

Analyzing Oil Futures with a Dynamic Nelson-Siegel Model

Niels S. Hansen and Asger Lunde

CREATES Research Paper 2013-36

Analyzing Oil Futures with a Dynamic Nelson-Siegel Model*

Niels S. Hansen^{†‡} Asger Lunde^{†§}

Abstract

In this paper the dynamic Nelson-Siegel model is used to model the term structure of futures contracts on oil and obtain forecasts of prices of these contracts. Three factors are extracted and modelled in a very flexible framework. The outcome of this exercise is a class of models which describes the observed prices of futures contracts well and performs better than conventional benchmarks in realistic real-time out-of-sample exercises.

Keywords: Oil futures, Nelson-Siegel, Normal Inverse Gaussian, GARCH, Copula.

JEL classification: G17, C32, C53.

*We acknowledge the support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. Comments from participants of the Conference of Energy Finance 2012 in Trondheim and the CREATES seminar series are highly appreciated. We are grateful for the comments from Thomas Quistgaard Pedersen and for the comments from two anonymous referees which have improved the quality of the paper significantly.

[†]Department of Economics and Business, School of Business and Social Sciences, Aarhus University and CREATES. Fuglesangs Allé 4, DK-8210 Aarhus V.

[‡]Corresponding author, e-mail: nhansen@econ.au.dk, phone: (45) 8716 5285.

[§]Associate Member at Oxford-Man Institute of Quantitative Finance

1 Introduction

In the aftermath of the financial crisis commodity markets have received a lot of attention. Caballero et al. (2008) argue that commodity markets presented a more reliable and attractive form of investment when the financial crisis spread throughout the financial markets. The growth in economic activity in Brazil, Russia, India and China has, according to Geman (2005), contributed to the increased popularity and to increased prices across commodity markets. Predicting prices of commodity futures contracts is very interesting from an academic perspective and very valuable for producers, speculators and risk managers. We analyze a large dataset consisting of prices and time to maturity of futures contracts on oil. We focus on oil because it is the most traded commodity in the world. It is used in the production of many products most importantly for petroleum. Oil also serves as an important indicator for the overall state of the world economy. In this paper we take on some of the challenges presented by data from commodity markets. We obtain very valuable insights about the prices of commodity futures and how to forecast them.

The objective in this paper is to develop a class of models which explains data well and performs well out of sample. This paper makes several contributions. First, we show that the dynamic Nelson-Siegel model can be used to extract a set of underlying factors which describe the prices of futures contracts on oil. Second, we model the factors using the class of GARCH models with Normal Inverse Gaussian innovations in a very flexible copula framework. Third, we successfully forecast the prices of futures contracts on oil based on both a conventional mean squared error criterion and a directional criterion. Finally, we show how our model can be used to calculate value at risk for portfolios consisting of the factors.

Several papers consider the modelling of oil futures. The traditional approach is to specify a model for the underlying spot price and derive a term structure for futures prices based on a no arbitrage argument, see Schwartz (1997) and Geman (2005) for examples. In this paper we do not consider spot prices or their relationship to the prices of futures contracts. We focus exclusively on modelling and forecasting the term structure of futures contracts.

The data in this market, as well as data from other commodity markets, has a structure which resembles data from fixed income markets. Therefore models from the fixed income literature, in particular interest rate models, are often applied to analyze commodity futures contracts. In this paper we apply a classical interest rate model to prices of oil futures. The model is the dynamic Nelson-Siegel model, Nelson and Siegel (1987) and Diebold and Li (2006), and the modelling approach is inspired by Noureldin (2011).

The analysis is based on the assumption that the relationship between futures prices and time to maturity can be described by three latent factors. The validity of this assumption will be determined by how well our models perform out of sample. By focussing the analysis on these factors the dynamic Nelson-Siegel model allows for a substantial dimension reduction which facilitates the analysis. The dynamic Nelson-Siegel model is also used with great success in areas other than yield curve modelling, see for example Chalamandaris and Tsekrekos (2011) and Guo et al. (2014) where implied volatility is modelled in this framework. In the commodity literature West (2011) uses the dynamic Nelson Siegel model to estimate futures prices for long dated contracts on agricultural products. The methodology is, however, different from the one we apply. West (2011) relies on different versions of the Nelson-Siegel model to take seasonality and other features into account. Instead we estimate the latent factors and obtain multivariate time series. The idea is then to model and forecast these factors in order to forecast futures prices. We show that models based on the dynamic Nelson-Siegel model performs well in realistic real-time exercises as forecasting and value at risk analysis.

Using techniques from the copula framework in Patton (2009) and Patton (2012), we apply a decomposition which allows us to model three univariate time series and a dependence structure individually. We may then draw on the vast literature on modelling of univariate time series. In particular, the very flexible class of Normal Inverse Gaussian GARCH (NIG-GARCH) models presented in Jensen and Lunde (2001), is shown to describe the data well and at the same time offer accurate forecasts of the factors. We consider two different models for the dependence..

We leave out a part of our sample for real-time forecast evaluation. In this period we forecast the factors and hence the term structure of futures prices using our model and compare the results to two other models. We show that we forecast more precisely than our benchmarks in terms of mean squared error. Practitioners may not, however, be interested in the mean squared error of a model but rather the models ability to accurately predict the direction of the changes of the futures prices. We carry out an analysis to investigate this desirable property of the different models. Finally, we show that portfolios of oil futures can be constructed from the estimated factors and that we can successfully calculate value at risk for these portfolios in our framework.

The rest of this paper is organized as follows. In section 2 we provide a thorough description of the data set analyzed in this paper. Section 3 contains a description of the model. In section 4 we present the results of the in-sample analysis. Section 5 contains the results of the out-of-sample forecast analysis. Section 6 is devoted to the calculation

and backtesting of value at risk. Finally, some concluding remarks are presented in section 7.

2 Data

Data from the markets for commodity futures contracts is very different from many other financial data sets. The unique features of the data make analyzing and forecasting challenging tasks. This section serves to illustrate the features of the data and highlight potential challenges which we have to overcome.

The data set we consider consists of daily closing prices of monthly futures contracts on oil from Reuters.¹ The contracts are on light sweet crude oil (WTI), more details can be found on the CME Group homepage.² Every day a number of futures with different time to maturity are traded. The maturities are approximately one month apart. Each contract expires on the third trading day prior to the 25th calendar day in the month before delivery. The first observations are from June 1st 2000 and the sample ends on December 31st 2012 meaning that our sample consists of 55.123 observations of prices and maturities.

It is important to understand the nature of the data in order to understand the problems we face in this paper. In order to grasp the structure and complexities of the data it is useful to consider an example of actual data. In Table 1 we have presented a small example of what the data looks like. This is only a very brief example, but there are several important things to notice. First, note that only a limited number of contracts exists on a given day. Consider for example the first row of the table, on this day a contract with 495 days to delivery does not exist. Such a contract exists on the next day, though. Secondly, the data contains a number of holes. On 06.09.2010 it was possible to trade in a contract with 494 days to maturity, but no one did. This means that we have two kinds of holes in the data set contracts that do not exist and contracts which exist but are not traded. Both pose problems for a traditional time series analysis which is not well suited for dealing with situations in which the number of observations changes from day to day. Assume now that on 06.10.2010 we are interested in forecasting the price of a contract with 492 days to delivery which is potentially traded on 06.11.2010. Futures contracts are characterized by their time to maturity, so a good way of forecasting is to consider a time series of prices of contracts with the same time to maturity. Such a time

¹The data is kindly provided by Oxford-Man Institute of Quantitative Finance and obtained from Thomson Reuters Tick History with help from Kevin Sheppard.

²http://www.cmegroup.com/trading/energy/crude-oil/light-sweet-crude_contract_specifications.html

Table 1: Data example.

Contract name	Price (time to maturity (trading days))					
	2012M	2012Z	2013M	2013Z	2014Z	
06.07.2010	... 81.50 (496)	82.48 (622)	-	85.00 (875)	85.65 (1127)	...
06.08.2010	... 81.37 (495)	83.10 (621)	83.70 (746)	84.60 (874)	86.22 (1126)	...
06.09.2010	... -	84.70 (620)	-	85.95 (873)	87.60 (1125)	...
06.10.2010	... 83.70 (493)	85.60 (619)	-	86.85 (872)	87.91 (1124)	...

Example of actual data. A subsample of the prices and corresponding maturities for contracts traded between 06.07.2010 and 06.10.2012 for five different contracts. The first part of the name for each contract is the year of delivery. The letter in the contract name denotes the month of delivery. M is June and Z is December.

series is not observed, but it could be constructed by interpolation, see Diebold and Li (2006) or Noureldin (2011) for examples from the interest rate literature. Another possibility is to consider the time series of prices for this particular contract, 2012M (delivery in June 2012) and fit an autoregressive model. The drawbacks here are that the previous observations of prices are based on another time to maturity. Furthermore, we have to decide what to do about missing values. If the contract was introduced to the market recently this approach might be infeasible due to the limited number of observations. Note, that we always know the maturities of the contracts which are potentially traded on the next day.

The model we apply in this paper is introduced by Diebold and Li (2006). They apply the model to monthly yields on U.S. Treasuries. We notice, that our data resembles the data from their analysis. Therefore, we apply the same model to model the prices of futures contracts on oil. Diebold and Li (2006) present the most important stylized facts of the yield curve dynamics and argue that their model in theory should be able to account for all these. We will argue, that the stylized facts presented by Diebold and Li (2006) are all present in our current data set. Diebold and Li (2006) consider the five following characteristics:

- (1) The average yield curve is increasing and concave.
- (2) The yield curve assumes a variety of shapes through time including upward sloping, downward sloping, humped and inverted humped.
- (3) Yield dynamics are persistent, and spread dynamics are much less persistent.
- (4) The short end of the yield curve is more volatile than the long end.
- (5) Long rates are more persistent than short rates.

We are going to present our data set and adapt the above characteristics to our framework and argue that prices of oil futures exhibit the same characteristics. To get an idea of which contracts we have available, we first present a plot of the observed maturities over time in Figure 1. Each dot in this plot indicates an observed price. Several things are worth noticing. Generally it seems that we have a lot of observations in the short end throughout the sample, while observations in the long end are more scarce. Time to maturity is measured in business days. Up until 2007 we have observations of contracts with maturities around 1700 days. From 2007 we have contracts with long time to maturity around 2200 days. It seems that, for long maturities, trading is concentrated in contracts where maturities are about 12 months apart. This is represented by the diagonal "lines". Transactions for medium long maturities are concentrated in contracts where maturity is 6 months apart. This can be seen as the diagonal lines are closer to each other for maturities between 300 and 500 days.

Diebold and Li (2006) analyses a data set of monthly yields and fix maturities by linear interpolation. This means that they analyze time series of 17 different contracts. We do not use linear interpolation which means that we have different maturities on different days. To enable a comparison of our data to the stylized facts we pool the data into 23 groups such that we have all prices of contracts with less than 100 days to maturity in the first group. Similarly all prices for contracts with 100 to 200 days to maturity are collected in the next group and so on. This division into groups is only used to highlight some of the stylized facts of Diebold and Li (2006) and not in the actual analysis.

The first of the characteristics presented above corresponds to a situation in which the average term structure of the futures contracts is increasing and concave. In the top panel of Figure 2 we present the average term structure. Generally, prices are increasing in maturity and it seems that averaged over time the term structure can be approximated reasonably well by a smoothly increasing function. The fourth characteristic states that the prices of contracts in the short end of the term structure are more volatile than the prices of contracts in the long end. To investigate this, we take a look at the standard deviation of the different maturities. In the middle panel of Figure 2 we have the standard deviations of the prices of contracts with different maturities. We see, that prices of contracts with short maturity are generally more volatile than contracts with long maturities. The standard deviations are quite high for some contracts in the medium range of time to maturity. Diebold and Li (2006) observes the same pattern. In the lower panel in Figure 2 we have plotted the natural logarithm of the number of observations for each group of maturities. Clearly, we have a fewer observations for

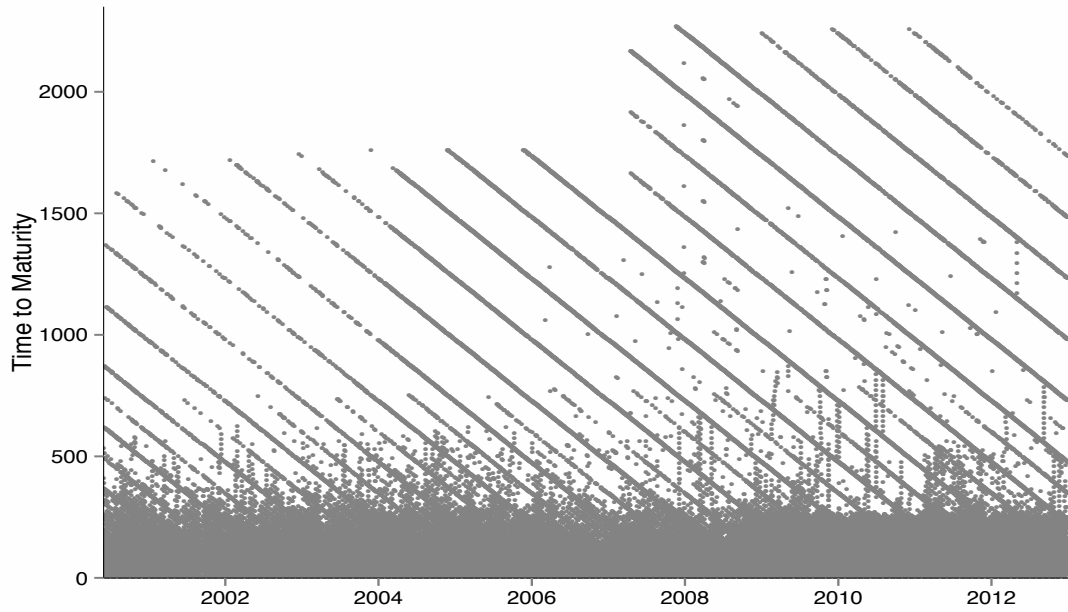


Figure 1: The data. Time to maturity of the observed contracts as a function of time. Each dot indicates an observation of a price.

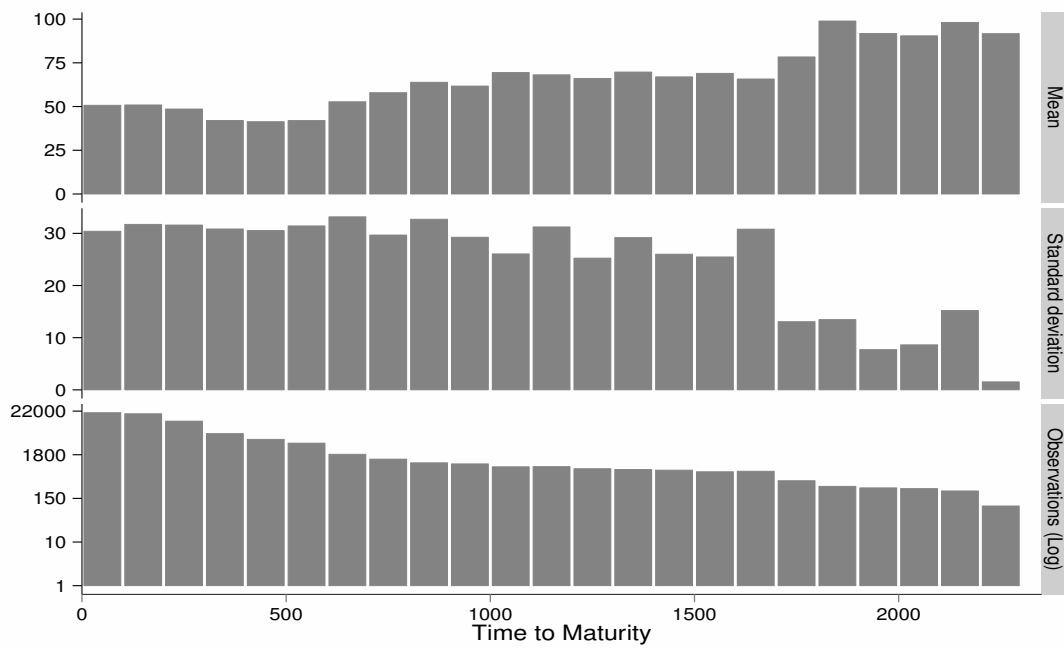


Figure 2: Stylized facts. Top panel: The average term structure of futures contracts on oil. Middle panel: The standard deviation of prices for different maturities. Bottom panel: The natural logarithm of the number of observations for each of the different maturities. The number of observations in the last plot has been transformed by the exponential function for convenience.

longer maturities, a problem not encountered in Diebold and Li (2006) because of the use of linear interpolation.

The term structure of futures contracts takes a variety of shapes through time. Four different examples are presented in Figure 3. In the top left panel of the figure we have a hump-shaped term structure. In the top right panel we have an increasing and concave term structure. The lower left panel is an example of a downward sloping term structure. Finally, the bottom right panel presents an example of a day, where the term structure has no clear shape.

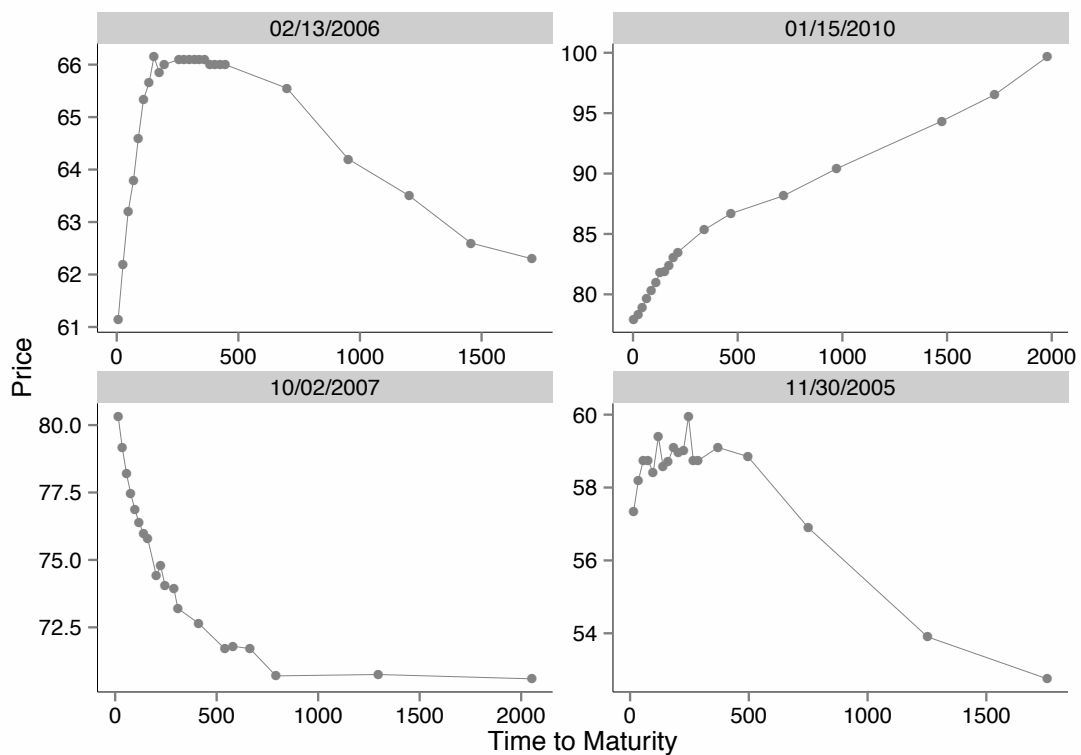


Figure 3: Different term structures. Term structures of futures contracts on oil for four different days. Top left panel: 07/29/2011. Top right panel: 08/12/2011. Bottom left panel: 05/02/2011. Bottom right panel: 05/23/2011.

Facts 3 and 5 above are difficult to handle in our setting. The structure of the data makes it infeasible to calculate the autocorrelations as in Diebold and Li (2006). Also notice, that these autocorrelations are calculated based on data which has been constructed by linear interpolation.

3 Modelling the Term Structure

The model we use is the Dynamic Nelson-Siegel model. The Nelson-Siegel model is introduced to model yield curves by Nelson and Siegel (1987). A dynamic version of the model is introduced by Diebold and Li (2006). It states that at any time, t , the k different yields can be explained by three factors according to the following model

$$\begin{pmatrix} Y_t(\tau_1) \\ Y_t(\tau_2) \\ \vdots \\ Y_t(\tau_{k_t}) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_{k_t}}}{\lambda\tau_{k_t}} & \frac{1-e^{-\lambda\tau_{k_t}}}{\lambda\tau_{k_t}} - e^{-\lambda\tau_{k_t}} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} e_t(\tau_1) \\ e_t(\tau_2) \\ \vdots \\ e_t(\tau_{k_t}) \end{pmatrix}$$

$$\mathbf{Y}_t = \mathbf{Z}_t \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \mathbf{e}_t \quad (1)$$

$\begin{matrix} k_t \times 1 & k_t \times 3 & 3 \times 1 & k_t \times 1 \end{matrix}$

We use this model to describe the relationship between futures prices and their time to maturity. \mathbf{Z}_t is the matrix of factor loadings. The dimensions of \mathbf{Z}_t , k_t , change over time depending on the number of contracts in the market on time t . Diebold and Li (2006) note that L_t , S_t and C_t can be interpreted as level slope and curvature effects, respectively. The interpretation comes from the resulting effects on the yield curve from changes in the factors. From (1) we see, that changes in L_t are constant across all maturities and that the loading on L_t determines the level of the curve. The loading on S_t determines the overall slope of the yield curve. In particular $Y_t(\infty) - Y_t(0) = -S_t$. Finally, C_t is interpreted as a curvature effect because the loading is zero on very short and very long maturities. Like Diebold and Li (2006) we extract the factors by a series of cross-section regressions.

λ determines the exponential decay rate and how well we fit the short end of the yield curve. A high value of λ gives a good fit in the short end, but it also causes an identification issue at longer maturities as S_t and C_t will become very difficult to distinguish. The value of λ also determines for which maturity the curvature factor attains its maximum. Time-varying specifications for λ are considered in the literature. Koopman et al. (2010) treat λ as a fourth latent factor and model it along with the three other factors. Hautsch and Ou (2012) also consider a time-varying λ , but conclude that this is not important for the models ability to fit the data. We do not consider a time-varying version of λ in this paper.

In order to model the factors we turn to the framework of copula modelling. Let the factors be represented by time series $x_{i,t}$ for $i = 1, 2, 3$, and assume that these are real-

izations of the random variables $\mathbf{X} = (X_1, X_2, X_3)'$. Now, we want to model their joint density, $f_{\mathbf{X}}(\mathbf{x})$. Copula modelling is based on the Theorem of Sklar, Sklar (1959). The theorem states that for any joint distribution function, $F_{\mathbf{X}}$, where \mathbf{X} is a g -dimensional random vector and \mathbf{x} its realizations, there exists a copula, C , such that

$$F_{\mathbf{X}}(\mathbf{x}) = C(F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_g}(x_g)),$$

where $F_{X_i}(x_i)$ denote the marginal cdfs for the g factors. The theorem is very useful as explained in Patton (2009), as it allows us to express the joint pdf as

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^g f_{X_i}(x_i) c \left(F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_g}(x_g) \right) \\ &= \prod_{i=1}^g f_{X_i}(x_i) c(u_1, u_2, \dots, u_g), \end{aligned} \quad (2)$$

where $f_{X_i}(x_i)$ for $i = 1, 2, \dots, g$ denote the pdfs for X_1, X_2, \dots, X_g , respectively. $c(\cdot)$ denotes the copula density. u_i are realizations from uniform distributed random variables, U_1, U_2, \dots, U_g . In our setting we work with three factors, hence $g = 3$.

Having established this, we turn to specifying the marginal distributions, $f_{X_i}(x_i)$, and the copula density, $c(\cdot)$.

3.1 Marginal Models

We must specify models for the conditional marginal distributions. To do this, we first introduce the information set available at time t , \mathcal{F}_t , which consists of all previous observations of the prices of futures contracts, maturities and hence factors. Each of the models are conditioning on a subset of \mathcal{F}_t , these are denoted $\mathcal{Z}_{i,t}$. The marginal models will have the following structure

$$\begin{aligned} x_{i,t} &= \mu_{i,t}(\mathcal{Z}_{i,t-1}) + \sigma_{i,t}(\mathcal{Z}_{i,t-1})\varepsilon_{i,t} \quad i = 1, 2, 3 \quad \text{and} \quad \mathcal{Z}_{i,t-1} \in \mathcal{F}_{t-1} \quad (3) \\ \varepsilon_{i,t} | \mathcal{F}_{t-1} &\sim F_{X_i}(0, 1) \quad \forall t. \end{aligned}$$

$F_{X_i}(0, 1)$ denotes the distribution of a standardized random variable. In order to keep the copula decomposition valid we must make sure that $\mathcal{Z}_{i,t}$ meet the requirements in Patton (2006) for $i = 1, 2, 3$. Informally speaking this requirement states that the distribution of x_i conditional on $\mathcal{Z}_{i,t-1}$ must be independent from the distribution of x_i conditional on $\mathcal{Z}_{j,t-1}$ for $j \neq i$ in the sense that

$$F_{X_i | \mathcal{Z}_{i,t-1}}(x_i | \mathcal{Z}_{i,t-1}) = F_{X_i | \mathcal{Z}_{i,t-1}, \mathcal{Z}_{j,t-1}}(x_i | \mathcal{Z}_{i,t-1}, \mathcal{Z}_{j,t-1}) \quad \text{for} \quad j \neq i \quad (4)$$

3.2 NIG-GARCH Model

In this analysis we use the NIG-GARCH class of models by Jensen and Lunde (2001) because of the great flexibility it offers. The Normal Inverse Gaussian (NIG) distribution and its properties are presented in Barndorff-Nielsen (1997) and Rydberg (1997). This distribution allows us to model both skewness and excess kurtosis and it nests the t-distribution as a special case and has the normal distribution as a limiting case. We assume that conditional on the partition $\mathcal{Z}_{i,t-1}$, $X_{i,t}$ follows a NIG distribution. That is

$$X_{i,t} | \mathcal{Z}_{i,t-1} \sim \text{NIG}(\bar{\alpha}, \bar{\beta}, m_{i,t}(\mathcal{Z}_{i,t-1}), \sigma_{i,t}(\mathcal{Z}_{i,t-1}))$$

$\bar{\alpha}$ and $\bar{\beta}$ are shape parameters. $\bar{\alpha}$ determines steepness and $\bar{\beta}$ determines symmetry. $m_{i,t}(\mathcal{Z}_{i,t-1})$ is a location parameter and $\sigma_{i,t}(\mathcal{Z}_{i,t-1})$ is a scale parameter. $\bar{\alpha}$ and $\bar{\beta}$ are specified as in Jensen and Lunde (2001) to ensure that the distribution is location and scale invariant. Introducing the notation $\bar{\gamma} = (\bar{\alpha}^2 - \bar{\beta}^2)^{1/2}$ we can write up the pdf for the NIG distribution

$$f_{X_i | \mathcal{Z}_{i,t-1}}(x_i | \mathcal{Z}_{i,t-1}) = \frac{\bar{\alpha}}{\pi \sigma_{i,t}(\mathcal{Z}_{i,t-1})} \exp \left[\bar{\gamma} + \bar{\beta} \frac{x_i - m_{i,t}(\mathcal{Z}_{i,t-1})}{\sigma_{i,t}(\mathcal{Z}_{i,t-1})} \right] q \left(\frac{x_i - m_{i,t}(\mathcal{Z}_{i,t-1})}{\sigma_{i,t}(\mathcal{Z}_{i,t-1})} \right)^{-1} \\ \times K_1 \left(\bar{\alpha} q \left(\frac{x_i - m_{i,t}(\mathcal{Z}_{i,t-1})}{\sigma_{i,t}(\mathcal{Z}_{i,t-1})} \right) \right), \quad (5)$$

where $q(x) = \sqrt{1+x^2}$ and $K_1(\cdot)$ is a modified Bessel Function, third order index 1.

The first two conditional moments are

$$E[X_{i,t} | \mathcal{Z}_{i,t-1}] = m_{i,t} + \sigma_{i,t} \frac{\bar{\beta}}{\bar{\gamma}} \quad \text{and} \quad V[X_{i,t} | \mathcal{Z}_{i,t-1}] = \sigma_{i,t}^2 \frac{\bar{\alpha}^2}{\bar{\gamma}^3},$$

where the dependence on $\mathcal{Z}_{i,t-1}$ of $m_{i,t}$ and $\sigma_{i,t}$ has been dropped to shorten notation. In order to formulate the model according to the structure in (3) we standardize $\varepsilon_{i,t}$ in the following way

$$\varepsilon_{i,t} \sim \text{NIG} \left(\bar{\alpha}, \bar{\beta}, -\frac{\sqrt{\bar{\gamma}} \bar{\beta}}{\bar{\alpha}}, \frac{\bar{\gamma}^{3/2}}{\bar{\alpha}} \right).$$

Now $E[\varepsilon_{i,t}] = 0$ and $V[\varepsilon_{i,t}] = 1$. We can now specify the model as in (3).

$$x_{i,t} = m_{i,t} + \frac{\sqrt{\bar{\gamma}} \bar{\beta}}{\bar{\alpha}} \sigma_{i,t} + \sigma_{i,t} \varepsilon_{i,t}. \quad (6)$$

From (6) it follows, that

$$X_{i,t} | \mathcal{Z}_{i,t-1} \sim \text{NIG} \left(\bar{\alpha}, \bar{\beta}, m_{i,t}, \sigma_{i,t} \frac{\bar{\gamma}^{3/2}}{\bar{\alpha}} \right)$$

which is very convenient because it implies that

$$E[X_{i,t} | \mathcal{Z}_{i,t-1}] = m_{i,t} + \sigma_{i,t} \frac{\sqrt{\bar{\gamma}} \bar{\beta}}{\bar{\alpha}} \quad \text{and} \quad V[X_{i,t} | \mathcal{Z}_{i,t-1}] = \sigma_{i,t}^2,$$

such that we can specify models for the conditional variance in terms of $\sigma_{i,t}$ alone. Models for the conditional mean are specified by choosing an appropriate model for m_t . Note, that the GARCH-in-mean effect can be removed by including $-\frac{\sqrt{\gamma}\bar{\beta}}{\bar{\alpha}}\sigma_{i,t}$ in $m_{i,t}$.

In terms of (3) we have

$$\begin{aligned}\mu_{i,t}(\mathbf{Z}_{i,t-1}) &= m_{i,t} + \sigma_{i,t} \frac{\sqrt{\gamma}\bar{\beta}}{\bar{\alpha}} \\ \sigma_{i,t}(\mathbf{Z}_{i,t-1}) &= \sigma_{i,t}\end{aligned}\tag{7}$$

We return to the different specifications of $\mu_{i,t}(\mathbf{Z}_{i,t-1})$ in Section 4.1. For the conditional variance we consider a GARCH(1,1) model for $\sigma_{i,t}$. Such that

$$\sigma_{i,t}^2 = \bar{\sigma}_i + \delta_i \mathcal{I}_{FC} + a_i(\sigma_{i,t-1}\varepsilon_{i,t-1})^2 + b_i\sigma_{i,t-1}^2,\tag{8}$$

where δ is the coefficient on a dummy variable, \mathcal{I}_{FC} , which takes the value zero for all days before 2008 and one after. This allows us to capture increases in volatility resulting from the financial crisis. Determining the starting point of the financial crisis is a matter of some debate, but the results are robust to different starting times.

3.3 Copula Density

The next step is to specify the copula density. In this exercise we consider two alternatives, the normal and the t copula. The density function for the normal copula is given as

$$c(\mathbf{u}; \boldsymbol{\Sigma}) = \frac{1}{\sqrt{|\boldsymbol{\Sigma}|}} \exp \left\{ -\frac{(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_3))' (\boldsymbol{\Sigma}^{-1} - I_3) (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_3))}{2} \right\},$$

where $\mathbf{u} = (u_1, u_2, u_3)'$, $u_i = F_{X_i}(x_i)$ for $i = 1, 2, 3$ and $\Phi^{-1}(\cdot)$ denotes the inverse cdf of a standard normal variable. $\boldsymbol{\Sigma}$ is a correlation matrix and given as

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}.$$

For the t copula the density function with degree of freedom, ν , is given as

$$c(\mathbf{u}; \boldsymbol{\Sigma}, \nu) = \frac{\Gamma((\nu + k)/2)(\Gamma(\nu/2))^{k-1}}{\sqrt{|\boldsymbol{\Sigma}|}(\Gamma((\nu + 1)/2))^k} \left(1 + \frac{\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W}}{\nu}\right)^{-(\nu+k)/2} \prod_{i=1}^k \left(1 + \frac{w_i^2}{\nu}\right)^{(\nu+1)/2},$$

where $\mathbf{W} = (w_1, \dots, w_k)'$, $w_i = t_\nu^{-1}$, and $\Gamma(\cdot)$ is the gamma function.

In order to choose between the two copulas we use a t-test as presented in Rivers and Vuong (2002). The test is based on the difference in the log likelihood of the two models. If this difference is not statistically significant we choose the simpler of the two models. We are only considering two different copula densities in this analysis.

3.4 Maximum Likelihood Estimation

After specifying the marginal models and the copula density we turn to parameter estimation. The log likelihood can be written as

$$l_{x,t}(\boldsymbol{\theta}; \mathbf{x}) = \sum_{i=1}^3 l_{x_i,t}(\boldsymbol{\psi}_i; \mathbf{x}_i) + l_{c,t}(\boldsymbol{\kappa}; \mathbf{u}), \quad \boldsymbol{\theta} = (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\psi}_3, \boldsymbol{\kappa}) \quad (9)$$

where $\boldsymbol{\psi}_i$ contains the parameters in marginal model i . $\boldsymbol{\kappa}$ contains the parameters from the copula density and $\mathbf{u} = (u_1, u_2, u_3)'$.

The maximization of (9) is carried out in two steps. In the first step we maximize the likelihood functions associated with the marginal models, $l_{x_i,t}(\boldsymbol{\psi}_i; \mathbf{x}_i)$. In the second step we use the estimates, $\hat{\boldsymbol{\psi}}_i$, to construct $\hat{u}_{i,t} = F_{X_i}(x_{i,t}; \hat{\boldsymbol{\psi}}_i)$. Then we maximize $l_{c,t}(\boldsymbol{\kappa}; \hat{\mathbf{u}})$ to estimate $\boldsymbol{\kappa}$. We do not have a closed form expression for the cdf of the NIG distribution, so the probabilities are found using numerical integration of the pdf in (5),

Optimization and numerical integration are carried out in Ox, see Doornik (2007). Optimization relies on the MaxSQP function with numerical derivatives and numerical integration is carried out using the QAGS function. MATLAB codes for estimation of parameters in the copula densities are available from Andrew Patton's website these have been adopted to the Ox language in this analysis.³

Statistical inference in this case is based on a block bootstrap as suggested by Patton (2012). The theoretical background and justification for this approach are presented in Gonçalves and White (2004). A bootstrap sample of the same length as our sample is generated and the model is estimated using the same approach as for the real data. This process is repeated 1.000 times and confidence intervals for the estimated parameters are constructed from the quantiles of the bootstrap parameter distribution.

3.5 Tests for Misspecification

It is important to test for possible misspecification in the marginal models. The decomposition in (2) is only valid if the marginal models are well specified.

We test for autocorrelation in the estimated probability integral transforms using the Breusch-Godfrey test with 12 lags, different lag lengths were tried with no discernible difference in the conclusions. We include enough lags in each model to make sure that we have adequately captured the autocorrelation. We carry out two tests for the density misspecification as suggested in Patton (2012). Both tests consider potential misspecification in the estimated and sorted probability integral transforms, $\tilde{u}_{i,t}$. These

³<http://public.econ.duke.edu/~ap172/>

are constructed based on the estimated parameters, $\hat{\psi}_i$ as

$$\tilde{u}_{i,t} = F_{X_i}(x_{i,t}; \hat{\psi}_i),$$

where \tilde{u}_i is sorted in ascending order. The Kolmogorov-Smirnov (KS) and the Cramer-von Mises (CvM) test statistics are calculated as

$$KS = \max_t \left| \tilde{u}_{i,t} - \frac{t}{T} \right| \quad \text{and} \quad CvM = \sum_{t=1}^T \left(\tilde{u}_{i,t} - \frac{t}{T} \right)^2$$

In the case of NIG distributed error terms we have to estimate the parameters of the distribution. This means that we cannot rely on the asymptotic distribution of these test statistics in our analysis. Instead we follow Patton (2012) and use a simulation based method to calculate the p-values of the tests. Based on the estimated parameters we can simulate a sample of the factor, x_i , based on the estimated parameters. Simulation from the normal and t copulas is quite easy and algorithms are presented in Schmidt (2007). Next, we estimate the model on the simulated data and compute the values of KS and CvM . We repeat this $S = 1.000$ times in order to obtain simulated distributions of the two test statistics.

4 Empirical Analysis

In this analysis we consider all observations up to and including 2010 as our in-sample period. The observations for 2011 and 2012 are saved for out-of-sample evaluation. The first step in the analysis is to estimate the factors. We do this in the following way. At each point in time we regress the observed futures prices on the matrix \mathbf{Z}_t for a fixed value of λ according to (1). From a purely mathematical perspective we only need three traded contracts per day. In order to eliminate outliers in the time series of the factors we choose to exclude trading days where less than 10 different contracts are traded. Only a very small fraction of the trading days in our sample does not meet this requirement. The parameter estimates are estimates of the factors, $(L_t, S_t, C_t)'$ at time t . A grid search is carried out on the interval $[0.001, 0.15]$ to find the value of λ which minimizes the sum of the squared error terms over time, $\sum_{t=1}^T e_t' e_t$. The sum of the error terms is minimized for $\lambda = 0.005$. The estimated factors are presented in Figure 4. The figure shows the estimated factors for both the in-sample and out-of-sample period. Note that λ is calculated based entirely on the in-sample period.

The level factor is increasing throughout the sample. This indicates that prices in general have increased over time. This is in contrast to the evidence from the interest rate market presented in Diebold and Li (2006) and Noureldin (2011). The slope factor

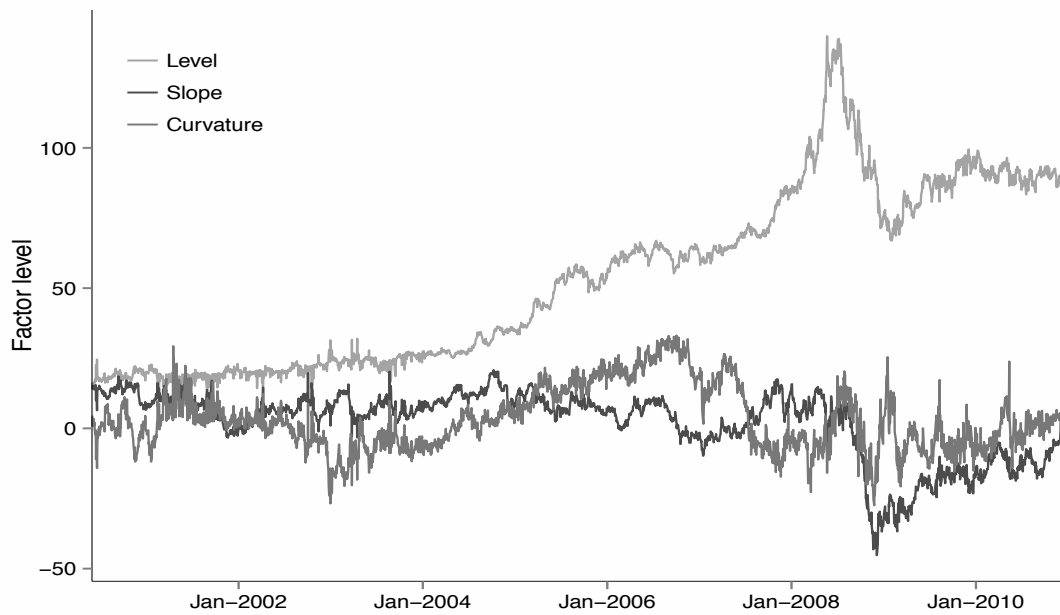


Figure 4: Estimated factors. The estimated factors. L_t : Level (light gray), S_t : Slope (dark gray) and C_t : Curvature (gray).

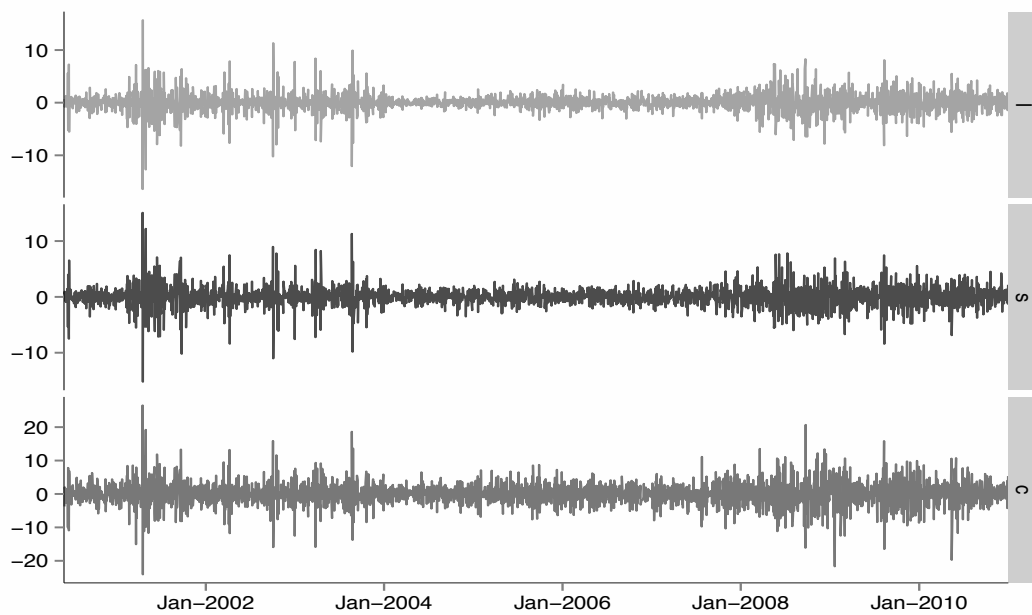


Figure 5: Estimated factors in first differences. $l_t = \Delta L_t$ (light gray), $s_t = \Delta S_t$ (dark gray), and $c_t = \Delta C_t$ (gray).

is predominantly positive for the first part of the sample until 2008-09 and then it turns negative. This indicates a downward sloping term structure in the first part of the sample and an upward sloping term structure in the last part of the sample. In the interest rate market Diebold and Li (2006) and Noureldin (2011) find that the slope factor is negative almost all the time. The curvature factor changes time several times throughout the sample. This pattern is also observed for interest rates by Diebold and Li (2006) and Noureldin (2011). We analyze the time series properties of the factors to determine how the model should be specified. Stationarity is rejected for L_t and S_t while C_t is borderline stationary. In the following we carry out the analysis on the first differences of all three factors. We introduce the following notation

$$\mathbf{x}_t = \begin{pmatrix} \Delta L_t \\ \Delta S_t \\ \Delta C_t \end{pmatrix} = \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix}.$$

4.1 Model Specifications

We are considering four different models. Two of the models are univariate in the sense that we only allow lags of the variable in question, thus effectively ignoring the requirement of independence in (4). The other two models also include lags of the other variables. In both cases we consider GARCH and GARCH-in-mean specifications. We refer to the four different models as uniNIG-GARCH, uniNIG-GARCHm, mvNIG-GARCH and mvNIG-GARCHm, respectively.

In each model we have included enough lags to remove the autocorrelation. For the univariate models we include three, two and three lags, in the models for ΔL_t , ΔS_t and ΔC_t , respectively. For the multivariate models we include three lags of each variable in each model. In each model we include a constant denoted ϕ_0 . The coefficient of lag j of factor i is denoted $\phi_{i,j}$.

The estimated parameters are presented in Table 2. Bold face is used to indicate significance at the 95% level of confidence. For some of the models we have some insignificant lags. They are, however, necessary to remove autocorrelation and to ensure well specified models. The GARCH parameters are significant for all the models. For some of the marginals the NIG parameter $\bar{\beta}$ is not significant. This indicates a symmetric density. For the remaining marginals $\bar{\beta}$ is significant indicating an asymmetric density. The copula parameters are significant for all marginals and models.

The results of the misspecification tests are presented in Panel A of Table 3. The Breusch-Godfrey (BG) test shows that, at a 5% level of significance, we have adequately captured the autocorrelation for all our models. The null-hypotheses of the Kolmogorov-

Smirnov (KS) and the Cramer-von-Mises (CvM) tests are that the residual density is specified correctly. We fail to reject this for all our models. Panel B in Table 3 provides the results of the test from Rivers and Vuong (2002), where the normal and t-copulas are compared based on the log likelihood. The null is, that the two copula specifications are equivalent. This is rejected for all the models and the t-copula is preferred in all cases.

In Figure 6 we present the estimated residual densities along with a NIG reference for the mvNIG-GARCH model. The corresponding plots for the other models are nearly identical to the ones presented here. The theoretical NIG density is constructed based on the estimated parameters from each of the marginal models. The NIG distribution offers a reasonable approximation for the residual density. To better assess the quality of the approximation in the tails we present the estimated log densities along with the theoretical log NIG densities. The theoretical log densities are based on the parameter estimates. We conclude that the NIG distribution offers a reasonable approximation of the residual densities for all the models.

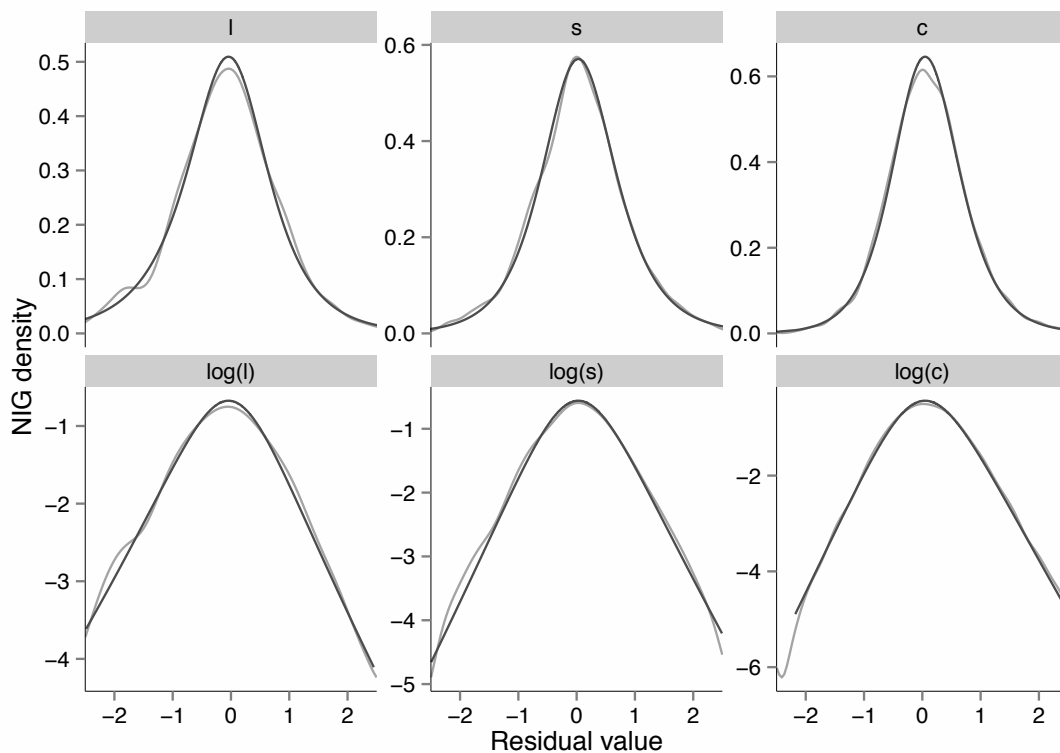


Figure 6: Residual densities for mvNIG-GARCH. The estimated residual densities (light gray) for each of the variables for each of the models are presented together with the theoretical NIG densities (gray). For each variable the log counterpart is also presented.

Table 2: Parameter estimates.

	uniNIG-GARCH			uniNIG-GARCHm			mvNIG-GARCH			mvNIG-GARCHm		
	$x_{1,t}$	$x_{2,t}$	$x_{t,3}$	$x_{1,t}$	$x_{2,t}$	$x_{t,3}$	$x_{1,t}$	$x_{2,t}$	$x_{t,3}$	$x_{1,t}$	$x_{2,t}$	$x_{t,3}$
ϕ_0	0.038	0.040	0.040	0.123	0.024	-0.093	0.024	0.042	0.003	0.116	-0.034	-0.172
$\phi_{1,1}$	-0.347	-	-	-0.341	-	-	-0.105	-0.016	0.490	-0.101	-0.021	0.487
$\phi_{1,2}$	-0.121	-	-	-0.115	-	-	-0.027	0.005	0.221	-0.024	0.013	0.222
$\phi_{1,3}$	-0.045	-	-	-0.042	-	-	-0.019	0.005	0.060	-0.016	0.004	0.059
$\phi_{2,1}$	-	-0.258	-	-	-0.257	-	0.142	-0.233	0.328	0.141	-0.239	0.327
$\phi_{2,2}$	-	-0.063	-	-	-0.062	-	0.010	-0.025	0.125	0.007	-0.020	0.129
$\phi_{2,3}$	-	-	-	-	-	-	-0.015	0.021	0.034	-0.015	0.013	0.034
$\phi_{3,1}$	-	-	-0.532	-	-	-0.530	0.121	-0.120	-0.413	0.121	-0.121	-0.412
$\phi_{3,2}$	-	-	-0.270	-	-	-0.267	0.096	-0.097	-0.192	0.095	-0.095	-0.189
$\phi_{3,3}$	-	-	-0.139	-	-	-0.137	0.050	-0.055	-0.106	0.050	-0.056	-0.105
	GARCH											
$\bar{\sigma}$	0.082	0.150	0.375	0.069	0.142	0.367	0.076	0.378	0.377	0.069	0.122	0.363
a	0.172	0.158	0.131	0.156	0.151	0.127	0.173	0.194	0.122	0.165	0.143	0.117
b	0.788	0.770	0.828	0.811	0.779	0.832	0.792	0.585	0.832	0.803	0.790	0.838
δ	0.368	0.397	0.616	0.310	0.372	0.617	0.286	0.965	0.716	0.264	0.315	0.704
	NIG											
$\bar{\alpha}$	0.941	1.343	1.920	0.944	1.377	1.906	0.948	1.238	1.774	0.923	1.601	1.754
$\bar{\beta}$	-0.110	0.090	0.170	-0.062	-0.002	0.058	-0.125	0.156	0.175	-0.068	0.053	0.078
	Copula											
	ρ_{12}	ρ_{23}	ρ_{13}	ρ_{12}	ρ_{23}	ρ_{13}	ρ_{12}	ρ_{23}	ρ_{13}	ρ_{12}	ρ_{23}	ρ_{13}
Σ	-0.486	-0.703	0.250	-0.488	-0.703	0.251	-0.486	-0.761	0.267	-0.492	-0.760	0.266
ν		2.960			2.942			3.363			3.126	

Parameter estimates for the four different models. Significance at a 5% level is indicated with bold face.

Table 3: Misspecification tests.

	uniNIG-GARCH			uniNIG-GARCHIm			mvNIG-GARCH			mvNIG-GARCHIm		
	$x_{1,t}$	$x_{2,t}$	$x_{t,3}$	$x_{1,t}$	$x_{2,t}$	$x_{t,3}$	$x_{1,t}$	$x_{2,t}$	$x_{t,3}$	$x_{1,t}$	$x_{2,t}$	$x_{t,3}$
BG	14.963 [0.243]	6.452 [0.892]	17.908 [0.119]	14.921 [0.246]	6.282 [0.901]	17.392 [0.135]	11.245 [0.508]	4.533 [0.972]	13.871 [0.309]	10.571 [0.566]	3.250 [0.994]	13.492 [0.334]
KS	0.009 [0.935]	0.015 [0.160]	0.010 [0.780]	0.014 [0.343]	0.014 [0.262]	0.014 [0.242]	0.010 [0.999]	0.014 [0.995]	0.012 [1.000]	0.019 [0.997]	0.018 [0.987]	0.017 [1.000]
CvM	0.033 [0.853]	0.070 [0.262]	0.035 [0.735]	0.087 [0.191]	0.083 [0.156]	0.084 [0.149]	0.035 [0.999]	0.087 [0.994]	0.044 [1.000]	0.125 [0.997]	0.159 [0.986]	0.095 [1.000]
Panel B	uniNIG-GARCH			uniNIG-GARCHIm			mvNIG-GARCH			mvNIG-GARCHIm		
RV	-5.409 [0.000]			-5.442 [0.000]			-5.750 [0.000]			-5.600 [0.000]		

Panel A: Results for the Breusch-Godfrey (BG), Kolmogorov-Smirnov (KS) and the Cramer-von-Mises (CvM) tests for misspecification of the marginals in each model. Panel B: Results for the t-test from Rivers and Vuong (2002) (RV) comparing the normal and t copulas for each model. p-values in brackets.

5 Forecasting the Term Structure

The forecasting exercise is very ambitious. At every point in time we forecast the entire term structure of futures prices. We know which contracts may potentially be traded the next day and the time to maturity for these contracts, meaning that in terms of (1) we know the factor loadings, Z_{t+1} with certainty. This, together with forecasts of the factors will allow us to forecast the futures prices. For these contracts we calculate the squared forecast errors. We consider two different benchmarks in this section. The first benchmark is a simple autoregressive model using only previous observations of the prices of a contract to form the forecast. Time series of prices for the individual contracts are extracted from the dataset and AR(1) models are fitted to each of them. The second benchmark is a VAR model specified on $(\Delta L_t, \Delta S_t, \Delta C_t)'$. The factors are extracted from the dynamic Nelson-Siegel model just as in our framework but instead of the copula decomposition and subsequent modelling of the marginal models and the copula density a VAR model is fitted to the three series. The forecast performance is evaluated using the model confidence set developed in Hansen et al. (2011). The model confidence set is implemented using the Ox package MulCom 3.0, Hansen and Lunde (2014). All results are based on the Range statistic.⁴

Forecasts are constructed as follows. At each point in time t the models for the factors are estimated. Based on the estimated parameters we get forecasts of $x_{i,t+1}$ from its conditional mean in (7). For the two GARCHm models the forecast of the conditional mean requires forecasts of $\sigma_{i,t+1}^2$ which are obtained from (8). Using the definition of the dynamic Nelson-Siegel model in (1) we can transform the forecasts of the factors into forecasts of prices for any maturity we wish. Notice, that the copula density plays no role in the construction of the forecasts. The forecast of the term structure can be expressed as a linear combination of the forecasts of the three factors and thus their correlation will not matter.

To evaluate the performance of the forecasts we have chosen to split the contracts into groups according to their maturity. This enables us to evaluate the accuracy of our forecasts for different parts of the term structure. First we calculate the mean squared error across maturities on every day in the out-of-sample period for each of the models. The model confidence set is estimated and the p-values are presented in Table 4 along with the average mean squared errors for a given range of maturities. For all the dynamic Nelson-Siegel models the mean squared error is lowest for medium term maturities, between 401 and 500 days to maturity. These models fit the entire term structure

⁴The package can be downloaded from http://mit.econ.au.dk/vip_htm/alunde/mulcom/mulcom.htm

and this explains the good fit for the medium term maturities. The mvNIG-GARCH model is best overall based on the estimated model confidence set. The AR model is excluded from the 95% model confidence set estimated on all maturities. None of the other benchmarks can be excluded from the 95% model confidence set. The mvNIG-GARCH model is very good for maturities shorter than 500 days but not for longer maturities. The AR model has very bad performance for shorter maturities but performs really well for maturities longer than 500 days. The good performance is not surprising considering Figure 1, where we see that trade is concentrated in very few contracts for maturities longer than 500 days. In this situation the price history of the individual contract should be expected to be very informative. The VAR model is quite good for some maturities, but it performs very poorly for other maturities.

Mean squared forecast errors may not be the best measure of the quality of our model. It might be very interesting to consider directional forecasts. That is, we predict whether the price of a contract will increase or decrease in the next day. The accuracy of the directional forecast is assessed in the following way. At every point in time, t , and for each contract, j , we construct the following variables.

$$\begin{aligned} d_{j,t} &= \text{sign}(Y_t(\tau_j) - Y_{t-1}(\tau_{j+1})) \quad \text{and} \\ \hat{d}_{j,t} &= \text{sign}(\hat{Y}_t(\tau_j) - Y_{t-1}(\tau_{j+1})). \\ \tilde{d}_{j,t} &= \begin{cases} 1 & \text{if } d_{j,t} = \hat{d}_{j,t} \\ 0 & \text{if } d_{j,t} \neq \hat{d}_{j,t} \end{cases}. \end{aligned}$$

$d_{j,t}$ indicates the direction of the actual change and $\hat{d}_{j,t}$ indicates the direction predicted by our model. If the two are equal the forecast is successful and if they are not then the forecast is poor. In order to estimate the model confidence set we specify the following loss function.

$$c_t = -\frac{1}{k_t} \sum_{j=1}^{k_t} \tilde{d}_{j,t}$$

With the directional forecast there are only two possible outcomes. Either we are right or we are wrong. That means that we would like to be right more than 50% of the time. To check whether we achieve this goal we include $c_t = -0.5$ as a benchmark before estimating the model confidence set. The model confidence set is estimated and the results are presented in 5. Again we see, that the best performance of the dynamic Nelson-Siegel models is for the medium term maturities, where the highest fractions of correct predictions are found. Based on all maturities we see the the uniNIG-GARCH and uniNIG-GARCHm models are superior compared to the other models, which are excluded from the 90% model confidence set. Furthermore, the predictive performance

is significant in the sense that the benchmark is excluded from the model confidence set. Once again the benchmarks perform well for some maturities but across maturities the dynamic Nelson-Siegel models with NIG-GARCH specifications outperform the benchmark models.

6 VaR for Portfolios of Factors

The purpose of this section is to illustrate the possibilities for practical applications within this framework. We choose a spread between L_t and S_t from Figure 4 as our portfolio of interest. We show how we can calculate value at risk in this situation.

We assume that it is possible to buy the linear combinations of the traded assets at any point in time. At any point in time, t , we forecast the difference in the factors one step ahead. We know the distributions of the first differences of the factors under each of our model specifications so we can simulate a large number of possible portfolio values for time $t + 1$. We will denote the α quantile of this distribution as VaR_α . We consider three different values for α , 0.01, 0.05 and 0.1. We calculate VaR_α for each day of the out-of-sample period and compare to the actual changes in the portfolio value. We do this for the Dynamic Nelson-Siegel models with NIG marginals and for the Dynamic Nelson-Siegel VAR model we leave out the AR benchmark in this analysis as it does not involve the factors.

It is important to assess the quality of the VaR calculations. In the following we introduce several models for backtesting and present the corresponding results. Backtesting in this analysis consists of four different tests. The first test is from Kupiec (1995) and compares the unconditional probability of VaR violations to the theoretical value, α . The test from Christoffersen (1998) takes into account that the VaR violations should be independent. A combination of the two tests is carried out as explained in Campbell (2005). The out-of-sample dynamic quantile test, DQ_{OOS} , presented in Engle and Manganelli (2004) is applied. The DQ_{OOS} test is based on the following instruments, a constant, the VaR forecast and four lags of the hit variable, Hit_{t+1} . The hit variable is defined as

$$Hit_{t+1} = \mathcal{I}_{\{y_{t+1} < \text{VaR}_{\alpha,t}\}} - \alpha$$

and takes the value $1 - \alpha$ if there is a VaR violation and the value $-\alpha$ if no violation occurs in a period. MATLAB code for this test is available from Simone Manganelli's website and has been adopted to the Ox language.⁵

Finally, we also compare the models using the model confidence set. We rely on the

⁵<http://www.simonemanganelli.org/Simone/Home.html>

Table 4: Forecasting with mean square criterion. p-values for the model confidence set.

	uniNIG-GARCH	uniNIG-GARCHm	mvNIG-GARCH	mvNIG-GARCHm	AR	VAR
1-100	3.18 [0.1653]	3.16 [0.5670]	3.07 [0.9864]	3.07 [0.9864]	4.55 [0.0002]	3.07 [1.0000]
101-200	2.81 [0.0983]	2.83 [0.0983]	2.64 [1.0000]	2.65 [0.1400]	4.88 [0.0000]	2.67 [0.0983]
201-300	2.49 [0.0641]	2.50 [0.0455]	2.26 [1.0000]	2.27 [0.1073]	4.32 [0.0000]	2.29 [0.0641]
301-400	1.71 [0.1765]	1.71 [0.1765]	1.54 [1.0000]	1.54 [0.6876]	3.10 [0.0508]	1.56 [0.6651]
401-500	0.93 [0.1645]	0.92 [0.2295]	0.80 [0.3224]	0.80 [0.9808]	1.57 [0.0069]	0.80 [1.0000]
501-1000	2.17 [0.0012]	2.13 [0.1487]	2.06 [0.0000]	2.03 [0.1487]	1.80 [1.0000]	2.07 [0.1487]
1001-1500	1.57 [0.0271]	1.56 [0.0271]	1.50 [0.0271]	1.50 [0.0271]	1.34 [1.0000]	1.52 [0.0233]
>1501	1.19 [0.0005]	1.21 [0.0005]	1.12 [0.0012]	1.13 [0.0005]	0.78 [1.0000]	1.10 [0.0012]
All Maturities	2.57 [0.2680]	2.57 [0.2680]	2.44 [1.0000]	2.44 [0.7383]	3.79 [0.0000]	2.45 [0.3534]

Mean squared error along with p-values for the estimated model confidence set for the different models and for different ranges of maturities.

Table 5: Forecasting with the directional criterion. p-values for the model confidence set.

	uniNIG-GARCH	uniNIG-GARCHm	mvNIG-GARCH	mvNIG-GARCHm	AR	VAR	0.5
1-100	0.548 [1.0000]	0.547 [0.8529]	0.517 [0.0587]	0.520 [0.0711]	0.505 [0.1144]	0.508 [0.0282]	0.500 [0.0282]
101-200	0.538 [1.0000]	0.531 [0.1826]	0.508 [0.1826]	0.500 [0.1151]	0.498 [0.1826]	0.488 [0.0545]	0.500 [0.1515]
201-300	0.543 [1.0000]	0.541 [0.6959]	0.524 [0.6959]	0.524 [0.6959]	0.487 [0.1048]	0.517 [0.6560]	0.500 [0.1475]
301-400	0.531 [0.9894]	0.530 [0.9894]	0.533 [0.9894]	0.536 [1.0000]	0.494 [0.7186]	0.518 [0.7320]	0.500 [0.7186]
401-500	0.607 [0.9803]	0.611 [1.0000]	0.603 [0.9803]	0.603 [0.9803]	0.477 [0.0122]	0.590 [0.8743]	0.500 [0.0018]
501-1000	0.579 [0.2678]	0.586 [1.0000]	0.563 [0.2678]	0.559 [0.2678]	0.509 [0.0078]	0.558 [0.2678]	0.500 [0.0000]
1001-1500	0.560 [1.0000]	0.553 [0.6676]	0.547 [0.6479]	0.553 [0.7273]	0.517 [0.6479]	0.546 [0.7273]	0.500 [0.0454]
>1501	0.523 [1.0000]	0.521 [0.9146]	0.520 [0.9687]	0.517 [0.9687]	0.481 [0.9372]	0.522 [0.9745]	0.500 [0.9373]
All Maturities	0.553 [1.0000]	0.551 [0.3285]	0.530 [0.0588]	0.529 [0.0588]	0.500 [0.0129]	0.520 [0.0281]	0.500 [0.0001]

Fraction of correct predictions along with p-values for the estimated model confidence set for the different models and for different ranges of maturities. 0.5 is the benchmark for directional forecasting.

framework of Giacomini and Komunjer (2005) and use the 'check' loss function

$$\mathcal{L}^\alpha(e_{t+1}) = (\alpha - \mathcal{I}_{\{e_{t+1} < 0\}}) e_{t+1},$$

where $e_{t+1} = y_{t+1} - VaR_{\alpha,t}$. We also consider the linex loss function

$$\mathcal{L}^\alpha(e_{t+1}) = \exp(\alpha e_{t+1}) - \alpha e_{t+1} - 1.$$

The results from backtesting are presented in Table 6. All the models fail the Kupiec test for $\alpha = 0.01$. This is not surprising as the Kupiec test has low power and the sample size is relatively small. The models pass all the other tests for $\alpha = 0.01$. All the NIG-GARCH models pass all the backtesting on a 5% level of significance for both $\alpha = 0.05$ and $\alpha = 0.10$. The VAR model fails the Kupiec test, the combined Kupiec and Christoffersen test and the DQ_{005} tests for $\alpha = 0.05$ and $\alpha = 0.1$. The p-values for the model confidence set show that the mvNIG-GARCHm model is doing well for $\alpha = 0.01$ and $\alpha = 0.05$ while the mvNIG-GARCH model is very good for $\alpha = 0.1$. The uniNIG-GARCH model performs well for $\alpha = 0.1$ and the uniNIG-GARCHm model is good for all values of α . The VAR model performs poorly and is excluded from the model confidence set for all values of α .

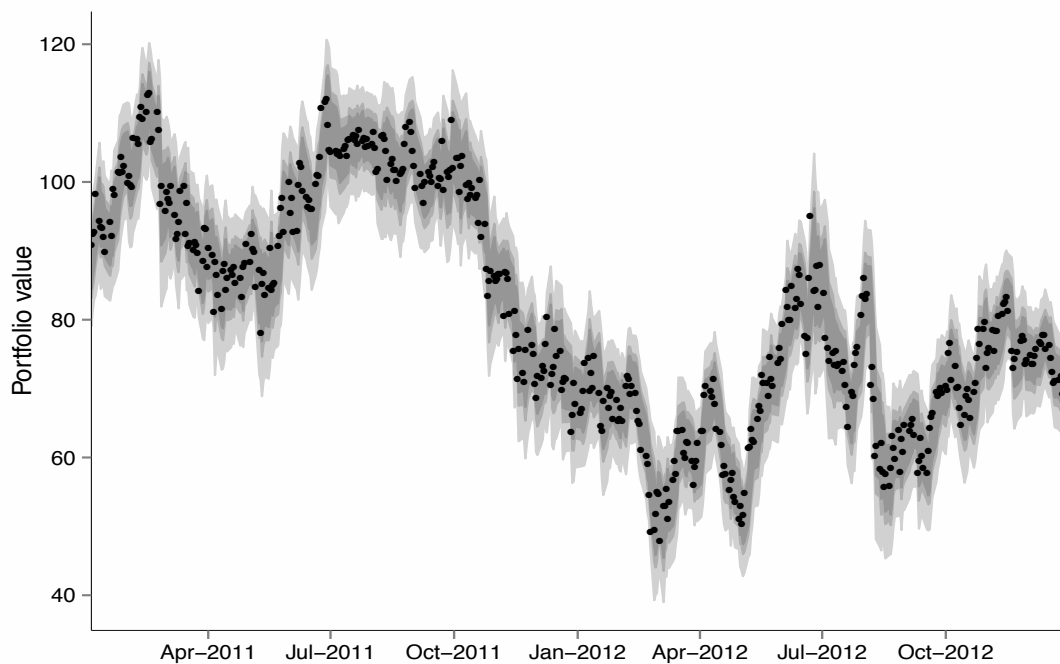


Figure 7: Value at risk for mvNIG-GARCH. Black dots indicate portfolio value. The gray area is the value at risk suggested by the different models. Darker gray indicates higher α .

Table 6: Backtesting Value at Risk.

α	uniNIG-GARCH	uniNIG-GARCHm	mvNIG-GARCH	mvNIG-GARCHm	VAR
<i>Kupiec test</i>					
1%	0.0275	0.0275	0.0275	0.0275	0.0275
5%	0.8127	0.9755	0.9755	0.8127	0.0000
10%	0.9173	0.8015	0.5864	0.9173	0.0000
<i>Christoffersen 1998 test</i>					
1%	0.9496	0.9496	0.9496	0.9496	0.9496
5%	0.1205	0.1054	0.8115	0.8828	0.7032
10%	0.6946	0.4540	0.3289	0.9290	0.2472
<i>Combined Kupiec and Christoffersen 1998 test</i>					
1%	0.0878	0.0878	0.0878	0.0878	0.0878
5%	0.2912	0.2695	0.9715	0.9618	0.0000
10%	0.9208	0.7320	0.5355	0.9907	0.0000
<i>Out of sample dynamic quantile test</i>					
1%	0.7798	0.7796	0.7797	0.7796	0.7796
5%	0.7776	0.8475	0.9794	0.9921	0.0053
10%	0.8985	0.6338	0.8011	0.9508	0.0004
<i>Model Confidence Set, Giacomini & Komunjer 2005, check loss</i>					
1%	0.0000	0.3404	0.0000	1.0000	0.0000
5%	0.0000	0.7556	0.7556	1.0000	0.0000
10%	0.0596	0.2197	1.0000	0.0000	0.0000
<i>Model Confidence Set, Giacomini & Komunjer 2005, linex</i>					
1%	0.0000	0.4739	0.0000	1.0000	0.0000
5%	0.0000	0.8553	0.8553	1.0000	0.0000
10%	0.4885	0.4885	1.0000	0.0005	0.0000

p-values for different backtesting methodologies for VaR_α calculated by the Dynamic Nelson-Siegel model with NIG marginals and for VaR_α calculated by the Dynamic Nelson-Siegel model with VAR specification. The last two panels contain the p-values for the model confidence set where the models are compared according to the 'check' loss function and the linex loss function, respectively.

The results of the VaR analysis for the mvNIG-GARCH model are presented graphically in Figure 7. The plots for the other models are very similar to this one. The black dots represent the actual portfolio value. We see, that the value of the portfolio has changed a lot during the sample period. Some of the changes are very large and sudden. This makes VaR calculation a challenging task but our class of models still performs well. The lower parts of three gray bands in Figure 7 represent the VaR_α . We see that the portfolio values are higher than the VaR bands for most of the sample. More importantly, and in accordance with the test result from Table 6, there is no systematic pattern in the violations. This means that we have successfully calculated the VaR throughout this sample.

7 Concluding Remarks

In this paper we successfully use the dynamic Nelson-Siegel model to analyze and forecast the prices of futures contracts on oil. We use the very flexible class of GARCH models with NIG innovations in a copula framework. We demonstrate that this class of models generally leads to good in-sample fit and well specified models. Out of sample the class of models provide very good forecasts. The forecast performance is evaluated based on both a conventional mean squared error criterion as well as a directional criterion. The model confidence set is used to show that our class of models performs better than conventional benchmarks. Finally, we show that we can construct portfolios based on the factors and calculate value at risk of these portfolios. Various backtesting methods confirm that our class of models performs really well in a realistic real time out-of-sample analysis.

Oil is the most traded commodity in the world and the prices of the futures contracts are relatively well-behaved compared to prices of other commodities. Data sets from other commodities might exhibit features which the benchmark models will not be able to handle. Our model is expected to perform well even for ill-behaved data sets because the combination of NIG-GARCH models and the copula framework allow for extreme flexibility and we will be able to handle large sudden price movements and a wide range of correlation structures.

There is room for improvement in the forecasting exercise. Our class of models is shown to be better than the other models and performs especially well in the short end of the term structure. A combination of these forecasts with the forecasts of a model which performs well in the long end will presumably lead to even better forecasts. This presents an interesting route for future research.

8 References

- Barndorff-Nielsen, O. E., 1997. Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*.
- Caballero, R. J., Farhi, E., Gourinchas, P.-O., 2008. Financial crash, commodity prices and global imbalances. *Brookings Papers on Economic Activity*, 1–55.
- Campbell, S. D., 2005. A review of backtesting and backtesting procedures. *Finance and Economics Discussion Series, Division of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C.*

- Chalamandaris, G., Tsekrekos, A., 2011. How important is the term structure in implied volatility surface modeling? evidence from foreign exchange options. *Journal of International Money and Finance* 30, 623–640.
- Christoffersen, P. F., 1998. Evaluating interval forecasts. *International Economic Review* 39 (4), 841–862.
- Diebold, F. X., Li, C., 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337–364.
- Doornik, J. A., 2007. *Object-Oriented Matrix Programming Using Ox*, 3rd Edition. London: Timberlake Consultants Press and Oxford: www.doornik.com.
- Engle, R. F., Manganelli, S., 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* 22 (4), 367–381.
- Geman, H., 2005. *Commodities and commodity derivatives: modeling and pricing for agriculturals, metals, and energy*. Wiley.
- Giacomini, R., Komunjer, I., 2005. Evaluation and combination of conditional quantile forecasts. *Journal of Business and Economic Statistics* 23 (4), 416–431.
- Gonçalves, S., White, H., 2004. Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics* 119, 199–219.
- Guo, B., Han, Q., Zhao, B., 2014. The nelson-siegel model of the term structure of option implied volatility and volatility components. *Journal of Futures Markets* (Forthcoming).
- Hansen, P., Lunde, A., 2014. *Mulcom 3.00*, econometric toolkit for multiple comparisons.
- Hansen, P., Lunde, A., Nason, J., 2011. The model confidence set. *Econometrica* 2 (79), 453–497.
- Hautsch, N., Ou, Y., 2012. Analyzing interest rate risk: Stochastic volatility in the term structure of government bond yields. *Journal of Banking & Finance* 36, 2988–3007.
- Jensen, M. B., Lunde, A., 2001. The NIG-S & ARCH model: a fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *Econometrics Journal* 4, 319–342.

- Koopman, S. J., Mallee, M. I. P., der Wel, M. V., 2010. Analyzing the term structure of interest rates using the dynamic nelson-siegel model with time-varying parameters. *Journal of Business & Economic Statistics* 28 (3), 329–343.
- Kupiec, P. H., 1995. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* 3 (2), 73–84.
- Nelson, C. R., Siegel, A. F., 1987. Parsimonious modeling of yield curves. *The Journal of Business* 60 (4), 473–489.
- Noureldin, D., 2011. Forecasting changes in the term structure of interest rates. Ph.D. thesis, Oxford University.
- Patton, A., 2006. Modelling asymmetric exchange rate dependence. *International Economic Review* 47 (2).
- Patton, A., 2009. Copula-based models for financial time series. In: Andersen, T. G., Davis, R. A., Mikosch, T. (Eds.), *Handbook of Financial Time Series*. Springer-Verlag, pp. 767–785.
- Patton, A., 2012. Copula methods for forecasting multivariate time series. In: Elliott, G., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*. Vol. 2B. Springer Verlag.
- Rivers, D., Vuong, Q., 2002. Model selection tests for nonlinear dynamic models. *Econometrics Journal* 5, 1–39.
- Rydberg, T. H., 1997. The normal inverse gaussian lévy process: Simulation and approximation. *Stochastic Models* 13 (4), 887–910.
- Schmidt, T., 2007. Coping with copulas. In: Rank, J. (Ed.), *Copulas - From Theory to Applications in Finance*. Risk Books.
- Schwartz, E. S., 1997. The stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of Finance* LII (3).
- Sklar, A., 1959. Fonctions de répartition à n dimensions e leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris* 8, 229–231.
- West, J., 2011. Long-dated agricultural futures price estimates using the seasonal nelson-siegel model. *Griffith Business Scholl, Discussion Papers, Finance* 7.

Research Papers 2013



- 2013-17: Daniela Osterrieder: Interest Rates with Long Memory: A Generalized Affine Term-Structure Model
- 2013-18: Kirstin Hubrich and Timo Teräsvirta: Thresholds and Smooth Transitions in Vector Autoregressive Models
- 2013-19: Asger Lunde and Kasper V. Olesen: Modeling and Forecasting the Volatility of Energy Forward Returns - Evidence from the Nordic Power Exchange
- 2013-20: Anders Bredahl Kock: Oracle inequalities for high-dimensional panel data models
- 2013-21: Malene Kallestrup-Lamb, Anders Bredahl Kock and Johannes Tang Kristensen: Lassoing the Determinants of Retirement
- 2013-22: Johannes Tang Kristensen: Diffusion Indexes with Sparse Loadings
- 2013-23: Asger Lunde and Anne Floor Brix: Estimating Stochastic Volatility Models using Prediction-based Estimating Functions
- 2013-24: Nima Nonejad: A Mixture Innovation Heterogeneous Autoregressive Model for Structural Breaks and Long Memory
- 2013-25: Nima Nonejad: Time-Consistency Problem and the Behavior of US Inflation from 1970 to 2008
- 2013-26: Nima Nonejad: Long Memory and Structural Breaks in Realized Volatility: An Irreversible Markov Switching Approach
- 2013-27: Nima Nonejad: Particle Markov Chain Monte Carlo Techniques of Unobserved Component Time Series Models Using Ox
- 2013-28: Ulrich Hounyo, Sílvia Goncalves and Nour Meddahi: Bootstrapping pre-averaged realized volatility under market microstructure noise
- 2013-29: Jiti Gao, Shin Kanaya, Degui Li and Dag Tjøstheim: Uniform Consistency for Nonparametric Estimators in Null Recurrent Time Series
- 2013-30: Ulrich Hounyo: Bootstrapping realized volatility and realized beta under a local Gaussianity assumption
- 2013-31: Nektarios Aslanidis, Charlotte Christiansen and Christos S. Savva: Risk-Return Trade-Off for European Stock Markets
- 2013-32: Emilio Zanetti Chini: Generalizing smooth transition autoregressions
- 2013-33: Mark Podolskij and Nakahiro Yoshida: Edgeworth expansion for functionals of continuous diffusion processes
- 2013-34: Tommaso Proietti and Alessandra Luati: The Exponential Model for the Spectrum of a Time Series: Extensions and Applications
- 2013-35: Bent Jesper Christensen, Robinson Kruse and Philipp Sibbertsen: A unified framework for testing in the linear regression model under unknown order of fractional integration
- 2013-36: Niels S. Hansen and Asger Lunde: Analyzing Oil Futures with a Dynamic Nelson-Siegel Model