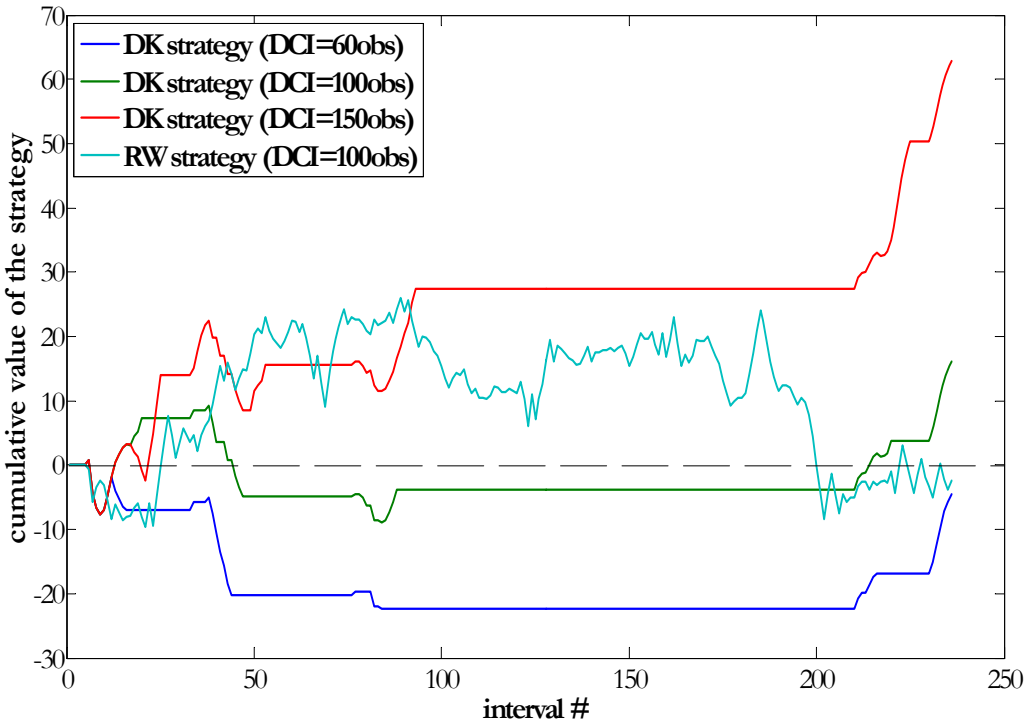


Figure 3.38. Brent (17/12/04-27/01/06, frequency 30 min, observations 2459-4919). Cumulative value of the DK trading band strategy for the threshold $r = 2.5$ that produces the highest outcomes: *n*(EMA)=110 obs, DK parameters - window=50 obs, interval=100 obs, DCI length=100 obs

The paths of the cumulative values for the optimal threshold for $r = 2.5$ that provides the maximum possible outcomes for DK strategy are given in Figure 3.38. We can notice that the optimal path for the DCI length of 150 observations is increasing and almost monotone; for other DCI lengths the outcomes goes well below zero. At the same time the path of the random-walk (RW) strategy that corresponds to one of the DCI values, exhibits random patterns and ends up in negative values.

Figure 3.39 presents the distribution of the profits per trade for the DK and RW strategies (threshold $r = 2.5$). As we can see the payoffs for RW strategy are close to symmetrical around 0, while the DK strategy exhibit more positive than negative profits. The mean of the profit per trade for the DK strategy is positive and larger than for the RW strategy (0.89 units versus -0.01 units); the median for the DK strategy is positive (1.25 units). More statistics for the strategies outcomes are given in Table 3.2, in Ch.6.

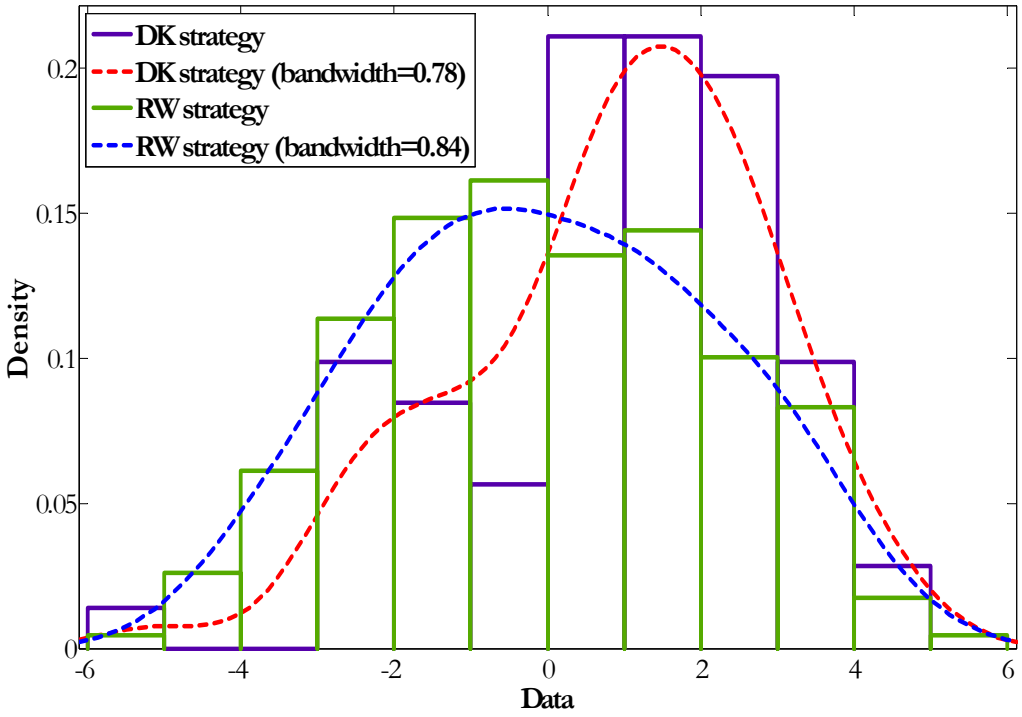


Figure 3.39. Brent (17/12/04-27/01/06, frequency 30 min, observations 2459-4919). Distribution of the payoffs for taken positions within DK and RW trading band strategies (kernel fit): *n*(EMA)=110 observations; threshold $r = 2.5$; DK parameters - window=50 obs, interval=100 obs, DCI length= 150 observations, RW parameters - window=50 obs, interval=100 obs, DCI length= 100 observations

The results of the trading strategies support the conclusion that the DK probability can improve the trading results for Brent comparing to the RW strategy.

5.3.4 X instrument

Short description of the X instrument, as well as the variogram model, used in the DK procedure are presented in Appendix E.

Contrary to previous instruments, the DK strategy does not give good results (see Figure 3.40). The random walk strategy as usual produce some random outcomes, without some consistency.

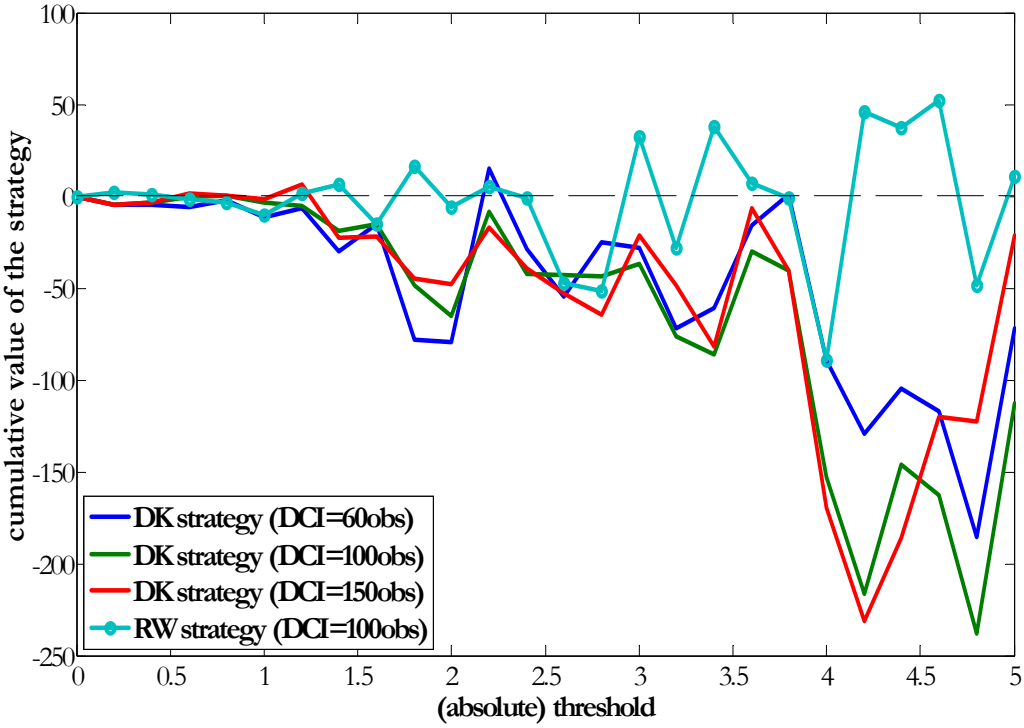


Figure 3.40. X instrument (frequency 1 hour, observations 2347-4693). Cumulative value (at the end of period) of the trading band strategy for different thresholds, based on DK and random walk (RW) approaches: *n*(EMA)=50 observations, DK parameters - window length=40 observations, interval length=100 observations, DCI length=100 observations

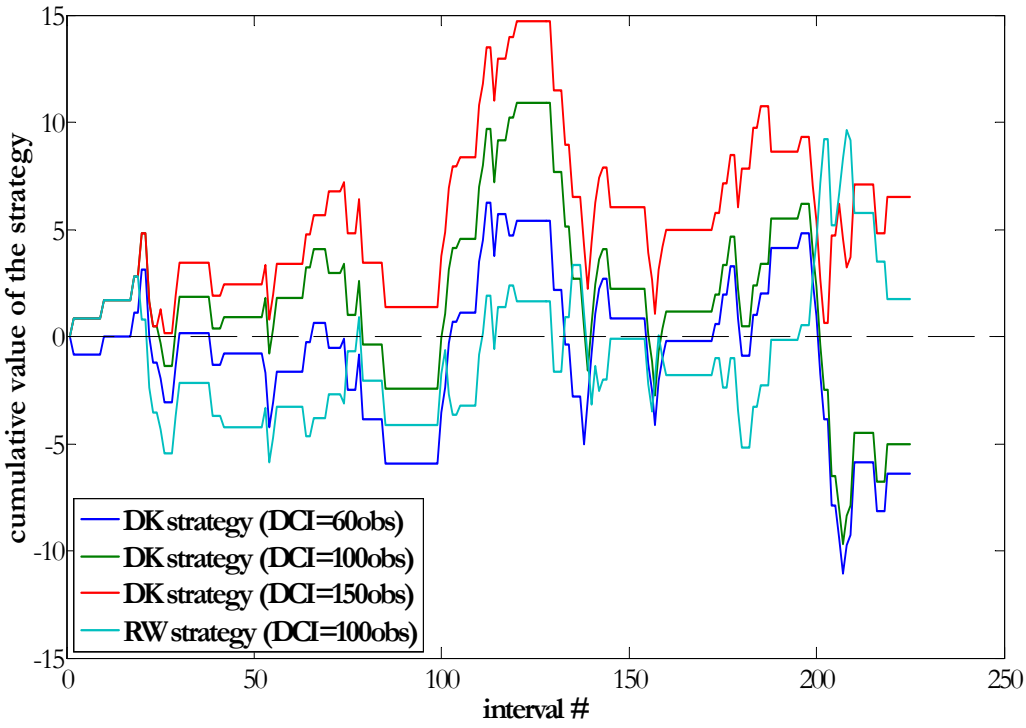


Figure 3.41. X instrument (frequency 1 hour, observations 2347-4693). Cumulative value (at the end of period) of the DK trading band strategy: *n*(EMA)=50 obs, thresholds r=0.8, DK parameters - window =40 obs, interval =100 obs

The highest profits are generated for the RW strategy, while DK strategy does not produce positive results at any thresholds levels. Change in the DCI length allows obtaining some low positive outcomes for DK strategy very short thresholds intervals. However, they seem to exhibit rather random nature; therefore, we do not present these results.

Figure 3.41 presents some of the paths of the cumulative values of the DK strategy; they exhibit a random patterns and ends up in negative values.

We can see that for the X instrument the DK method does not improve trading results.

6 Conclusions

This part discusses the disjunctive kriging (DK), an approach that is used to estimate the probability of breaching some threshold level. If predicted correctly, these estimates can improve the trading strategies, based on the trading bands. However, the application of the DK method to the financial data needs some adjustment due to the data peculiarities. One of these peculiarities is data non-stationarity; as the result the transform function, used in the DK procedure, should be recalibrated locally. We proposed the method of adjusting transformation function for local non-stationary volatility under assumption that the data follows the same distribution law. One of the directions for future research in the field can be the adjustment of the CDF not only for the volatility, but also for the change in local mean or skewness.

Two different probabilities types were considered in the part III: point and interval. The point DK probability $P_{DK}(t) = \Pr(Y(t+h) < y_c)$ shows convergence versus the unconditional normal CDF $F(y_c)$. Distance (h) at which this convergence takes place coincides with the range parameter of the variogram of the normal variable Y. The values of the DK probabilities at distances smaller than h depend on the path of the Y process before t. However, the point probabilities cannot be back-tested. Besides for the trading application it is often more important to know that the threshold will be breached in (near) future, then that the threshold will be breached at the precise moment of time t.

The interval DK probability $P_{DK}(t) = \Pr(Y_{[t+1;t+l]} < y_c)$ - the probability of breaching threshold during some interval of time in future - were analysed in depth, as they can be back-tested by the respective empirical frequencies calculated on the interval l. We have found that the values of the interval DK probabilities depend on the length of the distribution calibration interval (DCI), used for the adjustment of the transform function, and the length of the interval, for which the DK probabilities are calculated. The best probability estimators correspond to the medium DCI intervals and short and medium interval lengths. The interval probabilities do not depend on the window length (even for the instruments with the non-exponential variogram model) or bandwidth parameter, used in the kernel procedure for the estimation of the basic CDF function. Comparing to the empirical frequencies, the probability estimators exhibit almost no lag, while some over- or under-estimation in absolute values takes place.

The prediction power of the DK method was evaluated indirectly through the outcomes of the trading strategies, based on them. We have constructed two strategies: (1) DK strategy, where decision about entry position was made on the basis of DK probabilities; and (2) Random-walk strategy, where decision about entry position was made randomly. Our analysis shows that the DK strategy produces positive outcomes for the continuous interval of thresholds with almost monotone and increasing paths of the strategy value. For some instruments DK strategy

produces the highest possible outcomes (DAX, Bund, Brent). The only instrument for which DK does not produce any valid results is the X instrument. The RW strategy produces for all instruments the profits that have some random nature. The distributions of the profits per trade for the optimal DK strategy shows consistently more profitable and less non-profitable trades comparing with the results of the RW strategy. Main statistical characteristics for these distributions are summarized in Table 3.2.

Table 3.2
Statistical characteristics of profits per trade for the DK optimal strategy and respective RW strategy

	DAX		**Bund**		**Brent**	
	DK strategy	**RW strategy**	**DK strategy**	**RW strategy**	**DK strategy**	**RW strategy**
Mean	7.03	-0.40	0.05	0.00	0.88	-0.01
Median	26.16	-3.34	0.16	-0.03	1.25	-0.07
Standard Deviation	65.87	66.53	0.41	0.41	1.97	2.20

Table 3.3 summarizes the outcomes of DK and RW strategies for all instruments. We can see that the application of the DK method to strategies construction can improve trading results.

Table 3.3
Results of the application of the DK and RW strategies to different instruments

Instrument	Variogram ranges/ frequency	Optimal strategy parameters	optimal DK strategy					optimal RW strategy				
			Threshold value	Cumulative value*	Mean value*	Monotonicity	Part of the interval?	Threshold value	Cumulative value*	Mean value*	Monotonicity	Part of the interval?
Bund	a=55 freq=30 min	n(EMA)=115 *DK parameters:* window=55 interval=100 DCI=150	r=0.4	36.20	0.05	Yes	**Yes** [0.1; 0.5]	r=0.35	20.18	0.03	N/A	No
DAX	a(1)=50, a(2)=450 freq=30 min	n(EMA)=115 *DK parameters:* window=60 interval=100 DCI =150	r=60	4064.6	7.03	Yes	**Yes** (50; 80]	r=100	1300	1.40	N/A	No
Brent	a=50 freq=30 min	n(EMA)=110 *DK parameters:* window=50 interval=100 DCI =150	r=2.5	62.77	0.88	Yes**	**Yes** [2.4; 3.3]	r=2.1	39.82	0.18	N/A	No
X instrument	a(1)=40, a(2)=200 freq=30 min	n(EMA)=50 *DK parameters:* window=40 interval=100 DCI =150	r=1.2	6.53	0.09	N/A	N/A	r=4.6	51.87	0.25	N/A	N/A

* - in quotes units
** - after #300 observation

Appendices III

Appendix A

Estimation of the point probabilities for DAX

The DK analysis of the point probabilities (2.1) is performed according to the following algorithm:

1. The whole available sample S_i is used to estimate empirical CDF. The *point* DK probabilities are estimated for the same sample S_i. This approach creates the dilemma, when the obtained results are validated on the same sample, on which the parameters, used in the estimation procedure, are estimated. However, this allows us to analyse the results under the assumption that the obtained variogram and CDF estimates are the "true" estimates of the covariance and probability distribution.
2. The rolling windows of *n*-length are used as sub-samples for DK estimation procedure: $\{Y_i\}_{t-n+1\le i\le t}$, $t>0$, where Y is the normal standard variable from the analysed data sample S_Y: $Y_i \in S_Y$.
3. The DK probability $P_{DK}^{p}(t,h,y_c)=\Pr(Y(t+h)<y_c)$ is estimated for particular thresholds value $y_c = Y(t)$, i.e. threshold is equal to the last available observation in the window. This choice of the threshold value facilitates common sense validations of the obtained probabilities.

Appendix B

Estimation of the interval probabilities for Bund: global transformation function

The following algorithm lies behind the estimation of the interval probability with sole transformation function:

1. The whole available sample S_i is used to estimate the empirical CDF. The *interval* DK probabilities are estimated for the same sample S_i. This approach creates the dilemma, when the obtained results are validated on the same sample, on which the parameters, used in the estimation procedure, are estimated. However, this allows us to analyse the results under the assumption that the obtained variogram and CDF estimates are the "true" estimates of the covariance and probability distribution.
2. The rolling windows of *n*-length are used as sub-samples for DK estimation procedure: $\{Y_i\}_{t-n+1\le i\le t}$, $t>0$, where Y is the normal standard variable from the analysed data sample S_Y: $Y_i \in S_Y$.
3. The DK probability $P_{DK}^{p}(t,h,y_c)=\Pr(Y(t+h)<y_c)$ is estimated for the vector of thresholds.

The variogram of the transformed normal variable Y, used in the DK procedure is the same as in Figure C3.

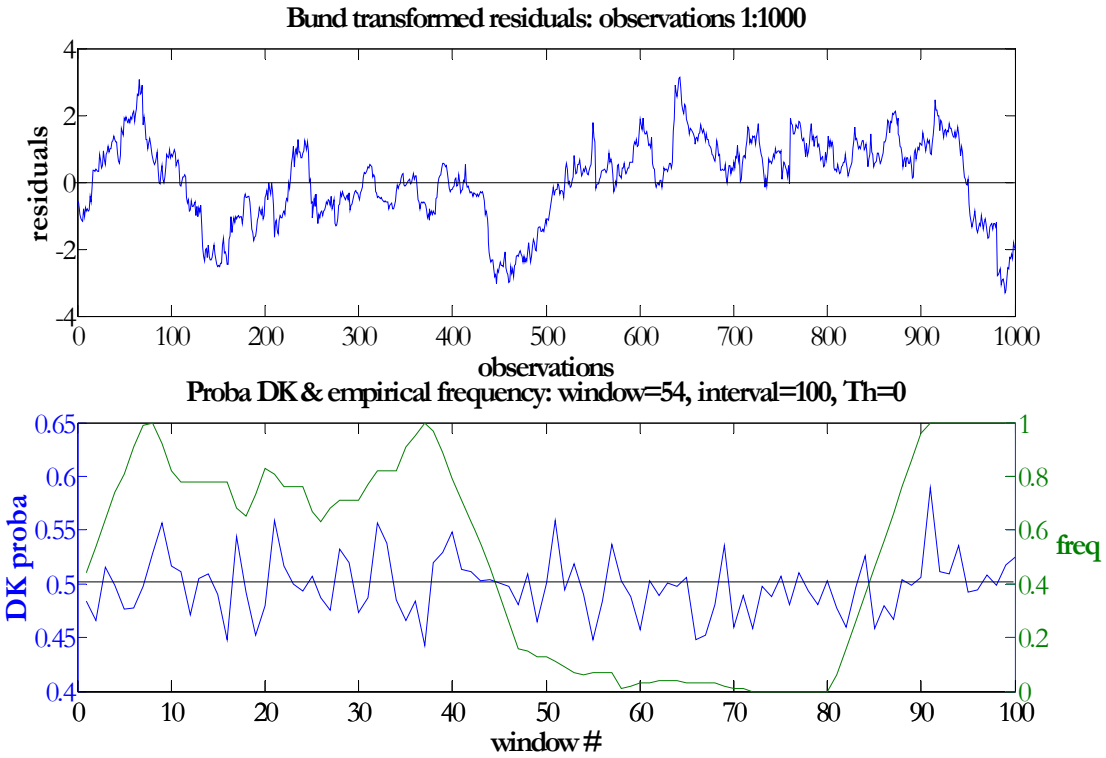


Figure B1. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations, observations 1:1000), DK probabilities and empirical frequencies (threshold value $y_c=0$, window length=54 observations, interval length=100 observations)

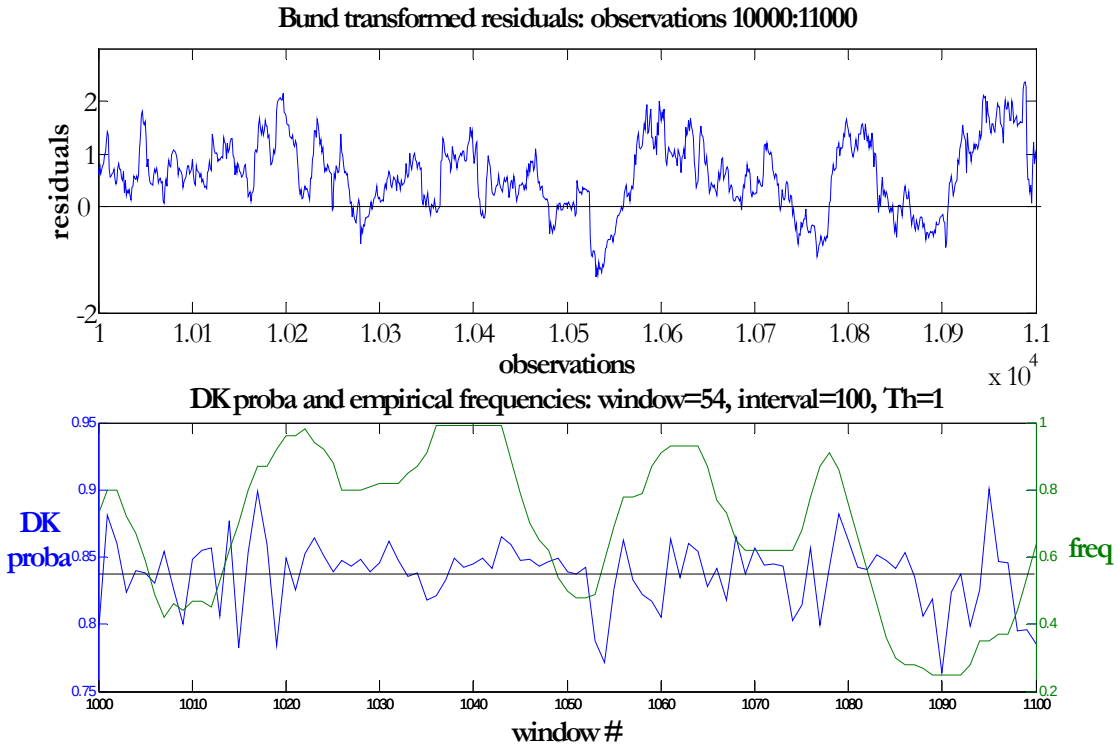


Figure B2. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations, observations 10000-11000) and estimated DK probabilities and empirical frequencies (threshold value $y_c = 1$, window length=54 observations, interval length=100 observations)

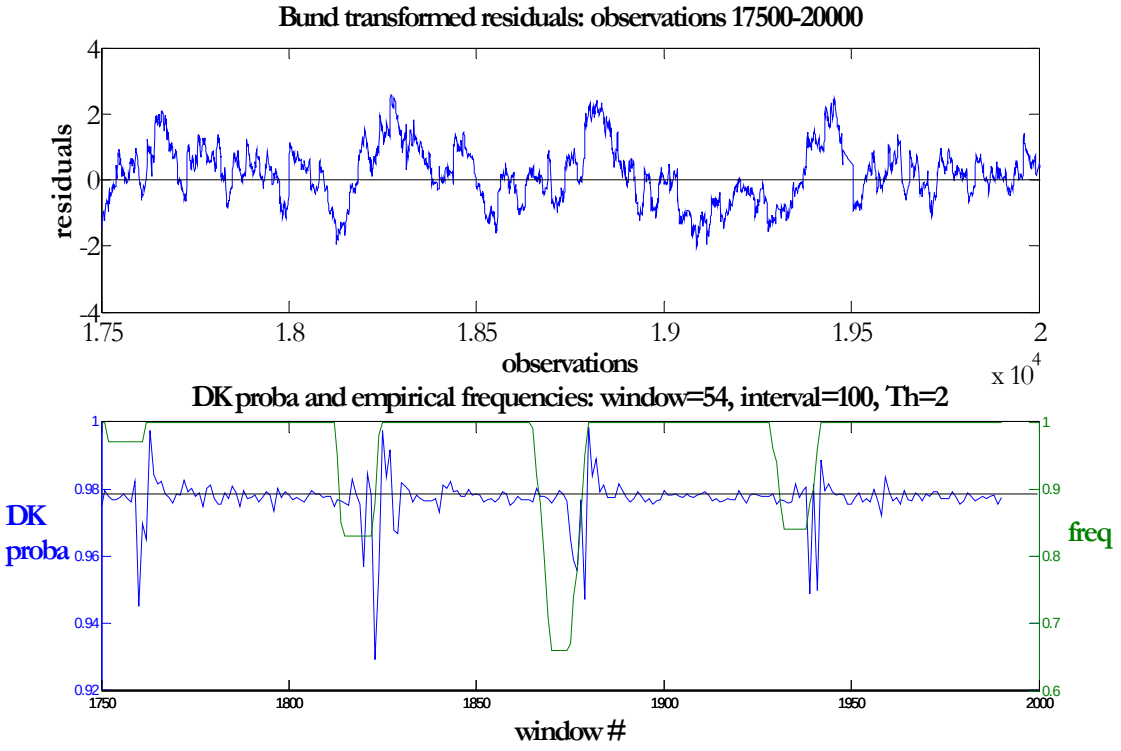


Figure B3. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations, observations 17500-20000) and estimated DK probabilities and empirical frequencies (threshold value $y_c = 2$, window length=54 observations, interval length=100 observations)

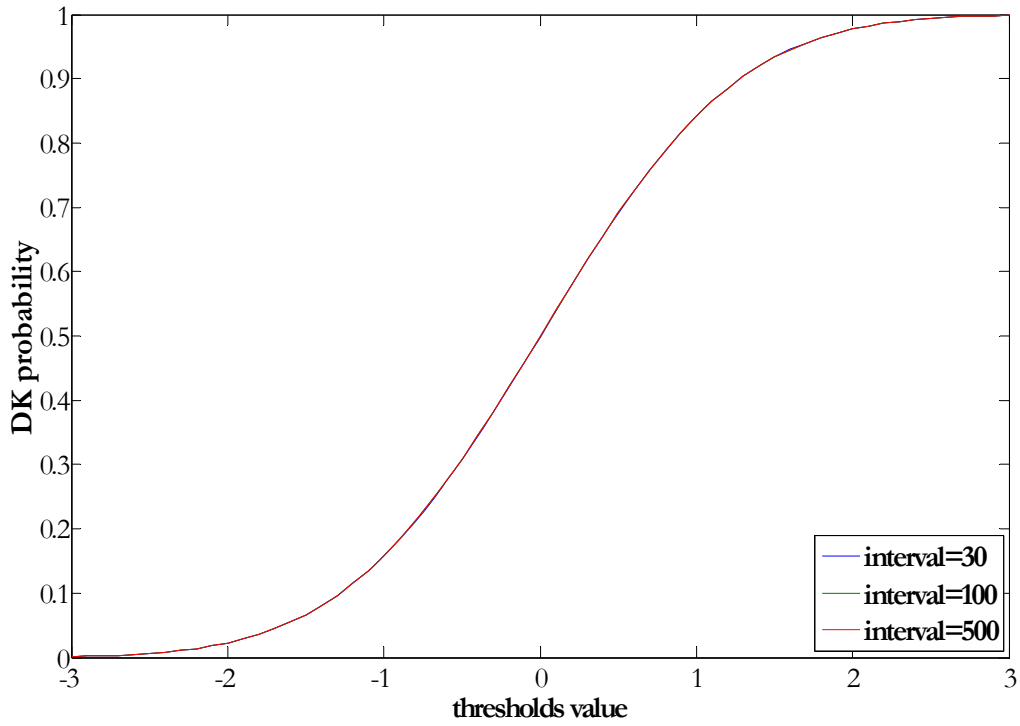


Figure B4. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): Mean of the DK probabilities $P_{DK}\left(Y_{[t+1;t+l]} < y_c\right)$ for different thresholds (y_c) and interval lengths: window length=54 observations, interval length=30, 100 and 500 observations.

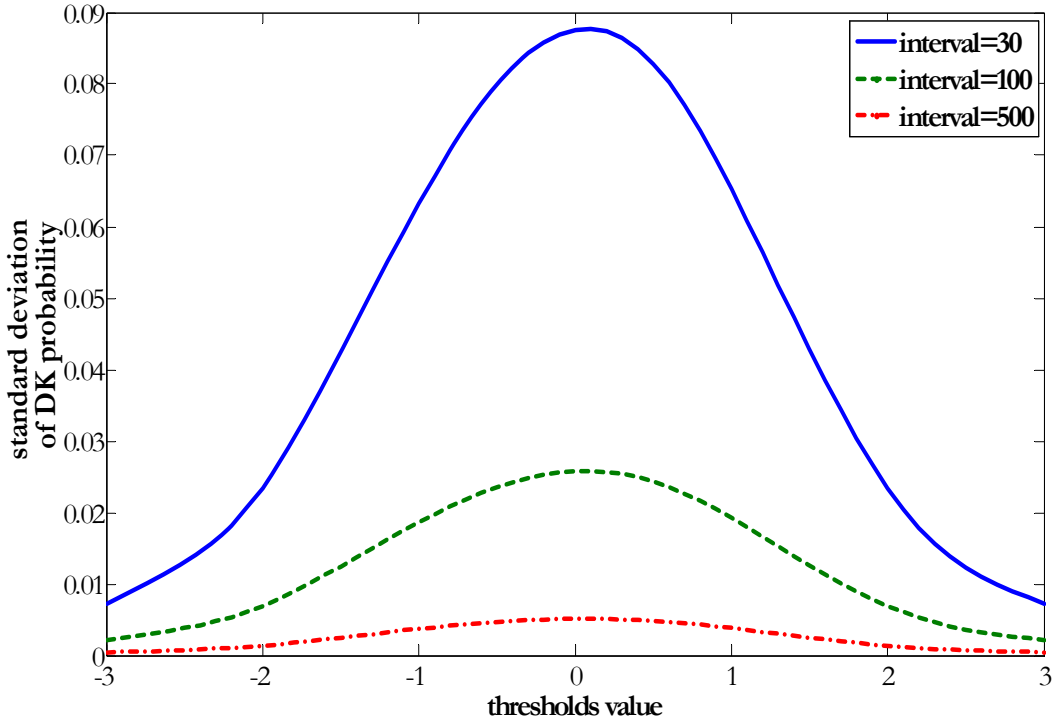
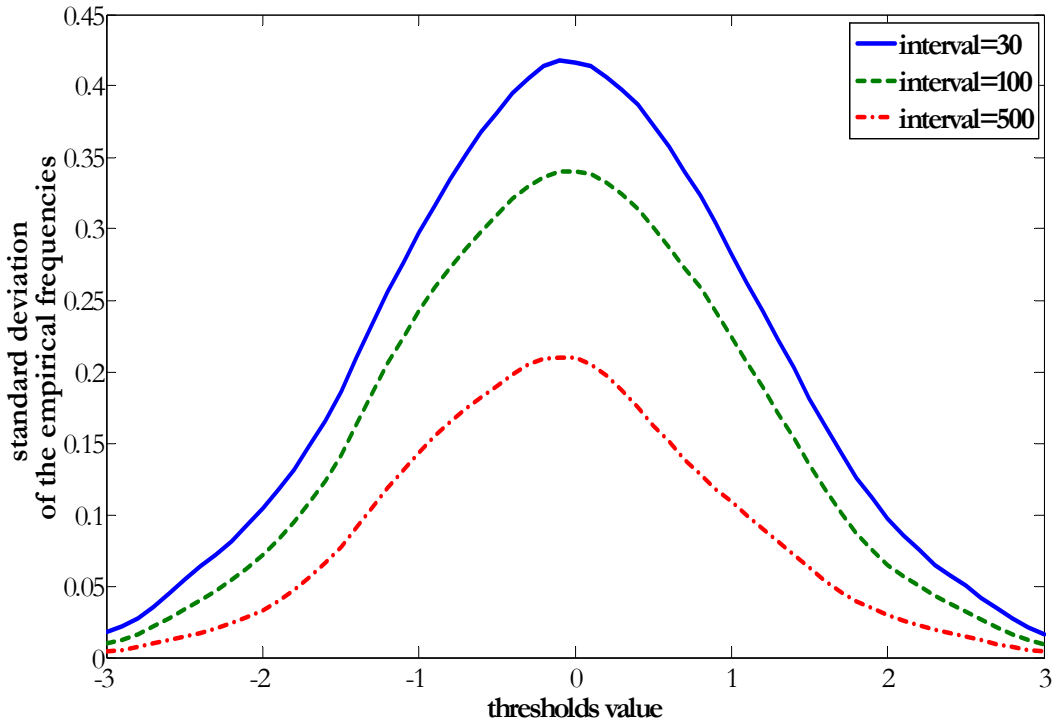


Figure B5. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): Standard deviation of the DK probabilities $P_{DK}\left(Y_{[t+1;t+l]} < y_c\right)$ for different thresholds (y_c) and interval lengths: window length=54 observations, interval length=30, 100 and 500 observations.



Picture B6. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): Standard deviation of the empirical frequencies for different thresholds (y_c) and interval lengths: interval length=30, 100 and 500 observations.

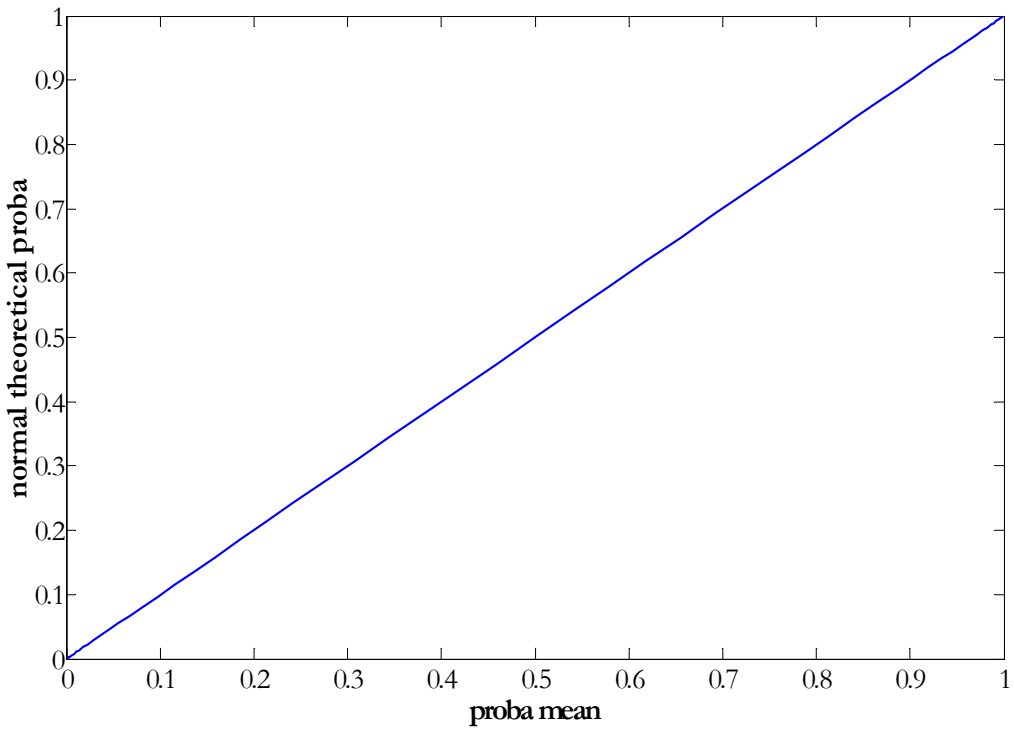


Figure B7. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): Mean of the DK probability versus normal theoretical probability for the same thresholds: window length=54 observations, interval length=100 observations.

Appendix C

Short description of the Bund instrument

Bund instrument (futures on Bund) represent the bond market. The data covers the period 30/7/2003 - 7/12/2006 with frequency 30 min (see Figure C1). We have eliminated the EMA of the length 115 observations to obtain residuals in Figure C2.

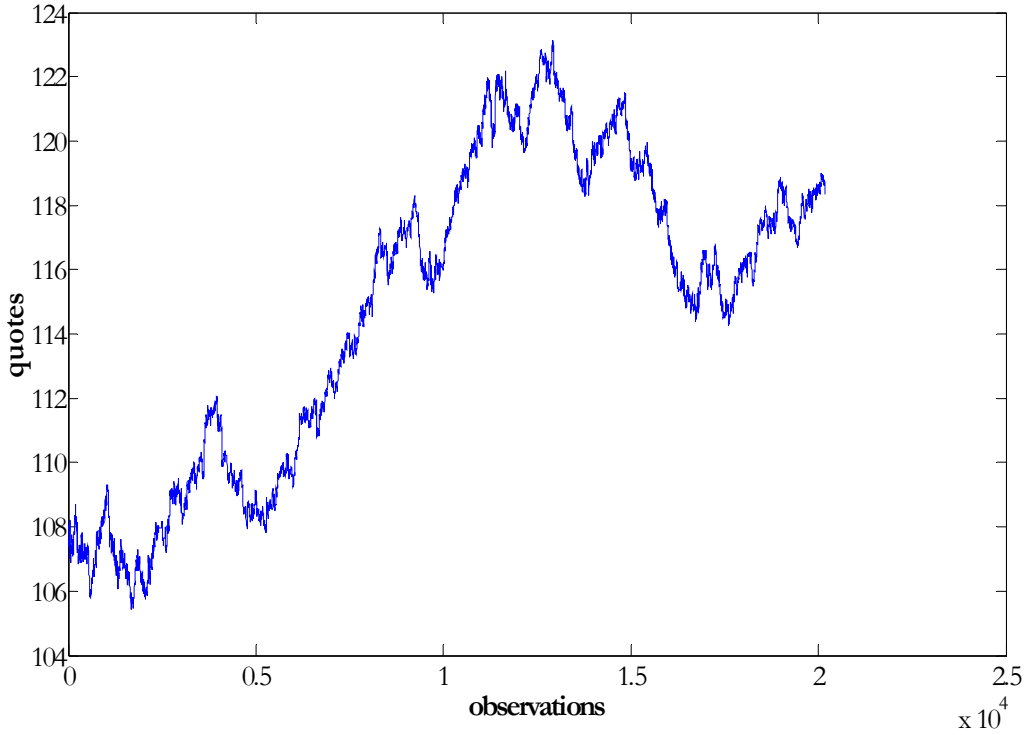


Figure C1. Bund quotes (30/7/2003 - 7/12/2006, frequency 30 min)

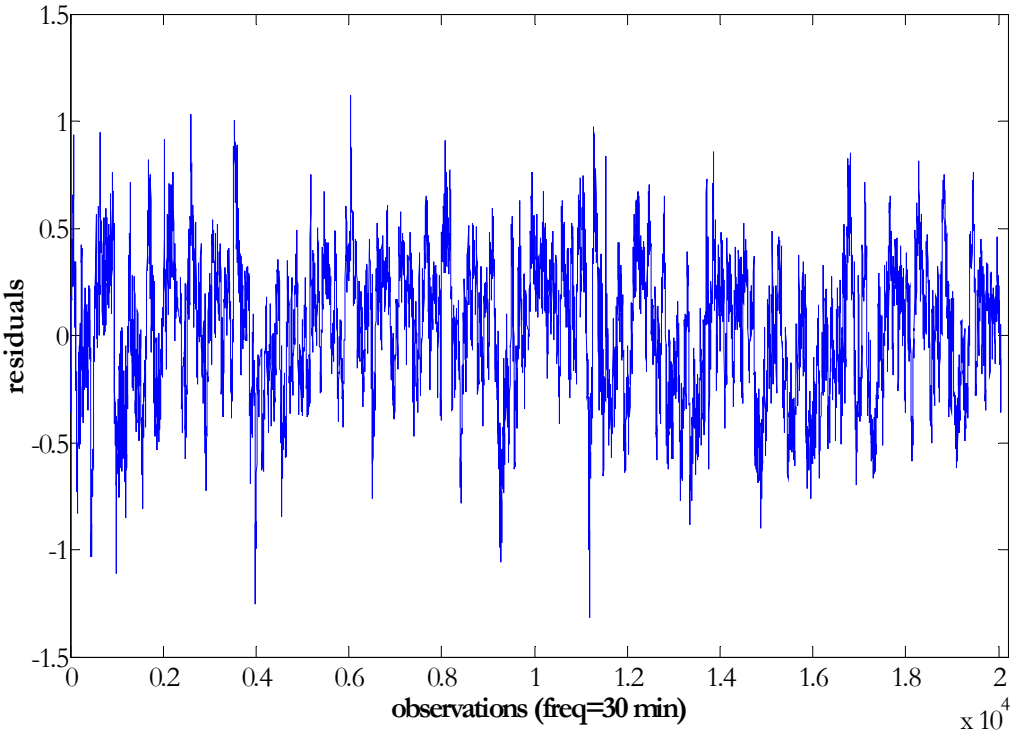


Figure C2. Bund (30/7/2003 - 7/12/2006, frequency 30 min): Residuals after extracting the EMA of the length *n*(EMA)=115 observations

As we can see from figure C2, the residuals do not exhibit any trend. At the same time data is not normal (see Figure C3). Thus, residuals transformation into normal variable is needed before application of the DK method.

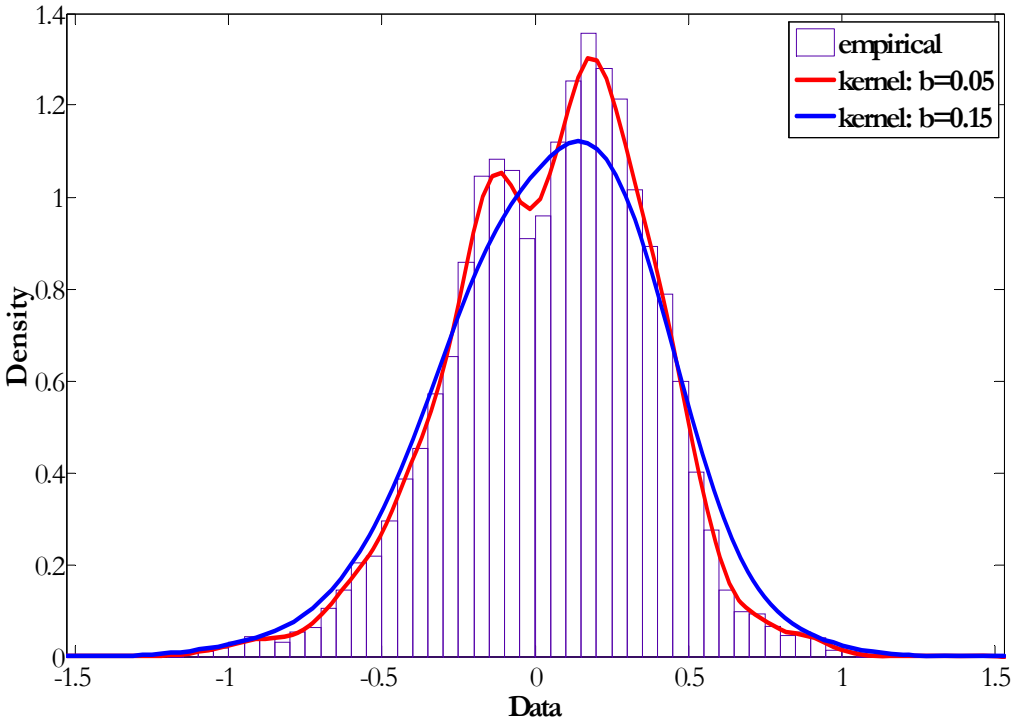


Figure C3. Bund (30/7/2003 - 7/12/2006, frequency 30 min, observations 1:10028):empirical histogram and kernel fits for different bandwidth parameters value

The variogram of the transformed residuals (see Figure C4) is evaluated on the sub-sample $[1;10028]$:

$$\gamma(h) = 1\left(1 - e^{-\left|\frac{h}{55}\right|}\right).$$

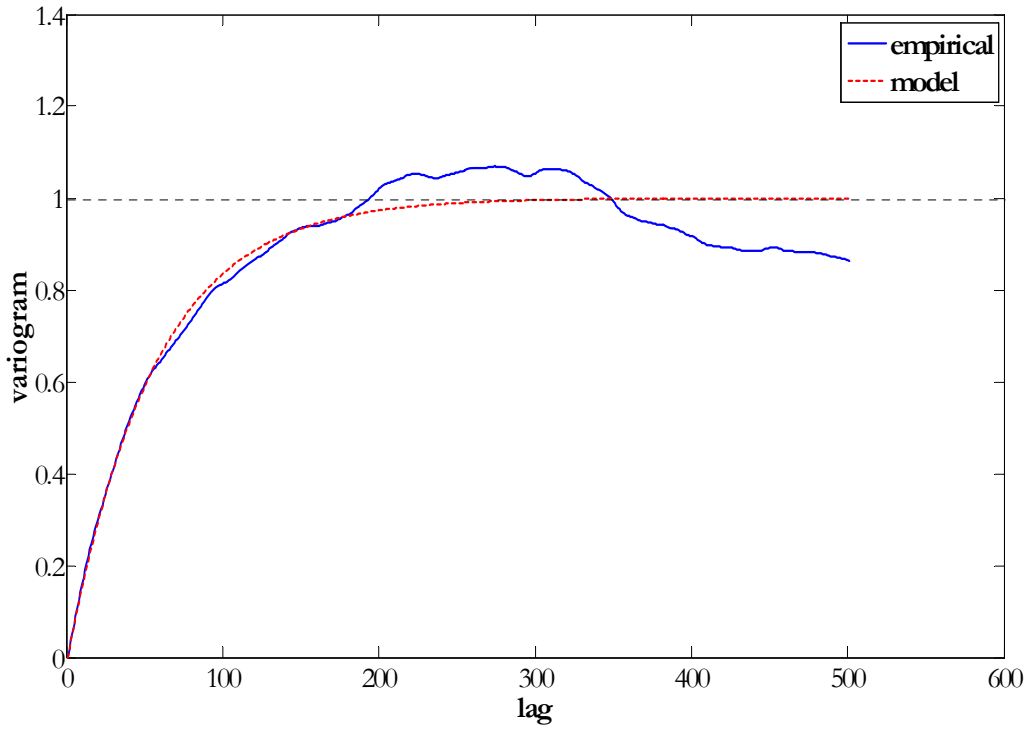


Figure C4. Transformed Bund residuals (30/7/2003 - 7/12/2006, frequency 30 min, *n*(EMA)=115 observations): Empirical (solid line) and estimated exponential model (dot line): $\gamma(h)=1\left(1-e^{-\left|\frac{h}{55}\right|}\right)$.

Appendix D

Short description of the Brent instrument

Brent is a futures on crude oil. Therefore, we consider this instrument as a representative of commodity markets. The peculiarity of this instrument is that it is highly volatile. Our data represents the period of 17/12/04-27/01/06 with 30 min frequency (see Figure D1).

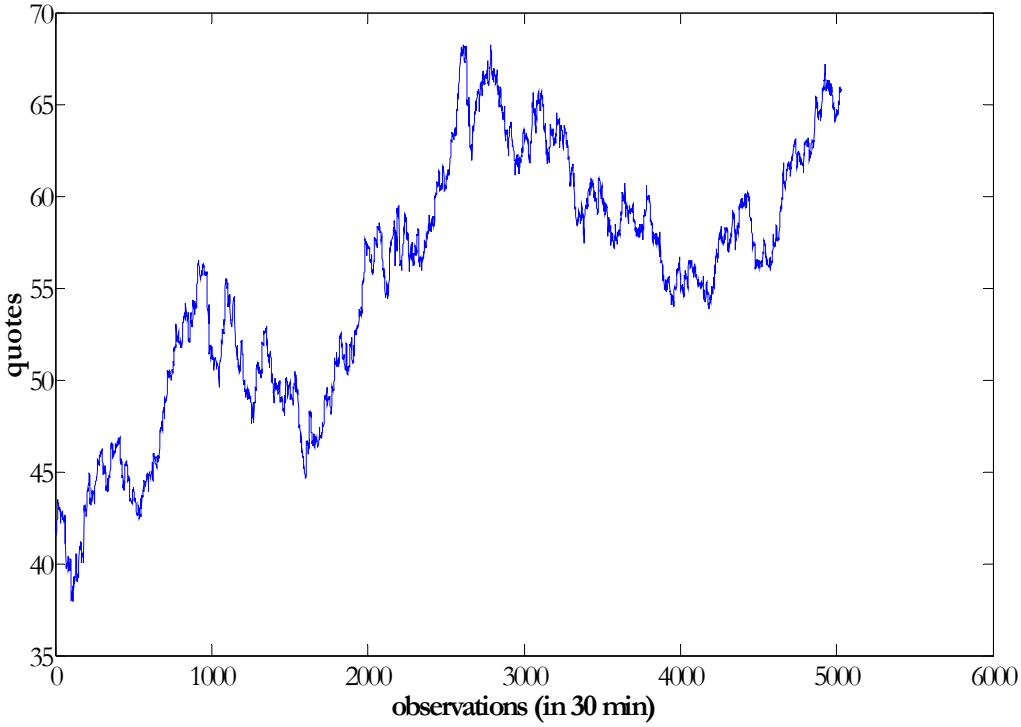


Figure D1. Brent quotes for period 17/12/04-27/01/06 (frequency 30 minutes).

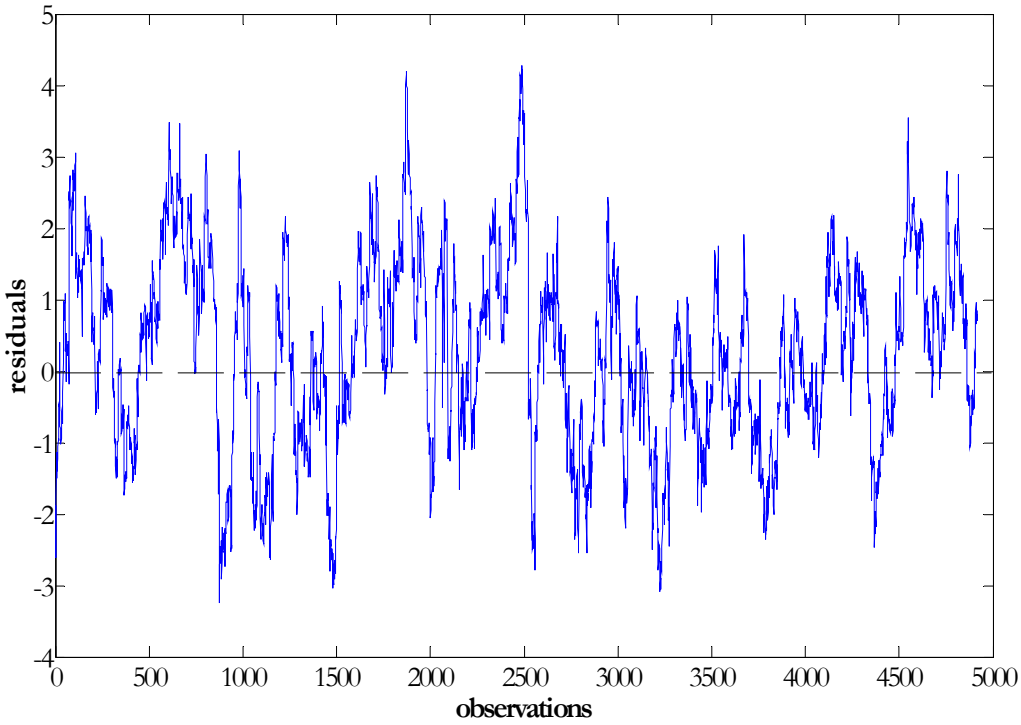


Figure D2. Brent residuals (17/12/04-27/01/06, frequency 30 minutes, n(EMA)=110 observations)

The residuals in Figure D2 are obtained by eliminating the EMA of the length 110 observations. Residuals data is not normal (see Figure D3). Thus, residuals transformation into normal variable is needed before application of the DK method.

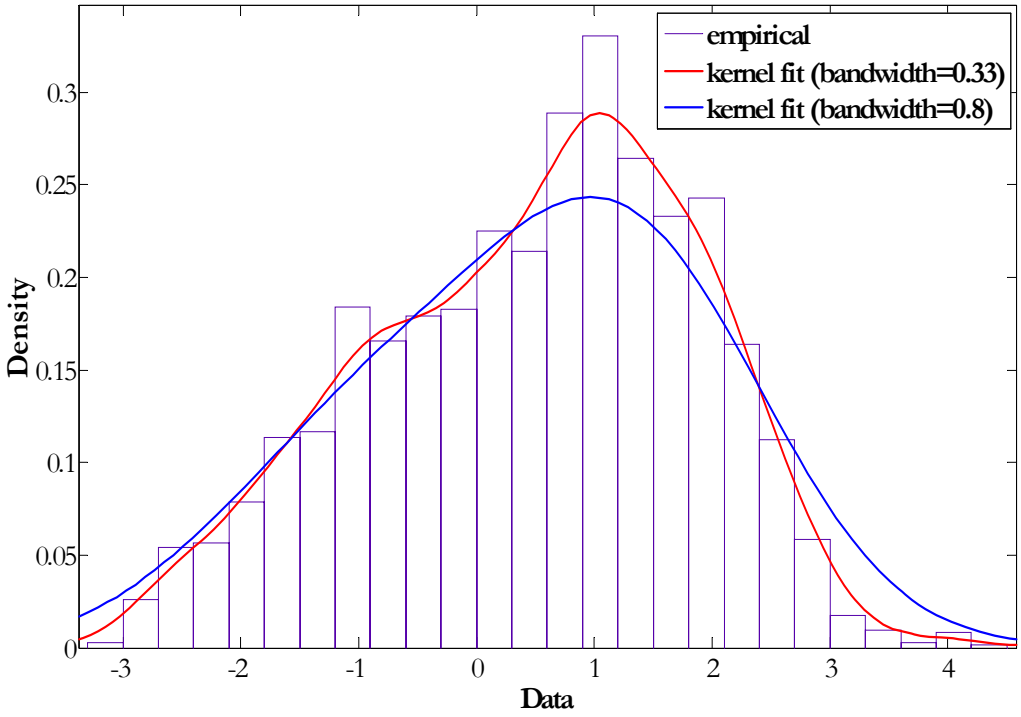


Figure D3. Brent residuals (17/12/04-27/01/06, frequency 30 minutes, n(EMA)=110 observations): Histogram and kernel fits for different bandwidth values

The variogram of the transformed residuals (see Figure C4) is evaluated on the sub-sample $[1;2459]$:

$$\gamma(h) = 1\left(1 - e^{-\left|\frac{h}{50}\right|}\right).$$

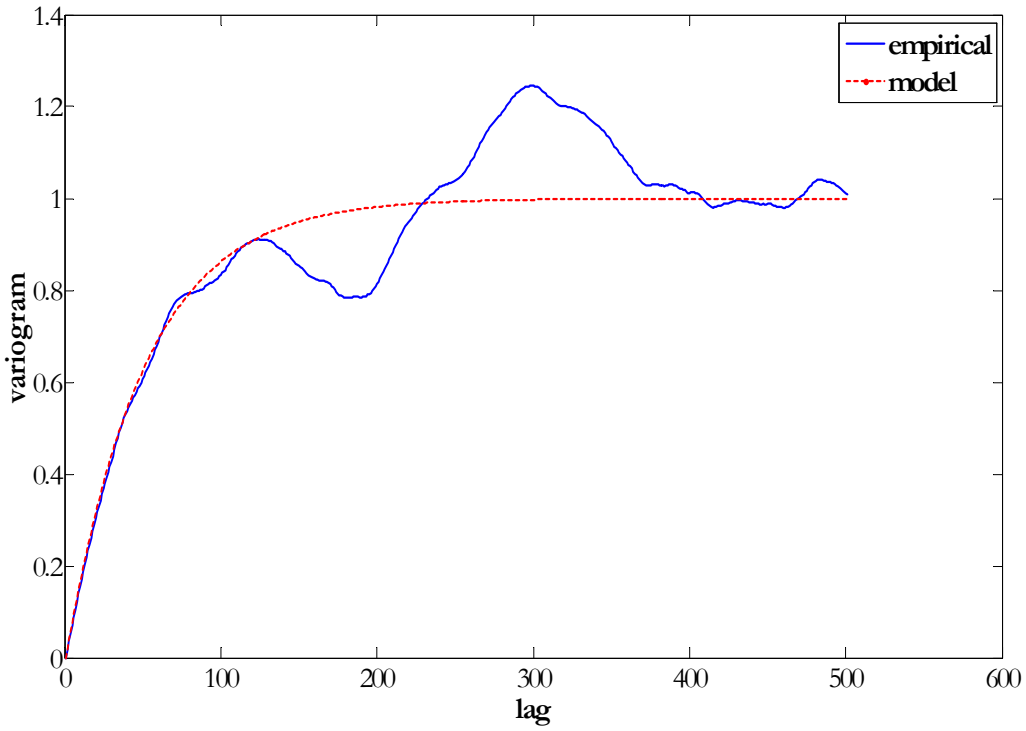


Figure D4. Brent residuals (17/12/04-27/01/06, frequency 30 minutes, n(EMA)=110 observations). Variogram of the transformed residuals for sub-sample [1;2459], as well as theoretical model fit to the data $\gamma(h)=1\left(1-e^{-\left|\frac{h}{50}\right|}\right)$

Appendix E

Short description of the X instrument

Instrument X represents an artificially created index, used by one bank for strategy constructions. Due to the confidentiality reason we cannot neither present its detail description, nor provide the information on its real quotes. That is why on Figure E1 that presents the quotes path during some period of time, there are no ticks on the Y-coordinate. The only information we can provide is that data frequency is 1 hour.

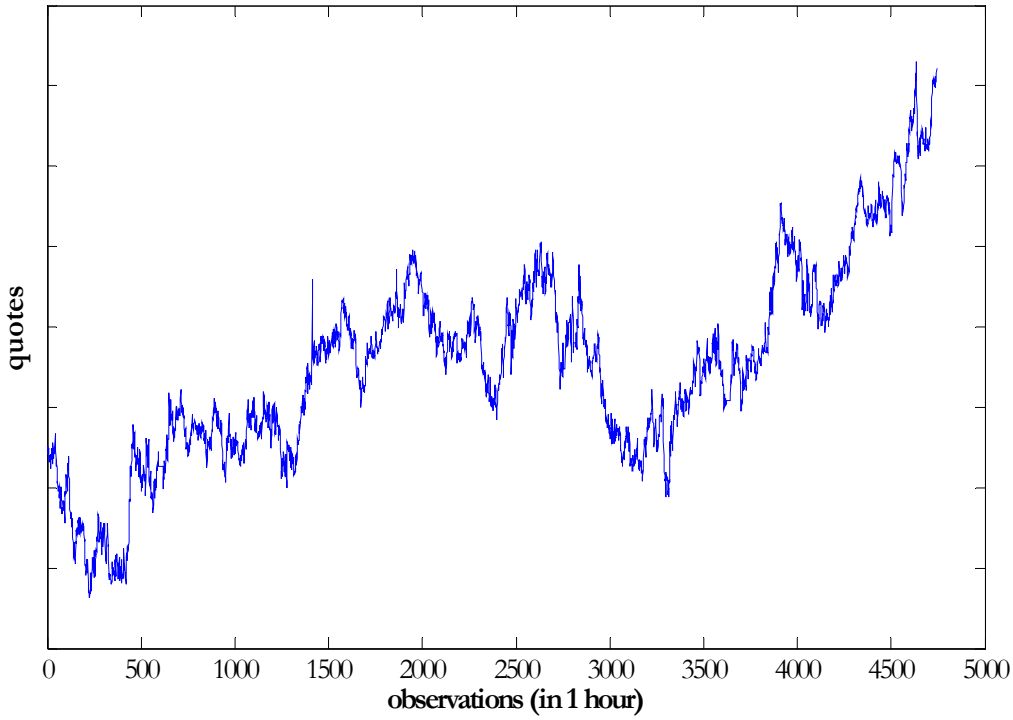


Figure E1. X instrument quotes (frequency – 1 hour)

We eliminate the EMA of the length of 50 observations to obtain residuals in figure E2. They are not normal (see Figure E3). Thus, residuals transformation into normal variable is needed before the application of the DK method.

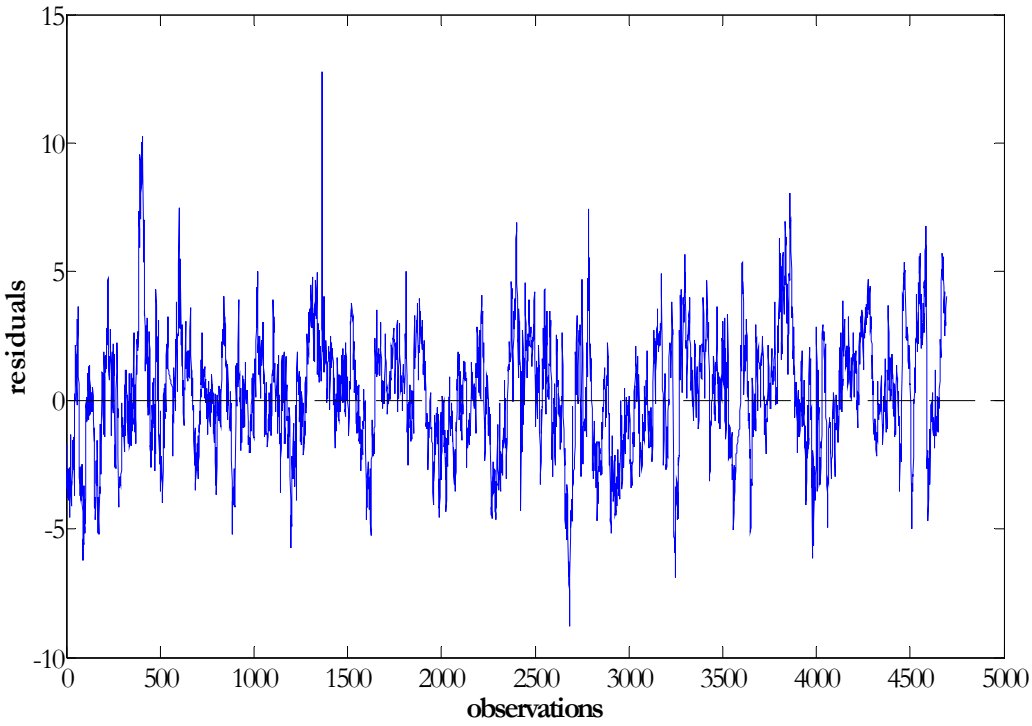


Figure E2. X instrument residuals (frequency – 1 hour, n(EMA)=50 observations)

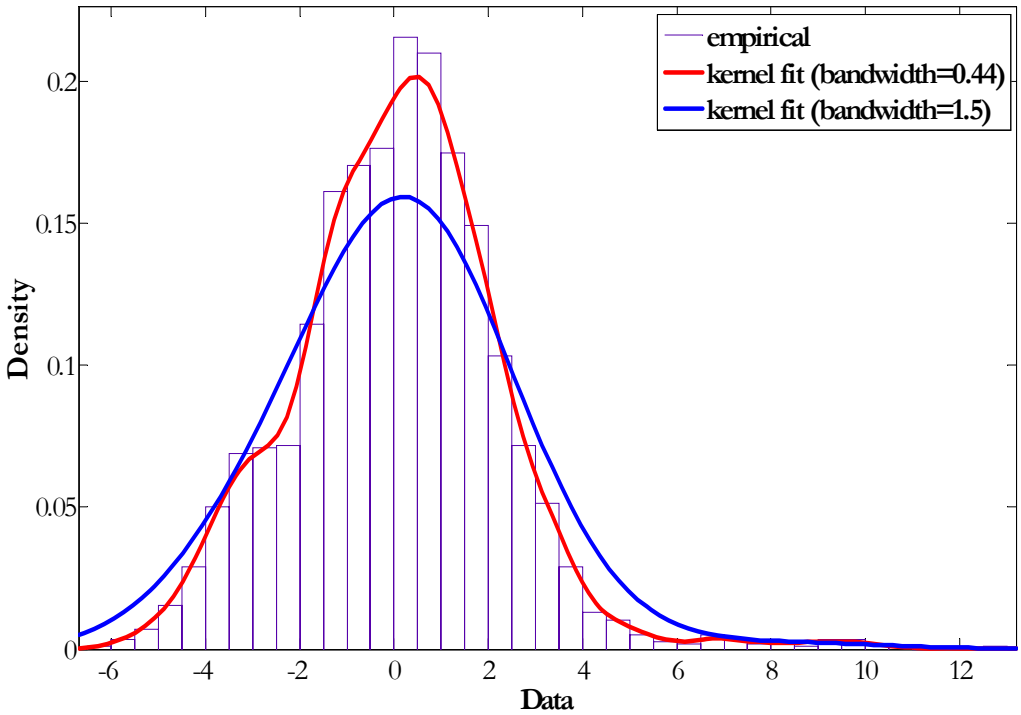


Figure E3. X instrument residuals (frequency 1 hour, n(EMA)=50 observations): Histogram and kernel fits for different bandwidth parameter values

The following vaiogram model, estimated on the sub-sample #1 $[1;2346]$ is used for further DK calculations (see Figure E2):

$$\gamma(h) = 0.85\left(\frac{3}{2}\cdot\left(\frac{h}{40}\right) - \frac{1}{2}\left(\frac{h}{40}\right)^3\right) + 0.15\left(1 - e^{-\left|\frac{h}{200}\right|}\right).$$

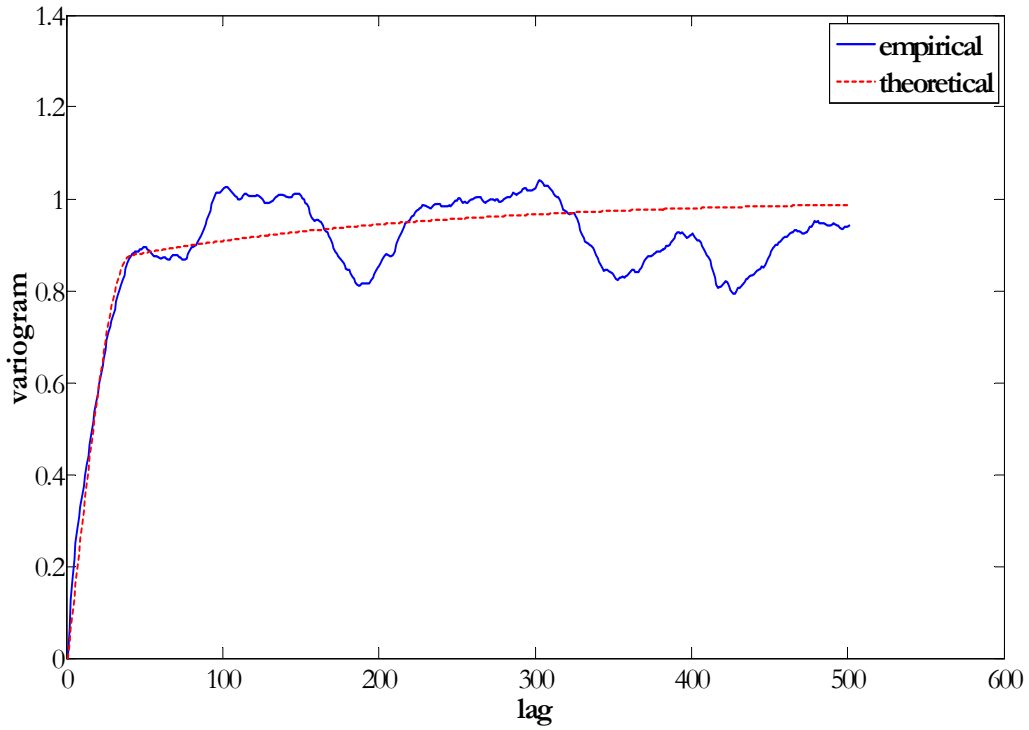


Figure E4. X instrument residuals (frequency - 1 hour, n(EMA)=50 observations): Variogram of the transformed residuals for sub-sample [1:2347], created on the base of the historic data, as well as theoretical model fit to the data.

General conclusions

In this thesis we have proposed several approaches to improve and optimize trading bands strategies. Parts I and part II concentrated on the optimization of the trading bands components: middle line (in the form of the moving average) and bands. Part III was dedicated to the improving of the process of the decision-making. Three parts of the thesis concentrated on slightly different trading bands. While first part produced the optimal middle line in the form of the kriged moving average (KMA), the second and the third parts used respectively simple moving average (SMA) and exponential moving average (EMA) as middle line components. Besides, while part II considered the trading bands defined by the data statistical characteristics (for example, variance), part III analyzed the trading strategies for the bands created by the parallel shift of their middle line. As for the first two parts, we wanted to avoid mixing the effects from the improvements of the trading bands components. Therefore, we did not introduce the optimal middle line (KMA) in the definition of the optimal bands values (DT bands). As for the third part, our strategy choice was explained by the huge popularity of this type of the bands in the trading applications. At the same time future research can address these questions and adjust the analyzed strategies.

In part I we considered the kriging method for the optimization of moving average (MA) weights. The kriging method is based on the statistical characteristics of data such as covariance (autocovariance) function. This allows obtaining optimal estimates that depend on the statistical characteristics of the data rather than on the historical data itself as in the case of the simulation studies. We have shown the examples of the method application to both equally and unequally sampled data. For the equally spaced data under assumption of the locally constant mean (within the moving window), the optimal weights follow some specific structure for certain covariance models stable over different window lengths: the largest weights in absolute value has the first and the last observation in the window, with smaller or relatively negligible weights for all other observations. As the result, KMA oscillates around the SMA curve. The volatility and amplitude of the oscillations is an indirect function of the KMA length: the longer the KMA the less it is volatile and it coincides more with the SMA curve. Therefore, trend-following strategies, based on KMA and SMA will take different positions at short window lengths and the same position at long lengths. We also saw that for the particular trend-following strategies, short KMA produces higher trading outcomes than the traditional MAs. It also seemed that the erratic nature of KMA curve did not necessary lead to more false signals, generated by these trading strategies.

As for the unequally spaced data we showed that the adjustment of nonregular sample to the regularly spaced one and calculation on them of the optimal MA, might lead to less effective trading strategies. Kriging results on such samples largely depend on the variability in the dependent variable.

Part II presented the data transformed (DT) bands as the alternative to the Bollinger bands. Traditional Bollinger method is statistically justified for the case of stationary, symmetric and known distributions. Departure from these assumptions complicates significantly applications and calculations of these bands. The DT bands under assumption of local stationarity provide simple, but powerful transformation of the Bollinger approach into better theoretical framework, which in addition is easier to optimize. The method is based on the transformation of raw data into normal random variable, for which the interval that contains some predefined percentage of data is known; then this interval is back-transformed into the bands for the raw data, which now contain the same percentage of data. We have considered particular residuals $R_i = P_i - SMA_t$,

$SMA_t = \frac{1}{n}\sum_i P_i$, $i \in [t-n+1;t]$ to calibrate transform function. Our main objective was to stay in the line with the Bollinger bands theory that uses these residuals for the calculation of the data standard deviation. However, this allowed us to obtain the bands, which are less dependent on the movements of the moving average. As the result, a specific stair-type form emerged for the DT bands, which change the level only when the large price innovations happen.

The numerical examples showed that the DT bands are not only the instruments that are better justified theoretically, but also can be a successful strategy component: in majority of cases confirmed DT bands strategies produced higher profits, trade efficiency and steeper P&L paths than confirmed (by momentum) strategies, based on the classical Bollinger bands. Moreover, the DT bands strategy was still profitable under non-zero transaction costs and slippage. Finally, the DT bands might be helpful in the definition of other technical rules – the Elliot waves and Support/Resistance levels.

Part III addressed the problem of the probability estimation of breaching some threshold levels. Another geostatistical approach, disjunctive kriging (DK) was used for this probability estimation. The application of the DK method to the financial data though needs some adjustment due to the data peculiarities. The main problem is its non-stationarity that demands local re-estimation of the DK parameters, in particular the transform function. We proposed the method to adjust the CDF for local volatility under assumption that the data followed the same distributional law, parameters of which are time dependent. The prediction power of the DK method was evaluated indirectly through the outcomes of the trading strategies, based on them. We constructed two strategies: (1) DK strategy, where decision about position entry was made on the basis of DK probabilities; and (2) Random-walk strategy, where decision about position entry was made randomly. Our analysis showed that the DK strategy produced positive outcomes for the continuous interval of thresholds with increasing strategy value over time. The RW strategy produced for all instruments the profits that had some random nature.

Future research

The results of the application of geostatistical and statistical methods to the improvement of the technical analysis techniques indicate the field for future research.

In this work, we have concentrated on the technical analysis applied to one instrument. The future research should consider the trading of a portfolio of instruments. In particular, another geostatistical multivariate methods, such as cokrging can be used to the estimation of the portfolio mean and prediction of its value.

In order to separate the effects of the improved middle line (by the introduction of the KMA) and improved bands (by the introduction of the DT bands), we have not considered the DT bands that incorporate the KMA as the middle line. We have also seen that at long MA lengths the DT bands, based on SMA and DT bands, based on KMA would be the same. However, it would be interesting to analyse the DT bands strategy, based on the short KMA.

As for the methods that have been considered in the thesis more attention should be devoted to the practical side of the application of the kriging method to subordinated random variables. In particular, the search of the optimal scaling factor that has impact on both the optimal weights structure as well as time-intensity of the calculation procedure should be studied in more details.

Besides, cokriging of the principal random variable and the variable, to which the prinicipal process is subordinated, can be considered as an alternative approach to the creation of a weighted MA.

The DT bands approach indicates the following directions of the research. The adjustment of the transform function to the local volatility, performed in the part III for the DK method, can be applied to the definition of the DT bands.

The analysis of the relationship between the strategy profitability and the value of K% section or k parameter of the DT bands will allow to create more successful trading strategies, based on these bands.

The DT approach creates the new opportunities to the improving of other technical analysis techniques. For example, the momentum based trading strategy might be ameliorated by the definition of optimum momentum thresholds on the basis of the data transformation approach. For this purpose momentum data should be transformed into normal, the *K%* section should be applied and the thresholds should be back-transformed to the real data. This might lead to the asymmetric momentum thresholds. The other example is the application of the DT bands to the definition of other technical indicators, such as Elliot waves and Support/Resistance trading rules.

Finally, the application of the DK method to the financial data can be improved further by the addressing the problem of adjustment of the CDF, used for data transformation, for the change in local mean or skewness.

Bibliography

1. Aan, P. "Weekly high/low moving average", Stock & Commodities, V.7:12, pp.431-432, Copyright © Technical Analysis Inc.

2. Acar, E., S.E. Satchell. 1997. "A theoretical analysis of trading rules: an application to the moving average case with Markovian returns", *Applied Mathematical Finance*, #4, pp.165-180.

3. Achelis, S. B. "Technical analysis from A to Z", *McGraw-Hill Professional*, 2000.

4. Ait-Sahalia, Y. 1996. "Non-parametric pricing of the interest rate derivative securities, *Econometrica,* #64, pp.527-560.

5. Ait-Sahalia, Y. 1999. "Transition densities for interest rate and other non-linear diffusions", *The Journal of Finance*, Vol. LIV, #4.

6. Ait-Sahalia, Y. 2002. "Maximum likelihood estimation of the discretely sampled diffusions: a closed form approximation approach", *Econometrica*, Vol. 70, #1, pp.223-262.

7. Alexander, S.S. 1961. "Price movements in speculative markets: Trends or random walks", *Industrial Management Review*, #2, pp.7-26.

8. Alexander, S.S. 1964. "Price movements in speculative markets: Trends or random walks, Number 2", *Industrial Management Review*, Spring, pp.25-46.

9. Alexander, C. "Trade with moving averages", Stock & Commodities, V.11:6, pp.257-260, Copyright © Technical Analysis Inc.

10. Allen, F. and Karjalainen. 1999. "Using genetic algorithms to find technical trading rules", *Journal of Financial Economics*, #51, pp. 245-271.

11. Arms, R.W., Jr. "Volume-adjusted moving averages", Stock & Commodities, V.8:3, pp.109-111, Copyright © Technical Analysis Inc.

12. Armstrong, M. "Basic linear geostatistics", *Springer,* 2004.

13. Armstrong, M. and P. Delfiner. 1980. "Towards a more robust variogram: A case study on coal", *Technical Report*, #671, Centre de Géostatistique, Fontainebleau, France.

14. Armstrong, M. and R. Jabin. 1981. "Variogram models must be positive-definite", *Journal of the International Association for Mathematical Geology*, #13(5), pp.455-459.

15. Arrington, G.R. "The basics of moving averages" Stock & Commodities, V.10:6, pp.275-278, Copyright © Technical Analysis Inc.

16. Arrington, G.R. "Building a variable length moving average" Stock & Commodities, V.9:6, pp.219-223, Copyright © Technical Analysis Inc.

17. Ausloos, M. 2000. "Statistical physics in foreign exchange currency and stock markets", *Physica A*, #285, pp.48-65.

18. Balsara, N., K. Carlson and N. V. Rao. 1996. "Unsystematic futures profits with technical trading rules: A case for flexibility", *Journal of Financial and Strategic Decision*, Vol.9, #1 (Spring, 1996), pp.57-66.

19. Bennett B.,K. 2001. "Using a moving average to determine cotton futures market entry dates", *The Journal of Cotton Science*, #5, pp.218-223.

20. Black, F. and M. Scholes. 1973. "The pricing of options and corporate liabilities, *Journal of Political Economy*, #81, pp.637-654.

21. Box, G. E. P, G. M. Jenkins and G.C. Reinsel. "Time series analysis: forecasting and control", *Wiley Series in Probability and Statistics*, 2008.

22. Blanchet-Scalliet, C., A. Diop, R. Gibson, D. Talay, E. Tanré, K. Kaminski. 2005. "Technical analysis compared to mathematical models based under misspecification", Working paper No.253, *National Centre of Competence in Research Financial Valuation and Risk Management*.

23. Blume L., D. Easley and M. O'Hara. 1994. "Market statistics and technical analysis: The role of volume", *The Journal of Finance*, Vol.49, #1 (March, 1994), pp.153-181.

24. Bollinger, J. "Bollinger on Bollinger bands", *McGraw-Hill*, 2002.

25. Brock, W., J. Laconishok and B. Lebaron. 1992. "Simple technical trading rules and the stochastic properties of stock returns", *The Journal of Finance*, Vol.47, #5 (December, 1992), pp.1731-1764.

26. Brown. D.P. and R.H. Jennings. 1989. "On technical analysis", *The Review of Financial Studies*, Vol. 2, #4, pp.527-551.

27. Carroll, R. J. Ruppert, D. "Transformation and weighting in regression", *CRC Press*, 1988.

28. Chande, T. S. "Adapting moving averages to market volatility", Stock & Commodities, V.10:3, pp.428-433, Copyright © Technical Analysis Inc.

29. Chande, T. S. "Beyond technical analysis: How to develop and implement a winning trading system", *John Wiley and Sons*, 2001.

30. Chilès, J.-P. 1977. "Géostatistique des phénomènes non stationnaires", *Doctoral Thesis*, Université de Nancy-I, France.

31. Chilès, J.-P. 1979a. "La dérive à la dérive", *Technical Report*, #591, Centre de Géostatistique, Fontainebleau, France.

32. Chilès, J.-P. 1979b. "Le variogramme généralisé", *Technical Report*, #612, Centre de Géostatistique, Fontainebleau, France.

33. Chilès, J.-P. and P. Delfiner. "Geostatistics. Modelling spatial uncertainty", *A-Wiley-Interscience publication*, *John Wiley & Sons, Inc.*, 1999.

34. Cressie N. A. C. «Statistics for Spatial Data», *A-Wiley-Interscience publication*, *John Wiley & Sons, Inc.*, 1991.

35. Cox, D. R., D. Oakes. "Analysis of Survival Data", *Chapman & Hall, London*, 1984.

36. Dai, L. H. Wei, and L. Wang. 2007. "Spatial distribution and risk assessment of radionuclides in soils around a coal-fired power plant : A case study from the city of Baoji, China", *Environmental Research*, Vol. 104, # 2, pp. 201-208.

37. Di Lorenzo, R. and V. Sciarretta. 1996. "Statistical evidence on a new method of trading the financial markets", published in *AF journal*, #24 (December, 1996).

38. Ehlers J. "Signal analysis concepts", http://www.mesasoftware.com/technicalpapers.htm, http://www.jamesgoulding.com/Research_II/Ehlers/Ehlers%20(Signal%20Analysis%20Concepts).doc, http://moving-averages.technicalanalysis.org.uk/Ehle.pdf.

39. Emery, X. 2006. "A disjunctive kriging program for assessing point-support conditional distributions", *Computers & Geosciences*, Vol. 32, # 7, pp. 965-983.

40. Fama, E.F. and M.E.Blume. 1966. "Filter rules and stock market trading", *Journal of Business*, #39, pp.226-241.

41. Fang, Y. and D. Xu. 2002. "The predictability of asset returns: an approach combining technical analysis and time series forecasts", *International Journal of Forecasting I.*

42. Fernandez-Rodriguez, F., S. Sosvilla-Rivero and J. Andrada-Felix. 1999. "Technical analysis in the Madrid stock exchange", *FEDEA Working paper (Documento de trabajo),* #99-05 (April, 1999).

43. Fernandez-Rodriguez, F., S. Sosvilla-Rivero and J. Andrada-Felix. 2000. "Technical analysis in foreign exchange markets: Linear versus nonlinear trading rules", *Documentos de Economia y Finanzas Internationalez,* DEFI00/02 (September, 2000), http://www.fedea.es/hojas/publicaciones.html.

44. Focardi, S. and F. J. Fabozzi. "The mathematics of financial modelling and investment management", *John Wiley and Sons*, 2004.

45. Gençay, R., Selçuk, F. and B. Whitcher. "An introduction to wavelets and other filtering Methods in Finance and Economics", *Academic Press an Imprint of Elsevier*, 2002.

46. Gray, A. and P. Thomson. 1997. "Design of moving average trend filters using fidelity, smoothness and minimum revisions criteria", *Bureau of the Census Statistical Research Division*, *Statistical Research Report Series*, #RR96/01.

47. Greene W.H. "Econometric analysis", *Prentice Hall*, 2007.

48. Hansen, L.P. and J.A. Scheinkman. 1995. "Back to the future: Generating moment implications for continous time Marokv processes, *Econometica*, #63, pp.767-804.

49. Hartle, T. "Sidebar: Variable length moving average", Stock & Commodities, V.13:10, pp.418-423, Copyright © Technical Analysis Inc.

50. Harvey, A. "Forecasting, structural time series models and the Kalman filter", *Cambridge University Press*, 1991.

51. Harvey, A. Koopman, S.J., Shephard, N. "State space and unobserved component models: Theory and applications", *Cambridge University Press*, 2004.

52. Honoré, P. 1997. "Maximum-likelihood estimation of non-linear continuous-time term structure models", *Working paper*, Aarhus University.

53. Hu, L.Y. et Ch. Lantuejoul. 1988. "Recherche d'une fonction d'anamorphose pour la mise en oeuve du krigeage disjonctif isofactoriel Gamma", *Etude Géostatistiques V – Séminaire C.F.S.G. sur la Géostatistique 15-16 Juin 1987, Fontainebleau. Sci. De la Terre, Sér. Inf., Nancy,1988,* #28, pp. 145-173.

54. Hudson, R., M. Dempsey and K .Keasey. 1996. "A note on the weak form efficiency of capital markets: The application of simple technical trading rules to UK stock prices – 1935-1994", *Journal of Banking and Finanace*, #20, pp.1121-1132.

55. Hutchinson, T. and P.G. Zhang. "Weighted moving averages", Stock & Commodities, V.11:12, pp.500-505, Copyright © Technical Analysis Inc.

56. James, F.E.Jr. 1968. "Monthly moving avearages – an effective investment tool?", *Journal of Financial and Quantitative Analysis*, (September, 1968), pp.315-326.

57. Jensen, M. C. and G. Benington. 1970. "Random walks and technical theories: Some additional evidence", *The Journal of Finance*, Papers and Proceedings of the Twenty-Eight Annual Meeting of the American Finance Association New-York, N.Y. December, 28-30, 1969, Vol.25, #2 (May, 1970), pp. 469-482.

58. Katz, J.O. and D. McCormick. "The encyclopedia of trading strategies", *McGraw-Hill Professional*, 2000.

59. Kavajecz K.A and E.R. Odders-White. 2004. "Technical analysis and liquidity provision", *The Review of Financial Studies*, Vol. 17, # 4 (Winter, 2004), pp.1043-1071.

60. Kennedy, P. "A guide to econometrics", *The MIT Press*, 1998.

61. Lang Chao-Yi, "Kriging interpolation", at www.nbb.cornell.edu.

62. Levich, R. and L. Thomas. 1993. "The significance of technical trading-rule profits in the foreign exchange market: A bootstrap approach", *Journal of International Money and Finance*, #12, pp.451-474.

63. Li, P. 2005. "Box-Cox transformations: An overview", presentation, http://www.stat.uconn.edu/~studentjournal/index_files/pengfi_s05.pdf.

64. Lien, K. "Day trading the currency market: Technical and fundamental strategies to profit from market swings", *John Wiley and Sons*, 2006.

65. Lo, A. W. 1988. "Maximum likelihood estimation of generalized Ito processes with discretely sampled data, *Econometric Theory*, #4, pp.231-247.

66. Lo, A. W. 2007. "Efficient market hypothesis" in "The New Palgrave: A Dictionary of Economics" by Blume, L. and S. Durlauf, *New York: Palgrave McMillan*, 2007.

67. Lo, A. W., Mamaysky, H. and J.Wang. 2000. "Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation", *The Journal of Finance*, Vol. LV, #4 (August, 2000), pp.1705-1765.

68. Lo, A. and J.Wang. 2000. "Trading volume: Definitions, data analysis and implications of portfolio theory", *The Review of Financial Studies*, Vol 13, #2 (Summer, 2000), pp.257-300.

69. Martinez, W. L. and A. R. Martinez. "Computational statistics handbook with MATLAB", *CRC Press*, 2001.

70. Matheron G. "Osnovy prikladnoi geostatistiki (Treatise of Geostatistics), *Mir, Moskow* 1968.

71. Matheron G. 1969. "Le krigeage universel", *Cahiers du Centre de Morphologie Mathématique de Fontainebleau*, Fasc.1, Ecole des Mines de Paris.

72. Matheron G. "La théorie des variables régionalisées, et ses applications", *Les Cahiers du Centre de Morphologie Mathématique de Fountainebleau*, 1970.

73. Matheron, G. 1973. "Le krigeage disjonctive", *Technical Report*, #360 Centre de Géostatistique, Fontainebleau, France.

74. Matheron, G. 1976. "A simple substitute for conditional expectation: the disjunctive kriging", in "Advanced Geostatistics in Mining Industry", 1976, *Reidel Publishing Company Dordrecht.*

75. Matheron, G. 1977. "Peut-on imposer des conditions d'universalite au krigeage disjonctif", *Technical Report*, Centre de Géostatistique, Fontainebleau, France.

76. Matheron, G. 1986. "Sur la positivité des poids de krigeage", *Technical Report*, #30/86/G, Centre de Géostatistique, Fontainebleau, France.

77. Murphy, John J. "Technical analysis of the financial markets", *New York Institute of Finance*, 1999.

78. Neely, C.J. 1997. "Technical analysis in the foreign exchange market: A Layman's guide", *Review*, September/October, 1997, pp.23-38.

79. Neely, C.J., P. Weller and R. Dittmar. 1997. "Is technical analysis in the foreign exchange market profitable? A genetic Programming Approach", *The Federal Reserve Bank of St. Louis Working Paper Series*, #96-006C, August, 1997.

80. Neftci, S. N. 1991. "Naïve trading rules in financial markets and Wiener-Kolmogorov predicition theory: A study of technical analysis", *Journal of Business,* Volume 64, Issue 4 (October, 1991), pp. 549-571.

81. Nikifork, R. "Trends and moving averages", Stock & Commodities, V.16:12, pp.583-587, Copyright © Technical Analysis Inc.

82. Orfeuil, J.P. 1977. "Une approche statistique du probleme de l'alerte en pollution atmospherique", Note de Centre Géostatistique Fonteaineblueau, *N-506*.

83. Osler C.L. 2003. "Currency Orders and Exchange Rate Dynamics: An Explanation for the Predictive Success of Technical Analysis", *The Journal of Finance*, Vol. 58, #5 (October, 2003), pp. 1791-1819.

84. Osler, C. L. and P.H. Kevin Chang. 1995. "Head and shoulders: not just a flaky pattern", *Federal Reserve Bank of New York Staff report*, #4, pp.1-65.

85. Pardo, R. "The Evaluation and Optimization of Trading Strategies", *John Wiley and Sons*, 2008.

86. Ratner, M. and R. Leal. 1999. "Tests of technical trading strategies in the emerging equity markets of Latin America and Asia", *Journal of Banking and Finance*, #23, pp.1887-1905.

87. Rode, D., Y. Friedman, S. Parikh and J. Kane. 1995. "An evolutionary approach to technical trading and capital market efficiency", *The Wharton School University of Pennsylvania*, May 1, 1995.

88. Rivoirard, J. "Disjuncive kriging and non-linear geostatistics", *Clarendon Press*, *Oxford University Press,* 1994.

89. Saacke, P. 2002. "Technical analysis and the effectiveness of central bank intervention", *Journal of International Money and Finance*, #21, pp.459-479.

90. Santa-Clara, P. 1995. Simulated likelihood estimation of diffusions with an application to the short term interest rate, *Working paper*, UCLA.

91. Scott, D. W. "Multivariate Density Estimation: Theory, Practice, and Visualization", *Wiley-Interscience*, 1992.

92. Shiryayev, I. "Probability", *New-York: Springer-Verlag,* 1985.

93. Sullivan, R., A. Timmermann and H. White. 1999. "Data-snooping, technical trading rule performance and the bootstrap", *The Journal of Finance*, Vol.54, #5 (October, 1999), pp. 1647-1691.

94. Theoret Raymond, Rostan Pierre. "Les Bandes de Bollinger comme technique de réduction de la variance des prix d'options sur obligations obtenus par la simulation de Monte Carlo", Les cahiers de la recherche, Research paper, Rouen School of Management Research, # 50-2003/2004, www.esc-rouen.fr

95. Tilley, D.L. "Moving averages with resistance and support, Stock & Commodities, V.16:9, pp.108-114, Copyright © Technical Analysis Inc.

96. Treynor, J.L. and R. Ferguson. 1985. "In defence of technical analysis", *The Journal of Finance*, Vol.40, #3 (July, 1985), Papers and Proceedings of the Forty-Third Annual

Meeting American Finance Association, Dallas, Texas, December 28-30 , 1984, pp.757-773.

97. Triantafilis, J. I., O. A. Odeh, B. Warr and M. F. Ahmed. 2004. "Mapping of salinity risk in the lower Namoi valley using non-linear kriging methods", *Agricultural Water Management*, Vol. 69, # 3, pp. 203-231.

98. Van Horne, J.C. and G.G.C. Parker. 1967. "The random walk theory: An empirical test", *Financial Analysts Journal*, #23, pp.87-92.

99. Vasicek, O. 1977. "An equilibrium characterization of the term structure, *Journal of Financial Economics*, #5, pp.177-188

100. von Steiger, B., R. Webster, R. Schulin and R. Lehmann. 1996. "Mapping heavy metals in polluted soil by disjunctive kriging", *Environmental Pollution*, Vol. 94, # 2, pp. 205-215.

101. Wand, M. P. and M. C. Jones. "Kernel Smoothing", *CRC Press*, 1995.

102. Williams, O.D. 2006. "Empirical optimization of Bollinger Bands for profitability", *MA Thesis,* Simon Fraser University