

**CHARLES UNIVERSITY IN PRAGUE**

**FACULTY OF SOCIAL SCIENCES**

Institute of Economic Studies



**Fractional Cointegration of Daily High  
and Low Stock Prices**

*Rigorous Thesis*

AUTHOR:

**MGR. SYLVIE DVOŘÁKOVÁ**

SUPERVISOR:

**PhDr. Jozef Baruník, Ph.D.**

ACADEMIC YEAR:

**2013/2014**

## **Declaration of authorship**

I hereby declare that I wrote this rigorous thesis on my own under the guidance of my supervisor and that all resources and literature I have used are referenced herein.

I hereby grant permission to reproduce and to distribute copies of this thesis document for study and research purposes provided I am credited as the author per generally accepted citation standards.

Prague, February, 12, 2014

---

Sylvie Dvořáková

## **Acknowledgement**

Foremost I would like to express my sincere gratitude to my thesis supervisor, PhDr. Jozef Baruník, Ph.D. for his patience, support, guidance and encouragement. Moreover, I would like to thank my friend Bc. Petr Pernikář for his help with programming and my friend Bc. Michael Janský for proofreading. Last but not least, I would like to thank my master thesis oponent RNDr. Michal Červinka, Ph.D., and to PhDr. Ladislav Krištofuk, Ph.D., for their valuable comments which have contributed to the better quality of this thesis.

# Abstract

In this thesis we provide unique empirical support for the fractional cointegration of daily high and low stock prices. The recently formalized fractionally cointegrated VAR model by Johansen and Nielsen (2012) is used due to its ability to capture both the cointegration between daily high and low stock prices and the long memory of their linear combination, the range. Daily high and low stock prices are of particular interest as they provide valuable information about range-based volatility, which is considered a highly efficient and robust estimator of volatility. We provide a comparison of the Czech PX index with the developed market indices (DAX, FTSE 100, S&P 500 and NIKKEI 225) during the 2003-2012 period as well as before and after the crisis. We find that the range of all indices displays long memory and is mostly in the non-stationary region (except for the ranges of the PX Index and NIKKEI 225 indices in the pre-crisis period). These findings provide evidence that volatility may not be a stationary process. No common pattern is detected among all five market indices and different behaviour is also observed in the pre-crisis and post-crisis periods.

**JEL Classification**

C32, C58, G15

**Keywords**Fractional cointegration, long memory, FCVAR  
model, daily high and low prices**Author's e-mail**

s.dvorakova@gmail.com

**Supervisor's e-mail**

barunik@fsv.cuni.cz

**Volume**

110 820 characters

# Abstrakt

Tato práce přináší jedinečnou empirickou analýzu podporující frakční kointegraci denních maximálních a minimálních cen akcií. Používáme zde frakčně kointegrovaný VAR model, který byl pouze nedávno zobecněn Johansenem a Nielsenem (2012). Tento model je schopen popsat jak frakční kointegraci mezi denními maximálními a minimálními cenami akcií, tak i dlouhou paměť jejich lineární kombinace, tzv. rozpětí (range). Denní extrémní ceny jsou obzvláště zajímavé, protože poskytují cenné informace o odhadu volatility pomocí rozpětí (range-based volatility). Tento odhad je v literatuře považován za vysoce eficientní a robustní odhad volatility. V práci přinášíme srovnání českého indexu PX s vyspělými tržními indexy (DAX, FTSE 100, S&P 500 a NIKKEI 225) v období 2003-2012. Tyto indexy také zkoumáme v období před a po krizi. Zjistili jsme, že rozpětí (range) všech indexů vykazuje dlouhou paměť a je většinou nestacionární (s výjimkou rozpětí indexů PX a NIKKEI 225 v období před krizí). Tyto poznatky naznačují, že volatilita nemusí být stacionární proces. Nenašli jsme žádné společné vlastnosti napříč všemi indexy, odlišné chování je také vyzorováno v obdobích před a po krizi.

**Klasifikace**

C32, C58, G15

**Klíčová slova**

Frakční kointegrace, dlouhá paměť, FCVAR model, denní maximální a minimální ceny

**E-mail autora**

s.dvorakova@gmail.com

**E-mail vedoucího práce**

barunik@fsv.cuni.cz

**Rozsah**

110 820 znaků

# Table of contents

List of Tables .....	vii
List of Figures .....	viii
List of Acronyms.....	ix
Rigorous Thesis Proposal .....	x
1 Introduction .....	1
2 Data description .....	4
3 Cointegrated VAR framework .....	10
3.1 Unit root tests .....	10
3.2 Cointegration tests.....	18
3.3 CVAR model.....	21
4 Long memory processes.....	23
4.1 Exact local Whittle estimator.....	26
4.2 GPH estimator.....	30
4.3 Testing the equality of integration orders .....	34
5 Fractionally cointegrated VAR.....	37
5.1 Derivation of FCVAR .....	37
5.2 FCVAR estimation .....	39
5.2.1 Maximum likelihood estimation.....	39
5.2.2 Imposing restrictions on parameters .....	40
5.2.3 Initial values .....	41
5.3 Cointegration rank tests.....	41
5.3.1 Cointegration rank test by Nielsen and Shimotsu (2007).....	42
5.3.2 Cointegration rank test by Johansen and Nielsen (2012).....	45
5.4 FCVAR estimation results.....	48
6 Conclusion .....	57
7 References .....	60
8 Appendix .....	64

# List of Tables

Table 1: P-values for an ADF test for variables in levels and first-differences.....	12
Table 2: P-values for KPSS test for variables in levels with two lag specifications.....	17
Table 3: Exact local Whittle estimator of the fractional degree of integration .....	29
Table 4: GPH estimator of the fractional degree of integration .....	32
Table 5: Test statistics for the equality of integration orders.....	35
Table 6: Cointegration rank test by Whittle and Shimotsu (2007).....	44
Table 7: Cointegration rank test by Johansen and Nielsen (2012).....	47
Table 8: FCVAR estimation results (no restrictions).....	50
Table 9: FCVAR estimation results (with restrictions).....	53
Table 10: Comparison of integration orders of range .....	55

# List of Figures

Figure 1: High and Low Log-Prices of PX Index .....	6
Figure 2: Range of PX Index .....	7
Figure 3: Comparison of High Log-Prices for All Indices.....	8
Figure 4: ACF of DAX Range in All Periods .....	14
Figure 5: ACF of PX Index Range in All Periods .....	15
Figure 6: High and Low Log-Prices of DAX.....	64
Figure 7: Range of DAX .....	64
Figure 8: High and Low Log-Prices of FTSE 100 .....	65
Figure 9: Range of FTSE 100.....	65
Figure 10: High and Low Log-Prices of NIKKEI 225 .....	66
Figure 11: Range of NIKKEI 225.....	66
Figure 12: High and Low Log-Prices of S&P 500.....	67
Figure 13: Range of S&P 500 .....	67
Figure 14: ACF of Detrended High and Low Log-Prices of DAX (2003-2007).....	68
Figure 15: ACF of Detrended High and Low Log-Prices of FTSE 100 (2003-2007) .....	68
Figure 16: ACF of FTSE 100 Range in All Periods.....	69
Figure 17: ACF of S&P 500 Range in All Periods .....	69
Figure 18: ACF of NIKKEI 225 Range in All Periods .....	70

# List of Acronyms

ACF	Autocorrelation Function
ADF	Augmented Dickey-Fuller test
ARCH	Autoregressive Conditional Heteroscedasticity
ARFIMA	Autoregressive Fractionally Integrated Moving Average
ARMA	Autoregressive Moving Average
BIC	Bayesian Information Criterion
CVAR	Cointegrated Vector Autoregression
EGARCH	Exponential Generalized Autoregressive Conditional Heteroscedasticity
ELW	Exact Local Whittle
FIGARCH	Fractionally Integrated Generalized Autoregressive Conditional Heteroscedasticity
FCVAR	Fractionally Cointegrated Vector Autoregression
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
GPH	Geweke and Porter-Hudak
IV	Initial Values
KPSS	Kwiatkowski–Phillips–Schmidt–Shin test
LR	Likelihood Ratio
LW	Local Whittle
NBER	National Bureau of Economic Research
VAR	Vector Autoregression
VECM	Vector Error Correction Model

# Rigorous Thesis Proposal

<b>Institut ekonomických studií UK FSV</b> <b>Teze RIGORÓZNÍ práce</b> Tyto teze tvoří přílohu „Přihlášky ke státní rigorózní zkoušce“	
<b>TUTO ČÁST VYPLŇUJE ŽADATEL:</b>	
Informace o žadateli:	Razítko podatelny:  katedra: k rukám:
Příjmení a jméno: Dvořáková Sylvie	
E-mail: s.dvorakova@gmail.com	
Telefon (nejlépe mobilní): 604 259 353	
<b>Předpokládaný název rigorózní práce v češtině:</b> Frakční kointegrace denních maximálních a minimálních cen akcií	
<b>Předpokládaný název rigorózní práce v angličtině:</b> Fractional Cointegration of Daily High and Low Stock Prices	
<b>Předpokládaný termín předložení práce:</b> Únor 2014	
<b>Pedagog, s nímž byly teze konzultovány :</b> PhDr. Jozef Baruník, Ph.D.	
<b>Charakteristika tématu a jeho dosavadní zpracování žadatelem (rozsah do 1000 znaků):</b>  <p>In this thesis we analyse daily high and low stock prices as they provide valuable information about range-based volatility, which is not included in the open and close prices. These prices are especially interesting, because as Parkinson (1980) shows the estimator of volatility based on the daily price range, defined as the difference between daily high and low prices, is a highly efficient and robust estimator of volatility. In finance, volatility is considered as one of the most important concepts measuring the risk of financial assets. This thesis thus brings new analysis of volatility in the fractional framework which contributes to the understanding of the volatility's behaviour and to the improvement of the predictability of the extreme prices. "</p> <p>This thesis has already been elaborated and defended as diploma thesis with excellent grade. Comments of the opponent and other improvements of the master thesis will be incorporated.</p>	
<b>Předpokládaný cíl rigorózní práce, původní přínos autora ke zpracování tématu, případně formulace problému, výzkumné otázky nebo hypotézy (rozsah do 1200 znaků):</b>	

There are two main hypotheses, first, daily high and low stock prices are fractionally cointegrated and second, their difference, the range, displays long memory and is potentially a non-stationary process. We analyse both the period before and after the recent financial crisis and the whole period 2003-2012. The main contribution of this thesis is an application of recently formalized fractionally cointegrated VAR and the contribution of our empirical results to the current literature.

We find significant evidence of long memory in daily range. In most cases, the range-based volatility estimate is in the non-stationary region. There are only two exceptions: the ranges of the PX Index and the NIKKEI 225 index in the 2003-2007 period, which are in the stationary region. The estimates of the long memory parameters are, however, quite sensitive to the selected methodology. We also show that the first period (2003-2007) is rather calm, as the estimates of long memory are generally the lowest, while the second period, 2008-2012, is more turbulent. Overall, the PX Index displays the lowest estimates of the order of price range integration, and its behaviour is very similar to that of the NIKKEI 225. The ranges of S&P 500, FTSE 100 and DAX indices display, on the other hand, relatively higher orders of integration.

**Předpokládaná struktura práce** (rozdělení do jednotlivých kapitol a podkapitol se stručnou charakteristikou jejich obsahu):

**1 Introduction**

**2 Data description**

In this section, we describe the used indices individually and also provide a comparison of the development in all indices studied in the thesis. We also discuss why daily high and daily low are of a particular interest.

**3 Cointegrated VAR framework**

**3.1 Unit root tests**

**3.2 Cointegration tests**

**3.3 CVAR model**

Here we present the original cointegrated framework, which does not account for the long memory. We test the series for unit root and describe the cointegration test and CVAR.

**4 Long memory processes**

**4.1 Exact local Whittle estimator**

**4.2 GPH estimator**

**4.3 Testing the equality of integration orders**

This section introduces the concept of long memory and presents two tests, namely the Exact Local Whittle and Geweke and Porter-Hudak, for a long memory in both daily prices and in the range. Since we are estimating the order of integration for two series (daily high and low log-prices) we also describe and conduct a test for the equality of integration orders.

**5 Fractionally cointegrated VAR**

**5.1 Derivation of FCVAR**

**5.2 FCVAR estimation**

**5.3 Cointegration rank tests**

**5.4 FCVAR estimation results**

This section presents the fractionally cointegrated VAR model and show how it can be estimated. We also

describe and conduct the cointegration rank tests which should be applied in the fractional framework. This chapter also contains the discussion of results and comparison of different methods applied to test the long memory in range.

## 6 Conclusion

**Základní literatura** (nejméně 10 nejdůležitějších titulů k tématu a metodě jeho zpracování; u všech titulů je nutné uvést stručnou anotaci na 2-5 řádků):

- Caporin, M., Rinaldo, A., and P. Santucci de Magistris (2011)

On the Predictability of Stock Prices: A Case for High and Low Prices, "Marco Fanno" Working Papers 0136, Dipartimento di Scienze Economiche "Marco Fanno".

*This paper applies a FCVAR methodology on Dow Jones Industrial Index and finds a cointegrating relationship between daily high and low prices and the long memory of their difference (range).*

- Cheung, Y. (2007)

An empirical model of daily highs and lows, International Journal of Finance and Economics, 12(1), 1–20.

*This paper models the relationship between daily high and low stock prices with the original cointegrated VAR model without taking into account the long memory and they find that the range is stationary.*

- Geweke, J., and S. Porter–Hudak (1983)

The estimation and application of long memory time series models, Journal of Time Series Analysis 4(4), 221–238.

*This study provides a methodology for the GPH test of long memory parameter, which we apply in our thesis. They propose to estimate the long memory parameter based on the linear regression of the log periodogram on a deterministic regressor. The estimator is then the OLS estimator of the slope parameter in the regression, which is formed using only the lowest frequency ordinates of the log periodogram.*

- Johansen, S. (2008)

A representation theory for a class of vector autoregressive models for fractional processes, Econometric Theory, 24(3), 651–676.

*This paper is the first to propose the fractionally cointegrated vector autoregressive model (FCVAR model) which is based on the fractional lag operator.*

- Johansen, S., and M. O. Nielsen (2010)

Likelihood inference for a nonstationary fractional autoregressive model, Journal of Econometrics 158(1), 51–66.

- Johansen, S., and M. O. Nielsen (2012)

Likelihood inference for a fractionally cointegrated vector autoregressive model, Econometrica, 80(6), 2667–2732.

*The two above mentioned papers by Johansen and Nielsen further analyse the FCVAR model proposed in Johansen (2008). They describe the estimation and testing of the model together with the asymptotic theory and also provide a generalization in terms of initial values, which are not assumed to be zero but instead bounded.*

- MacKinnon, J. G., and M. Ø. Nielsen (2012)

Numerical distribution functions of fractional unit root and cointegration tests, Journal of Applied Econometrics.

*This paper simulates numerical asymptotic distribution functions of likelihood ratio tests for fractional unit roots and cointegration rank. They make available a programme, which given the dimension of a problem calculates either critical values*

or asymptotic  $p$ -values for the LR test statistic.

- Nielsen, M., and L. Morin (2012)

FCVARmodel.m: A Matlab software package for estimation and testing in the fractionally cointegrated VAR model, QED working paper 1273, Queen's University.

*This paper provides a software package, which enables us to estimate the fractionally cointegrated VAR model together with a manual how to use it and an empirical example.*

- Nielsen, M. Ø., and K. Shimotsu (2007)

Determining the cointegration rank in non-stationary fractional system by the exact local Whittle approach, Journal of Econometrics, 141(2), 574–596.

*This paper presents a cointegration rank determination procedure which is suitable for both stationary and non-stationary fractionally integrated processes by applying the ELW analysis of Shimotsu and Phillips (2005).*

- Parkinson, M. (1980)

The Extreme Value Method for Estimating the Variance of the Rate of Return, The Journal of Business, 53(1), 61–65.

*In this study, Parkinson shows that the range-based volatility estimator (which is the extreme value approach) is an efficient and robust estimator of volatility.*

- Shimotsu, K., and P. Phillips (2005)

Exact local Whittle estimation of fractional integration, Annals of Statistics, 33(4), 1890–1933.

*This study provides another test for the long memory parameter, which is together with the GPH test used in our thesis to assess the long memory in both the daily prices and in their difference, the range. This test makes no assumptions about the presence of cointegration and is consistent both when cointegration is present and when it is absent. This estimator is also applicable to both the stationary and non-stationary case.*

**Diplomové a disertační práce k tématu** (seznam bakalářských, magisterských a doktorských prací, které byly k tématu obhájeny na UK, případně dalších oborově blízkých fakultách či vysokých školách za posledních pět let)

This topic is a very recent and advanced and I have been unable to find a similar topic.

**Datum / Podpis studenta**

10.07.2013

.....

TEZE JE NUTNO ODEVZDAT **VYTIŠTĚNÉ** (včetně části, kterou vyplňuje institut!), **PODEPSANÉ** A VE **DVOU** VYHOTOVENÍCH DO TERMÍNU UVEDENÉHO V HARMONOGRAMU PŘÍSLUŠNÉHO AKADEMICKÉHO ROKU, A TO PROSTŘEDNICTVÍM PODATELNY UK FSV.  
PŘIJATÉ TEZE JE NUTNÉ SI **VYZVEDNOUT** V SEKRETARIÁTU PŘÍSLUŠNÉ KATEDRY A **NECHAT VEVÁZAT** DO RIGORÓZNÍ PRÁCE.

<b>TUTO ČÁST VYPLŇUJE INSTITUT:</b>			
Vyjádření katedry:		Schváleno <input type="checkbox"/>	
		Neschváleno <input type="checkbox"/>	
Důvody případného neschválení práce		Téma je již zpracované <input type="checkbox"/>	
		Špatně formulované téma a cíl <input type="checkbox"/>	
		Špatně zvolená metoda práce <input type="checkbox"/>	
		Nedostatečná rešerše literatury <input type="checkbox"/>	
		Nevhodně zvolené prameny <input type="checkbox"/>	
		Nedostačující úroveň tezí <input type="checkbox"/>	
		Jiné .....	
		.....	
Navržený konzultant	Souhlas konzultanta navrženého vedením institutu	Příjmení a jméno	Datum /Podpis
		.....	.....
Návrhy na další konzultanty		Příjmení a jméno	Příjmení a jméno
		.....	.....
Schválené teze převzal/a žadatel/ka		Příjmení a jméno	Datum /Podpis
		.....	.....
Návrhy na oponenta:		Příjmení a jméno	
		.....	

# 1 Introduction

In this thesis we analyse daily high and low stock prices as they provide valuable information about range-based volatility, which is not included in the open and close prices. These prices are especially interesting, because as Parkinson (1980) shows, the estimator of volatility based on the daily price range, defined as the difference between daily high and low prices, is a highly efficient and robust estimator of volatility. In finance, volatility is considered as one of the most important concepts measuring the risk of financial assets. This thesis thus brings new analysis of volatility in the fractional framework which contributes to the understanding of the volatility's behaviour and to the improvement of the predictability of the extreme prices.

The stock prices in developed markets are generally considered to be unpredictable and are according to the Efficient Market Hypothesis assumed to follow a random walk. However, Caporin et al. (2013) argue that this rationale does not apply to the extreme prices (high and low prices). In most studies stock prices are shown to be integrated of order 1 (an  $I(1)$  processes)<sup>i</sup> and thus confirming the random walk and unpredictability hypothesis (see e.g. Cheung, 2007). Caporin et al. (2013) show that the strict choice between stationary  $I(0)$ <sup>ii</sup> and non-stationary  $I(1)$  processes can be too restrictive and instead argue that the degree of integration of daily highs and lows should be estimated in a fractional or long-memory framework, where the series are assumed to be integrated of order  $d$  and cointegrated of order less than  $d$ , say  $(d - b)$ , where  $d, b \in \mathbb{R}$  and  $0 < b \leq d$ . We thus say that the series are fractionally cointegrated of order  $(d - b)$ .

In order to determine the relationship between highs and lows we implement a fractionally cointegrated vector autoregressive model (FCVAR) as proposed by Johansen (2008) and further analysed by Johansen and Nielsen (2010, 2012). The motivation for using this methodology is twofold. First, daily highs and lows are assumed to be cointegrated, i.e. in the short term they may diverge, but in the long term they have an

---

<sup>i</sup> An  $I(1)$  process is a non-stationary process, where only first differencing is sufficient to obtain stationarity.

<sup>ii</sup> An  $I(0)$  process is a stationary process, where no differencing is needed to achieve stationarity.

embedded convergence path. Second, their specific linear combination, the range, is assumed to display a long memory. This property of range is well captured by the fractional autoregressive technique and the range is, in fact, the error correction term in the cointegrating relation between high and low log-prices.

In the thesis, the above-described methodology is applied in the following way. The high and low log-prices and range are first tested for the presence of a unit root based on the Augmented Dickey-Fuller test (ADF) and Kwiatkowski–Phillips–Schmidt–Shin test (KPSS). The results of these two tests for the range are not unambiguous, which confirms our initial hypothesis that the traditionally applied stationarity tests are not capable of properly determining the stationarity of the range. This is an indication of a potential long-memory nature of the range process, which we estimate through the Exact Local Whittle and Geweke Porter-Hudak estimators. Substantial evidence of the presence of long memory has been demonstrated in the literature not only in the volatility of asset prices, but also in the interest rate differentials, inflation rates, forward premiums, or exchange rates (Baillie, 1996). This concept has been widely examined in the literature and there is more or less a consensus that volatility is a long-memory process. However, not many studies report that volatility is a non-stationary process (e.g. when the order of integration of range is bigger than 0.5). Kellard et al. (2010) were one of the first who reported volatility as a non-stationary process.

When long memory is present in the range, the general well-known tests for cointegration are not suitable. We thus use two cointegration estimators which account for the long memory. Namely, we apply the methodology proposed by Nielsen and Shimotsu (2007) and the cointegration rank test which is part of the FCVAR methodology described by Johansen and Nielsen (2012). From the very first concept of cointegration analysis proposed by Granger (1986) or Engle and Granger (1987), cointegration has been defined for a general case where the series are integrated of order  $d$  but are cointegrated of order  $(d - b)$ , where  $d, b \in \mathbb{R}$  and  $0 < b \leq d$ . However, a complete estimation method has been developed only for the case when the series are  $I(1)$  and cointegrated of order 0 (e.g. when  $d = b = 1$ ). This method has been predominantly used in the literature for the analysis of cointegrated series. Only recently, Johansen and Nielsen (2012)

formalized the statistical methodology for the general case. Hence, its application in this thesis is a valuable contribution to current research.

The analysis is performed on four of the global stock market's principal indices, namely, the American Standard and Poor's 500, the German DAX, the Japanese NIKKEI 225, and the UK's FTSE 100 index over 2003-2012 period. These results are compared to the Czech PX Index<sup>iii</sup> over the same period. For all five indices, logarithmic transformation of daily high and low stock prices is considered. The period of interest is divided into two sub-periods with December 2007 as the breakpoint, which enables us to compare both cointegration and estimated volatility before and during the recent financial crisis.

Our findings are as follows. We find significant evidence of long memory in daily range. In most cases, the range-based volatility estimate is in the non-stationary region. There are only two exceptions: the ranges of the PX Index and the NIKKEI 225 index in the 2003-2007 period, which are in the stationary region. The estimates of the long memory parameters are, however, quite sensitive to the selected methodology. We also show that the first period (2003-2007) is rather calm, as the estimates of long memory are generally the lowest, while the second period, 2008-2012, is more turbulent. Overall, the PX Index displays the lowest estimates of the order of price range integration, and its behaviour is very similar to that of the NIKKEI 225. The ranges of S&P 500, FTSE 100 and DAX indices display, on the other hand, relatively higher orders of integration. Furthermore, we show that the unrestricted FCVAR performs better in detecting the stationarity of range than the FCVAR specification with restrictions on the cointegrating vector.

The remainder of the study is organised as follows. Chapter 2 contains motivation for using daily high and low prices and data description. Chapter 3 deals with the regular CVAR framework, describes the unit root and cointegration tests, and defines the CVAR model. Chapter 4 discusses long-memory processes as well as the formal tests for long memory and the test for the equality of integration orders. Chapter 5 describes fractional cointegration and presents the FCVAR model and the results of our analysis. Chapter 6 concludes.

---

<sup>iii</sup> PX Index was introduced in March 2006 as a merge of two indices, PX-D and PX 50. PX Index took the historical prices from the PX 50 and continued continuously.

## 2 Data description

Stock prices have been widely examined in different fields of economic research. Most studies employ the close-to-close return analysis (as standard econometric tools require stationarity of the data) and frequently ignore the information stemming from daily high and low prices. By the high price we understand the maximum price observed during the day, and the low price is the minimum price achieved during that day. These prices contain additional information about the change of direction of excess demand (Cheung, 2007). Other reasons why daily high and low prices can be of great importance are summarised by Caporin et al. (2013). First, daily high and low prices can have a role of a reference level. Market agents use these reference levels to make assumptions and predictions about future developments and use daily highs and lows as reference values. Second, daily highs and lows might work as a stop-loss indicator and may contain information about liquidity provision and the price discovery process. Third, high and low prices are more probable to correspond to ask and bid quotes, respectively, implying they may be influenced by transaction costs and other frictions (e.g. price discreteness, stale prices, or tick size). Moreover, daily high and low prices tend to react to unanticipated public announcements or other unexpected shocks.

Daily high and low stock prices are mainly valuable as a measure of dispersion, i.e. measure of a deviation from the mean. In finance, dispersion measures a degree of uncertainty, and thus risk, associated with a particular asset. Parkinson (1980) shows that a variance estimator based on close-to-close returns is a far less efficient volatility estimator than the price range (defined as a difference between daily high and low log-prices). Alizadeh et al. (2002) also show that the range-based estimator of volatility is highly statistically efficient and robust with respect to many microstructure frictions. Also, they state that range-based volatility proxies are much less contaminated by measurement error and they explain not only the autocorrelation of volatility but also the volatility of volatility. Furthermore, Corwin and Schultz (2012) argue that since daily high and low prices are mostly buy and sell trades, respectively, the price range thus represents a fundamental volatility as it reflects both the stock's variance and its bid-ask spread.

Alizadeh et al. (2002) note the range as a volatility proxy<sup>iv</sup> has a “long and colourful” history in finance (e.g. Garman and Klass, 1980; Parkinson, 1980 or Andersen and Bollerslev, 1998). More recently, Caporin et al. (2013) find evidence of long memory in the range of all 30 components of the DJIA index during the 2003-2010 period.

This thesis contains analysis of daily high and low log-prices of four major global indices over the 2003-2012 period. This sample period is quite representative, since the 10 years cover the calm and liquid markets of the last boom period as well as the recent financial crisis. We consider 4 of the world indices: the American S&P 500, the German DAX, the Japanese NIKKEI 225, and the UK's FTSE 100, available from Yahoo! Finance, which are examined and compared to the Czech stock market index (available at [stocktrading.cz](http://stocktrading.cz)). Each index is examined during the whole 10-year period from January 2003 to December 2012 as well as during two sub-periods. The first sub-period covers the pre-crisis years from January 2003 till December 2007. This break has been chosen based on the statement of the National Bureau of Economic Research (NBER), which has identified December 2007 as the peak of the pre-crisis economic activity. As rationale the NBER states that the subsequent decline in economic activity was large enough to be qualified as a recession. The second sub-period (from January 2008 till December 2012) thus covers the crisis and post-crisis periods. According to the NBER, the recession that started at the end of 2007 ended in June 2009, which is the last trough of the economic activity as well as the last business cycle reference date. This breakpoint is, however, not considered in our analysis. There were no more breakpoints announced by the NBER during our study sample.

For the United States we analyse the Standard and Poor's 500 index (S&P 500). This index is one of the most closely watched American indices, along with the DJIA or the NASDAQ Composite (NASDAQ). It is quoted in USD and contains 500 leading companies publicly traded in the US. As it covers stocks from all major American exchanges, this index is considered by many to be the best representation of the US market. FTSE 100, quoted in GBP, is a share index of the 100 largest companies by

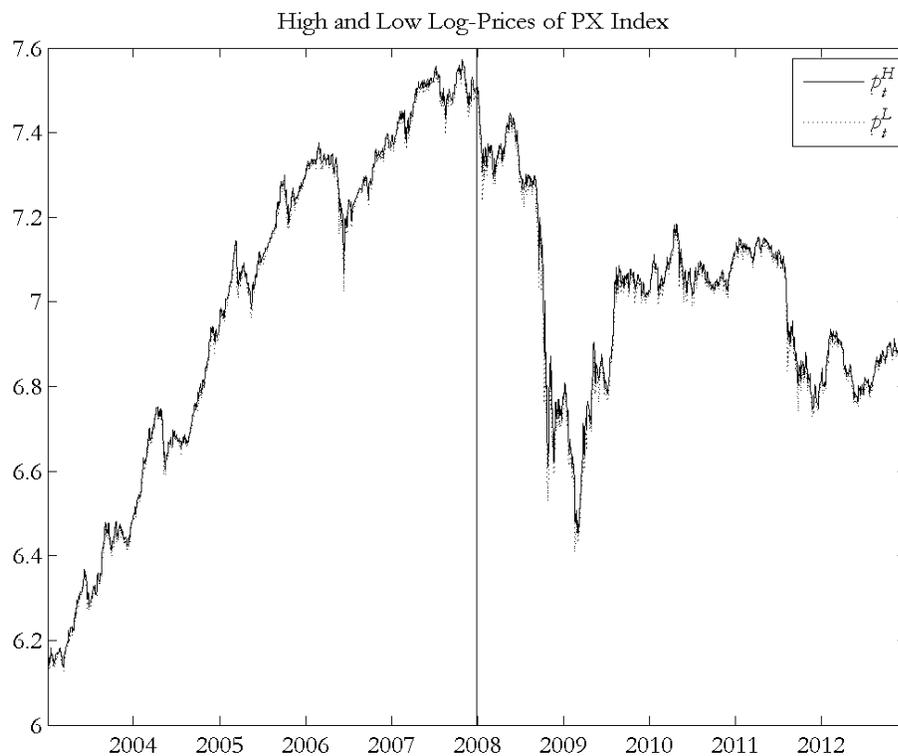
---

<sup>iv</sup> A proxy stands for an indirect measure which provides an approximation of a variable when a direct measure is absent.

market capitalization listed on the London Stock Exchange. DAX is a German stock market index composed of 30 major companies traded on the Frankfurt Stock Exchange and is quoted in EUR. NIKKEI 225 is a stock market index for the Tokyo Stock Exchange quoted in JPY and is the most extensively quoted average of Japanese equities. And finally, PX index is the official index of the major stocks trading on the Prague Stock Exchange and is quoted in CZK.

In the two following figures (Figure 1 and Figure 2) we can see the development of daily high and low log-prices of the PX index and their difference – the range. The vertical line marks the NBER peak in December 2007.

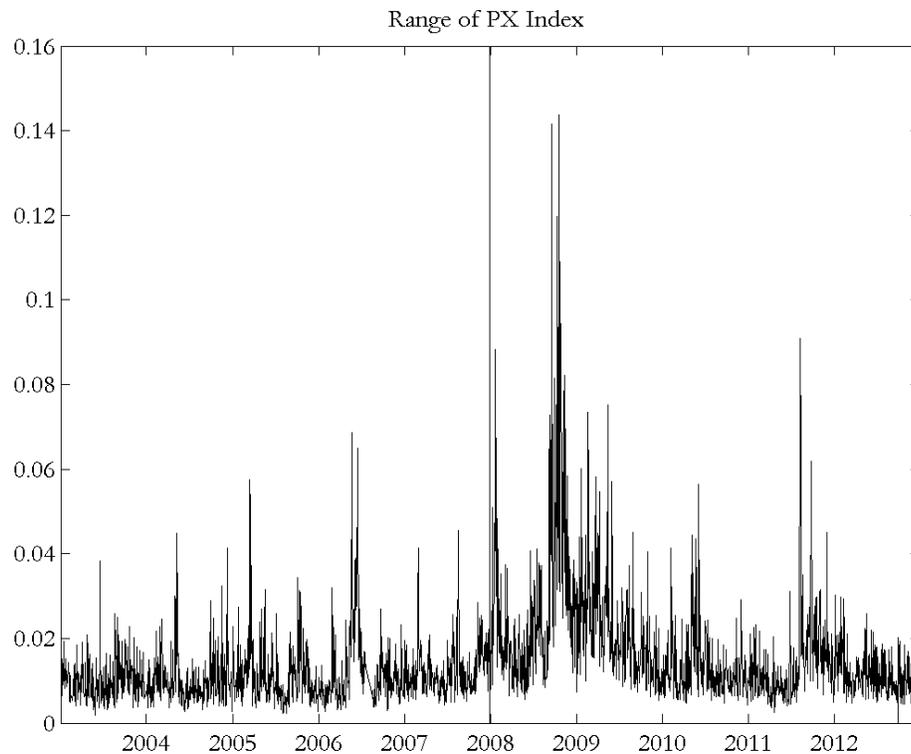
**Figure 1: High and Low Log-Prices of PX Index**



From Figure 1 we can see that the PX index was experiencing steep growth in the first period, but at the end of 2007 it was severely hit by the crisis. After a stable period from mid-2009 through mid-2011, another drop followed; however it was less dramatic than the one seen at the end of 2007. In Figure 2 we can see that range-based volatility

(measured as the difference between daily high and low log-prices) is significantly higher after the outbreak of the crisis and reaches its maximum at the end of 2008; then it gradually returns to its pre-crisis values.

**Figure 2: Range of PX Index**



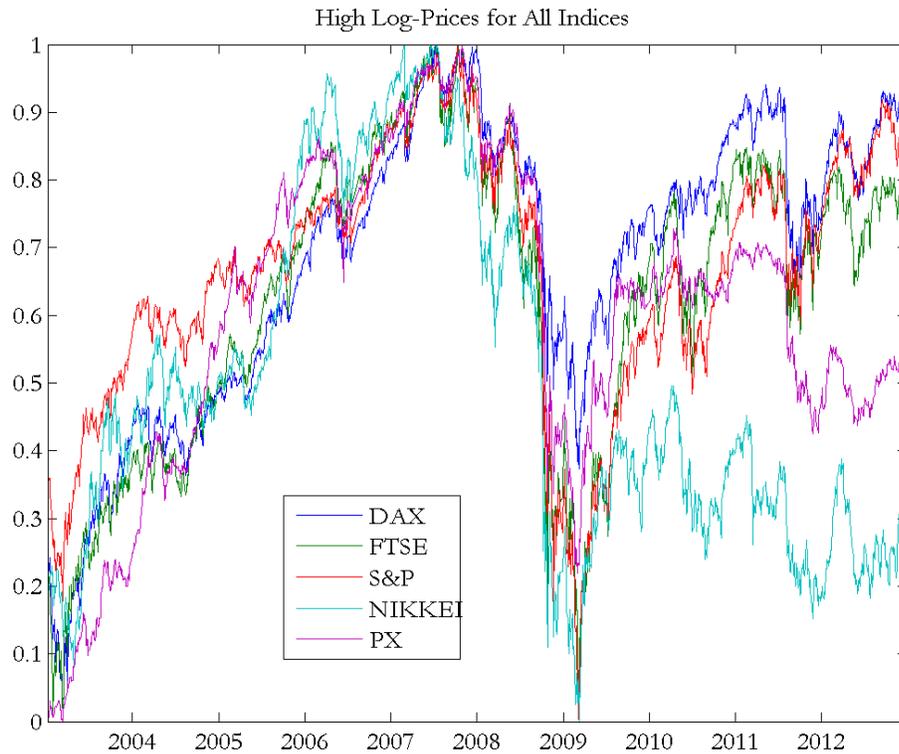
Similar figures for all remaining series can be found in the Appendix (see Figure 6 and Figure 7 for the DAX, Figure 8 and Figure 9 for the FTSE 100, Figure 10 and Figure 11 for the NIKKEI 225 and finally, Figure 12 and Figure 13 for the S&P 500). With the DAX we can see a slightly different behaviour, mainly after the crisis, when it has grown steadily since the drop in 2007, with the exception of a short period of decline at the end of 2011. This was also the case with the S&P 500. However, the FTSE 100 displayed only a very slight or no growth after the crisis, and the NIKKEI 225 has actually declined.

As for volatility, the DAX started the period with relatively high levels of volatility, which then slightly decreased during the first period, after which two peaks followed, one at the end of 2008 and another at the end of 2011. The ranges of FTSE 100 and S&P 500

exhibit a similar pattern. The range of NIKKEI 225 is quite stable in the first period, and then a peak follows at the end of 2008, similarly to the other indices. However, there is another high peak at the beginning of 2011 caused by a drop in the daily low log-price.

Comparison of high log-prices for all five indices is displayed in Figure 3.

**Figure 3: Comparison of High Log-Prices for All Indices**



We present only the development of high log-prices since the development of the low log-prices is very similar, as can be seen in the individual plots of high and low log-prices of the indices. In this figure, all high log-prices are standardised to make the comparison possible. We apply the following transformation:

$$\frac{p_t^H - \min(p_t^H)}{\max(p_t^H) - \min(p_t^H)}$$

We only include dates which are present in all series in order to maintain constant data length. Because each index comes from a different country, some observations might

otherwise be missing for some of the indices, for example due to different public holidays creating gaps in the time-series of each index. To avoid inconsistencies, even if only one index is missing the observation on a given day, the whole day is excluded from all five series.

All indices seem to be influenced by similar common factors; however, the reaction to the crisis is quite different for each of them. All indices experienced a rapid growth during the 2003-2007 period. At the end of 2007, a steep downturn followed. This figure shows that the DAX and PX indices were the two indices least hit by the crisis. The DAX was also the first index in our sample to recover from the crisis and remains the best-performing one. On the other hand, the PX index was the second-least hit by the crisis, but is the second worst-performing index as of the end of 2012. Until mid-2009 it was closely following the DAX, but then it lost pace and fell at the end of 2011. The other indices also experienced this fall, but contrary to the PX Index and NIKKEI 225 they were able to resume their previous growth. Even though NIKKEI 225 outperformed the remaining indices slightly before the crisis, this was no longer true after the crisis, when its performance has been the worst; it never quite recuperated from the crisis. Since the beginning of 2012 the S&P 500 has been catching up with DAX, and the FTSE 100 follows close behind. Overall, we can see that before the crisis, all indices were growing together, but the reactions to the crisis quite varied.

## 3 Cointegrated VAR framework

Following the intuition that daily high and low log-prices are non-stationary and share a common trend, we may use the original cointegrated VAR (CVAR) as it is used for example in Cheung (2007). Later in the study we show that the CVAR is not the most appropriate model for our analysis and thus do not provide the practical application of this model. We, however, find defining CVAR model very useful for the further analysis, which is building on this model. First, the stationarity of the data is checked by the Augmented Dickey-Fuller test (ADF test) and the Kwiatkowski–Phillips–Schmidt–Shin test (KPSS test). Then the possible cointegrating relation is discussed and the CVAR is defined.

### 3.1 Unit root tests

There is a common belief that stock market prices are non-stationary. The assumption that the best estimate of future price is today's price leads to the hypothesis of prices following a random walk. Having this in mind, we first test the data for stationarity.

Daily data on high stock prices ( $P_t^H$ ) and low stock prices ( $P_t^L$ ) is available. We employ the logarithmic transformation  $p_t^H = \log(P_t^H)$  and  $p_t^L = \log(P_t^L)$  following Cheung (2007) or Caporin et al. (2013). We define the following variable  $X_t \equiv (p_t^H, p_t^L)'$ , which will be used throughout the remainder of the study. It is a  $(2 \times 1)$  vector; however, the methodology is developed for a general case, where  $X_t$  is an  $(n \times 1)$  vector. Based on the simple visualization of the daily high and low log-prices (see Figure 3 in the previous chapter), a strong non-stationarity is apparent.

For preliminary unit-root testing, we first apply the ADF test proposed by Dickey and Fuller (1981) with the null hypothesis that each component of the series  $X_t$  has a unit root (i.e. is an  $I(1)$  process) against the alternative that each component of  $X_t$  is a stationary  $I(0)$  process, with the assumption that the data contains an ARMA structure. This test is influenced by the presence, but not by the value, of the deterministic terms in

the regression, meaning that different critical values should be applied if a constant, a trend, or both are present.

The ADF test can be formally written in the following way (e.g. Cipra, 2008):

$$\Delta X_t = c + \delta t + \Pi X_{t-1} + \sum_{j=1}^p \Psi_j \Delta X_{t-j} + \varepsilon_t,$$

$$\mathcal{H}_0: \Pi = 0, c = 0, \delta = 0,$$

$$\mathcal{H}_1: \Pi \neq 0, c \neq 0, \delta \neq 0$$

where  $c$  is an  $n$ -dimensional vector of drift coefficients,  $\delta$  is an  $n$ -dimensional vector of deterministic time trend coefficients,  $\Pi$  and  $\Psi_j$  are diagonal matrices of parameters,  $\varepsilon_t$  is an  $n$ -dimensional vector of random shocks and  $\Delta X_{t-j}$  captures the remaining serial correlation in the residuals, where  $j = 1, 2, \dots, p$ . The value of  $p$  is selected to minimize the Bayesian information criteria (BIC). For most indices and sub-periods we consider only the inclusion of a constant as the sole deterministic element; the trend is not included as it is found insignificant in the regressions.

The hypotheses then are in the following form:

$$\mathcal{H}_0: \Delta X_t = \sum_{j=1}^p \Psi_j \Delta X_{t-j} + \varepsilon_t,$$

$$\mathcal{H}_1: \Delta X_t = c + \Pi X_{t-1} + \sum_{j=1}^p \Psi_j \Delta X_{t-j} + \varepsilon_t,$$

The only exceptions are the tests of unit root for the daily high and low log-prices of DAX, FTSE 100 and S&P 500 in the first period, where we include both a constant and a trend. It is readily apparent from the visualisation of the series that a trend is present in these series in the first period. The trend is not, however, included when testing the stationarity of daily highs and lows of PX Index and NIKKEI 225 in the first period, as the trend is found insignificant in the regressions.

The hypotheses for these three particular cases (DAX, FTSE 100, S&P 500 in the first period) are in the following form:

$$\mathcal{H}_0: \Delta X_t = \sum_{j=1}^p \Psi_j \Delta X_{t-j} + \varepsilon_t,$$

$$\mathcal{H}_1: \Delta X_t = c + \delta t + \Pi X_{t-1} + \sum_{j=1}^p \Psi_j \Delta X_{t-j} + \varepsilon_t.$$

The p-values of this test are presented below in Table 1.

**Table 1: P-values for an ADF test for variables in levels and first-differences**

		$ADF_H$		$ADF_L$		$ADF_R$
		level	first-differences	level	first-differences	level
<b>S&amp;P 500</b>						
$c$	2003-2012	0.2947	0.001	0.2697	0.001	0.001
$c, t$	2003-2007	0.6369	0.001	0.2830	0.001	0.001
$c$	2008-2012	0.5718	0.001	0.5333	0.001	0.001
<b>FTSE 100</b>						
$c$	2003-2012	0.2683	0.001	0.1467	0.001	0.001
$c, t$	2003-2007	<b>0.0225</b>	-	0.0180	-	0.001
$c$	2008-2012	0.2863	0.001	0.1917	0.001	0.001
<b>DAX</b>						
$c$	2003-2012	0.3666	0.001	0.3380	0.001	0.001
$c, t$	2003-2007	<b>0.0419</b>	-	0.0329	-	0.001
$c$	2008-2012	0.3246	0.001	0.1799	0.001	0.001
<b>NIKKEI 225</b>						
$c$	2003-2012	0.4489	0.001	0.3779	0.001	0.001
$c$	2003-2007	0.4578	0.001	0.4349	0.001	0.001
$c$	2008-2012	0.1224	0.001	0.0979	0.001	0.001
<b>PX</b>						
$c$	2003-2012	0.1107	0.001	0.1162	0.001	0.001
$c$	2003-2007	0.2578	0.001	0.3114	0.001	0.001
$c$	2008-2012	0.2488	0.001	0.1570	0.001	0.001

*Note:* The letter “ $c$ ” denotes the inclusion of a constant only. Letter “ $t$ ” denotes additional inclusion of a trend for daily high and low log-prices in levels only. The reported p-value of 0.001 is the smallest reported p-value; the actual p-value is even smaller.

Table 1 presents the p-values of the ADF test for daily high ( $ADF_H$ ) and low ( $ADF_L$ ) log-prices as well as for their difference, range ( $ADF_R$ ). The stationarity of range is tested because we believe that a stationary linear combination of daily high and low stock prices exists. The ADF test is first applied on levels and if we are not rejecting the unit root, the

test is then applied on first-differences. Thus we decide whether a time series is stationary or difference-stationary.

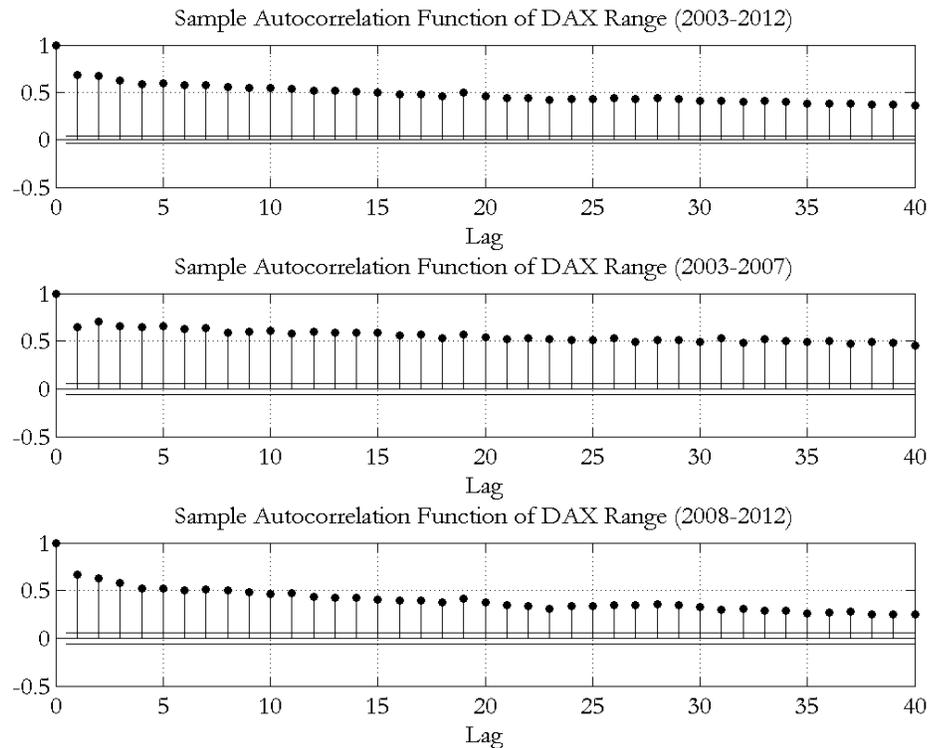
What we can see from Table 1 is that the range is always stationary and thus in this context is an  $I(0)$  process and that the daily highs and lows are mostly  $I(1)$  processes. There are, however, two exceptions. When we include the trend into the regression of DAX and FTSE 100 we find that the daily high and low log-prices in the first period are stationary on the 5% level of significance. This means that they are trend-stationary. This can be a slight problem for further application of the fractional methodology since it is a possibility that the long memory is caused by the trend. If we detrend these series, then based on the ADF test, we get a stationary process; however, the autocorrelation functions (ACF) of the detrended series decline at a very slow rate, which is definitely not consistent with a stationary  $I(0)$  process. The plots of the ACF of both detrended series in the first period are in the Appendix (see Figure 14 and Figure 15).

The shocks to a stationary  $I(0)$  process decay at an exponential rate, while no dissipation of shocks occurs in a non-stationary  $I(1)$  process. However, it may happen that based on the standard unit-root test the series is stationary, but the autocorrelation function displays a slow (non-exponential) decay, which is not consistent with a stationary process. In such a case we can assume the presence of long memory and consider the series to be fractionally integrated of order  $d$ , meaning that the  $d$ 'th difference is a stationary  $I(0)$  process, where  $d$  can be any real number. These fractionally integrated processes allow for more flexibility than the extreme assumption of a unit root and the corresponding implications for shock persistence in the system.

To illustrate the possibility of the presence of long memory in range we examine its autocorrelation functions. In Figure 4 we show the ACF of DAX range in all periods. We can see that despite the stationarity of range in all periods being implied by the ADF test, the ACF of the range is in no period declining exponentially, but rather displays a high degree of persistence. This can be an indication of long memory. On the other hand, the range of the PX Index in the first period, reported in the second panel of Figure 5 does not display a long memory, as its ACF decays exponentially. All remaining ACFs are reported in the Appendix (see Figure 16, Figure 17, and Figure 18); and all ranges display

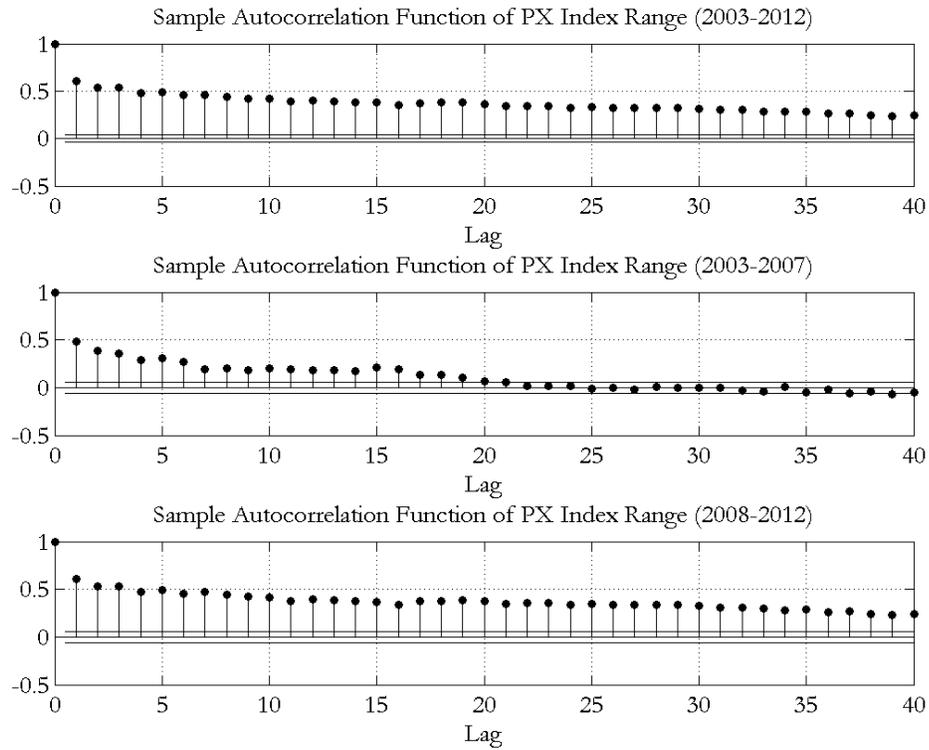
long memory. All autocorrelations are significant even after 40 lags, except for the PX Index in the first period, displayed below in Figure 5.

**Figure 4: ACF of DAX Range in All Periods**



The overall conclusion from the dependence in the autocorrelation functions is that the highest dependence is present either in the full study period of 2003-2012 or in the second period of 2008-2012. The first period, 2003-2012, displays the lowest degree of persistence, especially in the case of NIKKEI 225 and PX Index, where we might even say that long memory is not present. On the other hand, the behaviour of DAX is quite an exception. The highest degree of persistence can be seen in the first period and the lowest in the second, which is the opposite of the other series' behaviour.

**Figure 5: ACF of PX Index Range in All Periods**



Having seen that the implied stationarity of the range by the ADF test is not consistent with the inference from the autocorrelation functions we shall also apply an alternative to the ADF test, the KPSS test proposed by Kwiatkowski et al. (1992). This test is a reaction to the low power of ADF test when series is stationary but close to being a unit root. KPSS test thus has, contrary to the ADF test, stationarity as the null hypothesis. It is recommended to use both tests and take as significant such results, when the null of ADF test is rejected and at the same time the null of KPSS test cannot be rejected. In this case, the stationarity is confirmed. Or, when the null of ADF test cannot be rejected and at the same time the null of KPSS is rejected, we confirm the non-stationarity of the series.

The KPSS test can be formally written in following way (Greene, 2008):

$$X_t = c + \delta t + \Gamma Z_t + \varepsilon_t,$$

$$\mathcal{H}_0: \Gamma = 0, c \neq 0, \delta \neq 0,$$

$$\mathcal{H}_1: \Gamma \neq 0, c = 0, \delta = 0$$

where  $c$  is an  $n$ -dimensional vector of drift coefficients,  $\delta$  is an  $n$ -dimensional vector of deterministic time trend coefficients,  $\Gamma$  is a diagonal matrix of parameters,  $Z_t$  is an  $n$ -dimensional vector of  $\sum_{i=1}^t z_i^l$ ,  $t = 1, \dots, T$  for each  $l = 1, \dots, n$ , where  $z_i^l$  is an i.i.d. stationary series with zero mean and variance one,  $\varepsilon_t$  is an  $n$ -dimensional vector of random shocks.

If  $\Gamma = 0$ , then the process is stationary if  $\delta = 0$  and trend-stationary if  $\delta \neq 0$ . Following the discussion for ADF test we only include the trend into tests for daily highs and lows of DAX, FTSE 100 and S&P 500 in the first period. For all other cases we consider the inclusion of a constant only.

For the KPSS test we set the truncation lag parameter to  $\frac{3\sqrt{T}}{13}$  for the short-lag specification and to  $\frac{10\sqrt{T}}{14}$  for the long-lag specification and round it to the nearest integer,  $T$  is total number of observations for each index in a given period. We use these two specifications in order to get a more complex idea about how the stationarity changes with different number of lags used.

The test statistics under the null follows nonstandard distribution, even asymptotically. Kwiatkowski et al. (1992) use Monte Carlo simulations to tabulate asymptotic critical values for a standard set of significance levels between 0.01 and 0.1 for both the model with and without a trend. The p-values reported in Table 2 are interpolated from these tables using the MATLAB programme. Table 2 presents the p-values of the KPSS test for daily high ( $KPSS_H$ ) and low ( $KPSS_L$ ) log-prices as well as for their difference, range ( $KPSS_R$ ) for two lag length specification – denoted as short lag and long lag.

The inference from the KPSS test is the following. In case of daily high and low log-prices in all three periods the results confirm the ADF test results for both lag specifications. In the ADF test we were rejecting the unit-root null for DAX and FTSE 100 in the first period and then discussing their autocorrelation functions. In this test we, however, reject the null hypothesis of stationarity and confirm that even these series are not stationary. We can thus conclude that daily high and low log-prices are non-stationary.

The situation is, as expected, more complicated for the range. The ADF test was overall suggesting that the range was stationary. However, here in most cases we are rejecting the null hypothesis of stationarity in favor of the alternative of non-stationarity. Only when we use the long-lag parameter specification for ranges of DAX and NIKKEI 225 in the 2003-2012 period and PX Index in the 2003-2007 period, we are not rejecting the null. The ACF of the PX Index identified the first period as rather stationary, but the stationarity of ranges of the two other indices is not supported by the ACF functions, but is by the ADF test. This conclusion is achieved only for the long-lag parameter specification in the KPSS test for these indices. When the short-lag specification is used we are again rejecting the stationarity.

**Table 2: P-values for KPSS test for variables in levels with two lag specifications**

		$KPSS_H$		$KPSS_L$		$KPSS_R$	
		Short lag	Long lag	Short lag	Long lag	Short lag	Long lag
<b>S&amp;P 500</b>							
$c$	2003-2012	0.01	0.01	0.01	0.01	0.01	0.0131
$c, t$	2003-2007	0.01	0.01	0.01	0.01	0.01	0.01
$c$	2008-2012	0.01	0.01	0.01	0.01	0.01	0.01
<b>FTSE 100</b>							
$c$	2003-2012	0.01	0.01	0.01	0.01	0.01	0.0143
$c, t$	2003-2007	0.01	0.01	0.01	0.01	0.01	0.001
$c$	2008-2012	0.01	0.01	0.01	0.01	0.01	0.001
<b>DAX</b>							
$c$	2003-2012	0.01	0.01	0.01	0.01	0.01	<b>0.0784</b>
$c, t$	2003-2007	0.01	0.01	0.01	0.0188	0.01	0.001
$c$	2008-2012	0.01	0.01	0.01	0.01	0.01	0.001
<b>NIKKEI 225</b>							
$c$	2003-2012	0.01	0.01	0.01	0.01	0.01	<b>0.0943</b>
$c$	2003-2007	0.01	0.01	0.01	0.01	0.01	0.01
$c$	2008-2012	0.01	0.01	0.01	0.01	0.01	0.01
<b>PX</b>							
$c$	2003-2012	0.01	0.01	0.01	0.01	0.01	0.01
$c$	2003-2007	0.01	0.01	0.01	0.01	0.01	<b>0.0772</b>
$c$	2008-2012	0.01	0.01	0.01	0.01	0.01	0.01

*Note:* The letter “ $c$ ” denotes the inclusion of a constant only. Letter “ $t$ ” denotes additional inclusion of a trend for daily high and low log-prices in levels only. P-value 0.01 is the minimal reported p-value; the actual p-value is even smaller.

For the range, we are now in a situation where ADF test rejects the null of unit root and KPSS test mostly rejects the null of stationarity. This situation is thus inconclusive, which confirms our initial hypothesis that these tests are not capable of properly determining whether the range is stationary or not and that a long memory framework should be applied instead.

## 3.2 Cointegration tests

Based on the ADF and KPSS test results and the inspection of the autocorrelation functions of the detrended DAX and FTSE 100 in the first period (which were identified as trend-stationary by the ADF test and non-stationary by the KPSS test) we can conclude that all series are non-stationary. Engle and Granger (1987) state that a linear combination of two or more non-stationary series can be stationary. If this is the case, we say that the non-stationary series are cointegrated and the stationary linear combination is called a cointegrating equation. This equation is then interpreted as a long-run equilibrium relationship among the variables. If daily highs and lows are cointegrated, it means that in the short term they may diverge, but in the long term they have an embedded convergence path. In the previous section we have tested the stationarity of range, which is a linear combination of two non-stationary series.

The stationarity of range was confirmed by the ADF test but rejected by the KPSS test. We can postulate that daily high and low log-prices are cointegrated in the classical framework only if their linear combination, the range, is stationary. As we were not able to draw a decisive conclusion about the stationarity of the range, this is thus an indication that the classical notion of cointegration is not suitable for the data in question.

In order to be able to apply a proper model which takes the cointegration in the long memory framework into account, we first start by defining the original Vector Autoregression (VAR) model. We then extend the VAR model with cointegration to the cointegrated VAR model. After discussing the long memory processes in Section 4 we then provide the final model, the fractionally cointegrated VAR model, in Section 5.

The original VAR model requires the series to be stationary. If the series are non-stationary, the usual method of dealing with this issue is to use first differences. Even though this transition is statistically correct, it omits the long-run relationship between the original (non-differenced) variables if the variables are cointegrated. In the presence of cointegration the basic VAR is transformed into the cointegrated VAR (CVAR) model, which enables us to describe the short-term relationships between the growth components and also corrects for short-term deviations from the long-run equilibrium.

Consider the VAR model of order  $p$ , (see e.g. Cipra, 2008)

$$X_t = \mu + \sum_{i=1}^p \Pi_i X_{t-i} + \varepsilon_t, \quad t = 1, \dots, T,$$

where  $X_t$  is an  $(n \times 1)$  vector, in our case it is containing only the series of daily high and low logarithmic stock prices,  $\mu$  is an  $n$ -dimensional vector of constants,  $\Pi_i$  is an  $(n \times p)$  matrix of autoregressive coefficients,  $i = 1, 2, \dots, p$ , and  $\varepsilon_t$  is an  $n$ -dimensional *i. i. d.*  $(0, \Omega)$  vector of random shocks, where  $\Omega$  is a positive-definite variance matrix. The VAR model is frequently used by econometricians, as it provides a simple tool for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model. Among other advantages, all variables are considered endogenous, so there is no need to specify which variables are to be considered endogenous and which exogenous. Each variable depends on its own lagged values, white noise, and on the lagged values of the other variables included in the model.

We can rewrite this VAR model as

$$\Delta X_t = \mu + \Pi X_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t, \quad t = 1, \dots, T,$$

where  $\Pi = \sum_{i=1}^{p-1} \Pi_i - I$  and  $\Gamma_i = -\sum_{j=i+1}^p \Pi_j$  for  $i = 1, 2, \dots, p-1$ .

Applying the Granger representation theorem (Engle and Granger, 1987), if the coefficient matrix  $\Pi$  of dimension  $(n \times n)$  is of reduced rank  $r < n$ , then there exist matrices  $\alpha$  and  $\beta$ , both of dimensions  $(n \times r)$  with a full column rank  $r$ , such that

$\Pi = \alpha\beta'$  and  $\beta'X_t$  is a stationary  $I(0)$  process;  $r$  is then the cointegration rank representing the number of cointegrating relations. Furthermore, each of the columns of  $\beta$  is a cointegrating vector and  $\alpha$  is a vector of adjustment parameters.

In our particular case where  $n = 2$ , we are mostly interested in the situation  $r = 1$ . In this case, out of the two eigenvalues of matrix  $\Pi$  exactly one eigenvalue is nonzero, meaning that there is exactly one unit root in the system of two time series.

To formally test for cointegration the Johansen procedure is commonly used. This procedure consists of estimating the  $\Pi$  matrix by unrestricted VAR and testing whether we can reject the restrictions implied by the reduced rank of  $\Pi$ .

The Johansen procedure (Johansen, 1991) tests for significant canonical correlations measuring the partial dependency between the  $n$ -dimensional vectors  $\Delta X_t$  and  $X_{t-1}$  when  $\Delta X_{t-1}, \dots, \Delta X_{t-p+1}$  are fixed. These canonical correlations are the squared roots of the eigenvalues  $\lambda_1, \dots, \lambda_n$  of a positive semi-definite matrix closely related to  $\Pi$ . In particular, the number of positive values among  $\lambda_1, \dots, \lambda_r$  is equal to the rank  $r$  of the matrix  $\Pi$ . There are two tests which can be applied as a preliminary analysis of cointegration (for more details see Johansen, 1991): the Johansen Trace Test and the Johansen Maximum Eigenvalue Test. These tests are essentially based on likelihood ratios (LR), whose critical values are simulated, and test the null of  $\lambda$  being zero.

Consider the Johansen Trace Test with the test statistic:

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i),$$

which has the joint null hypothesis of no more than  $r$  cointegrating relations, against the alternative of more than  $r$  cointegrating relations. The test is performed successively for  $r = 0, 1, \dots, n - 1$ . (In our case where  $n = 2$ , it would only be performed for  $r = 0, 1$ .)

The Johansen Maximum Eigenvalue Test has the test statistic:

$$\lambda_{max}(r) = -T \ln(1 - \hat{\lambda}_{r+1}),$$

with the null hypothesis of  $r$  cointegrating relations against the alternative of  $(r + 1)$  cointegrating relations. The test is conducted in the same manner as the Johansen Trace Test.

We do not provide the results of the cointegration tests here, because these tests are not applicable for fractionally cointegrated variables. The definition of fractional cointegration is provided later in this study. However, the presence of cointegration is quite foreseeable since the daily high and low log-prices are driven by common trends and are the prices of one underlying index.

The results of ADF tests of range indicate that there might exist a combination of two non-stationary variables (the daily high and low stock prices) which is stationary. The stationarity of range was however questioned by the KPSS test. This supports the fact that the classical notion of cointegration is not sufficient here.

In the preliminary analysis, we have tested the presence of cointegration and we found the cointegrating vector was mostly  $(1, -1)$ . These findings are, however, only valid as long as we do not consider the presence of long memory (e.g. the presence of fractional integration/cointegration). Thus a more general approach, which takes long memory into account, shall be discussed later and the results of this generalized test will be provided.

### 3.3 CVAR model

When the variables are non-stationary and cointegrated, one can apply the cointegrated VAR model (or the vector error correction model – VECM), in which information about the long-run equilibrium is not lost. For an  $n$ -dimensional non-stationary time series  $X_t$  we define the CVAR in the following form (see e.g. Johansen and Nielsen, 2010):

$$\Delta X_t = \alpha(\beta' X_{t-1} + \rho') + \sum_{i=1}^p \Gamma_i \Delta X_{t-i} + \varepsilon_t, \quad t = 1, \dots, T, \quad (1)$$

where  $\Delta X_{t-i} = X_{t-i} - X_{t-i-1}$  and where  $\varepsilon_t$  is an  $n$ -dimensional vector of *i.i.d.*  $(0, \Omega)$  random shocks, where  $\Omega$  is a positive-definite variance matrix.  $\alpha$  and  $\beta$  are  $(n \times r)$  matrices with  $0 \leq r \leq n$ . The model is used to describe the long-run economic relations

given by the stationary combinations  $\beta'X_t$ . The matrices  $(\Gamma_1, \dots, \Gamma_p)$ , each with dimension  $(n \times n)$ , govern the short-run dynamics. The parameter  $\rho$  is the restricted constant term (since the constant term in the model is restricted to be of the form  $\mu = \alpha\rho'$ ) and is interpreted as the mean level of the long-run equilibrium.

We do not estimate the cointegrated VAR, because this methodology is not suitable for fractionally cointegrated processes with an error correction term which is potentially non-stationary (as indicated by the autocorrelation functions and the KPSS test). This model needs some generalization in order to be able to take into account the fractional cointegration of daily high and low log-prices and the potential long memory of their linear combination. This generalization and the definition of the long memory process are described in the following chapters.

## 4 Long memory processes

In this chapter we analyse the so called long memory processes, for whose description the fractional models are very useful. So far, we have assumed that a series can be integrated of order 0 or 1; it is, however, possible that the series is instead integrated of order  $d$ , where  $d$  can be any real number. When  $d < 0$ , we say that the process is anti-persistent. When  $d > 0$  we refer to the process as having long memory. The process is said to be stationary when  $d < 1/2$  and non-stationary when  $d > 1/2$ .

Long-memory models have been used by the natural sciences (specifically, hydrology and climatology) since 1950's. They drew the attention of econometricians around 1980 when Granger (1980) and Granger and Joyeux (1980) developed the autoregressive fractionally integrated moving average (ARFIMA) and Geweke and Porter-Hudak (1983) proposed a technique for estimating the  $d$  parameter. Later on, the fractionally integrated GARCH( $p, d, q$ ) (FIGARCH) was proposed by Baillie (1996). The long memory processes are characterized by slowly decaying autocorrelation functions, with the decay rate being hyperbolic. This decay of the ACF is, for example, slower than the exponential decay associated with ARMA models. The persistence in the ACF is consistent neither with an  $I(1)$  process nor with an  $I(0)$  process. If the series displays long memory, the ACF may appear to be non-stationary while at the same time the differenced series appears over-differenced.

Substantial evidence of the presence of long memory has been demonstrated in the literature on financial data such as the volatility of asset prices, interest rate differentials, inflation rates, forward premiums, or exchange rates (Baillie, 1996). Evidence of long-range dependence in asset price volatility was also found for example by Ding, Granger and Engel (1993), Andersen and Bollerslev (1997), Breidt et al. (1998), Kellard et al. (2010) or Garvey and Gallagher (2012). Breidt et al. (1998) show that this type of persistence cannot be appropriately modelled by the classic autoregressive conditional heteroskedastic models (ARCH) and their commonly used variations the generalized ARCH (GARCH) or exponential generalized ARCH (EGARCH). These models are short-memory models and the only way to incorporate long-memory persistence is by

unit root. Both ARFIMA and FIGARCH thus allow  $d$  to take any value between 0 and 1. In this study, we are dealing with a multivariate case, which is why we later on describe the fractionally cointegrated vector autoregressive (FCVAR) model, which is capable of capturing both the long memory and the fact that the series are cointegrated.

The literature also mentions some challenges to the long-memory findings. Lobato and Savin (1997) note that the presence of long memory can be spurious due to the effect of aggregation. This means that the individual stocks may not exhibit long memory, but the index constructed from them does due to aggregation. Granger and Hyung (2004) propose structural breaks as another challenge to the presence of long memory. They show that a structural break in volatility can reproduce a slowly decaying ACF and other properties of long-memory processes.

So far, we have not been able to decide whether daily range is stationary or not. The ADF test was suggesting stationarity of the range, whilst the KPSS test was rejecting the stationarity of the range. Moreover, the high dependency in most of the ACF functions of range for the five indices rather indicated the non-stationarity. All in all, this evidence was pointing towards the presence of long-memory in the range.

We can define a long-memory process as in Brockwell and Davis (1991): A weakly stationary process has long memory if its autocorrelation function ACF  $\rho(\cdot)$  has a hyperbolic decay,

$$\rho(h) \sim C \cdot h^{2d-1} \text{ as } h \rightarrow \infty,$$

where  $C \neq 0, d < 0.5$ .

In contrast, a weakly stationary process has short memory if its autocorrelation function ACF  $\rho(\cdot)$  is geometrically bounded.

$$|\rho(h)| \leq C \cdot r^{|h|} \text{ for some } C > 0, 0 < r < 1.$$

The evidence of long-range dependence in asset price volatility means that a shock affecting the evolution of volatility will persist for a long time.

In the literature, for example, Caporin et al. (2013) analyse the degree of integration of daily high and low stock prices and of their difference (range) in a fractional or long-memory framework as opposed to the classical cointegration, which is considered only in the “ $I(1)/I(0)$ ” context. In the previous section, we have discussed that the traditional notion of cointegration, requiring the cointegrating relation to be stationary, was not consistent with the existence of long memory in the range. We shall thus now characterize the range by a fractional degree of integration.

We have already mentioned that an  $(n \times 1)$  vector  $X_t$  with  $I(1)$  elements is cointegrated if there exists a linear combination  $\beta'X_t$  which is an  $I(0)$  process. Robinson and Yajima (2002) note that the possibility of existence of a long-run, stable relationship among non-stationary series  $X_t$  does not depend on whether or not the series are  $I(1)$ . There can exist a cointegration with  $\beta'X_t$  which is stationary but not necessarily  $I(0)$ . This need for a flexible approach is solved by considering an  $I(d)$  series, i.e. series integrated of order  $d$ , with a real-valued  $d$ .

According to Robinson and Yajima (2002), for  $d < \frac{1}{2}$ , we say that  $X_t$  is  $I(d)$  if

$$u_t = (1 - L)^d X_t$$

is  $I(0)$ , where  $L$  is the lag operator and

$$(1 - L)^d = \sum_{i=0}^{\infty} \frac{\Gamma(i - d)}{\Gamma(-d)\Gamma(i + 1)} L^i,$$

where  $\Gamma(\cdot)$  is the *gamma* function defined as  $\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$  for  $a > 0$ . It can be extended for all but non-positive integers (with a proper use of complex analysis), while for the non-positive integers we are left with singularities. We then say that  $X_t$  is fractionally integrated of order  $d < 1/2$  and is covariance-stationary.

For  $d \geq \frac{1}{2}$  we define a non-stationary  $I(d)$  series  $X_t = (1 - L)^{-d} u_t I\{t \geq 1\}$ ,  $t = 0, \pm 1, \pm 2, \dots$ , where  $I(\cdot)$  is the indicator function.

Shimotsu and Phillips (2005) show that the representation can also be rewritten in the following form, giving a valid linear representation of  $X_t$ :

$$u_t I\{t \geq 1\} = (1 - L)^d X_t = \sum_{k=0}^{t-1} \frac{(-d)_k}{k!} X_{t-k},$$

so

$$X_t = (1 - L)^{-d} u_t I\{t \geq 1\} = \sum_{k=0}^{t-1} \frac{(d)_k}{k!} u_{t-k},$$

where  $(d)_k = \frac{\Gamma(d+k)}{\Gamma(d)} = d(d+1) \dots (d+k-1)$  is the forward factorial function.

If  $d > 0$  we say that the process has long memory, and if  $d < 0$  we say that the process is anti-persistent. One can easily see that if  $d = 1$  then the process represents a random walk and if  $d = 0$  the process is stationary. The parameter  $d$  is called the fractional differencing parameter, fractional degree of persistence or fractional order of integration and describes the memory properties of  $X_t$  (Robinson and Yajima, 2002).

There are several methods of estimating the fractional degree of persistence  $d$ . It can be estimated within the FCVAR framework as it is described in Sub-section 5.2.1, as well as separately. Here we present two semi-parametric methods, a univariate exact local Whittle estimator (ELW) proposed by Shimotsu and Phillips (2005) and a procedure proposed by Geweke and Porter-Hudak (GPH) (1983). Since we are estimating the order of integration for two series (daily high and low log-prices) we also describe and conduct a test for the equality of integration orders proposed by Nielsen and Shimotsu (2007), as this is a condition for further FCVAR analysis.

## 4.1 Exact local Whittle estimator

The exact local Whittle estimator proposed by Shimotsu and Phillips (2005) makes no assumptions about the presence of cointegration and is consistent both when cointegration is present and when it is absent. This estimator is also applicable to both the stationary and non-stationary case, which removes the limitation of the original tests proposed by Robinson and Yajima (2002) which only allow stationary data. It also improves the original local Whittle (LW) estimator whose asymptotic theory was discontinuous for  $d = \frac{3}{4}$  and again for  $d = 1$ .

We consider the fractional process  $X_t$  generated by the model

$$(1 - L)^d X_t = u_t, \quad t = 1, 2, \dots, T,$$

where  $u_t$  is stationary with zero mean and spectral density  $f_u(\lambda)$ .

Next we define the discrete Fourier transform of a time series  $x_t$  evaluated at frequency  $\lambda$  as

$$\omega_x(\lambda_j) = (2\pi T)^{-\frac{1}{2}} \sum_{t=1}^T x_t e^{it\lambda},$$

and the periodogram of a time series  $x_t$  computed at the Fourier frequency  $\lambda_j$  as

$$I_x(\lambda_j) = |\omega_x(\lambda_j)|^2, \quad \lambda_j = \frac{2\pi j}{T} \quad (j = 1, \dots, m < T).$$

Following Shimotsu and Phillips (2005), we minimize the (negative) Whittle likelihood function of stationary innovations  $u_t$  based on frequencies up to  $\lambda_m$  and up to scale multiplication. By applying the local approximation of spectral density,  $f_u(\lambda) \sim G$  for  $\lambda \rightarrow 0$ , where  $G$  is a finite and non-zero matrix with strictly positive diagonal elements, and an algebraic transformation, the likelihood function  $Q_m(G, d)$  becomes data-dependent:

$$Q_m(G, d) = \frac{1}{m} \sum_{j=1}^m \left[ \log(G \lambda_j^{-2d}) + \frac{1}{G} I_{\Delta^d x}(\lambda_j) \right],$$

where  $I_{\Delta^d x}(\lambda_j)$  is the periodogram of  $\Delta^d X_t = (1 - L)^d X_t$  at the Fourier frequency  $\lambda_j$  of the fractionally differenced series  $u_t$ . The number  $m = m(T)$  denotes a bandwidth parameter determining the number of periodogram ordinates used in the equation. In our estimation we use two bandwidth parameter specifications;  $m = T^{0.5}$  and  $m = T^{0.6}$ , which are the same bandwidth parameters as those chosen by Nielsen and Shimotsu (2007) for their empirical application.

We obtain the estimates of  $d$  and  $G$  by minimizing  $Q_m(G, d)$ , hence,

$$(\hat{G}, \hat{d}) = \arg \min_{G \in (0, \infty), d \in [\Delta_1, \Delta_2]} Q_m(G, d),$$

where  $\Delta_1$  and  $\Delta_2$  are the lower and upper bounds, respectively, of the admissible values of  $d$  such that  $-\infty < \Delta_1 < \Delta_2 < \infty$ .

Furthermore, if we concentrate  $Q_m(G, d)$  with respect to  $G$ , the Exact Local Whittle estimator of  $d$  reduces to

$$\hat{d}_i = \arg \min_{d \in [\Delta_1, \Delta_2]} R_i(d),$$

where

$$R_i(d) = \log \hat{G}_{ii}(d) - 2d \frac{1}{m} \sum_{j=1}^m \log \lambda_j \text{ and } \hat{G}_{ii}(d) = \frac{1}{m} I_{\Delta^d x_i}(\lambda_j).$$

In our case,  $i = H, L, i = R$ , respectively. Shimotsu and Phillips (2005) show that this exact local Whittle estimator is consistent and has a  $N(0, \frac{1}{4})$  limit distribution for all values of  $d$  if the optimization covers an interval of a width lesser than  $\frac{9}{2}$  and if the initial values of the process are assumed to be known. The word “exact” in the estimator’s name is used to distinguish this estimator from the original LW estimator. It is “exact” because the estimator relies on an exact algebraic manipulation, rather than on an approximation as the conventional LW estimator does.

In Table 3 we present the estimates of the long memory parameter  $d$  based on the 2-step ELW estimator. For the possibly trend-stationary series (DAX in the first period and FTSE 100 in the first period) we also include the results of the 2-step ELW estimator with prior detrending. The section “Bandwidths” contains three bandwidth figures: the first is the total number of observations ( $T$ ) of each index in each period, and the two remaining figures are the bandwidth parameters used ( $T^{0.5}$  and  $T^{0.6}$ ). Then, we report the estimates of three long memory parameters for each bandwidth; the first one is for the series of high log-prices ( $\hat{d}_H$ ), the second one is for the series of low log-prices ( $\hat{d}_L$ ), and the third one is for their difference, the range ( $\hat{d}_R$ ). The estimates of the long memory parameters are provided for both bandwidths considered in order to show that the choice of bandwidth may, to a certain extent, influence the results. It is also possible to calculate the standard errors of the estimates, which are defined as  $(4m)^{-1/2}$ , where  $m$  is equal to either  $T^{0.5}$  or  $T^{0.6}$ , depending on the bandwidth chosen. This means that for each

bandwidth the standard errors are the same for all three integration orders ( $\hat{d}_H, \hat{d}_L, \hat{d}_R$ ), as it only depends on the number of observations used for the calculation of the estimates.

**Table 3: Exact local Whittle estimator of the fractional degree of integration**

	Bandwidths			$ELW_{m=T^{0.5}}$				$ELW_{m=T^{0.6}}$			
	$T$	$T^{0.5}$	$T^{0.6}$	$\hat{d}_H$	$\hat{d}_L$	$\hat{d}_R$	s.e.	$\hat{d}_H$	$\hat{d}_L$	$\hat{d}_R$	s.e.
<b>S&amp;P 500</b>											
2003-2012	2517	50	109	1.0972	1.0709	<b>0.7626</b>	0.0707	1.0442	1.0036	<b>0.6216</b>	0.0479
2003-2007	1258	35	72	0.9094	0.8547	<b>0.5870</b>	0.0845	0.9265	0.8932	<b>0.5740</b>	0.0589
2008-2012	1259	35	72	1.1428	1.0853	<b>0.6643</b>	0.0845	1.0512	1.0130	<b>0.6867</b>	0.0589
<b>FTSE 100</b>											
2003-2012	2526	50	110	0.9990	0.9763	<b>0.6575</b>	0.0707	0.9673	0.9369	<b>0.6324</b>	0.0477
2003-2007	1264	35	72	0.8518	0.8182	<b>0.6220</b>	0.0845	0.8539	0.8357	<b>0.6028</b>	0.0589
detrending				0.5765	0.5877			0.8138	0.8011		
2008-2012	1262	35	72	1.0377	1.0026	<b>0.6278</b>	0.0845	0.9799	0.9568	<b>0.6169</b>	0.0589
<b>DAX</b>											
2003-2012	2556	50	110	1.0221	1.0055	<b>0.5884</b>	0.0707	1.0434	1.0058	<b>0.5887</b>	0.0477
2003-2007	1274	35	72	0.9711	0.9158	<b>0.6142</b>	0.0845	0.9817	0.9461	<b>0.5968</b>	0.0589
detrending				0.9076	0.8451			0.9630	0.9289		
2008-2012	1282	35	73	1.0913	1.0766	<b>0.7001</b>	0.0845	1.0237	1.0057	<b>0.6229</b>	0.0585
<b>NIKKEI 225</b>											
2003-2012	2543	49	108	1.0572	1.0372	<b>0.4833</b>	0.0714	1.0564	1.0428	<b>0.6493</b>	0.0481
2003-2007	1229	35	71	0.9909	0.9630	<b>0.4955</b>	0.0845	0.9755	0.9447	<b>0.3916</b>	0.0593
2008-2012	1224	34	71	1.1082	1.1018	<b>0.6131</b>	0.0857	1.0697	1.0544	<b>0.6044</b>	0.0593
<b>PX</b>											
2003-2012	2494	49	109	1.0990	1.0876	<b>0.5158</b>	0.0714	1.1659	1.1367	<b>0.5197</b>	0.0479
2003-2007	1235	35	71	0.8911	0.8601	<b>0.2811</b>	0.0845	1.0174	0.9659	<b>0.4455</b>	0.0593
2008-2012	1259	35	72	1.2337	1.2189	<b>0.5723</b>	0.0845	1.1216	1.0986	<b>0.5165</b>	0.0589

The results in Table 3 support two our initial hypotheses. First, that daily highs and lows are not stationary. The order of integration is generally close to 1; however, in the full period and in the second sub-period unitary integration is substantially exceeded. Only in 25 cases out of 60 is the integration order of daily highs and lows smaller than 1. The lowest values of integration of daily highs and lows are achieved in the first period. In the case of the DAX and the FTSE 100, which, based on the ADF test, are assumed to be trend-stationary in the first period, we can see that even if the series are detrended first, they remain non-stationary as their order of integration is greater than 0.5. For the smaller bandwidth ( $m = T^{0.5}$ ), the level of integration of FTSE 100 is closer to the marginal value

of 0.5, but when the higher bandwidth is used the order of integration is around 0.8. In the case of DAX, both bandwidth specifications suggest non-stationarity of the detrended series. We thus conclude that based on the ELW estimates both series are non-stationary in all periods rather than trend-stationary. This confirms the results from the KPSS test.

Our second hypothesis was that the difference between high and low log-prices (range) is not stationary and displays long memory. We can see that  $\hat{d}_R$  is significantly different from zero and in most cases it is in the region of non-stationarity (i.e.,  $\hat{d}_R > 0.5$ ). Moreover, the order of integration of range is higher in the second period than in the first one, implying that the second period is more non-stationary. This supports our initial findings from the visualisation of the autocorrelations of range (except for the DAX).

The PX Index is quite different from the other indices. In the first period, the range is in the stationary region; this is also true for the Japanese NIKKEI 225, but not for the other indices, as their range is not stationary in any of the three periods. We have also noted this weak dependence in the range of the PX Index and NIKKEI 225 in the visualisation of the ACF functions in Section 3.1. Moreover, the range of NIKKEI 225 displays the highest variation in the estimates of the order of integration. For example, when the  $m = T^{0.5}$  bandwidth is used the range in the full period is in the stationary region, but when the  $m = T^{0.6}$  bandwidth is used the range is in the non-stationary region. The only conclusion drawn from the ELW estimator which differs from the conclusions drawn from the autocorrelation functions is in the case of the DAX index. The ELW estimator suggests that the highest dependence is present in the second period, but based on the ACF function of range this period should display the lowest dependence.

## 4.2 GPH estimator

The above-discussed procedure by Shimotsu and Phillips (2005) provides estimators for the fractional degree of persistence of daily high ( $\hat{d}_H$ ) and low ( $\hat{d}_L$ ) log-prices and their difference – range ( $\hat{d}_R$ ). Similar procedure used to obtain estimates of the long-memory parameters is proposed by Geweke and Porter-Hudak (GPH) (1983). They prove that

provided a large enough sample (more than 50 observations), the asymptotic theory is reliable.

GPH (1983) again consider the fractional process  $(1 - L)^d X_t = u_t$ , where  $u_t$  is a stationary linear process with spectral density  $f_u(\lambda)$ . The spectral density function is considered to be finite, bounded away from zero and continuous on the interval  $(-\pi, \pi)$ . They also define the spectral density function of  $\{X_t\}$  in the following form:  $f(\lambda) = \frac{\sigma^2}{2T} \{4 \sin^2(\lambda)\}^{-d} f_u(\lambda)$ .

The estimator is based on the following linear regression (taken from Garvey and Gallagher, 2012) of the log periodogram on a deterministic regressor:

$$\ln I(\lambda_j) = c - d \ln\left(4 \sin^2\left(\frac{\lambda_j}{2}\right)\right) + \epsilon(\lambda_j),$$

where  $I_x(\lambda_j) = |\omega_x(\lambda_j)|^2$  is the periodogram of the data and  $\lambda_j$ ,  $j = 0, \dots, m < T$  is the Fourier frequency close to zero,  $c$  is a constant term and  $\epsilon$  denotes the residuals. This estimator is semi-parametric, only assuming near the zero frequencies and treating spectral density non-parametrically when away from the origin.

The least-squares estimator of the slope parameter in the regression ( $\hat{d}$ ) formed by only the lowest frequency ordinates  $\lambda_j = \frac{2\pi j}{T}$  ( $j = 1, \dots, m < T$ ) of the log periodogram  $\ln I(\lambda_j)$  is then the estimator of our long memory parameter. The purpose of this estimator is to capture any fractional structure in the lower frequencies, i.e. it is looking for the correlation structure that is  $I(d)$ , where  $d$  is to be estimated. As noted by Kanzler (1998), the lower frequencies are covered exclusively, so no short-memory process specification is needed. The estimation only makes use of a relatively small number of the lowest frequencies; thus, only a restricted number of observations is included into the regression. If too many frequencies are considered, high-frequency cycles can be introduced, which would disrupt the analysis. On the other hand, if we do not include enough frequencies, the reliability of the OLS regression's results may become suspect. In the method proposed by Kanzler (1998), the first  $T^{-1}$  frequencies are excluded from the regression and all the remaining lower frequencies are included, up to the ( $m = T^{0.5}$ )-th frequency.

Kanzler (1998) argues that 0.5 is the standard value used in the literature. In our analysis we consider two values of this bandwidth parameter,  $m = T^{0.5}$  and  $m = T^{0.6}$ , in the same manner as in the ELW estimation. The distribution of the long memory parameter is asymptotically normal with standard errors  $\frac{\pi}{(6T)^{\frac{1}{2}}}$  provided certain conditions are fulfilled (for more details see GPH, 1983, pp. 226-227). GPH (1983) also demonstrate that this estimator of the long memory parameter is biased and inefficient when the regression residuals are substantially autocorrelated.

In Table 4 we present the results obtained with the GPH estimator.

**Table 4: GPH estimator of the fractional degree of integration**

	Bandwidths			$GPH_{m=T^{0.5}}$			$GPH_{m=T^{0.6}}$			s.e.
	$T$	$T^{0.5}$	$T^{0.6}$	$\hat{d}_H$	$\hat{d}_L$	$\hat{d}_R$	$\hat{d}_H$	$\hat{d}_L$	$\hat{d}_R$	
<b>S&amp;P 500</b>										
2003-2012	2517	50	109	1.0261	1.0202	<b>0.7646</b>	1.0819	1.0593	<b>0.6742</b>	0.0256
2003-2007	1258	35	72	1.0921	1.0819	<b>0.5204</b>	1.0914	1.0945	<b>0.6778</b>	0.0362
2008-2012	1259	35	72	1.2206	1.2483	<b>0.7526</b>	1.0390	1.0398	<b>0.7211</b>	0.0361
<b>FTSE 100</b>										
2003-2012	2526	50	110	1.0187	1.0095	<b>0.7095</b>	1.0524	1.0306	<b>0.7123</b>	0.0255
2003-2007	1264	35	72	1.0299	1.0157	<b>0.5145</b>	1.1522	1.1606	<b>0.5960</b>	0.0361
detrending				0.6543	0.6425		0.6674	0.6732		
2008-2012	1262	35	72	1.1200	1.0774	<b>0.7387</b>	0.9274	0.9292	<b>0.7010</b>	0.0361
<b>DAX</b>										
2003-2012	2556	50	110	1.0756	1.0647	<b>0.6518</b>	1.1018	1.1023	<b>0.7059</b>	0.0254
2003-2007	1274	35	72	1.0695	1.0623	<b>0.8031</b>	1.0768	1.0819	<b>0.8264</b>	0.0359
detrending				0.7396	0.7131		0.8819	0.8758		
2008-2012	1282	35	73	1.1732	1.1692	<b>0.7099</b>	0.9912	0.9828	<b>0.7384</b>	0.0358
<b>NIKKEI 225</b>										
2003-2012	2543	49	108	1.0696	1.0841	<b>0.4989</b>	1.0753	1.0869	<b>0.6589</b>	0.0254
2003-2007	1229	35	71	1.0403	1.0310	<b>0.4627</b>	1.0703	1.0590	<b>0.4013</b>	0.0366
2008-2012	1224	34	71	1.2168	1.1852	<b>0.7009</b>	1.0606	1.0344	<b>0.7841</b>	0.0367
<b>PX</b>										
2003-2012	2494	49	109	1.0653	1.0673	<b>0.5765</b>	1.0871	1.0859	<b>0.6167</b>	0.0257
2003-2007	1235	35	71	1.0166	1.0159	<b>0.2106</b>	1.0158	1.0128	<b>0.3155</b>	0.0365
2008-2012	1259	35	72	0.8935	0.9182	<b>0.7603</b>	0.8478	0.8336	<b>0.5805</b>	0.0361

We again list the bandwidths used for the estimation alongside the estimated degrees of fractional integration of daily high ( $\hat{d}_H$ ) and low ( $\hat{d}_L$ ) log-prices and their difference ( $\hat{d}_R$ )

(range). Standard errors are displayed in the last column. This time the standard errors do not depend on the bandwidth parameter but only on the total number of observations ( $T$ ), thus only one standard error is provided for all the estimates of each index in each period. In the case of the two trend-stationary series (DAX and FTSE 100 in the first period), as implied by the ADF test, we detrend the series first, then we apply the GPH estimator and we find that they remain non-stationary. The order of integration for FTSE 100 is in both specifications around 0.65 and in the case of DAX it is even higher. This supports the results from the ELW estimator. Thus, for the remainder of the study we consider these two series non-stationary and use them in their original form (with no detrending).

The results presented in Table 4 generally support the results of estimation using ELW in two ways. First, the integration orders of daily highs and lows are close to 1 and second, range is not integrated of order 0 and displays long memory. However, we can see some differences between the two estimation procedures. The GPH estimates of the integration orders of daily highs and lows using each bandwidth are much closer to each other than the ELW estimates. However, if we look at the first (calmer) period, ELW estimates of the integration orders of daily highs and lows are below 1, but GPH estimates are mostly greater than 1 (overall, in only 8 cases out of 60 are the GPH integration orders of daily highs and lows smaller than 1). Surprisingly, from the GPH results with  $m = T^{0.6}$  bandwidth one would believe that the log-prices in the second (after-crisis) period are less non-stationary than in the first (pre-crisis) period. If we compare this result with results using the  $m = T^{0.5}$  bandwidth we arrive at opposite conclusion.

The estimates of integration orders of range vary substantially between the two bandwidth specifications. Similarly to the ELW estimator, in the case of NIKKEI 225 in the full period the choice of bandwidth changes the conclusion about the stationarity of range. In 24 cases out of 30 the GPH estimator provides higher estimates of the long memory of the range than the ELW estimator. Among indices, the integration orders of range vary substantially, from 0.21 for the PX Index in the first period up to 0.83 for DAX in the same period. The range of the PX Index and of NIKKEI 225 is definitely

stationary in the first period and non-stationary in the others. The range of the other indices is non-stationary in all periods. One can also see the large difference between the ELW and GPH estimates of the integration order of the range of DAX in the first period; the GPH estimate is more than 30% higher than the ELW estimate with the same bandwidth specification. The GPH estimate is more in line with the findings from the inspection of the DAX autocorrelation function than the ELW estimate.

Overall, one can see that the results are sensitive to both the selected estimator (ELW or GPH) and to the bandwidth parameter. However, it is clear that the daily high and low log-prices are not stationary and display long memory. Range also displays long memory, but in some cases it is in the stationary region.

In the following section we describe and provide a test statistic for testing the equality of integration orders of daily high and low log-prices, which is a necessary condition for the further FCVAR analysis.

### 4.3 Testing the equality of integration orders

Nielsen and Shimotsu (2007) present the possibility of testing the equality of integration orders. This approach is based on the fact that the presence or absence of cointegration is not known when we estimate the fractional integration orders. They propose a test statistic which takes this into account.

They design two test hypotheses: one is the pairwise equality of the integration orders; the other is the equality of all integration orders. For each hypothesis a test statistic is defined. However, in our bivariate case, these two possible hypotheses collapse into one, allowing us to only focus on the hypothesis of equality of all integration orders:

$$\mathcal{H}_0: d_H = d_L = d_*$$

The test statistic is the following:

$$\hat{\tau}_0 = m(S\hat{d})' \left( S \frac{1}{4} \hat{D}^{-1} (\hat{G} \circ \hat{G}) \hat{D}^{-1} S' + h(T)^2 \right)^{-1} (S\hat{d}),$$

where  $\circ$  denotes the Hadamard product,  $\hat{d} = (\hat{d}_H, \hat{d}_L)$ ,  $D = \text{diag}(G_{11}, G_{22})$ ,  $S = (1, -1)'$  and  $h(T) = \log(T)^{-k}$  for  $k > 0$ . Under the null hypothesis ( $\mathcal{H}_0$ ) Nielsen and Shimotsu (2007) prove that if the variables are cointegrated (i.e.,  $r = 1$ ) then the test statistic  $\hat{T}_0$  should converge in probability to 0. However, if they are not cointegrated ( $r = 0$ ) then under the null hypothesis the test statistic  $\hat{T}_0$  should converge in distribution to the  $\chi^2(1)$  distribution<sup>v</sup>. This means that if the value of the test statistic  $\hat{T}_0$  is significantly large with respect to the  $\chi^2(1)$  density, we can take this as evidence that the null of the equality of integration orders is rejected.

**Table 5: Test statistics for the equality of integration orders**

	Bandwidths			$\hat{T}_0$ statistics	
	$T$	$T^{0.5}$	$T^{0.6}$	$\hat{T}_0(m = T^{0.5})$	$\hat{T}_0(m = T^{0.6})$
<b>S&amp;P 500</b>					
2003-2012	2517	50	109	0.2636	1.3361
2003-2007	1258	35	72	0.7113	0.5400
2008-2012	1259	35	72	0.8033	0.7043
<b>FTSE 100</b>					
2003-2012	2526	50	110	0.1979	0.7689
2003-2007	1264	35	72	0.2716	0.1596
2008-2012	1262	35	72	0.3005	0.2635
<b>DAX</b>					
2003-2012	2556	50	110	0.1041	1.1610
2003-2007	1274	35	72	0.7085	0.6007
2008-2012	1282	35	73	0.0525	0.1604
<b>NIKKEI 225</b>					
2003-2012	2543	49	108	0.1493	0.1517
2003-2007	1229	35	71	0.1919	0.4660
2008-2012	1224	34	71	0.0096	0.1129
<b>PX</b>					
2003-2012	2494	49	109	0.0488	0.6988
2003-2007	1235	35	71	0.2329	1.2668
2008-2012	1259	35	72	0.0538	0.2520

Table 5 presents the test statistics for testing the equality of integration orders of daily high and low log-prices. For each index and time period we present two test statistics; the

<sup>v</sup> Nielsen and Shimotsu (2007) show, that the degrees of freedom is  $(p - 1)$ , where  $p$  is the number of integration orders, whose equality is tested. As we are testing equality of integration orders of daily highs and lows,  $p = 2$ .

first is estimated using the  $T^{0.5}$  bandwidth, and the second using the  $T^{0.6}$  bandwidth. In our case we use  $k = 1$ , and thus  $h(T) = 1/\log(T)$ . Since the maximum test statistic is 1.3361 and the lowest critical value of the  $\chi^2(1)$  distribution is 2.71 (for the 90% confidence interval), we cannot reject the null hypothesis of equality of the integration orders. This implies that we can perform the FCVAR estimation, which requires the integration orders to be the same (i.e.,  $d_H = d_L$ ).

## 5 Fractionally cointegrated VAR

Having confirmed that range displays long memory and that the integration orders of daily high and low log-prices are the same, we can now define the fractionally cointegrated VAR. The fractionally cointegrated vector error correction model (FVECM) or fractionally cointegrated VAR (FCVAR) was first proposed by Johansen (2008) based on an idea of Granger (1986), and further analysed by Johansen and Nielsen (2010, 2012). In this chapter we introduce the fractionally cointegrated VAR model and show how it can be estimated. We also describe the cointegration rank tests which should be applied in the fractional framework.

### 5.1 Derivation of FCVAR

The cointegrated VAR was already described in Section 3.3; however, in order to explain the generalization to the fractional form we recall the equation (1) for the cointegrated VAR for an  $n$ -dimensional non-stationary time series  $X_t$  once more:

$$\Delta X_t = \alpha(\beta' X_{t-1} + \rho') + \sum_{i=1}^p \Gamma_i \Delta X_{t-i} + \varepsilon_t, \quad t = 1, \dots, T,$$

We want to arrive at a VAR model which is able to capture fractional processes characterized by slowly decaying autocorrelation functions. The generalized model should also allow  $X_t$  to be fractional of order  $d$  and cofractional of order  $d - b$ ; that is,  $\beta' X_t$  should be fractional of order  $d - b \geq 0$ , thereby extending the model (1) to fractional processes. In other words, fractional cointegration assumes the existence of a common stochastic trend which is integrated of order  $d$ , and the short-term departures from the long-run equilibrium being integrated of order  $(d - b)$  (Caporin et al., 2013).

The following methodology is taken from Johansen and Nielsen (2012) and Nielsen and Morin (2012).

We derive the model in two steps. First, the usual lag operator  $L = 1 - \Delta$  and the difference operator  $\Delta$  are replaced by the fractional lag operator and fractional difference

operator,  $L_b = 1 - \Delta^b$  and  $\Delta^b = (1 - L)^b$ . The fractional difference operator is defined by the binomial expansion  $\Delta^b Z_t = \sum_{n=0}^{\infty} (-1)^n \binom{b}{n} Z_{t-n}$ . Second, the resulting model is applied to  $Z_t = \Delta^{d-b} X_t$ . Thus we define the fractionally cointegrated VAR<sub>d,b</sub>(p) by

$$\Delta^d X_t = \Delta^{d-b} L_b \alpha \beta' X_t + \sum_{i=1}^p \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (2)$$

where  $\varepsilon_t$  is an  $n$ -dimensional *i.i.d.*  $(0, \Omega)$ ,  $\Omega$  is a positive-definite variance matrix and  $\alpha$  and  $\beta$  are  $(n \times r)$  matrices,  $0 \leq r \leq n, d \geq b > 0$ .

If we want to model data with a non-zero mean, e.g.  $Y_t = \mu + X_t$  where  $X_t$  is given by (2), then  $\Delta^a Y_t = \Delta^a (\mu + X_t) = \Delta^a X_t$  because  $\Delta^a 1 = 0$  for  $a > 0$ .  $Y_t$  thus satisfies the same equations as  $Y_t = \mu + X_t$ . This means that the model with  $d > b$  is invariant to inclusion of a restricted constant term  $\rho$ , when this term is included into the model in a similar way as in (1). Therefore we consider the inclusion of a constant term only in the model with  $d = b$ :

$$\Delta^d X_t = L_d \alpha (\beta' X_t + \rho') + \sum_{i=1}^p \Gamma_i \Delta^d L_d^i X_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (3)$$

Both models include the original cointegrated VAR as the special case  $d = b = 1$ . The interpretation is similar as before. The model also describes cointegration and adjustment towards equilibrium, but it is more general, as it incorporates fractional integration and cointegration.  $X_t$  are integrated of order  $d$ , and  $b$  is the strength of the cointegrating relations (a higher  $b$  means less persistence in the cointegrating relations;  $b$  can also be called the cointegration gap). Moreover, if  $(d - b) < 1/2$  then  $\beta' X_t$  in (2) is asymptotically a zero-mean stationary process. If we write  $\Pi = \alpha \beta'$ , where the  $(n \times r)$  matrices  $\alpha$  and  $\beta$  with  $r \leq n$  are assumed to have full column rank  $r$ , the columns of  $\beta$  are then the  $r$  cointegrating (cofractional) relations determining the long-run equilibria. The rank  $r$  is called the cointegration or cofractional rank. The parameter  $\alpha$  determines the speed of adjustment towards the equilibria. The parameters  $\Gamma = (\Gamma_1, \dots, \Gamma_p)$  govern the short-run dynamics. The parameter  $\rho$  is the restricted constant term (since the constant term in the model is restricted to be of the form  $\mu = \alpha \rho'$ ) and is interpreted as the mean level of the long-run equilibria. In the special case when  $d = b$ ,  $(\beta' X_t + \rho')$  is a zero-mean

process of fractional order zero. Having defined the fractional cointegrated VAR, we shall now take a look at its estimation.

## 5.2 FCVAR estimation

### 5.2.1 Maximum likelihood estimation

According to Johansen and Nielsen (2012), the model (2) is estimated by conditional likelihood (given initial values) by maximizing the function

$$\log L_t(\lambda) = -\frac{T}{2} \log \det(T^{-1} \sum_{t=1}^T \epsilon_t(\lambda) \epsilon_t(\lambda)'),$$

where

$$\epsilon_t(\lambda) = \Delta^d X_t - \Delta^{d-b} L_b \alpha \beta' X_t - \sum_{i=1}^p \Gamma_i \Delta^d L_b^i X_t, \quad \lambda = (d, b, \alpha, \beta, \Gamma),$$

for model (2) and

$$\epsilon_t(\lambda) = \Delta^d X_t - L_d \alpha (\beta' X_t + \rho') - \sum_{i=1}^p \Gamma_i \Delta^d L_d^i X_t, \quad \lambda = (d, b, \alpha, \beta, \rho, \Gamma),$$

for model (3). Johansen and Nielsen (2012) state that the model is relatively easy to estimate, since for fixed  $(d, b)$  the model is estimated by reduced rank regression, thus reducing the numerical problem to the optimization of a function of two variables. The parameters  $(\alpha, \beta, \rho, \Gamma)$  can be concentrated out of the likelihood function and the profile likelihood function is optimized over the two fractional parameters  $d$  and  $b$  using numerical optimization. The existence, uniqueness and consistency of the maximum likelihood estimator are also shown in Johansen and Nielsen (2012), which allows the application of standard likelihood theory.

Johansen and Nielsen (2012) also examine the asymptotic theory. They prove that for i.i.d. errors with suitable moments conditions, the estimated cointegrating vectors  $(\hat{\beta}, \hat{\rho})$  are locally asymptotically mixed normal when  $b > 1/2$ . The remaining conditional

maximum likelihood parameter estimates  $(\hat{d}, \hat{b}, \hat{\alpha}, \hat{\Gamma}_1, \dots, \hat{\Gamma}_p)$  are asymptotically Gaussian. For this case of “strong cointegration” the asymptotic theory is non-standard and requires type-II fractional Brownian motion.

When  $b < 1/2$  the asymptotic theory is standard and all conditional maximum likelihood parameter estimates  $(\hat{d}, \hat{b}, \hat{\alpha}, \hat{\beta}, \hat{\rho}, \hat{\Gamma}_1, \dots, \hat{\Gamma}_p)$  are asymptotically Gaussian. When  $b$  is close to  $1/2$ , a large number of observations is needed to achieve the asymptotic results. Having in mind these results, asymptotically standard ( $\chi^2$ ) inference on all model parameters, including the orders of fractionality and the cointegrating relations, is achieved through the application of quasi-likelihood ratio tests.

### 5.2.2 Imposing restrictions on parameters

As Nielsen and Morin (2012) describe, it is possible to impose various restrictions, the most interesting being those on  $\alpha$  and  $\beta$ . We can impose either linear restrictions or restrictions to known values. This can be achieved through partitioning  $\alpha$  and  $\beta$  into sets of columns,  $\alpha = [\alpha_1, \alpha_2, \dots]$  and  $\beta = [\beta_1, \beta_2, \dots]$ , such that each  $\alpha_i$  and  $\beta_i$  are  $n \times r_i$  matrices with  $\sum_i r_i = r$ . For each  $\alpha_i$  and  $\beta_i$  we can specify a restriction to known values  $a_i$  and  $b_i$ . Imposing linear restrictions is achieved through specifying matrices  $A_i (n \times s_i)$  and  $H_i (n \times s_i)$  so that  $\alpha_i = A_i \psi_i$  and  $\beta_i = H_i \phi_i$  for conforming matrices  $\psi_i$  and  $\phi_i$  (for more details see Johansen, 1996). The switching algorithm proposed in Johansen (1996) is used for estimation. For each pair of submatrices  $\alpha_i$  and  $\beta_i$  the parameters are estimated by taking other submatrices as fixed. The estimation is repeated through the sequence of submatrices until convergence is achieved.

When a restricted constant term is present, we also have to impose restrictions on this constant term, because the two parameters  $\beta$  and  $\rho$  are estimated jointly through  $\beta^* = (\beta', \rho')'$  of dimension  $n_1 \times r$ , where  $n_1 = n + 1$ . If we want to specify some columns of  $\beta$  as known, we have to do the same for the corresponding columns of  $\rho$ . For linear restrictions, we make the modification  $\beta_i^* = H_i^* \phi_i$ , where  $H_i^*$  is  $(n_1 \times s_i)$ . We can impose any linear restrictions we wish on the fractional parameters  $d$  and  $b$ . For the matrix of short term adjustments  $\Gamma_i$ , only exclusion restrictions are allowed in the MATLAB

software package proposed by Nielsen and Morin (2012). If no restrictions are specified, the model is estimated unrestricted.

### 5.2.3 Initial values

Another issue one has to address when estimating the FCVAR is specifying the appropriate number of observations reserved for initial values that the model will be conditioned upon in the estimation. Johansen and Nielsen (2012) propose a generalization regarding initial values. The initial values are neither modelled nor assumed to be zero as it was done in the previous work, but are instead assumed to be bounded. The likelihood function depends on  $\Delta^a X_t$  for different values of  $a > 0$ , and since we do not know all the initial values of  $X_t$  we assume we have observations of  $X_t, t = -N_0 + 1, \dots, T$  and define initial values for the calculation as  $\tilde{X}_{-n} = X_{-n}, n = 0, \dots, N_0 - 1$  and  $\tilde{X}_{-n} = 0, n \geq N_0$ . Thus  $N_0$  observations are set aside for initial values, which are assumed to be uniformly bounded (Johansen and Nielsen, 2012). The choice of  $N_0$  carries a bias/efficiency trade-off: a smaller number of initial values introduces more bias, but leaves a greater number of observations for the estimation. This was the case in Nielsen and Morin (2012), who together with the MATLAB software package for estimation of FCVAR provide an example in which they have only 55 observations, so they use only five observations as the initial values. In our case, we have more than a thousand observations, so the number of initial values is chosen so as to optimize the performance and achieve convergence. The exact number of initial values used for estimation is listed together with the results from the cointegration rank test by Johansen and Nielsen (2012) in Sub-section 5.3.2.

## 5.3 Cointegration rank tests

As we have already noted, the traditional tests for cointegration proposed by Johansen (1991) are not applicable in the presence of long memory. Instead, more recent tests allowing for fractional cointegration should be applied. Fractional cointegration is defined as follows:  $X_t$  is said to be fractionally cointegrated  $CI(d, b)$  if  $X_t$  has  $I(d)$  elements and for some  $b > 0$  there exists  $\beta$  such that  $\beta' X_t$  is integrated of order  $(d - b)$ . Originally, we

assumed that  $d = b = 1$ ; however, whenever either  $d$  or  $b$  (or both) are different from 1 we are dealing with fractional cointegration.

We define here two cointegration rank tests, the first one proposed by Nielsen and Shimotsu (2007) and the second by Johansen and Nielsen (2012).

### 5.3.1 Cointegration rank test by Nielsen and Shimotsu (2007)

The cointegration rank determination procedure proposed by Nielsen and Shimotsu (2007) extends the one found in Robinson and Yajima (2002) since it allows for both stationary and non-stationary fractionally integrated processes. Robinson and Yajima (2002) allow only stationary data, which is quite limiting for financial data. The ability to consider any value of the fractional differencing parameter  $d$  follows from the application of the exact local Whittle analysis of Shimotsu and Phillips (2005). The exact local Whittle estimate of  $d$  is then used to examine the rank of the spectral density matrix of the  $d$ 'th differenced process around the origin in order to provide a consistent estimate of the cointegration rank. This semi-parametric method only requires information about the behaviour of the spectral density matrix around the origin, but it also relies on the choice of the bandwidth and threshold parameters. On the other hand, this approach does not require the cointegrating vectors to be estimated in order to determine the cointegration rank.

The estimate of the cointegration rank  $r$  of  $X_t$  is achieved through estimating the matrix  $G$  and its eigenvalues. For simplicity, Nielsen and Shimotsu (2007) assume the equality of integration orders, in our case  $d_H = d_L = d_*$ . In this study we do not have to assume the equality, because we have provided a formal test for the equality of integration orders.

Nielsen and Shimotsu (2007) define

$$\hat{G}(d_*) = \frac{1}{m_1} \sum_{j=1}^{m_1} \text{Re}[I_{\Delta(L, d_*, d_*)x}(\lambda_j)], \quad (4)$$

where  $I_{\Delta(L,d_*,d_*)x}(\lambda_j)$  is in our bivariate case the periodogram of  $(\Delta^{d_*} p_t^H, \Delta^{d_*} p_t^L)'$ , and  $\text{Re}(\cdot)$  denotes the real part of the periodogram. The estimator  $\hat{G}(d_*)$  depends on a new bandwidth parameter  $m_1(T)$  as some complications may arise when  $d_*$  is estimated. Since  $d_*$  is unknown, it has to be substituted with an estimate, but as Nielsen and Shimotsu (2007) state, we cannot use the multivariate version of the exact local Whittle estimator to obtain  $d_*$ , because when  $p_t^H$  and  $p_t^L$  are cointegrated, the matrix  $G$  is not of full rank. That is why the matrix  $G$  is estimated by (4) based on  $m_1$  periodogram ordinates and each  $d_i$  is estimated by  $\hat{d}_i = \arg \min_{d \in [\Delta_1, \Delta_2]} R_i(d)$  using  $m$  ordinates such that  $\frac{m}{m_1} \rightarrow 0$ .

Similarly to Nielsen and Shimotsu (2007) we also consider two combinations of bandwidth parameters  $m$  and  $m_1$ : first,  $m = T^{0.6}$  and  $m_1 = T^{0.55}$ , and second,  $m = T^{0.5}$  and  $m_1 = T^{0.45}$ . However, since both specifications yield similar results, for the sake of brevity, we only present here the first combination.

Denoting  $\hat{\delta}_i$  the  $i$ 'th eigenvalue of  $\hat{G}$ , it is possible to apply a model selection procedure to determine the cointegration rank  $r$ .

$$\hat{r} = \arg \min_{u=0,1} L(u),$$

where the loss function is defined as

$$L(u) = v(T)(2 - u) - \sum_{i=1}^{2-u} \hat{\delta}_i,^{vi}$$

for a  $v(T) > 0$  such that  $v(T) + \frac{1}{m_1^{1/2} v(T)} \rightarrow 0$ .

Table 6 summarizes the results from the cointegration rank test proposed by Nielsen and Shimotsu (2007). We only provide here the results for the combination of  $m = T^{0.6}$  and  $m_1 = T^{0.55}$ , but the same analysis was also conducted for  $m_1 = T^{0.5}$  and  $m = T^{0.45}$ . Also, for the rank estimates we only show the results for the smallest  $v(T) = m_1^{-0.45}$  and for the largest  $v(T) = m_1^{-0.05}$ , but we have also considered all of the following values of

---

<sup>vi</sup> In the original definition there is  $(p - u)$  instead of  $(2 - u)$ , but in their notation  $p$  is the dimension of  $X_t$ , which in our notation is  $n = 2$ .

$v(T)$ :  $m_1^{-0.40}$ ,  $m_1^{-0.35}$ ,  $\dots m_1^{-0.10}$ . These all yielded the same result of one cointegrating vector. The first column of Table 6 represents the common integration order  $\bar{d}_*$ , which is used in the fractional cointegration analysis and is simply computed as an average of the estimated integration orders of daily high and low log-prices from the ELW estimator based on a given bandwidth. The other two columns stand for the estimates of the eigenvalues of matrix  $\hat{G}(\bar{d}_*)$  from (4). Their values indicate that one eigenvalue of  $\hat{G}(\bar{d}_*)$  could be zero. This implies that there could be one cointegrating relation.  $L(0)$  and  $L(1)$  are the values of the loss function evaluated with regard to a cointegration rank of 0 or 1. The cointegration rank  $\hat{r}$  is then determined by  $\text{argmin}$  of  $L(u)$ . In all cases the loss function  $L(1)$  is smaller than  $L(0)$ , implying that there is exactly one cointegrating relation between daily high and low log-prices.

**Table 6: Cointegration rank test by Whittle and Shimotsu (2007)**

	$\bar{d}_*$	Eigenvalues of $\hat{G}(\bar{d}_*)$		Rank estimates					
		$\delta_1$	$\delta_2$	$v(T) = m_1^{-0.45}$		$\hat{r}$	$v(T) = m_1^{-0.05}$		$\hat{r}$
				$L(0)$	$L(1)$		$L(0)$	$L(1)$	
<b>S&amp;P 500</b>									
2003-2012	1.0239	0.3169	0.0010	-1.7117	-1.8491	1	-0.3872	-1.1868	1
2003-2007	0.9099	0.2151	0.0009	-1.6561	-1.8197	1	-0.3553	-1.1694	1
2008-2012	1.0321	0.5076	0.0022	-1.6561	-1.8189	1	-0.3553	-1.1686	1
<b>FTSE 100</b>									
2003-2012	0.9521	0.4360	0.0010	-1.7117	-1.8511	1	-0.3872	-1.1889	1
2003-2007	0.8448	0.3286	0.0014	-1.6561	-1.8194	1	-0.3553	-1.1691	1
2008-2012	0.9684	0.6313	0.0018	-1.6561	-1.8222	1	-0.3553	-1.1718	1
<b>DAX</b>									
2003-2012	1.0246	0.5136	0.0017	-1.7117	-1.8491	1	-0.3872	-1.1869	1
2003-2007	0.9639	0.4063	0.0025	-1.6591	-1.8172	1	-0.3569	-1.1662	1
2008-2012	1.0147	0.7520	0.0029	-1.6591	-1.8217	1	-0.3569	-1.1706	1
<b>NIKKEI 225</b>									
2003-2012	1.0496	0.4847	0.0011	-1.7099	-1.8504	1	-0.3861	-1.1885	1
2003-2007	0.9601	0.4551	0.0010	-1.6561	-1.8235	1	-0.3553	-1.1731	1
2008-2012	1.0621	0.6491	0.0021	-1.6529	-1.8199	1	-0.3537	-1.1703	1
<b>PX</b>									
2003-2012	1.1513	0.3563	0.0009	-1.7099	-1.8496	1	-0.3861	-1.1877	1
2003-2007	0.9917	0.4132	0.0016	-1.6561	-1.8200	1	-0.3553	-1.1696	1
2008-2012	1.1101	0.5811	0.0035	-1.6561	-1.8159	1	-0.3553	-1.1656	1

### 5.3.2 Cointegration rank test by Johansen and Nielsen (2012)

In this sub-section, we return to the methodology proposed by Johansen and Nielsen (2012) as it was summarised by Nielsen and Morin (2012) and we describe another approach to estimating the cointegration rank  $r$ . We again consider model (2)

$$\Delta^d X_t = \Delta^{d-b} L_b \Pi X_t + \sum_{i=1}^p \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t, \quad t = 1, \dots, T,$$

where  $\Pi = \alpha\beta'$ , and we want to test the hypothesis  $\mathcal{H}_r : \text{rank}(\Pi) = r$  against  $\mathcal{H}_n : \text{rank}(\Pi) = n$ . Let  $L(d, b, r)$  be the profile likelihood function given a rank  $r$ , where  $(\alpha, \beta, \Gamma)$  have been concentrated out by regression and reduced rank regression (see Johansen and Nielsen, 2012, p. 23). In the case of the model with a constant we test  $\mathcal{H}_r : \text{rank}(\Pi, \mu) = r$  against  $\mathcal{H}_n : \text{rank}(\Pi, \mu) = n$  and the profile likelihood function given rank  $r$  is then  $L(d, r)$ , where again the parameters  $(\alpha, \beta, \rho, \Gamma)$  have been concentrated out.<sup>vii</sup>

We maximize the profile likelihood function under both hypotheses  $\mathcal{H}_r$  and  $\mathcal{H}_n$ ; the likelihood ratio (LR) statistic is  $LR_T(q) = 2 \log(L(\hat{d}_n, \hat{b}_n, n)/L(\hat{d}_r, \hat{b}_r, r))$ , where  $q = n - r$  and

$$L(\hat{d}_n, \hat{b}_n, n) = \max_{d,b} L(d, b, n), \quad L(\hat{d}_r, \hat{b}_r, r) = \max_{d,b} L(d, b, r),$$

similarly for the model with a constant. The asymptotic distribution of  $LR_T(q)$  depends both qualitatively and quantitatively on the parameter  $b$  and on  $q = n - r$ . This dependence on the unknown parameter  $b$  makes the empirical analysis more complicated; however, a computer program exists, made available by MacKinnon and Nielsen (2012), which, given the dimension of the problem,  $q = n - r$  (the number of variables of interest less the cointegration rank), and the value of  $b$ , computes the asymptotic critical values and asymptotic p-values for the LR rank test in this fractional framework.

---

<sup>vii</sup> The model with an inclusion of a constant is considered only when  $d = b$  and that is why the profile likelihood function depends only on the given rank  $r$  and the parameter  $d$ .

MacKinnon and Nielsen (2012) deal with the model's uncertainty in the specification of response surface regressions by using model averaging instead of model selection.

As we have already discussed, in the case of “weak cointegration”, when  $0 < b < 1/2$ ,  $LR_T(q)$  has a standard asymptotic distribution,  $LR_T(q) \xrightarrow{D} \chi^2(q^2)$ . This means that we can use the standard critical values of the  $\chi^2$  distribution to determine the statistical significance of the cointegration rank. The situation is, however, different when  $1/2 < b \leq d$ . Then the asymptotic theory is nonstandard and

$$LR_T(q) \xrightarrow{D} \text{Tr} \left\{ \int_0^1 dW(s)F'(s) \left( \int_0^1 F(s)F'(s)ds \right)^{-1} \int_0^1 F(s)dW'(s) \right\},$$

where the vector process  $dW$  is the increment of ordinary vector Brownian motion of dimension  $q = n - r$ . The vector  $F$  depends on deterministics in a similar way as the CVAR model in Johansen (1996). If we do not include any deterministic terms into the model, then  $F(u) = W_b(u)$ . If the restricted constant term is included into the model, then  $F(u) = (W_b'(u), 1)'$ , where  $W_b(u) = \Gamma(b)^{-1} \int_0^u (u-s)^{b-1} dW(s)$  is vector fractional type-II Brownian motion.

If we are in the situation where  $b > 1/2$ , we use the computer program proposed by MacKinnon and Nielsen (2012) to determine the appropriate critical value given the parameters  $q$  and  $b$ , and based on this critical value we decide on the significance of the cointegration rank. In particular, values of  $b$  in the interval  $1/2 < b \leq 2$  are of interest. However, the lowest value of  $b$ , which the procedure permits, is 0.51 rather than 0.5, because there is no theory for the case when  $b = 1/2$ .

The results of this cointegration test are presented in the following Table 6. The first column,  $IV$ , stands for the number of initial values used in the estimation. Then, for each rank  $r = 0, 1, 2$  we present the estimates of the parameter of the fractional order of integration ( $\hat{d}$ ), the parameter of the cointegration gap ( $\hat{b}$ ), and the corresponding likelihood ratio statistic ( $LR$ ) and its critical value at 5% level of significance. When  $b$  is smaller than 0.5, it follows the  $\chi^2(q^2)$  distribution; this means that for cointegration rank  $r = 0$ ,  $q = 2$  and  $\chi_{0.95}^2(4) = 9.49$ . When  $r = 1$ , then  $q = 1$  and  $\chi_{0.95}^2(1) = 3.84$ . If  $b$  is

greater than 0.51, we calculate the corresponding critical value using the programme provided by MacKinnon and Nielsen (2012).

**Table 7: Cointegration rank test by Johansen and Nielsen (2012)**

	<i>IV</i>	<i>rank = 0</i>				<i>rank = 1</i>				<i>rank = 2</i>	
		$\hat{d}$	$\hat{b}$	<i>LR</i>	$CV_{5\%}$	$\hat{d}$	$\hat{b}$	<i>LR</i>	$CV_{5\%}$	$\hat{d}$	$\hat{b}$
<b>S&amp;P 500</b>											
2003-2012	70	0.681	0.372	271.232	9.490	<b>0.998</b>	<b>0.302</b>	0.058	3.840	1.002	0.295
2003-2007	80	0.517	0.517	117.535	9.362	<b>0.999</b>	<b>0.467</b>	1.338	3.840	0.984	0.481
2008-2012	60	0.719	0.329	136.629	9.490	<b>0.981</b>	<b>0.334</b>	1.699	3.840	1.033	0.233
<b>FTSE 100</b>											
2003-2012	80	0.623	0.416	254.914	9.490	<b>0.970</b>	<b>0.357</b>	2.368	3.840	0.954	0.387
2003-2007	70	0.497	0.497	118.642	9.490	<b>0.984</b>	<b>0.380</b>	2.510	3.840	0.968	0.424
2008-2012	30	0.667	0.361	120.148	9.490	<b>0.987</b>	<b>0.463</b>	2.035	3.840	1.009	0.416
<b>DAX</b>											
2003-2012	60	0.609	0.439	233.689	9.490	<b>0.987</b>	<b>0.388</b>	3.808	3.840	0.966	0.408
2003-2007	50	0.561	0.524	38.776	9.359	<b>1.042</b>	<b>0.476</b>	0.323	3.840	1.055	0.480
2008-2012	80	0.638	0.405	133.836	9.490	<b>0.967</b>	<b>0.336</b>	0.302	3.840	0.955	0.361
<b>NIKKEI 225</b>											
2003-2012	80	0.591	0.476	138.368	9.490	<b>1.004</b>	<b>0.517</b>	0.254	3.636	0.997	0.515
2003-2007	60	0.513	0.513	80.636	9.365	<b>1.019</b>	<b>0.635</b>	0.003	3.587	1.018	0.635
2008-2012	60	0.708	0.370	79.292	9.490	<b>0.996</b>	<b>0.010</b>	28.181	3.840	0.978	0.547
<b>PX</b>											
2003-2012	80	0.520	0.520	109.031	9.360	<b>1.007</b>	<b>0.475</b>	1.289	3.840	0.988	0.472
2003-2007	60	0.532	0.532	65.024	9.360	<b>1.021</b>	<b>0.680</b>	5.991	6.373*	0.977	0.666
2008-2012	60	0.511	0.511	53.246	9.367	<b>0.982</b>	<b>0.454</b>	1.092	3.840	1.010	0.451

*Note:* Asterisk (\*) denotes the 1% critical value rather than the 5% critical value. The cointegrating relation is not significant at the 5%, but it is at the 1% level of significance.

We can see from Table 7 that we find one significant cointegrating relation, except for the NIKKEI 225 in the second period, where no cointegrating vector is found. In the case of the PX Index in the first period, the LR statistic for one cointegrating vector (*rank* = 1) is significant only at the 1% level and not at the 5% level of significance, which we mark by an asterisk. When *r* = 0, the likelihood ratio (LR) statistic is significantly larger than the corresponding critical value, meaning that we reject the null hypothesis of zero cointegrating relations. When *r* = 1 the LR statistic is significantly smaller than the corresponding critical value and thus we do not reject the null of one cointegrating relation. This is not true for the NIKKEI 225 in the second period, where we do not find

a significant cointegrating vector even though its presence was hinted at by the results of the cointegration rank test proposed by Nielsen and Shimotsu (2007). One of the possible explanations may be that the optimization procedure did not converge for NIKKEI 225 in the second period as the lowest possible value for  $b$ ,  $b = 0.01$ , was achieved for one cointegrating relation.

## 5.4 FCVAR estimation results

The estimation of the cointegration rank presented above is performed together with the estimation of the FCVAR model using the MATLAB software package proposed by Nielsen and Morin (2012). Having in mind the result that in each specification there is one significant cointegrating vector (except for the NIKKEI 225 in the second period) we use this information for the FCVAR estimation. We also specify the lag length of the short-term deviations to be  $p = 1$ . We use only one lag as that is sufficient to capture the autocorrelation of residuals. MacKinnon and Nielsen (2012) state that in the fractional model a single lag is usually sufficient, which is in contrast with the standard cointegrated VAR, where several more lags are needed to capture the serial correlation in residuals. We can also specify the number of initial values used for the estimation, which we have discussed in Sub-section 5.2.3. The number of initial values used for the estimation of the FCVAR model is the same number as the number of initial values used for the estimation of cointegration rank in Table 7. You can see that the number of initial values varies from 30 to 80. We have estimated several models with various numbers of initial values and present here only those which have performed the best (in terms of low standard errors, finding a significant cointegrating vector, stationarity, and convergence). We may also want to impose restrictions, as we discussed in Sub-section 5.2.2. We are particularly interested in the order of integration of range. Since range is defined as the difference between the maximum and minimum daily log-prices, i.e.  $(p_t^H - p_t^L)$ , we would like the cointegrating vector to be  $(1, -1)$ . If the cointegrating vector is different from  $(1, -1)$ , we cannot interpret the difference  $(d - b)$  as the order of integration of range. That is why we first estimate the model without any restrictions imposed to see whether the model yields a significant cointegrating vector and significant estimates of  $\hat{d}$  and  $\hat{b}$ , and then we impose the  $(1, -1)$  restriction on the cointegrating vector. However, when a

restriction is imposed, the standard errors are not provided, thus we cannot make any inference on the significance of the estimated parameters. We estimate the model for the case when  $d \neq b$ ; however, the procedure is capable of detecting whether  $d$  and  $b$  are close to equality; if they are, we would re-estimate the model with the restriction  $d = b$ . This situation did, however, not occur; that is in line with our previous findings, as the equality of the  $d$  and  $b$  parameters would imply that the order of integration of range is 0, which we have rejected based on both the ELW and GPH estimator results.

We first present the results of the FCVAR estimation without any restrictions imposed, in Table 8. The model is estimated in the following form (it is derived from model (2), here with  $p = 1$ )

$$\Delta^d X_t = \Delta^{d-b} L_b \alpha \beta' X_t + \Gamma_1 \Delta^d L_b X_t + \varepsilon_t, \quad t = 1, \dots, T,$$

The first two columns of Table 8 contain the estimates of the two parameters of interest ( $\hat{d}$  and  $\hat{b}$ ). These estimates are the same as the results from the cointegration rank test presented in Table 7 for one cointegrating vector (i.e., when  $rank = 1$ ). Next we present the estimated cointegrating vector ( $\hat{\beta}$ ) and the remaining six columns list the FCVAR parameters for the speed of adjustment,  $\alpha = (\alpha_H, \alpha_L)'$ , and the estimates of the matrix of short-run dynamics,  $\Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$ .

In all specifications, the parameters of interest ( $\hat{d}$  and  $\hat{b}$ ) are significant (based on small standard errors) and different from each other (thus the specification  $d \neq b$  is correct). Also, the estimates of the cointegrating vector are very close to the desired vector of  $(1, -1)$ . All this suggests that a linear combination of the daily high and low log-prices (possibly the range) is integrated of a non-zero order. However, since the estimate of the cointegrating vector is not exactly  $(1, -1)$ , we cannot interpret the difference  $(d - b)$  as the order of integration of range.

Daily high and low log-prices are integrated of an order close to 1. Surprisingly, in the first period, which we consider the calmer period, daily log-prices are more non-stationary than in the second period or in the full period. Also, the orders of integration of daily log-prices are smaller than unity in 9 out of 14 cases.

**Table 8: FCVAR estimation results (no restrictions)**

		$\hat{d}$	$\hat{b}$	$\hat{\beta}$	$\alpha_H$	$\alpha_L$	$\gamma_{11}$	$\gamma_{12}$	$\gamma_{21}$	$\gamma_{22}$
<b>S&amp;P 500</b>										
<i>u</i>	2003-2012	0.998 (0.014)	0.302 (0.028)	(1,-1.007)	-0.195 (0.302)	3.986 (0.945)	-0.773 (0.310)	1.050 (0.316)	-1.614 (0.727)	2.619 (0.776)
	2003-2007	0.999 (0.009)	0.467 (0.038)	(1,-1.002)	-1.072 (0.254)	1.680 (0.363)	0.153 (0.178)	0.076 (0.193)	-0.282 (0.258)	0.924 (0.283)
<i>u</i>	2008-2012	0.981 (0.015)	0.334 (0.035)	(1,-1.008)	-0.013 (0.348)	3.081 (0.922)	-0.755 (0.363)	1.049 (0.370)	-0.935 (0.696)	1.890 (0.752)
<b>FTSE 100</b>										
	2003-2012	0.970 (0.011)	0.357 (0.024)	(1,-1.005)	-0.657 (0.218)	2.308 (0.466)	-0.263 (0.190)	0.569 (0.204)	-0.620 (0.358)	1.505 (0.386)
<i>u</i>	2003-2007	0.984 (0.009)	0.380 (0.027)	(1,-1.003)	-1.344 (0.331)	2.339 (0.513)	0.245 (0.252)	-0.009 (0.280)	-0.994 (0.426)	1.727 (0.442)
	2008-2012	0.987 (0.009)	0.463 (0.031)	(1,-1.004)	-0.429 (0.189)	1.344 (0.305)	-0.309 (0.162)	0.524 (0.171)	-0.049 (0.231)	0.669 (0.247)
<b>DAX</b>										
	2003-2012	0.987 (0.011)	0.388 (0.033)	(1,-1.004)	-0.436 (0.209)	1.891 (0.440)	-0.332 (0.195)	0.571 (0.202)	-0.433 (0.327)	1.101 (0.354)
	2003-2007	1.042 (0.013)	0.476 (0.064)	(1,-1.001)	-0.551 (0.259)	1.520 (0.429)	-0.145 (0.194)	0.319 (0.203)	-0.169 (0.279)	0.526 (0.305)
	2008-2012	0.967 (0.016)	0.336 (0.035)	(1,-1.007)	-0.016 (0.393)	2.651 (0.846)	-0.695 (0.413)	0.980 (0.415)	-0.910 (0.675)	1.718 (0.711)
<b>NIKKEI 225</b>										
	2003-2012	1.004 (0.012)	0.517 (0.043)	(1,-1.002)	-0.175 (0.134)	1.318 (0.271)	-0.295 (0.124)	0.535 (0.133)	-0.066 (0.183)	0.543 (0.205)
	2003-2007	1.019 (0.010)	0.635 (0.053)	(1,-1.001)	-0.361 (0.151)	1.144 (0.242)	-0.154 (0.120)	0.309 (0.125)	-0.171 (0.168)	0.522 (0.186)
	2008-2012	-	-	-	-	-	-	-	-	-
<b>PX</b>										
	2003-2012	1.007 (0.017)	0.475 (0.045)	(1,-1.003)	-0.682 (0.168)	1.281 (0.310)	-0.172 (0.127)	0.340 (0.136)	-0.447 (0.236)	0.861 (0.258)
	2003-2007	1.021 (0.015)	0.680 (0.065)	(1,-1.002)	-0.410 (0.119)	0.612 (0.168)	0.021 (0.095)	0.204 (0.096)	0.123 (0.120)	0.243 (0.130)
	2008-2012	0.982 (0.017)	0.454 (0.056)	(1,-1.003)	-0.732 (0.253)	1.346 (0.445)	-0.272 (0.194)	0.453 (0.211)	-0.600 (0.355)	1.031 (0.389)

*Note:* Standard errors in brackets, “*u*” denotes that model is unstable (some roots of the characteristic polynomial lie outside the unit root circle).

The adjustment coefficients  $\alpha_H$  and  $\alpha_L$  capture the speed of adjustment of  $p_t^H$  and  $p_t^L$  towards equilibrium. These coefficients are mostly significant, non-zero, and with

expected signs ( $\alpha_H$  is negative and  $\alpha_L$  is positive) meaning that they move in opposite directions to restore equilibrium after a shock to the system occurs. We can note that the absolute values of the estimates of  $\alpha_H$  are much smaller than of  $\alpha_L$ , which suggests that the correction in the equation for daily lows overshoots the long-run equilibrium. Caporin et al. (2013) obtained similar results when analysing the DJIA index. When interpreting the short-run dynamics parameters ( $\gamma_{11}, \dots, \gamma_{22}$ ) we may notice that the coefficients of lagged daily highs are mostly negative, whereas those of lagged daily lows are mostly positive. Cheung (2007) states that negative coefficients imply a regressive behaviour, whereas positive coefficients are an indication of spill-over effects. He argues that higher daily highs tend to fall to a lower level, lower daily highs tend to drift up to a higher level, and higher daily lows lead to higher daily highs.

In one case, namely the second period of NIKKEI 225, we were unable to estimate the FCVAR model without restrictions, the model did not converge for any specified number of initial values. Also, we have to note that in 3 cases (S&P 500 in the full period and in the second sub-period and FTSE 100 in the first sub-period) the model was not stable. We say that a model is stable when the roots of the characteristic polynomial are smaller than unity. In these three situations, the roots exceeding unity are 1.365, 1.374, and 1.101, respectively, and thus the model should be interpreted with caution in these cases. In Table 8 we mark this situation by the letter “u” set before the period specification of the affected periods. We also test the residuals for the remaining autocorrelation and heteroscedasticity. Based on the Ljung-Box Q-test we reject in most cases the null of no autocorrelation; however, the value of the statistic is rather marginal. Based on the visualization of the autocorrelation functions the dependency is quite weak and it disappears after the second lag. Also, based on the visualization of the autocorrelation function of squared residuals we can detect some heteroscedasticity; however, it is again very weak. Neither of these findings impacts the quality of our estimates: they are still unbiased and consistent; however, the reliability of inference may be slightly affected, as standard errors may be biased.<sup>viii</sup>

---

<sup>viii</sup> For the sake of brevity, this residual diagnostic is not presented here, but is available upon request.

Since the estimates of the cointegrating vector from the unrestricted model are very close to the desired vector of  $(1, -1)$ , we impose the restriction on the vector to be exactly  $(1, -1)$  in order to be able to interpret the cointegrating relation as range. Table 9 contains the results of this estimation. In three specifications we had to moreover impose the restriction  $d = 1$  in order to achieve convergence of the model (we mark this restriction by the letter “ $d$ ” set before the period specification of the affected periods in Table 9). The imposition of this restriction solved the instability of the model for FTSE 100 in the first period. However, this restriction was of no use in the case of S&P 500. In the case of S&P 500, the models in the full period and in the second sub-period remain unstable (which we mark by “ $u$ ”). The roots of the characteristic polynomial exceeding unity are 1.205 and 1.237, respectively.

We can see that the order of integration of daily log-prices is in all cases greater than 1 (which is different from the estimation without the restriction, where the order of integration was mostly below 1). The estimates of the cointegration gap  $\hat{b}$  are quite similar to the unrestricted specification, apart from the case of NIKKEI 225 and PX Index in the first period. In these two situations the difference  $(d - b)$  changed from 0.384 to 0.539 and from 0.341 to 0.533, respectively (see Table 10). The estimates of the order of integration of range have thus moved from the stationary region into the non-stationary region. We have already discussed these two indices when analysing the ACF functions, where we note that in the first period the presence of long memory is arguable; this was further supported by the results from the GPH and ELW estimators, where range was found stationary (but not integrated of order 0). In all other cases the maximum change in the estimate of the difference  $(d - b)$  was 0.05 and the implication for stationarity or non-stationarity remained unchanged. In the cases of NIKKEI 225 and PX Index in the first period we can also note the highest differences among the  $\alpha$  and  $\gamma$  parameters in the estimation with and without the restriction. In the other specifications, these six parameters vary slightly, but nowhere near as much as in these two cases. The adjustment coefficients  $\alpha_H$  are again negative, and the adjustment coefficients  $\alpha_L$  remain positive. When interpreting these signs, we can make use of the fact that the cointegrating vector is now the range. An increase in daily range is reduced in the next day by decreasing the high log-price and boosting the low log-price for that day. The short-run dynamics

parameters ( $\gamma_{11}, \dots, \gamma_{22}$ ) are again mostly negative for the lagged daily highs and mostly positive for the lagged daily lows; the interpretation remains the same as in the previous specification. We can conclude that the imposition of the  $(1, -1)$  restriction on the cointegrating vector may have impaired the results, even though the original estimates of the cointegrating vector were fairly close to the desired  $(1, -1)$ .

**Table 9: FCVAR estimation results (with restrictions)**

	$\hat{d}$	$\hat{b}$	$\hat{\beta}$	$\alpha_H$	$\alpha_L$	$\gamma_{11}$	$\gamma_{12}$	$\gamma_{21}$	$\gamma_{22}$
<b>S&amp;P 500</b>									
<i>u</i> 2003-2012	1.035	0.339	(1,-1)	-0.258	3.483	-0.692	0.846	-1.362	2.139
<i>d</i> 2003-2007	1.000	0.426	(1,-1)	-1.114	1.834	0.145	0.129	-0.396	0.997
<i>u</i> 2008-2012	1.024	0.366	(1,-1)	-0.154	2.788	-0.646	0.819	-0.827	1.583
<b>FTSE 100</b>									
2003-2012	1.013	0.390	(1,-1)	-0.623	2.135	-0.303	0.487	-0.595	1.286
<i>d</i> 2003-2007	1.000	0.419	(1,-1)	-1.055	1.890	0.078	0.140	-0.743	1.295
2008-2012	1.018	0.468	(1,-1)	-0.593	1.215	-0.224	0.369	0.021	0.515
<b>DAX</b>									
2003-2012	1.018	0.409	(1,-1)	-0.398	1.876	-0.392	0.543	-0.495	1.039
2003-2007	1.054	0.446	(1,-1)	-0.685	1.670	-0.103	0.253	-0.252	0.591
2008-2012	1.013	0.373	(1,-1)	-0.098	2.343	-0.647	0.795	-0.790	1.385
<b>NIKKEI 225</b>									
2003-2012	1.028	0.489	(1,-1)	-0.210	1.511	-0.342	0.539	-0.202	0.645
2003-2007	1.046	0.507	(1,-1)	-0.590	1.631	-0.138	0.261	-0.468	0.828
2008-2012	1.001	0.516	(1,-1)	-0.079	1.123	-0.295	0.542	0.188	0.297
<b>PX</b>									
2003-2012	1.039	0.455	(1,-1)	-0.753	1.510	-0.201	0.306	-0.642	0.993
2003-2007	1.053	0.520	(1,-1)	-0.624	1.049	0.041	0.160	-0.121	0.493
<i>d</i> 2008-2012	1.000	0.427	(1,-1)	-0.863	1.493	-0.252	0.399	-0.727	1.122

*Note:* “*u*” denotes that model is unstable (some roots of characteristic polynomial lie outside the unit root circle), “*d*” denotes that the restriction  $d = 1$  had to be imposed in order to achieve model convergence.

The residual diagnostics are very similar to the previous case. According to the Ljung-Box *Q*-test, there is some marginal autocorrelation in the residuals, which may bias the standard errors; also, there is some small heteroscedasticity in the residuals based on the visualization of the ACF of squared residuals; however, as this model does not provide any standard errors, neither of these shortcomings changes any of our inferences.<sup>ix</sup>

<sup>ix</sup> For the sake of brevity, this residual diagnostic is not presented here, but is available upon request.

Finally, having all the estimates we can make the final comparison of the level of integration of range, which was our prior interest. We have first examined the stationarity of range based on the ADF and KPSS tests. The conclusion of these tests was, as expected, not unanimous. ADF test indicated that the range was stationary, which was not consistent with the results of the KPSS test and our inspection of the autocorrelation functions of range. Except for the PX Index in the first period, and possibly the NIKKEI 225 in the first period, in all other specifications we could see a substantial amount of dependence in the autocorrelation function, which is an indication of the possibility of long memory in range. Formal tests for long memory were applied, namely the ELW and GPH estimators. Both estimators confirmed our hypothesis of long memory in range. These estimators have also confirmed the stationarity of range for both the NIKKEI 225 and PX Index in the first period and the non-stationarity of the remaining indices in all periods. Finally, we have estimated the long memory of range using the fractionally cointegrated VAR model, after imposing a restriction to the values  $(1, -1)$  on the cointegrating vector. Due to this restriction, the order of integration of range was represented by the difference between the order of integration of daily log-prices and the cointegration gap, i.e. the integration order of range was the difference  $(d - b)$ . All these results are summarised in Table 10. In the first four columns of Table 10, we present the integration orders of range from GPH and ELW estimators with both bandwidth specifications we considered earlier in the study. The integration orders of range from the FCVAR framework (both with and without the imposed restriction) are displayed in the last two columns of Table 10.

In Table 10 we can note that the estimates of long memory in range are quite sensitive to both the chosen methodology and the chosen bandwidth parameter. In the case of S&P 500, even though there were some problems with the stability of the FCVAR model, we can see that the FCVAR estimates of long memory in range are in line with both the GPH and ELW estimates (with the smaller bandwidth parameter). For the FTSE 100 index in the second period, the FCVAR estimates of long memory appear to be quite underestimated, as the values would imply that the second period is less non-stationary than the first one, which contrasts all other results (ELW and GPH estimates and inspection of ACF). In the case of DAX, we can note that each estimator chooses

different period to be the most and the least non-stationary, despite the fact that the estimates are in magnitude close to each other.

**Table 10: Comparison of integration orders of range**

	GPH		ELW		FCVAR	
	$m = T^{0.5}$	$m = T^{0.6}$	$m = T^{0.5}$	$m = T^{0.6}$	R	NR
<b>S&amp;P 500</b>						
2003-2012	0.765	0.674	0.763	0.622	0.696	0.696
2003-2007	0.520	0.678	0.587	0.574	0.574	0.532
2008-2012	0.753	0.721	0.664	0.687	0.658	0.647
<b>FTSE 100</b>						
2003-2012	0.710	0.712	0.658	0.632	0.623	0.613
2003-2007	0.515	0.596	0.622	0.603	0.581	0.604
2008-2012	0.739	0.701	0.628	0.617	0.550	0.524
<b>DAX</b>						
2003-2012	0.652	0.706	0.588	0.589	0.609	0.599
2003-2007	0.803	0.826	0.614	0.597	0.608	0.566
2008-2012	0.710	0.738	0.700	0.623	0.640	0.631
<b>NIKKEI 225</b>						
2003-2012	0.499	0.659	0.483	0.649	0.539	0.487
2003-2007	0.463	0.401	0.496	0.392	0.539	0.384
2008-2012	0.701	0.784	0.613	0.604	0.485	-
<b>PX</b>						
2003-2012	0.577	0.617	0.516	0.520	0.584	0.532
2003-2007	0.211	0.316	0.281	0.446	0.533	0.341
2008-2012	0.760	0.581	0.572	0.517	0.573	0.528

*Note:* “R” denotes model with restriction on the cointegrating vector; “NR” denotes model without restriction.

The FCVAR results for NIKKEI 225 are quite poor. Contrary to the ELW and GPH estimates, the FCVAR with restrictions does not confirm the stationarity of range in the first period, and it overestimates the order of integration. Moreover, in the second period the dependence is significantly underestimated. This may be caused by the fact that the FCVAR model does not work very well for NIKKEI 225 in the second period (the model does not converge when no restrictions are imposed). We can see that the FCVAR model also fails to confirm the stationarity of the range of the PX Index in the first period, and overestimates the dependence as well. The FCVAR results in the two

remaining periods are quite similar to both the ELW and GPH estimates (with the bigger bandwidth parameter).

Overall we can see that the FCVAR model with restrictions fails to detect the lower orders of integration of range and suggests that the range is in the non-stationary region, when it should be stationary according to the results from other applied tests. However, we should note that when the restriction on the cointegrating vector was not imposed, the FCVAR model did detect the stationarity of the “range”. However, without the restriction, interpreting the error correction term in the model as range is incorrect, even though the cointegrating vector was fairly close to the value required for the interpretation to be valid even in the unrestricted model. Also, when no restrictions are imposed, we obtain the standard errors of the estimated parameters, whereas no standard errors are provided with the restricted model, which makes it impossible to draw an inference about the significance of the estimates.

The most unanimous conclusion is that apart from the ranges of PX Index and NIKKEI 225 in the first period, which are in the stationary region, the remaining ranges are non-stationary and display long memory. The best results can be seen in the case of S&P 500, where all four different methods for examining long memory yield results closest to each other, despite the instability of the FCVAR model for this index.

## 6 Conclusion

In this thesis, we provide unique empirical support for fractionally cointegrated daily high and low stock prices. These maximum and minimum daily prices are examined because they provide valuable information about range-based volatility. The range, defined as a difference between daily high and low log-prices, is considered a highly efficient and robust estimator of volatility (Parkinson, 1980).

We apply the fractionally cointegrated VAR methodology formalized only recently by Johansen and Nielsen (2012) to model the high and low prices. This framework is able to capture both the cointegration between daily high and low log-prices and the long memory of their linear combination, the range. In this concept, the range is, in fact, the error correction term in the FCVAR model and is allowed to fall into a non-stationary region.

Since the original definition of cointegration by Granger (1986) and Engle and Granger (1987), the case when series are integrated of order 1 and their linear combination is integrated of order 0 have been predominantly considered in the literature. However, this may be too restrictive and that is why we apply here the very general case when the series can be integrated of order  $d$  and cointegrated of order  $(d - b)$ , where  $d, b \in \mathbb{R}$ ,  $0 < b \leq d$ . If the cointegrating relation has an interpretation of range, the order of the fractional cointegration between daily high and low log-prices determines the stationarity and long memory of range. If  $(d - b) > 0$ , we then say that range displays long memory, and if  $(d - b) > 0.5$ , we say that range is non-stationary.

The analysis is performed on four major global indices, namely, the American S&P 500, the German DAX, the Japanese NIKKEI 225 and the UK's FTSE 100, and the results are compared to the Czech PX Index. We consider three periods, the base period being 2003-2012, and its division into two sub-periods, with year 2007 as the breakpoint. The first sub-period captures the relatively calm time before the crisis, and the second period covers the outbreak of the crisis and the post-crisis turbulences.

Our analysis is conducted as follows. We first test the stationarity of daily high and low log-prices and of price range. The ADF test implies that range is stationary, which contradicts findings from the KPSS test and the inspection of autocorrelation functions, where a significant dependency is visible. We thus employ the Exact Local Whittle and Geweke Porter-Hudak estimators of long memory and confirm the existence of long memory in range. The evidence of long memory in asset price volatility was also found for example by Ding, Granger and Engel (1993), Andersen and Bollerslev (1997), Breidt et al. (1998), Kellard et al. (2010) or Garvey and Gallagher (2012).

Since the general well-known tests for cointegration proposed by Johansen (1991) are not suitable when series display long memory, the cointegration of daily high and low log-prices is tested in the fractional framework. We employ the cointegration rank test proposed by Nielsen and Shimotsu (2007) with the result of one cointegrating vector. This finding is also confirmed by the cointegration rank test, which is part of the FCVAR estimation procedure proposed by Johansen and Nielsen (2012). Two FCVAR specifications are considered, one without any restrictions imposed and one with a restriction on the cointegrating vector in order to get an interpretation of range.

We find significant evidence that range-based volatility estimate displays a long memory. Moreover, in most cases range is in the non-stationary region (i.e. is integrated of order greater than 0.5). Only the ranges of the PX Index and NIKKEI 225 in the first (2003-2007) period are in the stationary region. Overall, the estimates of the long memory are quite sensitive to the selected methodology. In general, the first period is rather the calmer period, in which the estimates of the long memory parameters are mostly the lowest. The second period (2008-2012) is more turbulent. We also show that in the FCVAR framework with restrictions imposed the results are slightly inferior to the original unrestricted FCVAR model, mainly in the situation when range should be in the stationary region based on other applied tests. Furthermore, the results for the PX Index are very similar to the results for the NIKKEI 225 as their ranges display the lowest estimates of integration orders. Integration orders of ranges of S&P 500, FTSE 100 and DAX indices are, on the other hand, relatively higher.

These results can be quite useful for the predictability of asset prices. The fact that we find a long memory in the range allows the predictability of the variance to be embedded in a model for the mean dynamics of high and low prices and to obtain better forecasts of the future extreme prices based on past values. Our results thus provide compelling evidence that daily high and low prices are predictable and can be modelled.

## 7 References

- Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002)  
Range-Based Estimation of Stochastic Volatility Models, *The Journal of Finance*, 57(3), 1047–1091.
- Andersen, T. G., and T. Bollerslev (1997)  
Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns, *The Journal of Finance*, 52(3), 975–1005.
- Andersen, T., and T. Bollerslev (1998)  
Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, *International Economic Review*, 39(4), 885–905.
- Baillie, R. T. (1996)  
Long Memory Processes and Fractional Integration in Econometrics, *Journal of Econometrics*, 73(1), 5–59.
- Breidt, F. J., N. Crato, and P. de Lima (1998)  
The detection and estimation of long memory in stochastic volatility, *Journal of Econometrics*, 83 (1–2), 325–348.
- Brockwell, P.J., and R.A. Davis (1991)  
*Time Series: Theory and Methods*, 2nd edition, Springer-Verlag, New York.
- Caporin, M., Ranaldo, A., and P. Santucci de Magistris (2013)  
On the Predictability of Stock Prices: A Case for High and Low Prices, *Journal of Banking & Finance* 37(12), 5132–5146.
- Cheung, Y. (2007)  
An empirical model of daily highs and lows, *International Journal of Finance and Economics*, 12(1), 1–20.
- Cipra, T. (2008)  
*Finanční ekonometrie, (Financial Econometrics)*, first edition, Ekopress 2008.
- Corwin, S. A., and P. H. Schultz (2012)  
A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices, *The Journal of Finance*, 67(2), 719–760.
- Dickey, D. A., and W. A. Fuller (1981)  
Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica: Journal of the Econometric Society*, 1057-1072.

- Ding, Z., Granger, C.W.J., and R.T. Engle (1993)  
A long memory property of stock market returns and a new model, *Journal Empirical Finance* 1(1), 83–106.
- Engle, R., and C.W.J. Granger (1987)  
Cointegration and error correction: representation estimation, and testing, *Econometrica*, 55(2), 251–276.
- Garman, M. B., and M. J. Klass (1980)  
On the Estimation of Security Price Volatilities from Historical Data, *The Journal of Business*, 53(1), 67–78.
- Garvey, J. F., and L.A. Gallagher (2012)  
The Realised–Implied Volatility Relationship: Recent Empirical Evidence from FTSE–100 Stocks. *J. Forecast.*, 31(7), 639–660.
- Geweke, J., and S. Porter–Hudak (1983)  
The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4(4), 221–238.
- Granger, C.W.J. (1980)  
Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics*, 14(2), 227–238.
- Granger, C.W. J. (1986)  
Developments in the study of cointegrated economic variables, *Oxford Bulletin of Economics and Statistics*, 48(3), 213–228.
- Granger, C.W.J., and N. Hyung (2004)  
Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of Empirical Finance*, 11(3), 399–421.
- Granger, C. W. J., and R. Joyeux, (1980)  
An introduction to long memory time series models and fractional differencing, *Journal of Time Series Analysis*, 1(1), 15–29.
- Greene, W. H. (2008)  
*Econometric Analysis*, Sixth Edition, Prentice Hall: New Jersey
- Johansen, S. (1991)  
Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* 59(6), 1551–1581.

- Johansen, S. (1996)  
Likelihood–Based Inference in Cointegrated Vector Autoregressive Models, 2nd edition, Oxford University Press, Oxford.
- Johansen, S. (2008)  
A representation theory for a class of vector autoregressive models for fractional processes, *Econometric Theory*, 24(3), 651–676.
- Johansen, S., and M. O. Nielsen (2010)  
Likelihood inference for a nonstationary fractional autoregressive model, *Journal of Econometrics* 158(1), 51–66.
- Johansen, S., and M. O. Nielsen (2012)  
Likelihood inference for a fractionally cointegrated vector autoregressive model, *Econometrica*, 80(6), 2667–2732.
- Kanzler, L. (1998)  
GPH: MATLAB module to calculate Geweke–Porter–Hudak long memory statistic, Statistical Software Components.
- Kellard, N., Dunis, C., and N. Sarantis (2010)  
Foreign exchange, fractional cointegration and the implied–realized volatility relation, *Journal of Banking & Finance*, 34(4), 882–891.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Y. Shin (1992)  
Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root, *Journal of Econometrics* 54(1), 159–178.
- Lobato, I.N., and N.E. Savin (1997)  
Real and spurious long–memory properties of stock–market data, *Journal of Business and Economic Statistics*, 16(3), 261–268.
- MacKinnon, J. G., and M. Ø. Nielsen (2012)  
Numerical distribution functions of fractional unit root and cointegration tests, *Journal of Applied Econometrics*.
- NBER  
US Business Cycle Expansions and Contractions, available at: <http://www.nber.org/cycles/>, last accessed on 15. 5. 2012.
- Nielsen, M., and L. Morin (2012)  
FCVARmodel.m: A MATLAB software package for estimation and testing in the fractionally cointegrated VAR model, QED working paper 1273, Queen’s University.

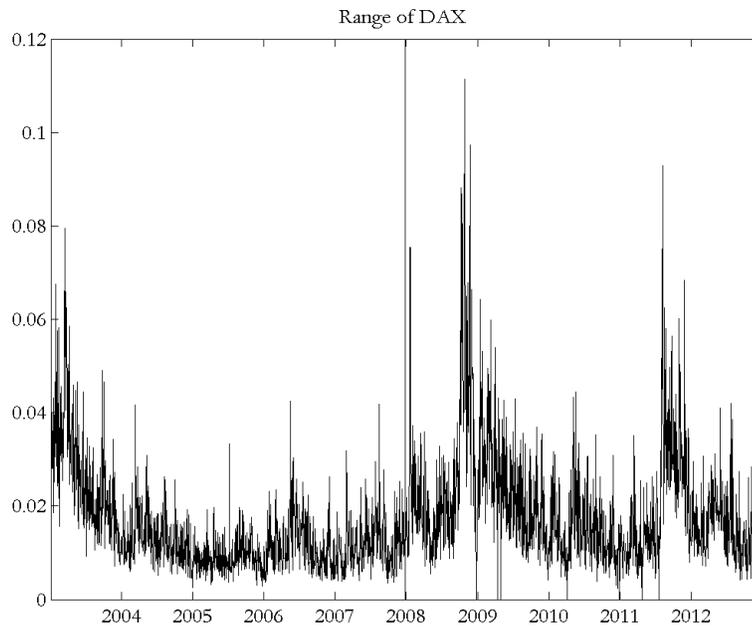
- Nielsen, M. Ø., and K. Shimotsu (2007)  
Determining the cointegration rank in nonstationary fractional system by the exact local Whittle approach, *Journal of Econometrics*, 141(2), 574–596.
- Parkinson, M. (1980)  
The Extreme Value Method for Estimating the Variance of the Rate of Return, *The Journal of Business*, 53(1), 61–65.
- Robinson, P. M., and Y. Yajima (2002)  
Determination of cointegrating rank in fractional systems, *Journal of Econometrics*, 106(2), 217–241.
- Shimotsu, K., and P. Phillips (2005)  
Exact local Whittle estimation of fractional integration, *Annals of Statistics*, 33(4), 1890–1933.

## 8 Appendix

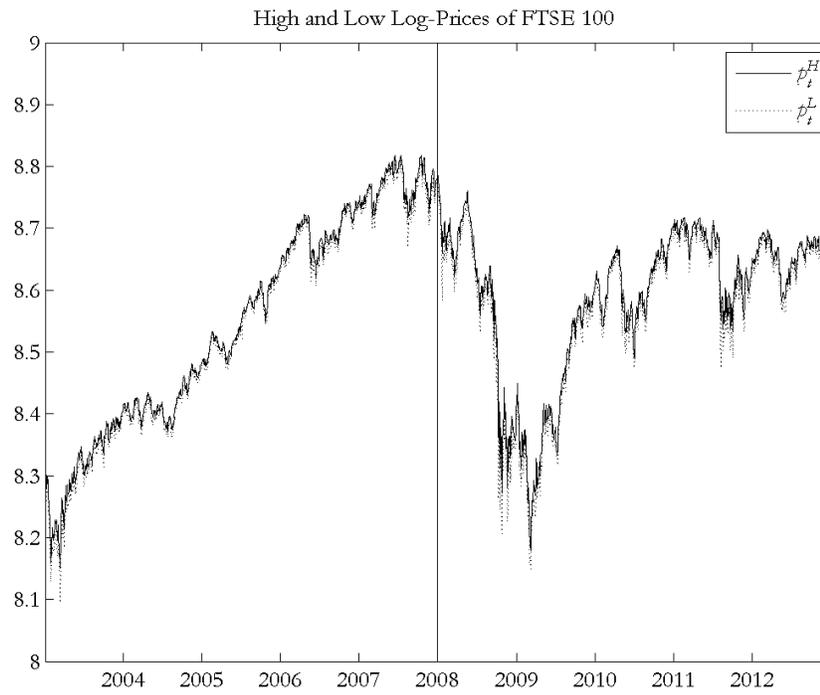
Figure 6: High and Low Log-Prices of DAX



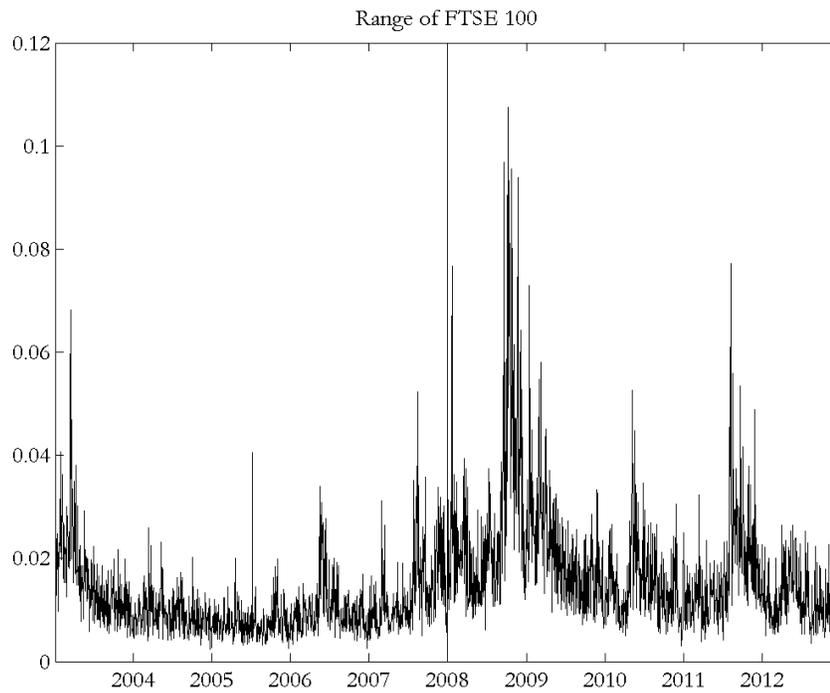
Figure 7: Range of DAX



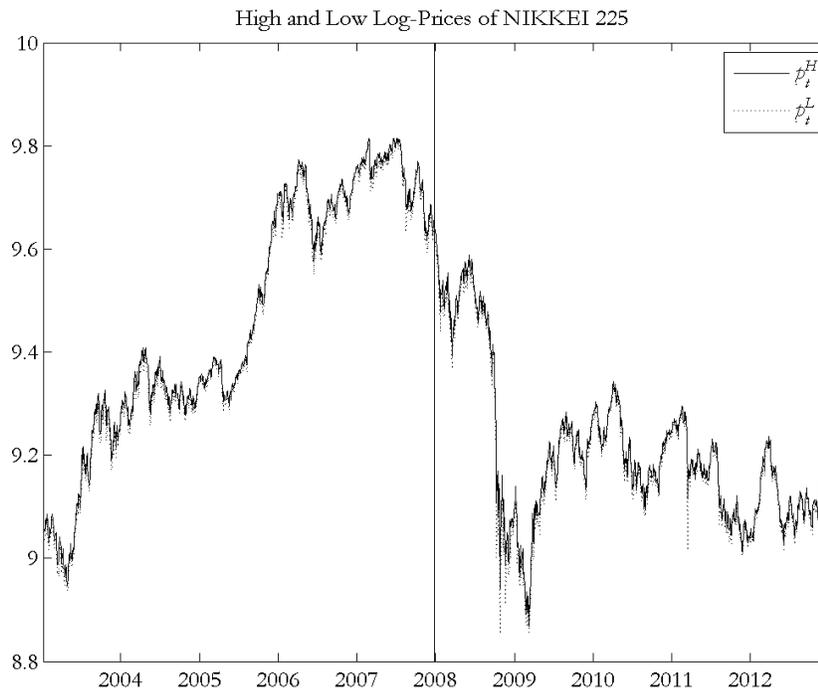
**Figure 8: High and Low Log-Prices of FTSE 100**



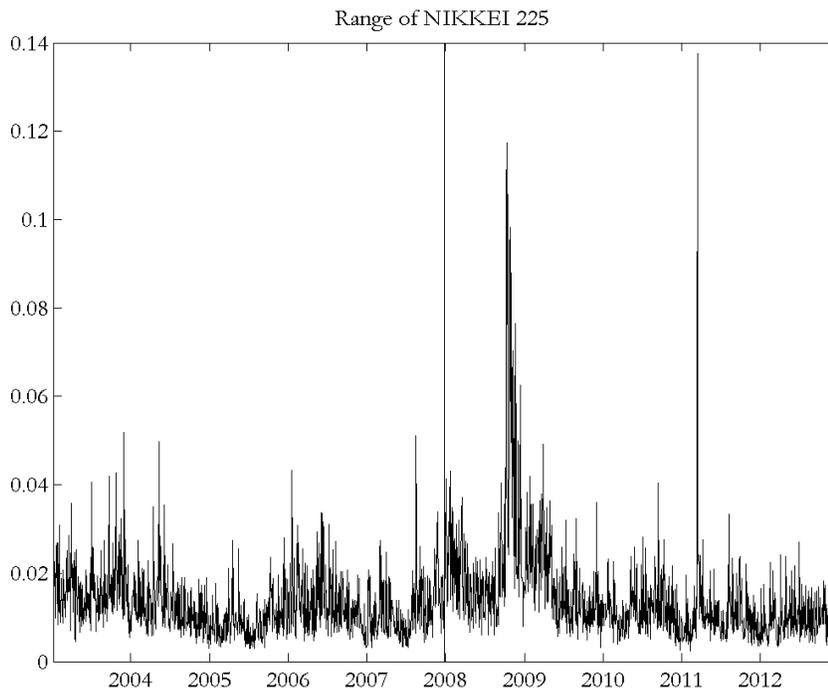
**Figure 9: Range of FTSE 100**



**Figure 10: High and Low Log-Prices of NIKKEI 225**



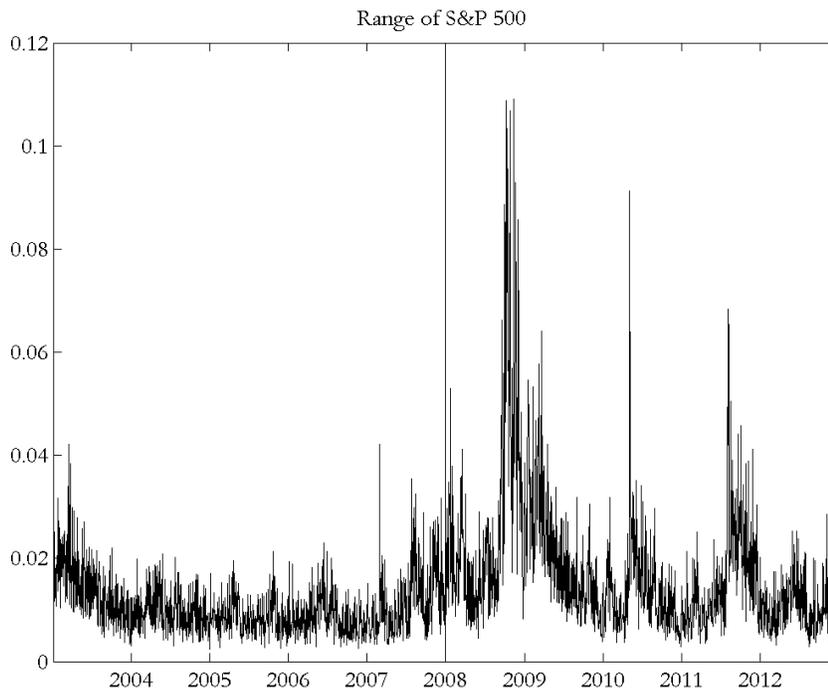
**Figure 11: Range of NIKKEI 225**



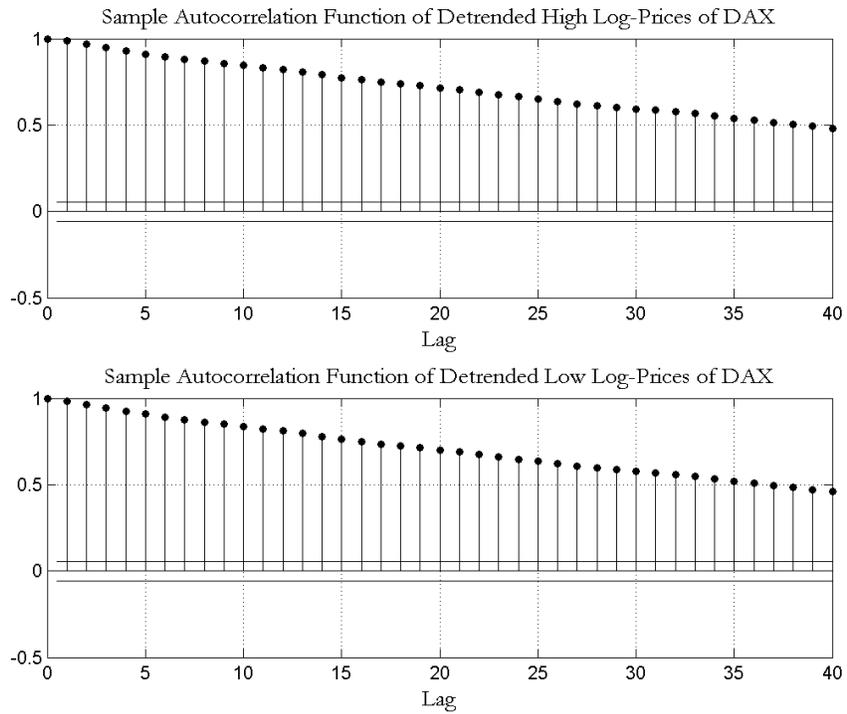
**Figure 12: High and Low Log-Prices of S&P 500**



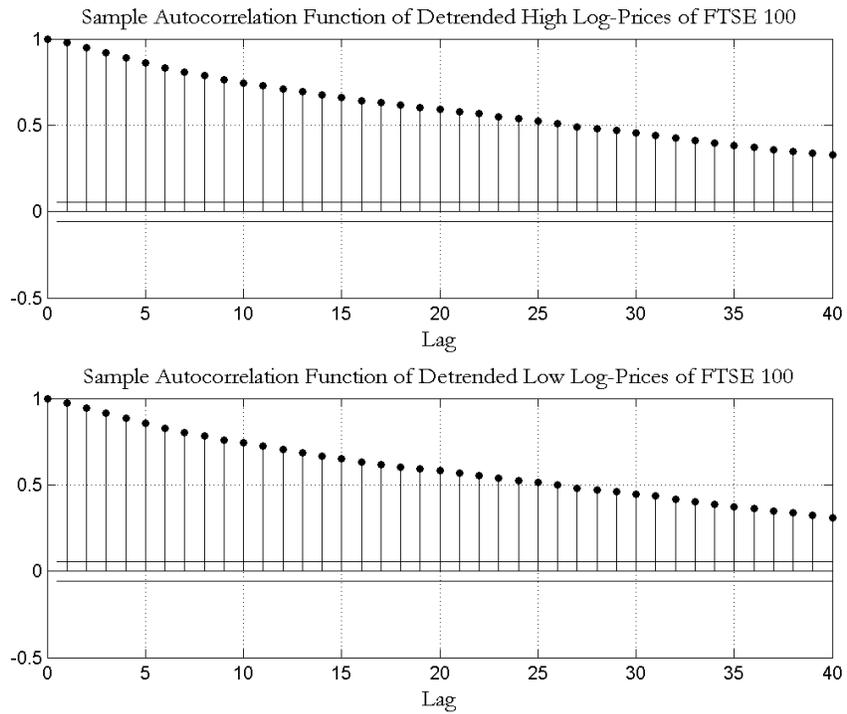
**Figure 13: Range of S&P 500**



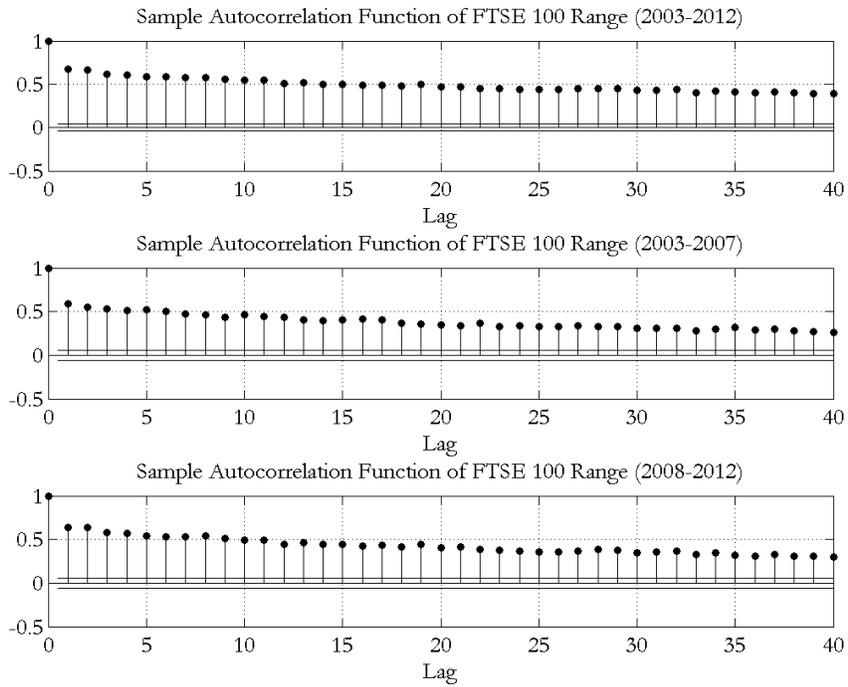
**Figure 14: ACF of Detrended High and Low Log-Prices of DAX (2003-2007)**



**Figure 15: ACF of Detrended High and Low Log-Prices of FTSE 100 (2003-2007)**



**Figure 16: ACF of FTSE 100 Range in All Periods**



**Figure 17: ACF of S&P 500 Range in All Periods**

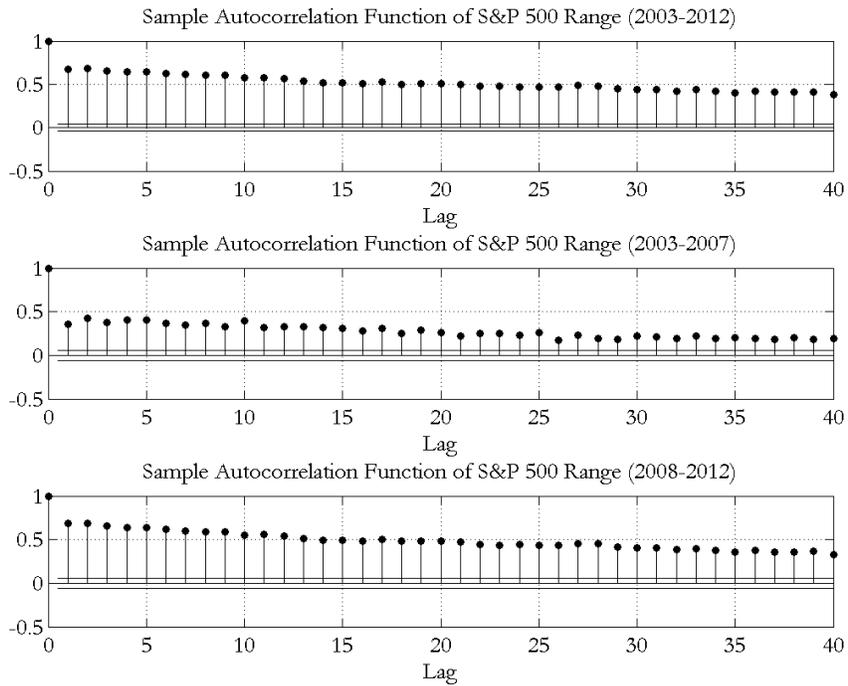


Figure 18: ACF of NIKKEI 225 Range in All Periods

