A Realized Dynamic Nelson-Siegel Model with an Application to Crude Oil Futures Prices*

Ilya Archakov
University of Vienna

Bo Laursen
Aarhus University, CREATES

1st December, 2017

Abstract

We extend the popular dynamic Nelson-Siegel framework (Diebold et al. (2006)) by introducing time-varying volatilities in the factor dynamics and incorporate realized measures to estimate the latent volatility process more accurately. The new model is able to effectively describe the conditional variance dynamics of the term structure of futures prices and can still be readily estimated with the Kalman filter. We apply our framework to model the crude oil futures prices. Using more than 150,000,000 intra-daily transaction prices we construct realized volatility measures corresponding to the latent Nelson-Siegel factors, estimate the model at daily frequency and evaluate it by forecasting the conditional variance and density of futures prices. We document that the time-varying volatility specification suggested in our model strongly outperforms the constant volatility benchmark. In addition, the use of realized measures provides moderate, but systematic gains in forecasting.

JEL classification: C32, C51, G17.
Keywords: Term Structure, Dynamic Nelson-Siegel, Realized GARCH, High Frequency Data, Oil Futures.

*We are thankful to Asger Lunde, Peter Hansen and Martin Andreasen for multiple discussions that helped to improve the paper significantly. All errors are ours.

1 Introduction

Throughout the last decade commodities have become a popular asset class among investors. The commodity markets have experienced a rapid growth in both trading activity and volume due to large inflows of investment capital. Such financialization has directly affected commodity price dynamics and volatility[1] and, as a consequence, has emphasized the role of interactions between commodity prices and macroeconomic variables.[2] Furthermore, the financial crisis spurred investors interest in commodities since they provide a natural alternative for financial securities.[3] Analyses of commodity prices have, therefore, become increasingly important both from an academic perspective and for practitioners.

The predominant amount of literature related to modeling and forecasting commodity prices focuses on point forecasts, which are informative only about the central tendency of predictions. In contrast to that, density forecasts provide rich information about uncertainty around the central tendency and allow to invoke many additional useful statistics such as interval predictions, upside and downside risk assessments, etc. This appealing feature recently caused a burst of interest in density forecasting of economic and financial variables among researchers and policymakers (see e.g. Hall and Mitchell (2007), Clark (2011)). The density dynamics is of special interest when dealing with commodity markets. The poor point predictability of commodity spot and futures prices enhances the importance of having reliable density forecasts. To accurately evaluate uncertainty associated with point forecasts is of great practical value for risk management purposes and might also be useful for central bankers who are interested in relations between the dynamics of commodity prices and macroeconomic variables.

In this paper, we analyze the role of time-varying volatility in modeling commodity futures prices from the standpoint of forecasting. For this purpose we develop a class of models that allow us to capture the futures price dynamics for the full spectrum of maturities. These models perform well both in and out of sample and simultaneously account for the time-varying volatility. The latter is the key element in our framework that allows capturing changes in the second moment of distribution dynamics. We examine to what extent time-varying volatility is able to improve the uncertainty prediction and whether we gain from the use of realized measures of volatility.

Commodities are by nature very different from financial securities. The vast majority of trading in commodities happens through exchange traded futures contracts. Analyzing commodity prices therefore often involve modeling the whole term structure of futures prices. A standard approach for modeling the term structure is to specify a dynamic model for the underlying spot price[4] and derive the futures price based on no arbitrage arguments (see Brennan and Schwartz (1985), Schwartz (1997), Geman (2005), etc.). In this paper, we make use of another approach. Instead of building a model based on the spot price dynamics, we explicitly assume a functional form for the term structure curve.

We apply the dynamic Nelson-Siegel (DNS henceforth) factor approach of Diebold and Li (2006). The approach is based on a dynamic version of the classical Nelson-Siegel interest rate model, which is specifically suited for forecasting yield curves. The data on commodity futures share many similarities with interest rate data. Hence, the DNS framework can be directly applied to commodity markets. In this case, the Nelson-Siegel approach is based on the assumption that the futures prices can be described by three unobserved factors interpreted as the level, slope,

[1] Empirical evidence on the impact of the commodity market financialization on the prices is given in Buyuksahin and Robe (2014), among others. An analysis of such impact from the theoretical perspective is suggested in Basak and Pavlova (2015) (see also references therein).

[2] The links between commodity prices and a real economic activity are discussed in Kilian (2008), Kilian (2009), Collier and Goderis (2012) as well as in many other papers.

[3] An attractive diversification potential of commodities was pointed out in Gorton and Rouwenhorst (2006), for example. However, recent empirical studies provide a cautious assessment of such diversification and hedging abilities identifying an increasing integration between commodity and equity markets around the financial crisis episode (see Ravazzolo and Lombardi (2012), Hansen et al. (2014), Bhardwaj et al. (2015), etc.).

[4] In addition, modeling commodity prices potentially involves specifying a certain stochastic variable, which reflects benefits/costs to the holder coming from the possession of a commodity. For example, the net convenience yield is a usual interpretation of such variable.

and curvature of the futures curve. The factors are assumed to follow particular dynamics that allow forecasting the whole term structure of futures prices.

The futures returns of the majority of commodities show a clear sign of time-varying volatility. The turbulence of energy prices around the financial crisis and the recent oil price downfall provide fresh examples of periods with excess volatility. The volatility dynamics of many commodity prices also tend to cluster, that is demonstrating a cyclical behavior where the moderately volatile time intervals alternate with the periods of high price variation.

We introduce time-variation directly into the conditional variance of the three dynamic factors. Therefore, both the conditional mean vector and the conditional covariance matrix of the futures term structure are determined by the time-varying mean and volatility components of the Nelson-Siegel factors respectively. This factor approach effectively reduces the dimensionality of the problem and allows to parsimoniously model the joint dynamics of highly multivariate futures prices. Moreover, such parsimony likely benefits the forecasting performance of the model. Compared with the constant volatility frameworks, time-varying aspects allow to assess more accurately an increased degree of the forecast uncertainty throughout the periods of price turbulence and not to overestimate such uncertainty during the calm periods. We note that an additional challenge comes from the fact that we model the volatility of unobserved variables (Nelson-Siegel factors) opposed to the more standard frameworks (e.g., VAR models with time-varying volatility) where only the volatility is unobserved.

Furthermore, we introduce realized measures of volatility into the model. While doing this we closely follow the Realized GARCH framework of Hansen et al. (2012). Namely, we not only insert the realized measures directly into the conditional dynamics of the volatility components, but also "close" our model by adding measurement equations for the realized measures. The use of precise volatility information extracted from high frequency data often provides a better signal about the current level of volatility than what could otherwise be extracted from the lower frequency data. Hence, models that exploit realized measures should be better at "catching up" with the hidden volatility dynamics and therefore more suited to periods where the volatility changes rapidly.

We apply our modeling framework to the futures prices of the most traded commodity from the energy sector - WTI light sweet crude oil. We demonstrate that time-varying volatility significantly improves variance and, more generally, distribution forecasting of the oil futures prices. On top of that, the use of realized volatility measures provides moderate in magnitude, but systematic improvements. We also find that including several dynamic components to control the conditional variance of the DNS factors is preferable to using just a single factor.

The rest of the paper is organized as follows. In the remaining part of this section we provide a brief review of the related literature. In Section 2 we introduce our main modeling framework. In Section 3 we describe the futures data used in the empirical analysis. We discuss the choice and construction of realized volatility measures from the high frequency intra-daily price observations in Section 4. We outline the model evaluation strategy in Section 5 and Section 6 provides the corresponding results. Section 7 concludes and all supplementary materials are found in the Appendices.

1.1 Related Literature

Our paper refers to several strands of the literature. First of all, it relates to the area of modeling and forecasting term structure curves with dynamic factor models. Pioneered by Diebold and Li (2006) and Diebold et al. (2006) DNS models have become a popular framework for modeling and forecasting term structures of interest rates. The original models primarily focused on point forecasting and assumed a constant conditional variance for the term structure curve.

Recently, several versions of the DNS model with time-varying volatility have been proposed in the literature. Koopman et al. (2010) added a time-varying volatility component to the idiosyncratic conditional variances of the term structure. Alternatively, Bianchi et al. (2009), Hautsch and Ou (2012), Shin and Zhong (2015) introduced stochastic volatility directly into the dynamics of the common Nelson-Siegel factors. Applied to the interest rate yield dynamics, these models confirmed the usefulness of a time-varying volatility for in-sample fit and revealed some gains for point and density predictability. In our paper, we suggest a modification of the DNS modeling framework, which introduces an observation driven[5] conditional volatility into the Nelson-Siegel factors and exploits intra-daily realized volatility measures.

The DNS model has previously been applied to commodity markets. Karstanje et al. (2015) examine comovements in different commodities by assuming that the factors of each commodity can be decomposed into a market, sector, and idiosyncratic components. Hansen and Lunde (2013) analyze futures prices on oil by modeling the factors using a GARCH model with Normal Inverse Gaussian innovations within a copula framework. The method provides a flexible model suitable for matching heavy tails in the conditional density of the commodity prices. Although we rely on a conditionally Gaussian density, we note that by simply allowing the volatility of the factor innovations to be time-varying the model may also replicate some of the heavy-tailed aspects of futures returns. The procedure in Hansen and Lunde (2013) demands extracting the latent factors in a first step and subsequently treating them as if they were observable. This method becomes problematic when the cross-sectional dimension of the futures data is sparse. In our approach the factors are extracted directly using the Kalman filter. Thus, while filtering the latent factors, we effectively exploit both cross-sectional and time-series dimensions of the data.

Secondly, we add to the spacious and still growing area of volatility models that exploit realized measures as an extra source of information about the latent volatility level. The idea originates from Andersen and Bollerslev (1998) where the realized variance based on intra-daily asset returns was used as an ex-post return variance proxy to assess the performance of the ARCH-type dynamic models. Shortly thereafter, realized variance measures based on high frequency data were directly introduced into the volatility dynamics in (G)ARCH models (Engle (2002)) and stochastic volatility models (Barndorff-Nielsen and Shephard (2002)). The more accurate estimation and the forecasting gains associated with the inclusion of realized measures into the modeling frameworks inspired a rapidly growing research activity in this area.[6]

The way of volatility modeling suggested in this paper is closely related to the Realized GARCH model of Hansen et al. (2012). We also incorporate realized measures into the GARCH dynamics of the latent conditional volatility and use measurement equations to directly establish a link between realized measures and the volatility process. Loosely speaking, our framework can be considered as a multivariate version of the Realized GARCH where the state-space DNS factor model plays a role of the return equation linking the dynamic volatility components with the conditional distribution of the low frequency price vector. Naturally, the presented framework can be easily extended for the case of a more general dynamic factor model in a place of the return equation.[7]

A closely related paper from a methodological perspective is Shin and Zhong (2015). They incorporated realized measures into the DNS framework with stochastic volatility and documented the corresponding gains for bond yield density prediction. Similarly to their framework, we introduce a time-varying conditional volatility into the Nelson-Siegel factor dynamics and link it with realized measures of volatility obtained from observations of a higher frequency. Our modeling strategy, however, is different in several aspects. Opposed to the prevailing practice in the literature, we model the conditional volatility dynamics not as a stochastic volatility process, but in an observation driven manner. The stochastic volatility models often lead to non-linear non-Gaussian state space models and thereby complicate the estimation dramatically. In contrast, our observation driven specification can be estimated

[5]According to the classification in Cox (1981) (see also a related discussion in Koopman et al. (2015)), an observation driven specification of a time-varying parameter implies that its current realization is a deterministic function of its lagged realizations and lagged observable variables (the most notable examples include (G)ARCH, ACD and Dynamic Conditional Score (DCS) models). In contrast, a parameter driven specification suggests that a time-varying parameter is driven by its own stochastic innovations (as in unobserved component models, stochastic volatility models, etc.).

[6]Some selected examples related to asset return modeling include observation driven MEM of Engle and Gallo (2006), HEAVY model of Shephard and Sheppard (2010) as well as parameter driven (stochastic volatility) models of Takahashi et al. (2009), Koopman and Scharth (2012), etc.

[7]From this perspective, our model is a hybrid of the Realized GARCH of Hansen et al. (2012) and the framework of Harvey et al. (1992) (or the Factor GARCH of Engle et al. (1990) where the common factors are latent and have to be extracted).

straightforwardly by maximum likelihood via the Kalman filter. Moreover, we suggest the use of *direct* realized volatility measures for the latent Nelson-Siegel factor innovations to avoid imposing structural non-linear relations between the vector of term structure realized volatility measures and the three volatility components related to the Nelson-Siegel factors, as in Shin and Zhong (2015).

Finally, we contribute to the empirical literature on the forecasting of commodity futures prices. In doing so, we apply our modeling framework to the prices of futures contracts and evaluate how well the model is able to predict the conditional density for the whole range of maturities. So far, the literature related to density forecasting of commodities is scarce and the corresponding findings are not very conclusive. Using the DNS framework Hansen and Lunde (2013) found some improvements from the time-varying volatility for the point forecasts of crude oil futures prices and for the corresponding value-at-risk forecasts. Lunde and Olesen (2014) used the Realized GARCH framework and documented only slight gains in the forecasting of return densities of univariate electricity forwards from the use of realized volatility measures. Despite that our framework comprises certain elements from these two settings, it differs from both. On one hand, we complement to Hansen and Lunde (2013) by focusing on the implications of time-varying volatility on the variance prediction and by integrating realized volatility measures into the framework. On the other hand, although we also exploit realized measures and invoke Realized GARCH dynamics in our study, we model the joint multivariate conditional distribution of the whole term structure, but not only the univariate dynamics of the shortest contract as in Lunde and Olesen (2014).

2 Modeling Strategy

2.1 Nelson-Siegel Framework

We model the term structure of futures contracts within a dynamic Nelson-Siegel framework. In their original paper Nelson and Siegel (1987) suggested a parsimonious and flexible three-component polynomial approximation for the term structure of Treasury bond yield curves. The choice of polynomial functions was associated with the solution to a second-order differential equation and provided an ability to fit a range of shapes typically revealed by yield curves including monotonic, hump-shaped and inverted hump-shaped patterns.

Subsequently, Diebold and Li (2006) interpreted the Nelson-Siegel framework as a factor model where the factor loadings are defined by the Nelson-Siegel polynomials and are treated as latent dynamic factors. Such extension made it possible to filter the dynamically evolving latent Nelson-Siegel factors and to produce out-of-sample forecasts of the entire yield curve. The class of dynamic Nelson-Siegel factor models has been then successfully applied to modeling and forecasting different types of assets and financial instruments with a term structure embedded.

We denote the logarithmic price at time period t of a contract with τ periods until expiration by $y_t(\tau)$. Assume that at time t there are n_t types of contracts in the market that differ in their times until expiration: $\tau_{t,1} < \tau_{t,2} < ... < \tau_{t,n_t}$. Note that n_t likely changes with t since the number of contracts available for trade is varying from day to day. We also introduce the $n_t \times 1$ vector $y_t(\tau_t)$, which collects the futures prices for all n_t contracts traded at time t, $y_t(\tau_t) = (y_t(\tau_{t,1}), y_t(\tau_{t,2}), ..., y_t(\tau_{t,n_t}))'$. In accordance with the dynamic Nelson-Siegel framework, a scalar log-price component $y_t(\tau)$ for some given τ obtains the following factor representation[8]

$$y_t(\tau) = f_{l,t} + \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau}\right) f_{s,t} + \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) f_{c,t} \tag{1}$$

where $f_{l,t}$, $f_{s,t}$ and $f_{c,t}$ are latent dynamic factors and $\lambda > 0$ is a parameter of the Nelson-Siegel factor loadings.[9]

[8] In equation (1) we use a slightly reparametrized version of the original Nelson-Siegel curve, as it is suggested in Diebold and Li (2006).

[9] We note that parameter λ can be considered as a time dependent parameter, λ_t. In most of related studies, however, λ is treated as a constant parameter. Although being convenient for estimation, such assumption is quite restrictive, especially if the time span of the analyzed period is sufficiently long. There are several papers where time-varying λ_t was examined in the context of term structure

The model in (1) links futures prices with the maturities by means of the three common factors and the associated factor loadings. The loadings to the first factor $f_{l,t}$ are constant across the whole range of maturities and a change in $f_{l,t}$ translates to the same change for all prices in y_t. Factor $f_{l,t}$ is therefore usually interpreted as a level of the term structure curve. The second factor, $f_{s,t}$, has loadings which are monotonically decreasing as time to maturity τ grows, so it is commonly interpreted as a slope of the term structure. Parameter λ regulates the steepness of this slope. The higher λ is, the faster decay is exhibited by the loading function. The loadings to the factor $f_{c,t}$ also depend on τ and λ. They converge to 0 for $\tau \to 0$ and $\tau \to \infty$ and reach a maximum at some intermediate value of τ. Parameter λ governs the location of the extremum. When λ gets larger, the peak of the loading function shifts towards lower values of τ. This hump-shaped loading function associated with $f_{c,t}$ leads to an interpretation of this factor as the curvature of the term structure. Whereas a change in the level factor $f_{l,t}$ affects all the contracts in y_t uniformly, an increase in the slope factor $f_{s,t}$ imposes an effect mainly on short-term contract prices and an increase in the curvature factor $f_{c,t}$ increases prices for mid-term contacts.

The dynamic Nelson-Siegel framework treats factors as time-varying processes, which explain the variation in the term structure. Most of the related literature, however, is focused exclusively on the dynamics of the conditional mean of the factor process. As a consequence, an associated analysis is usually limited to modeling a dynamic location of the term structure variable and its point forecasting, so the uncertainty around the extracted and predicted term structure curves is left uncovered. This paper shifts the focus to the dynamic conditional variance of the factors. Such aspect naturally allows to model and forecast the conditional density dynamics of the term structure curve, which is implied by the conditional density dynamics of Nelson-Siegel factors.

2.2 Dynamic Nelson-Siegel Model with Time-Varying Volatility and Realized Measures

2.2.1 Dynamics of Conditional Mean

Following Diebold et al. (2006) we cast the dynamic Nelson-Siegel model from (1) in a state-space representation.[10] The observation equation in a matrix form is

$$y_t = \Lambda_t f_t + \varepsilon_t \tag{2}$$

where $f_t = (f_{l,t}, f_{s,t}, f_{c,t})'$ is a 3×1 vector of dynamic latent factors, $\varepsilon_t = (\varepsilon_{t,1}, \varepsilon_{t,2}, ..., \varepsilon_{t,n_t})'$ is a $n_t \times 1$ vector of idiosyncratic components, and $\Lambda_t = \Lambda_t(\lambda; \tau_{t,1}, \tau_{t,2}, ..., \tau_{t,n_t})$ is $n_t \times 3$ matrix of factor loadings

$$\Lambda_t = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_{t,1}}}{\lambda\tau_{t,1}} & \frac{1-e^{-\lambda\tau_{t,1}}}{\lambda\tau_{t,1}} - e^{-\lambda\tau_{t,1}} \\ 1 & \frac{1-e^{-\lambda\tau_{t,2}}}{\lambda\tau_{t,2}} & \frac{1-e^{-\lambda\tau_{t,2}}}{\lambda\tau_{t,2}} - e^{-\lambda\tau_{t,2}} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_{t,n_t}}}{\lambda\tau_{t,n_t}} & \frac{1-e^{-\lambda\tau_{t,n_t}}}{\lambda\tau_{t,n_t}} - e^{-\lambda\tau_{t,n_t}} \end{pmatrix} \tag{3}$$

The state equation describes the dynamics of latent factors. We suggest a simple random walk specification[11]

modeling of interest rates (see Creal et al. (2008) (section 4.1.3), Koopman et al. (2010)). Nevertheless these studies do not provide an unambiguous answer on whether the gains from time-varying λ_t are significant for the model fit and forecasting. Due to the different focus of the present paper, we will treat λ as a constant parameter as it is common in the literature on term structure modeling.

[10] State-space representation of a dynamic Nelson-Siegel model is closely related to the traditional factor model representation (see Stock and Watson (2011) for a related survey). According to the factor model interpretation, as it is conventional in the literature, term $\Lambda_t f_t$ from (2) is called a common component and ε_t is an idiosyncratic component. A part of variation in y_t explained by the common component is due to the latent factors f_t which are common for all variables in y_t. The idiosyncratic component is a zero mean stationary process, which comprises variable specific variation of y_t that is not explained by the common factors and absorbs measurement errors.

[11] Note that a more general VARMA process (with proper identification restrictions) can be used to model transition dynamics within a linear state-space framework. In particular, a popular choice is to assume a first-order autoregressive process as in Diebold et al. (2006).

$$f_t = f_{t-1} + \eta_t \tag{4}$$

where η_t is a 3×1 shock to the state vector f_t orthogonal to ε_t. We select a random walk for several reasons. At first, it is an appealing choice due to its parsimony and is a pretty suitable assumption in our context, since we mainly focus on the time-varying volatility of the futures term structure and not on the dynamics of its conditional mean. Such simple specification not only facilitates the estimation process, but is also less prone to overfitting that may have an adverse effect on the out-of-sample model performance. At second, the Nelson-Siegel factors extracted from commodity futures prices on a daily basis are, in fact, close-to-unit-root processes (see Hansen and Lunde (2013) for an empirical evidence).

We assume that the idiosyncratic component ε_t is an autoregressive process which reads

$$\varepsilon_t = \beta C_t \varepsilon_{t-1} + w_t \tag{5}$$

where C_t is a selection matrix of dimensions $n_t \times n_{t-1}$ filled with zeros and ones. An entry (i, j) of matrix C_t is equal to 1 if some contract that is traded at $t-1$ (with maturity $\tau_{t-1,j}$) is also traded at t (with maturity $\tau_{t,i}$ which is equal to $\tau_{t-1,j} - 1$). If otherwise, an entry (i, j) of matrix C_t is equal to 0. Parameter $\beta \in (0, 1)$ is a constant scalar that determines the persistence of the price misspecification generated by the model. We then suppose that $w_t \sim$ i.i.d.$N(0, \Sigma_t)$, where Σ_t is a constant diagonal matrix with time-dependent dimensions $n_t \times n_t$. We use a simple parametrization $\Sigma_t = \sigma_w^2 I_{n_t}$, where I_{n_t} is the identity matrix of size n_t.[12]

The conditional mean dynamics of the futures prices suggested in this subsection have an appealing and intuitive interpretation. The common Nelson-Siegel term $\Lambda_t f_t$ from (2), which is governed by the random walk factors, can be treated as a persistent long-run dynamic component of the futures price vector y_t. The autoregressive idiosyncratic component ε_t, in turn, can be related to the transitory short-run fluctuations in y_t. As a result, our model implies that the futures prices are mean-reverting to its long-run (equilibrium) dynamics.[13]

2.2.2 Dynamics of Conditional Variance

Time-varying volatility is a central element in our analysis. We introduce it directly into the dynamics of the Nelson-Siegel factors

$$\eta_t \big| \mathcal{F}_{t-1} \sim N(0, \Omega_t) \tag{6}$$

where Ω_t is a 3×3 time-varying conditional covariance matrix of the factor increment $\Delta f_t = \eta_t$ and $\mathcal{F}_t$ is a filtration that contains all the relevant observable information up to day t inclusively. This way of modeling imposes a factor structure on the conditional variance of the multivariate price process, y_t

$$\mathrm{Var}\big(y_t \big| f_{t-1}; \mathcal{F}_{t-1}\big) = \Lambda_t \Omega_t \Lambda_t^{'} + \Sigma_t \tag{7}$$

Thus, the large $n_t \times n_t$ conditional covariance matrix of the whole term structure of futures prices $\mathrm{Var}\big(y_t \big| f_{t-1}; \mathcal{F}_{t-1}\big)$ is fully characterized by the deterministic matrices Λ_t and Σ_t as well as the small scale 3×3 time-varying conditional covariance matrix Ω_t.

A similar approach was used in Bianchi et al. (2009) and Hautsch and Ou (2012) among others, where stochastic volatility was introduced directly into the dynamics of the Nelson-Siegel factors. Koopman et al. (2010) pointed

[12] We note that the time dependence in Λ_t, C_t, and Σ_t is related to the number of contracts n_t traded at a given day t and their maturities $\{\tau_{t,1}, \tau_{t,2}, ..., \tau_{t,n_t}\}$. Since this information is exogenous in our model, matrices Λ_t, C_t and Σ_t change with time in a deterministic way.

[13] From this perspective our specification is conceptually similar to the modeling framework of Schwartz and Smith (2000).

out that this approach is somewhat restrictive since in this case the Nelson-Siegel loading matrix Λ_t serves as a weighting matrix for both conditional mean and variance of elements in y_t. As an alternative, they suggested to endow a covariance matrix of idiosyncratic components Σ_t with a time-varying volatility factor, so each element from y_t has a time-varying volatility component.

We, nonetheless, maintain the assumption from (6) and incorporate time-varying conditional volatility in the factor dynamics. In doing so, we are motivated by the argument that the total variation of the term structure variable y_t is mainly explained by the common factor dynamics f_t and only to a little extent by the idiosyncratic component ε_t. Therefore, uncertainty associated with the futures prices y_t is largely incorporated in f_t. As a result, we may expect that time-varying volatility of the factors is more relevant for density modeling and forecasting than time-varying volatility in the idiosyncratic term ε_t. Moreover, these two approaches have been compared in Shin and Zhong (2015) in a similar study in the context of bond yield density prediction. Their results confirm that whereas time-varying volatility in factor dynamics helps to sufficiently improve the model performance, dynamic components in the variance of the idiosyncratic term do not seem to play a significant role.

We use a constant conditional correlation (Bollerslev (1990)) structure to parametrize Ω_t and treat the conditional factor volatilities as time-varying processes. Thus, the conditional variance matrix can be decomposed as follows

$$\Omega_t = H_t R H_t = H_t \begin{pmatrix} 1 & \rho_{ls} & \rho_{lc} \\ \rho_{ls} & 1 & \rho_{sc} \\ \rho_{lc} & \rho_{sc} & 1 \end{pmatrix} H_t \tag{8}$$

where R is a constant correlation matrix and H_t is a diagonal matrix with conditional volatilities of the factor increments Δf_t on the main diagonal. Therefore, the dynamics of Ω_t is fully driven by the dynamics of the 3×3 diagonal matrix H_t. We will suggest two alternative dynamic specifications for H_t - with one and three time-varying parameters.

One of the key features of our framework is that we use realized measures of variance based on intra-daily high frequency data while formulating the dynamics for the time-varying parameters from Ω_t. The use of realized measures of volatility is commonly motivated by the precise information content of the high frequency data (e.g., Dobrev and Szerszen (2010)). We, thus, expect that the realized signals of the futures price volatility will help to formulate and extract the latent conditional variance dynamics more accurately and to improve the predictive performance of the model.

We incorporate realized measures into the model along the lines of Realized GARCH of Hansen et al. (2012). Thus, we formulate a transition (GARCH) equation that links the latent volatility with own lags and lagged realized measures. The realized measures, in turn, are linked to the latent volatility through the specially formulated measurement equation. The measurement equation provides a contemporaneous relationship between the latent dynamic variable and its realized proxy allowing for a stochastic measurement error. The role of such measurement equation, therefore, is to directly relate the latent volatility process to the corresponding measures and, thus, to incorporate the precise information about the conditional factor variance Ω_t into the model. Apart from that, this equation may account for a systematic measurement bias and can be accommodated with some stylized features of financial data, such as the leverage effect.

Note that another possibility could be to specify the dynamics for all six distinct elements of the 3×3 matrix Ω_t, thus reflecting a time variation in both conditional variances and correlations.[14] However, an approach with fully specified dynamics of Ω_t is hardly implementable in our application due to the difficulties with the extraction

[14]For example, a proper dynamics for all elements from Ω_t (which preserves a positive definiteness of Ω_t) can be formulated along the lines of Chiriac and Voev (2011) or Golosnoy et al. (2012). In the former case, the vector of Cholesky elements of Ω_t is modeled and then these elements can be related to the corresponding realized measures. In the latter case, Ω_t is treated as a matrix process and the realized measures of Ω_t are assumed to be distributed as a Wishart matrix random variable.

of a full realized matrix of Nelson-Siegel factor covariances at intra-daily frequency. We describe these challenges in details in the section related to the construction of realized volatility measures.

2.2.3 One-component Specification for the Conditional Variance Matrix

We start with a simple parametrization of the conditional covariance matrix Ω_t. We assume that the diagonal volatility matrix H_t from (8) depends on a single time-varying parameter controlling the volatilities of all 3 factors

$$H_t = \begin{pmatrix} \sqrt{\exp(h_t)} & 0 & 0 \\ 0 & \sqrt{a_s + b_s \exp(h_t)} & 0 \\ 0 & 0 & \sqrt{a_c + b_c \exp(h_t)} \end{pmatrix} \tag{9}$$

where a_s, b_s, a_c, b_c are non-negative scaling coefficients and h_t is a scalar time-varying parameter. Therefore, the variances of all three Nelson-Siegel factors are linear functions of the single time-varying variance component $\exp(h_t)$. Note that for identification we directly set $\exp(h_t)$ equal to the variance of the level factor. An exponential transformation simplifies the formulation of the transition and measurement equations for h_t since for any real values of h_t the positivity of the factor variances is preserved. Closely following the specification of the Realized GARCH model we set up the transition (GARCH) equation as an autoregressive process[15]

$$h_t = \gamma_0 + \gamma_1 s_{t-1} + \gamma_2 h_{t-1} \tag{10}$$

where s_t is an appropriately chosen realized measure for h_t. The measurement equation is formulated as

$$s_t = \xi + \phi h_t + \delta(z_t) + u_t \tag{11}$$

where $u_t \sim N\big(0, \sigma_u^2\big)$ is a stochastic measurement error, ξ and ϕ are constant parameters which determine the systematic bias of the realized measure s_t, $\delta(\cdot)$ is the so-called leverage function that introduces a dependence between variance and returns of futures prices, z_t is a suitable $\mathcal{F}_t$-measurable variable that reflects the price movement occurred at day t. We discuss the choice of z_t below.

An empirical usefulness of the leverage component introduced in such way was confirmed in Hansen et al. (2012) and Hansen and Huang (2016) in the context of modeling the conditional variance dynamics of asset returns. A standardized asset return plays the role of variable z_t there, so the measurement equation explicitly establishes the link between the conditional variance and price shocks. The specification for the leverage component $\delta(\cdot)$ suggested in Hansen et al. (2012) is the following polynomial function

$$\delta(z_t) = \delta_1 z_t + \delta_2 (z_t^2 - 1) \tag{12}$$

where δ_1 and δ_2 are constant parameters. Such functional form is convenient for several reasons. As long as z_t is a standardized random shock, the leverage component $\delta(z_t)$ has zero mean. It, therefore, does not induce an additional bias for s_t in the measurement equation (11). Also, the quadratic specification in (12) allows to capture asymmetry effects in the volatility responses to the price shocks.

The one-component specification of the conditional variance matrix Ω_t given by (8) and (9) implies that

$$\mathrm{Var}\big(y_t | f_{t-1}; \mathcal{F}_{t-1}\big) = \Lambda_t H(h_t) R H(h_t) \Lambda_t^{'} + \Sigma_t \tag{13}$$

Particularly, it means that conditional variances of all futures contracts are driven by the single dynamic volatility component h_t. As a consequence, we may expect that the realized variances of the futures prices can serve as good

[15] In this paper, we consider only the parsimonious (1,1) specification for lags. The lag order can be extended if needed.

signals about the unobservable volatility state and, thus, can be related to h_t. More specifically, for some selected $\tau \in \{\tau_{t,1}, \tau_{t,2}, ..., \tau_{t,n_t}\}$[16]

$$s_t \equiv \log RV\big(\Delta y_t(\tau)\big) \simeq \log \mathrm{Var}\big(y_t(\tau)|f_{t-1}; \mathcal{F}_{t-1}\big) \tag{14}$$

For such choice of s_t we suggest the following choice of variable z_t to establish a dependence between volatility component h_t and the futures price change

$$z_t = \frac{\Delta y_t(\tau) - \mathrm{E}\big(\Delta y_t(\tau)|f_{t-1}; \mathcal{F}_{t-1}\big)}{\sqrt{\mathrm{Var}\big(\Delta y_t(\tau)|f_{t-1}; \mathcal{F}_{t-1}\big)}} \tag{15}$$

so z_t is a conditionally standardized $\mathcal{F}_t$-measurable Gaussian random variable, which represents a price shock to the contract τ.

Summarizing, equations (9), (10), and (11), where s_t and z_t are defined as in (14) and (15) respectively, constitute the one-component dynamic specification for the conditional covariance matrix Ω_t.

2.2.4 Three-component Specification for the Conditional Variance Matrix

We also suggest a richer dynamic parametrization for Ω_t where three time-varying parameters are introduced. We link these dynamic components with the three conditional variances of the level, slope and curvature. The matrix H_t reads

$$H_t = \begin{pmatrix} \sqrt{\exp(h_{l,t})} & 0 & 0 \\ 0 & \sqrt{\exp(h_{s,t})} & 0 \\ 0 & 0 & \sqrt{\exp(h_{c,t})} \end{pmatrix} \tag{16}$$

where $h_{l,t}$, $h_{s,t}$ and $h_{c,t}$ are the three time-varying logarithmic conditional variances of the 3-variate shock η_t to the dynamic factor process f_t. We denote a 3×1 vector of dynamic volatility components as $h_t = \big(h_{l,t}, h_{s,t}, h_{c,t}\big)'$ and note that $\exp(h_t) = diag\big(\Omega_t\big)$. Similar to the previous specification, we write the transition equation as

$$h_t = \mathit{\Gamma}_0 + \mathit{\Gamma}_1 s_{t-1} + \mathit{\Gamma}_2 h_{t-1} \tag{17}$$

where $s_t = \big(s_{l,t}, s_{s,t}, s_{c,t}\big)'$ is a 3×1 vector of the proper realized measures for h_t and $\mathit{\Gamma}_0$, $\mathit{\Gamma}_1$, $\mathit{\Gamma}_2$ are constant parameter matrices of suitable dimensions. The measurement equation is defined as

$$s_t = \zeta + \Phi h_t + D(z_t) + u_t \tag{18}$$

where $u_t \sim N\big(0, \Sigma_u\big)$ is a 3×1 vector of measurement errors, $D(\cdot)$ is a multivariate leverage function, z_t is a $\mathcal{F}_t$-measurable 3×1 vector and ζ, Φ are constant matrices of proper dimensions. The leverage function in a multivariate version has the following form

$$D(z_t) = D_1 z_t + D_2(z_t \circ z_t - \iota_3) \tag{19}$$

where D_1, D_2 are 3×3 parameter matrices and ι_3 is the 3×1 vector of ones.

We attempt to directly extract realized measures s_t of the latent factor variances from intra-daily data by exploiting the recent advances in high frequency volatility estimation. The description of the estimation procedure

[16] Note that several realized measures related to distinct contracts $\tau \in \{\tau_{t,1}, \tau_{t,2}, ..., \tau_{t,n_t}\}$ can be used to extract h_t. In this case, in analogy with Hansen and Huang (2016), several measurement equations (11) which correspond to different realized measures $s_t(\tau)$ have to be formulated.

is referred to the next section. Thus, signals s_t in this case are the realized variance measures for Nelson-Siegel factor increments Δf_t

$$s_{i,t} \equiv \log RV\big(\Delta f_{i,t}\big) \simeq \log \mathrm{Var}\big(\Delta f_{i,t} | f_{t-1}; \mathcal{F}_{t-1}\big), \qquad i \in \{l, s, c\} \tag{20}$$

We define the corresponding 3×1 vector z_t, which is required for the leverage function, as

$$z_t = \mathrm{Var}\big(\mathrm{E}\big(f_t | \mathcal{F}_t\big) - \mathrm{E}\big(f_{t-1} | \mathcal{F}_{t-1}\big) | \mathcal{F}_{t-1}\big)^{-\frac{1}{2}} \big(\mathrm{E}\big(f_t | \mathcal{F}_t\big) - \mathrm{E}\big(f_{t-1} | \mathcal{F}_{t-1}\big)\big) \tag{21}$$

so z_t is a conditionally standard Normal random vector.

As a result, the three-component specification for the conditional covariance matrix Ω_t is represented by equations (16), (17) and (18) with s_t and z_t specified in (20) and (21) respectively. In our empirical analysis we stick to the parsimonious model parametrization. In particular, matrices $\mathit{\Gamma}_1$, $\mathit{\Gamma}_2$, Φ, D_1, D_2 from (18)-(20) are assumed to be diagonal.

2.3 Estimation

Suppose that we observe futures prices y_t and realized measures s_t within a sample period $t = 1, ..., T$. The joint conditional density function of observable variables for some given trading day t can be factorized as follows

$$p(y_t, s_t | \mathcal{F}_{t-1}; \theta) = p(y_t | \mathcal{F}_{t-1}; \theta) p(s_t | y_t, \mathcal{F}_{t-1}; \theta) \tag{22}$$

where $\mathcal{F}_t = \sigma(y_t, s_t, y_{t-1}, s_{t-1}, ...)$ is a σ-field generated by all observables up to t and θ is a vector of model parameters.

The first term in the decomposition, $p(y_t | \mathcal{F}_{t-1}; \theta)$, is the conditional density of observed futures prices. Equation (2) and the dynamic Nelson-Siegel specification imply that this conditional density is Gaussian and can be evaluated by means of the Kalman filter[17] given the sequence $\{\Omega_r,\ r \leq t\}$. The time-varying covariance matrix $\Omega_t = \Omega(h_t)$ from the state equation (6) is a $\mathcal{F}_{t-1}$-measurable variable as it follows from the assumptions of GARCH dynamics for h_t in (10) and (17). Therefore, for every t we have that $p\big(y_t | \Omega_t, \mathcal{F}_{t-1}; \theta\big) = p\big(y_t | \mathcal{F}_{t-1}; \theta\big)$ and the Kalman filter can be applied to evaluate the density $p(y_t | \mathcal{F}_{t-1}; \theta)$.

The second term in the factorization (22), $p(s_t | y_t, \mathcal{F}_{t-1}; \theta)$, relates to the conditional density of the realized measures s_t. From (11) and (18) it follows that such density is Gaussian. According to (15) and (21), the leverage variable $z_t = z_t(y_t, \mathcal{F}_{t-1}; \theta)$ for all t is expressed through the variables from the Kalman forward recursions applied to $p(y_t | \mathcal{F}_{t-1}; \theta)$, so the conditional densities for s_t satisfy $p(s_t | y_t, h_t, z_t, \mathcal{F}_{t-1}; \theta) = p(s_t | y_t, \mathcal{F}_{t-1}; \theta)$.

Finally, the log-likelihood function of the model can be written as

$$\log\ p(y, s | \theta) = \sum_{t=1}^{T} \log\ p\big(y_t | \mathcal{F}_{t-1}; \theta\big) + \sum_{t=1}^{T} \log\ p\big(s_t | y_t, \mathcal{F}_{t-1}; \theta\big) \tag{23}$$

The function in (23) is effectively a sum of conditional Gaussian log-densities. It can be maximized by means of standard numerical optimization routines and the corresponding quasi-maximum likelihood (QML) parameter estimates, $\hat{\theta}$, can be obtained. While estimating the model we rely on the QML argument in sense of White (1982). In particular, in case of certain distribution misspecifications we refer to (23) as to the quasi-likelihood function, so the corresponding QML estimator is supposed to retain its consistency.

[17] See comprehensive material on the state-space models and estimation in Harvey (1991), Durbin and Koopman (2012).

2.4 Discussion

Conceptually our modeling framework consists of two blocks. The first block relates to the observed term structure of futures prices. It is described by the dynamic Nelson-Siegel model with time-varying volatility imposed on the factor dynamics. The Nelson-Siegel factors govern the dynamics of the futures price vector y_t and the time-varying volatility embedded in these factors allows to explicitly model the uncertainty associated with y_t. The second block is auxiliary to the dynamic Nelson-Siegel model and links the latent time-varying factor volatility to the realized measures extracted from the high frequency data. This part of the model is unusual in the literature on term structure modeling and we expect that it improves the identification and forecasting the conditional volatility of factors.

The time-varying volatility component h_t is modeled as an observation driven process and fully determines dynamics of the factor conditional variance matrix Ω_t. Important to note that the process h_t incorporates information from both blocks of the framework, i.e. from both low and high frequency data. First, the daily increments of the latent Nelson-Siegel factors Δf_t (which are extracted from the daily prices y_t) can provide (noisy) signals about the latent conditional variance of f_t. Second, the realized measures s_t (based on intra-daily price observations) directly contribute information about the latent factor variance and these signals are expected to be more precise.

3 Data description

3.1 Transaction data

The raw dataset is obtained from Tick Data Inc. It consists of transactions on light crude oil (CL) futures traded on the Chicago Mercantile Exchange (CME). The data is analyzed using a period ranging from August 9, 2004 to January 16, 2015.

All futures trading are managed by the CME group and through their clearing house they act as a counterpart to all transactions. The transactions are made through a fully electronic trading system that allows market participants to trade around the clock Monday-Friday except for a 45 minutes trading break each day from 5:15p.m.-6.00p.m. EST. A special feature of these futures contracts is that the settlement at maturity is not by cash but via physical delivery. However, the vast majority of contracts never go to delivery. Most are canceled beforehand by taking the offsetting position. This allows for a huge amount of transactions to take place since it is possible to be a market participant without taking part in the physical delivery of the commodity. The availability of large amounts of transactions data facilitates computation of intra-daily measures, which form the basis of our analysis on volatility of oil futures.

Table 1 presents statistics for the transaction data. Generally the number of transactions has been increasing through most of the sample period. The average number of transactions per contract shows a similar pattern suggesting that the increase in transactions is not solely due to the introduction of additional contracts. The majority of the transactions are in futures with a short time to maturity while contracts with medium and long time to maturity are less traded.

The raw data series are cleaned using the algorithm in Barndorff-Nielsen et al. (2009). Specifically we delete the observation if the transaction price is more than five mean absolute deviations from a rolling centered median of the 25 preceding and the 25 subsequent observations. On rare occasions observations may also be reported with a time stamp that is not consistent with the corresponding trading day. These observations are also deleted. In total we delete 732,881 out of 153,883,976 transactions corresponding to nearly 0.5% of the initial sample.

Table 1: Average number of transactions

	Average number of transactions											
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total
Shortest	4,892	5,130	7,865	23,348	33,002	33,341	35,396	85,002	98,906	102,910	88,186	49,907
Medium	33	78	85	415	844	674	718	8,662	13,399	16,847	13,855	5,406
Longest	0	10	15	167	1,152	321	258	2,747	5,803	6,228	4,752	2,084
All	4,926	5,217	7,964	23,931	34,998	34,336	36,373	96,412	118,108	125,984	106,794	57,398

	Average number of transactions per contract											
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total
Shortest	699	735	1,148	3,343	4,700	4,741	5,048	12,046	13,951	14,586	12,494	7,080
Medium	10	15	23	80	153	112	125	884	1,322	1,666	1,373	560
Longest	NaN	9	13	51	269	97	40	211	411	450	364	200
All	480	395	652	1,548	2,042	2,072	1,902	3,348	3,776	4,063	3,494	2,273

The first part of the table presents the average number of transactions per day for each year in the sample period (rounded to nearest integer). *Shortest* refers to all futures contracts with time to maturity less than 149 business days. *Medium* contracts have between 149 and 360 business days to maturity and *Longest* refers to all contracts with time to maturity larger than 360. The second part shows the average number of transactions per contract per day (rounded to nearest integer). The year 2014 includes the period January 1, 2014 to January 16, 2015.

3.2 Daily data

The cleaned high-frequent transaction data is used to construct a dataset with daily observations. For all available contracts the closing price, assumed to be the last transaction of the day, is used as the daily futures price. This procedure results in a daily panel dataset consisting of 56,412 observations.

Table 2 presents some descriptive statistics for the constructed dataset. The futures prices are all reported as per barrel of oil. The number of traded contracts pr. day generally increases throughout the sample period. The dataset is therefore an unbalanced panel where the cross-sectional dimension varies with time. As evident from the statistics on time to maturity, the increase in the cross-sectional dimension generally comes from the introduction of new contracts with a very long time until expiration. The *Vol.* columns show that the volatility of the futures returns is generally decreasing with time to maturity. This effect is known as the Samuelson effect (Samuelson (1965)). The argument is that the arrival of news will have a direct impact on the shortest futures while long-term contracts tend to remain unchanged since adjustment of fundamentals is likely to take place before the contracts come to delivery at maturity.

Figure 1 presents the daily dataset visually. There are two types of plots. In both types each dot represents an observation. The first plot shows the evolution of all the futures prices as a function of time. The next plot shows the evolution of the corresponding time to maturity (measured in business days). The diagonal “lines” in the maturity plot represent specific contracts and arise since the time to maturity decreases linearly with time. The holes in the “lines” represent days where the contract is not traded and are more common for long contracts.

Generally it seems there are a lot of observations in the short end, while observations in the very long end are more scarce and generally not present before 2010. The long contracts are typically also issued with a longer time span between them, reflected by the longer distance between the diagonal “lines” in the maturity plots.

We divided the dataset into three periods. The first 100 days represent an initialization period used to initialize the Kalman filter (August 9, 2004 to December 30, 2004). The second period is the estimation period (January 3, 2005 to December 31, 2013) and the third period is used for forecasting (January 2, 2014 to January 16, 2015). The three periods are separated by vertical dashed lines in figure 1.

Figure 2 gives an example of what typical term structures may look like. Each dot represents a futures price

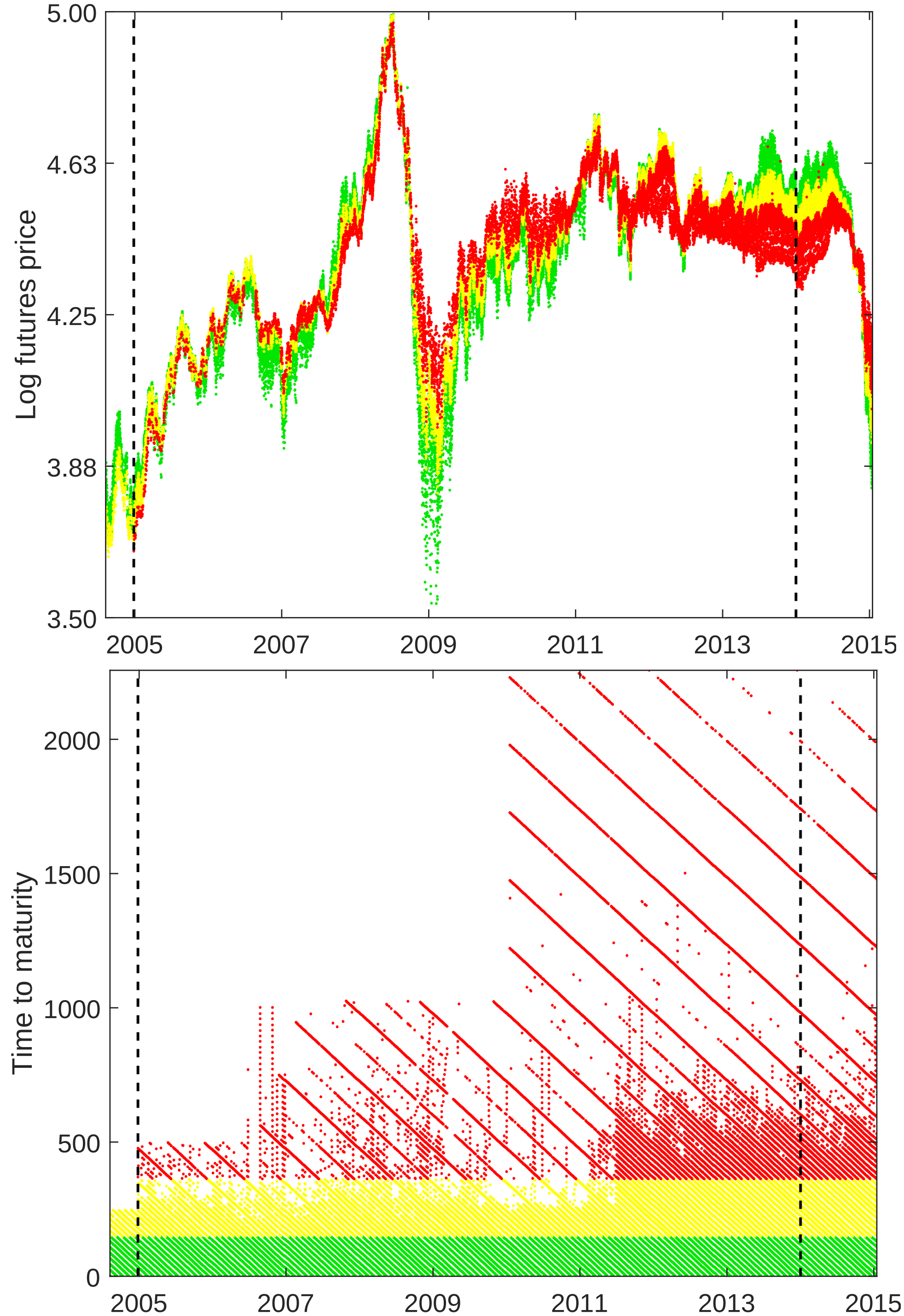


Figure 1: The daily dataset. Daily futures prices and time to maturity as functions of time. Each dot represents an observation. For illustration all observations are divided into three groups and colored with respect to time to maturity. Each group consists of approximately one-third of the observations. The first group represents observations with a maturity less than 149 business days and is presented in green while observations with a maturity longer than 360 are colored in red. The second group consists of observations in between and is colored in yellow.

Table 2: Descriptive statistics for daily futures data

			Shortest futures				Median futures				Longest futures			
Period	#Cont.	τ	Mean	Min.	Max.	Vol.	Mean	Min.	Max.	Vol.	Mean	Min.	Max.	Vol.
2004	10	108	47.3	40.7	55.2	0.40	43.2	38.6	49.5	0.28	43.2	38.6	49.5	0.27
2005	13	141	56.7	42.2	69.9	0.32	57.5	40.8	69.4	0.24	56.0	39.1	68.2	0.23
2006	12	140	66.3	55.9	77.4	0.27	70.9	64.2	80.8	0.21	70.6	64.5	80.0	0.20
2007	15	168	72.3	50.4	98.3	0.29	72.7	56.1	91.4	0.20	71.4	58.4	88.5	0.18
2008	17	191	99.9	33.2	145.4	0.57	100.8	49.4	146.3	0.40	102.0	64.8	142.2	0.34
2009	16	177	61.9	34.4	81.3	0.51	70.2	48.0	86.0	0.32	77.1	54.3	92.4	0.26
2010	19	205	79.5	69.2	91.4	0.29	84.3	73.9	94.4	0.24	91.5	77.0	100.3	0.18
2011	26	288	95.1	76.4	113.7	0.34	97.7	79.0	114.1	0.28	97.2	87.3	107.1	0.21
2012	31	326	94.2	78.0	109.6	0.25	96.0	81.6	109.2	0.21	87.1	82.0	94.1	0.17
2013	31	325	98.0	86.4	110.2	0.18	93.1	85.5	97.4	0.13	81.5	77.0	86.4	0.11
2014	30	319	91.1	45.7	107.3	0.24	86.4	53.7	98.1	0.18	80.8	66.1	89.8	0.13
Total	19	233	80.4	33.2	145.4	0.35	81.6	38.6	146.3	0.25	80.2	38.6	142.2	0.23

The descriptive statistics are based on the daily data set with sample period August 9, 2004 to January 16, 2015. The three time series *Shortest futures*, *Median futures*, and *Longest futures* are constructed by each day taking the shortest contract, the contract with time to maturity closest to 233 (the median number of business days to maturity in the full sample) and the longest contract respectively. *#Cont.* refers to the median number of contracts available per day during the given period. τ shows the median number of business days to maturity for all contracts available. *Vol.* measures the annualized volatility of futures return. The returns are calculated by excluding periods where the shortest contract moves into maturity, a new contract becomes the median contract or a new longest contract is issued.

observation and they are linearly interpolated for a better visualization. All the futures curves are normalized by the price of the shortest futures. It is clear that the futures curves are able to take very different forms. They can be both increasing (contango), decreasing (backwardation), hump-shaped, smooth, and irregular.

4 Intra-daily Data and Realized Measures

In the preceding section we built a framework for the term structure modeling that introduces time-varying conditional volatility in the dynamics of the Nelson-Siegel factors. The key element in our approach is the use of realized volatility measures constructed with intra-daily futures prices to approximate the latent volatility components in the factor dynamics. In this section, we briefly discuss the features of the high frequency futures data and outline how we use it to retrieve appropriate realized measures of factor volatility.

4.1 Intra-daily Data and Practical Challenges

For most business days the futures contracts considered in this paper are traded 24 hours with a short break between 5:15p.m. and 6:00p.m. EST. Thus, we conventionally associate the intra-daily trading period for business day t with the period between 6:00p.m. of $t-1$ and 5:15p.m. of t. Suppose that n_t contracts with times until maturity $\{\tau_{t,1}, \tau_{t,2}, ..., \tau_{t,n_t}\}$ are traded at a given day t. We now assume that there exists n_t-variate process $p_t(v) = \big(p_t^{(\tau_{t,1})}(v), p_t^{(\tau_{t,2})}(v), ..., p_t^{(\tau_{t,n_t})}(v)\big)'$, where $v \in [0,1]$ is an intra-daily time index and $p_t^{(\tau)}(v)$ denotes an intra-daily efficient log-price process of the contract with time to maturity $\tau \in \{\tau_{t,1}, \tau_{t,2}, ..., \tau_{t,n_t}\}$ (contract τ, henceforth). The daily closing price $y_t(\tau)$ which we consider in our main model is, therefore, associated with $p_t^{(\tau)}(1)$.[18]

In practice, an efficient intra-daily price process is never observable. Instead, we observe $p_t(v)$ only at the discrete moments of transactions and these observations are contaminated by microstructural noise. In addition

[18]Practically, however, we use $p_t^{(\tau)}(\nu_{last})$, where ν_{last} is the time of the last observed transaction for the contract τ at day t.

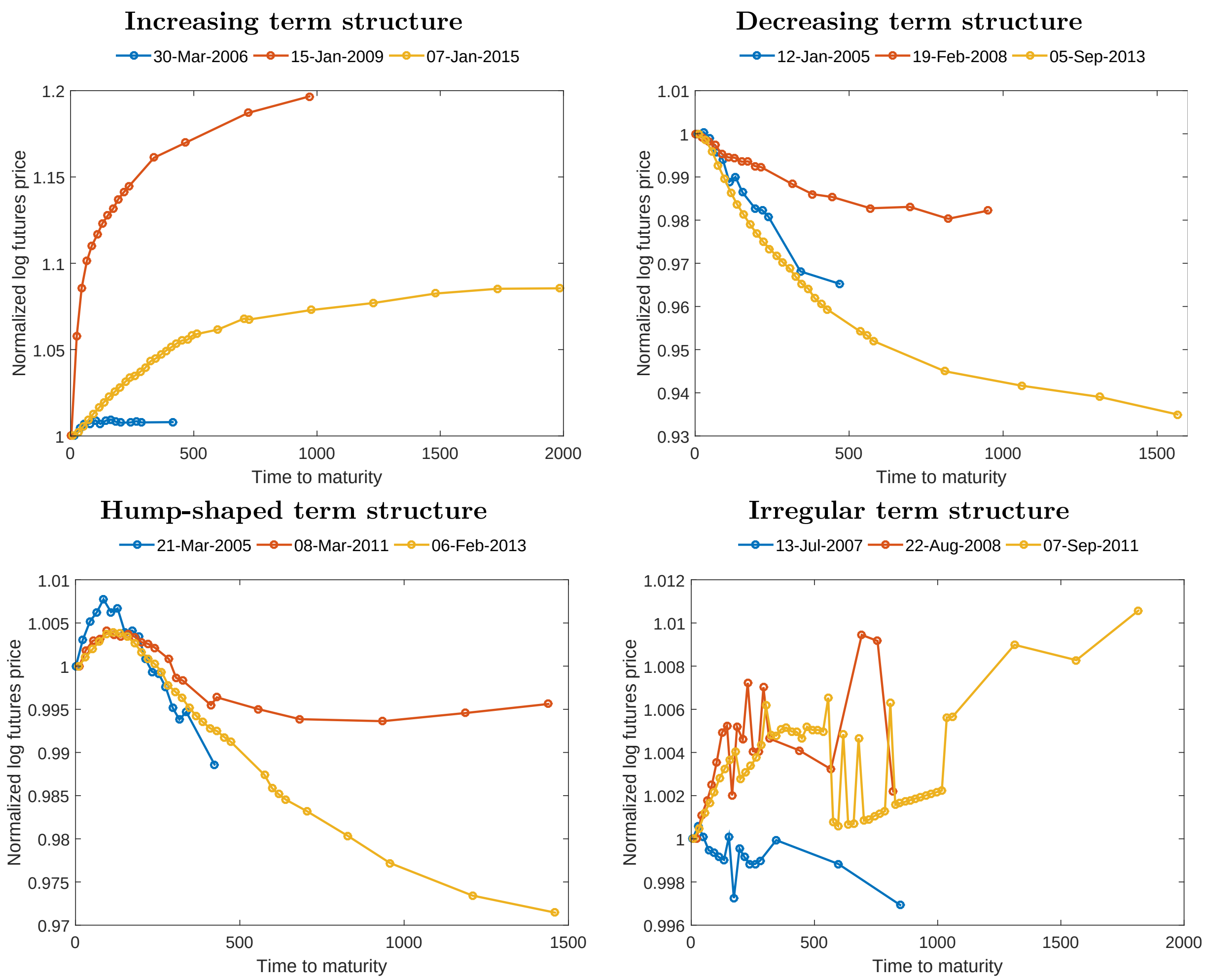


Figure 2: Different term structures of futures prices.

to the market microstructural effects,[19] a multivariate nature of intra-daily futures data suggests several other challenges. In particular, some of these challenges are illustrated in Figure 3, which provides examples of typical transaction patterns for the crude oil futures:

- transactions are irregular
- on average, the number of transactions is lower for contracts with higher time to maturity
- transactions are asynchronous across contracts

Active trading mostly occurs within a relatively short intra-daily sub-period (the distributions of intra-daily trading intensities for the futures contracts are given in Figure 4). As a consequence, transactions appear very irregularly over the trading day. In particular, between-trade durations range from milliseconds to several hours depending on the intra-daily period and the time to maturity of the contract.

Another implication of intra-daily transaction data relates to the unbalancedness of observed trades across contracts with distinct maturities. Short-maturity contracts are the most traded ones, whereas long-maturity contracts are traded less frequently. This feature is especially pronounced in the first years of our sample. The

[19] Market microstructural effects are inevitable attributes of financial data observed at high frequency. Among the sources of the microstructural noise researchers emphasize a bid-ask spread, fixed minimal price increment (tick size), etc. Theoretical and empirical implications of market microstructural effects on the observed price process and the analysis of the realized variance are considered in Bandi and Russell (2008) and Hansen and Lunde (2006b).

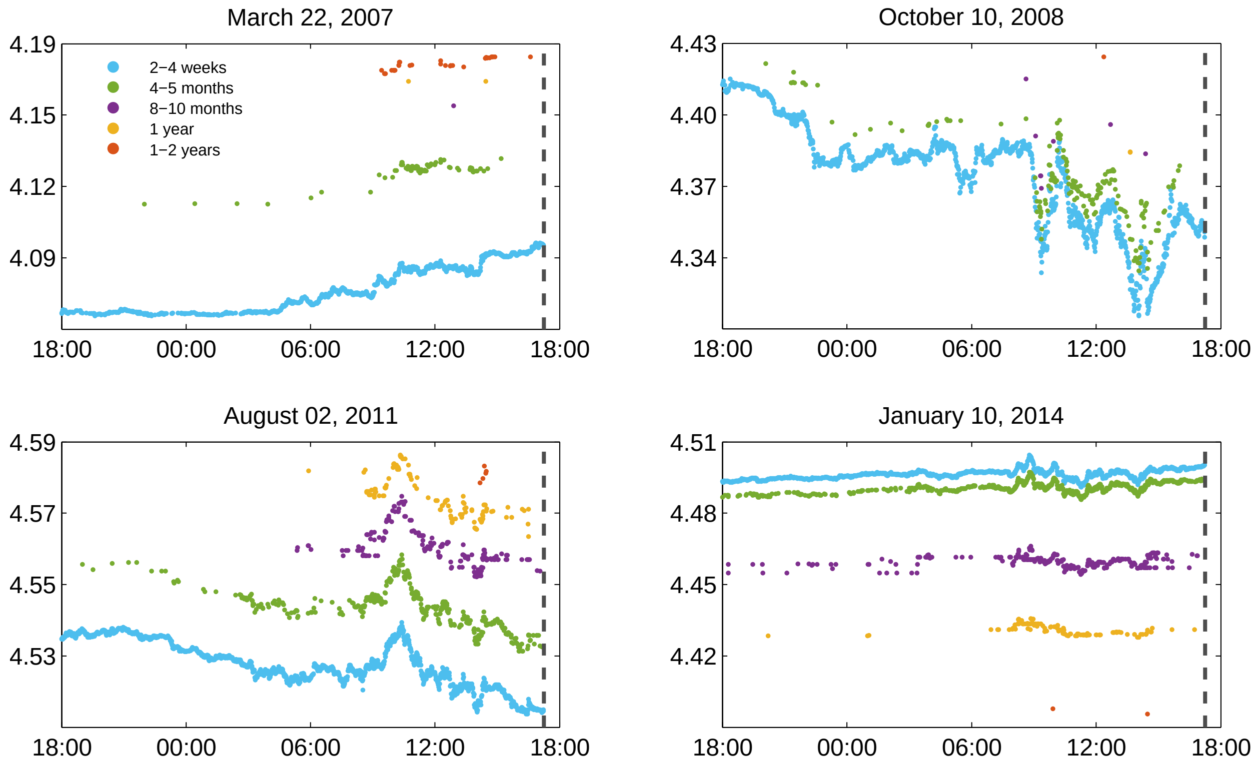


Figure 3: Examples of intra-daily transaction data for a 24-hour period which starts at 6p.m. the day before the date specified in the header and lasts until 6p.m. of the specified date. Plots contain log futures prices observed at the moments of transactions for 4 selected trading days from the analyzed sample. For expositional convenience we plot the data for only 5 selected contracts with different times until expiration. Dashed vertical lines correspond to the start of the trading break at 5:15p.m.

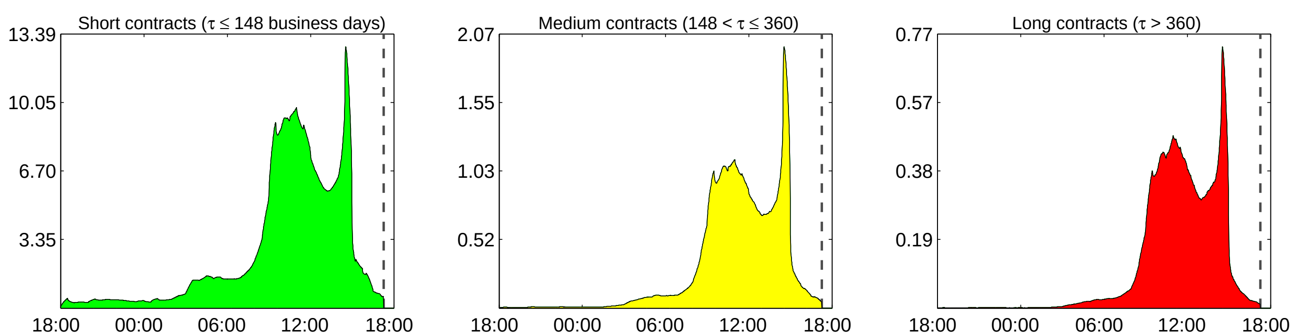


Figure 4: Intra-daily distributions of transactions observed over the period 01/2005-12/2014. The left plot corresponds to the futures contracts which expire no later than in 148 business days. The middle plot displays contracts with maturities between 148 and 360 days. The right plot refers to the futures contracts with more than 360 trading days until maturity. On the vertical axis there is a total number of observed transactions (in millions) for the specified contracts from the sample period 01/2005-12/2014 computed over 30-minute rolling intra-daily intervals. Dashed vertical lines correspond to the start of the trading break at 5:15p.m.

scarcity of transaction data for long-maturity contracts naturally complicates the extraction of realized volatility signals related to the long end of the term structure.

Transactions on distinct contracts arrive non-synchronously. Partially the asynchronicity can be attributed to the liquidity issues. Namely, we may expect that the fundamental news associated with the underlying commodity are getting incorporated into the futures prices not simultaneously, but with the lags for less liquid long contracts. Another contribution to non-synchronous trading stems from the situations where some information relevant only in the short run does not lead to the trading of long-maturity contracts and conversely. Such situation is typical for multivariate high frequency asset prices and leads to the so-called Epps effect (see Epps (1979)) in the context of covariance estimation.

Irregular, highly unbalanced and asynchronous transaction data substantially complicates the intra-daily analysis of the multivariate factor structure in the futures contracts as well as the construction of the corresponding realized measures of factor volatilities. Therefore, the expected degree of improvements in modeling and forecasting the term structure density from the use of high frequency realized volatility measures should be assessed cautiously. In the following subsections we discuss our approach to the construction of realized measures related to the conditional variance of the Nelson-Siegel factors.

4.2 Volatility Measures from Intra-daily Data

The aim of our intra-daily developments is to obtain realized measures s_t for the time-varying component h_t from the conditional factor variance matrix $\mathrm{Var}(\Delta f_t | f_{t-1}; \mathcal{F}_{t-1}) = \Omega_t$. As it is outlined in the model description, we consider two specifications of Ω_t based on the one- and three-component dynamic parametrizations. We now suggest and discuss suitable realized measures for each specification. An empirical justification of using few dynamic volatility components for modeling Ω_t is given in Appendix A.

4.2.1 Realized Measures for the One-component Specification

According to (14) the realized variance of a close-to-close futures return on some contract $\tau \in \{\tau_{t,1}, \tau_{t,2}, ..., \tau_{t,n_t}\}$ can serve as a signal of h_t, since all contracts $y_t(\tau_t)$ contain information about the common latent volatility component h_t. The unbalancedness of trading intensity across contracts with different times until expiration suggests that the front contract with the shortest time to maturity is the best candidate for measuring h_t. Such contract is systematically traded more frequently than others and, what is more important, the corresponding transactions appear in the course of the whole trading day. Therefore, the realized variance computed on the basis of intra-daily transactions on the shortest contract $y_t(\tau_{t,1})$ is expected to be a more efficient measure of h_t rather than a realized variance computed using some other contract with longer time to maturity. The realized variance of the shortest contract, thus, can be treated as a state variable which controls all three Nelson-Siegel factor volatilities.

To minimize the influence of microstructural noise effects we suggest to use a sparse sub-sampled 5-minute realized variance estimator.[20] For this purpose we partition a trading period $v \in [0, 1]$ with a 1-minute grid. We denote the set of equidistant gridpoint times as $\{v_i\}_{i=0}^{m}$, so $0 = v_0 < v_1 < ... < v_m = 1$. Although futures contracts with the shortest maturity $\tau_{t,1}$ are liquid enough and traded frequently, they do not necessarily occur exactly at the moments v_i for all i. Thus, we refer to the previous tick interpolating scheme. In particular, let $\tilde{p}(v_i)$ denote the transaction price in case we observe it at v_i. If we do not observe a transaction at that moment, then $\tilde{p}(v_i)$ stands

[20] We prefer model-free realized variance estimator since the almost 24-hours trading span provides sufficiently many sparsely sampled intra-daily returns. Also, Liu et al. (2015) demonstrated in an extensive empirical study that a sparsely sampled realized variance estimator exhibits an accuracy which is not significantly outperformed by any of the competing realized estimators. As an alternative, noise robust estimators of the quadratic variation based on more frequently sampled intra-daily returns can be exploited, such as in Barndorff-Nielsen et al. (2008), Jacod et al. (2009).

for the price observed for the nearest transaction prior to v_i.[21] Then, the sub-sampled 5-minute realized variance estimator based on intra-daily prices for contract τ reads

$$RV_t^{ss}(\tau) = \frac{1}{5}\sum_{j=1}^{5}\frac{m}{m-4}\sum_{i=j}^{m-5+j}\left(\tilde{p}_t^{(\tau)}(v_{5i}) - \tilde{p}_t^{(\tau)}(v_{5(i-1)})\right)^2 \tag{24}$$

and we define a realized measure for the one-component dynamic variance specification as $s_t \equiv \log RV_t^{ss}(\tau_{t,1})$.

The sub-sampled estimator in (24) is essentially an average of realized variances computed for 5 sparse 5-minute grids shifted by a 1-minute interval with respect to each other. Such sub-sampling of the realized variance was proposed in Zhang et al. (2005) and later considered in Andersen et al. (2011), Liu et al. (2015), among others[22]. The sub-sampling is aimed to improve efficiency of sparsely sampled estimators since more intra-daily information is used in this case.

The estimator in (24) is consistent for the quadratic variation of an asset return for the open-to-close trading period. Note, that despite s_t is linked in (11) with the close-to-close return variance, we do not account explicitly for the "overnight" (and "overweekend") variation while specifying s_t. Firstly, the length of the overnight period for the considered commodities is very short (45 minutes). Secondly, the specification of the measurement equation (11) is able to capture a systematic bias in s_t stemming from the lack of overnight variation embedded in the realized measures.[23]

4.2.2 Realized Measures for the Three-component Specification

When we formulate the realized measure for the one-component specification of Ω_t, we discard much of the available intra-daily data. In particular, we use the high frequency price observations of the contract with the shortest maturity, whereas the intra-daily data from all of the traded contracts is available.

According to the three-component dynamic specification of Ω_t, the time-varying elements are the conditional volatilities of the Nelson-Siegel factors, i.e. the components of h_t of the diagonal matrix H_t from (16). We now suggest a way to retrieve a proxy for h_t from the intra-daily data on multiple traded contracts with different expiration times τ. Namely, we directly assess factor volatilities by evaluating parametrically the dynamic Nelson-Siegel models intra-daily.

Estimation of the multivariate state-space models with high frequency data closely relates to the recent developments of Shephard and Xiu (2012) and Corsi et al. (2015) in the parametric QML estimation of realized covariances in equity prices. In particular, these methods consider an intra-daily multivariate price process as a local level model and provide consistent QML estimates of the covariance matrix with irregular asynchronous price observations under weak assumptions on the underlying price process. In our case we treat an intra-daily multivariate price process $p_t(v)$ as a factor (dynamic Nelson-Siegel) model. As a consequence, our intra-daily state-space model has an additional challenge of latent factor identification. A typical pattern of the observed high frequency futures prices provides a very harsh field for an intra-daily implementation of such state-space methods. Thus, we reorganize slightly the structure of high frequency observations and impose several simplified assumptions on the intra-daily Nelson-Siegel framework.

First of all, similar to case with the one-component measure, we partition a trading day with a 5-minute grid and define the vector of observed prices as $\tilde{p}_t(v_i) = \left(\tilde{p}_t^{(\tau_{t,1})}(v_i), \tilde{p}_t^{(\tau_{t,2})}(v_i), ..., \tilde{p}_t^{(\tau_{t,n_t})}(v_i)\right)'$ for $i = 1, ..., m$ using the

[21] Previous tick interpolation was introduced in Wasserfallen and Zimmermann (1985). See also a related discussion in Hansen and Lunde (2006b).

[22] Note that estimator (24) can also be considered as a case of the realized kernel estimator where Bartlett kernel is used.

[23] Alternatively, it is possible to combine the open-to-close realized variance and the overnight squared return to obtain a consistent measure for a close-to-close quadratic variation, as in Hansen and Lunde (2005). This approach was used in Christoffersen et al. (2014) in a similar context to measure the variation in commodity futures prices.

previous tick interpolation scheme.[24] As a result, we obtain a panel of observed prices $\{\tilde{p}_t(v_i)\}_{i=0}^m$. The fact that we use a sparse 5-minute grid is helpful for mitigating microstructural noise effects and the improved balancedness of an intra-daily observations panel enhances a stability of the model identification and estimation.

We then assume that the intra-daily price dynamics of $\tilde{p}_t(v_i)$ is compatible with the constant volatility dynamic Nelson-Siegel model. Thus, the intra-daily dynamic Nelson-Siegel model for a trading day t is given by

$$\tilde{p}_t(v_i) = \Lambda_t g_t(v_i) + \epsilon_t(v_i) \qquad \epsilon_t(v_i) \sim iidN\big(0, \sigma_\epsilon^2 I_{n_t}\big) \tag{25}$$

$$g_t(v_i) = g_t(v_{i-1}) + \omega_t(v_i) \qquad \omega_t(v_i) \sim iidN\left(0, \frac{1}{m} H_t R H_t\right) \tag{26}$$

where $g_t(v_i)$ is a 3×1 vector of the latent intra-daily Nelson-Siegel factors. For any given trading day t the parameters are fixed and do not depend on the intra-daily index i. Note, that in the intra-daily model specification we do not endow the idiosyncratic term $\epsilon_t(v_i)$ with the autocorrelation effect as we do it in the daily model. The model in (25)-(26) is a discrete time linear Gaussian state-space model that can be estimated by means of the standard Kalman filter. We also note, that the discrete state-space specification (25)-(26) complies with the Brownian semimartingale assumption for the latent efficient futures prices $p_t(v)$ which is a standard assumption in the high frequency econometrics literature.

We facilitate the estimation of (25)-(26) by presetting parameter values for λ, σ_ϵ^2 and R with the corresponding estimates from the constant volatility dynamic Nelson-Siegel model evaluated at a daily frequency. Therefore, only the intra-daily factor covariance matrix $H_t = H(h_t)$ is left for estimation in (25)-(26). We then define the required realized measure as $s_t \equiv \hat{h}_t$. Vector $\hat{h}_t$ here is the corresponding factor variance estimate obtained from the intra-daily dynamic Nelson-Siegel models (25)-(26), which are evaluated using the high frequency price observations $\{\tilde{p}_t(v_i)\}_{i=0}^m$ of contracts traded at day t.

Of course, the realized measure s_t is hardly a consistent measure of the conditional variances h_t of the latent Nelson-Siegel factors from the daily model at least due to the difference in specifications between the daily and the simplified intra-daily models (25)-(26). What is important, we suppose that these measures correlate well with the unobservable factor volatility processes. Thus, they still can be considered as valid and the possible systematic bias can be captured by the parameters of the measurement equations (18) in the main model. Informally, we will refer to these measures of factor variance as realized measures.

4.3 Remarks

Table (3) contains descriptive statistics of the estimated realized measures for the one- and three-component model specifications. Figure (5) provides the corresponding time series plots.

								ACF		
Realized measure	Mean	Std.	Skew.	Kurt.	Min	Med	Max	$\rho(1)$	$\rho(5)$	$\rho(22)$
Shortest contract	0.306	0.163	2.823	16.006	0.048	0.275	2.046	0.857	0.803	0.713
Level factor	0.113	0.078	1.590	6.609	0.007	0.095	0.553	0.556	0.437	0.379
Slope factor	0.184	0.102	1.848	7.858	0.021	0.163	0.784	0.646	0.529	0.444
Curvature factor	0.371	0.272	1.890	8.199	0.031	0.308	2.004	0.631	0.553	0.494

Table 3: Descriptive statistics of the realized measures of volatility (annualized) for the sample period which starts in January, 2005 and lasts until December, 2014.

The realized measures for the one-component specification based on the non-parametric realized variance estimator demonstrate relatively smooth and highly persistent dynamics. On contrary, parametric volatility measures for

[24] In addition, if some contract τ that is traded at t was traded at $t-1$ with time to maturity $\tau+1$, we use the close price of the contract $\tau+1$ at $t-1$ in the interpolation scheme when we obtain $\tilde{p}_t^{(\tau)}(v_i)$ for day t.

the three-component specification reveal noisy and weakly persistent dynamics, which suggests that these measures are likely less accurate.

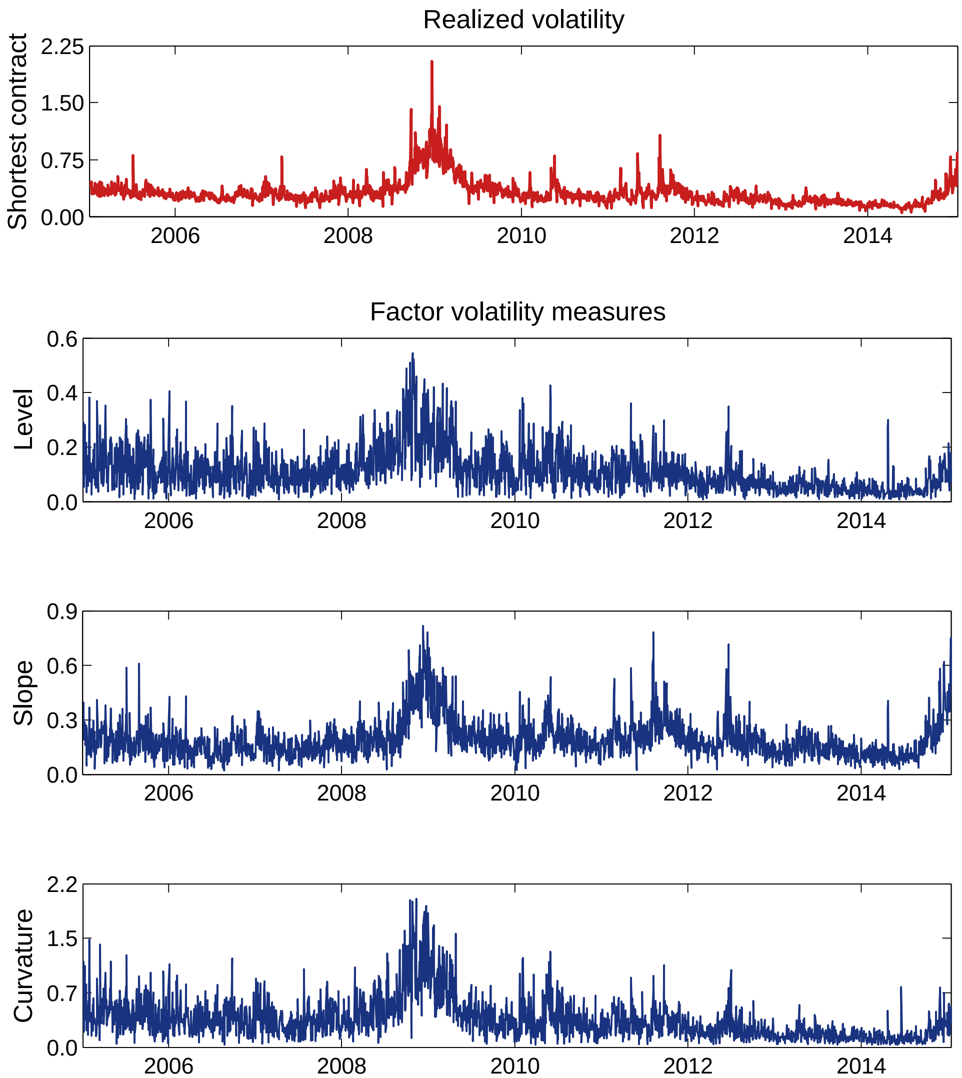


Figure 5: In the top panel: annualized subsampled 5-min realized volatility for the front futures contracts. In the 3 bottom panels: annualized volatility measures of the Nelson-Siegel factors estimated parametrically using intra-daily state-space models for each trading day of the sample. The sample period starts in January, 2005 and lasts until December, 2014.

There are several potential reasons of why the three-component measures are relatively noisy. First of all, since high frequency price observations are unbalanced and we usually do not observe a sufficient number of transactions on long-term contracts, the factor identification in the intra-daily Nelson-Siegel models is weak and unstable.[25] We also have to recognize that the data interpolation may lead to the distortions in the intra-daily estimation of $\hat{h}_t$. We argue, nonetheless, that an adverse impact from these causes is expected to be limited since we preset factor

[25]In practical implementation we use the following method to improve the factor identification. When estimating an intra-daily Nelson-Siegel model for a trading day t we supply it with the interpolated 5-minute prices from 2 consecutive trading days $t-1$ and t, $\{\tilde{p}_{t-1}(v_i); \tilde{p}(v_i)\}_{i=0}^{m}$. Although only the data related to day t is used for likelihood maximization, additional data from the previous trading day significantly improves the ability of the model to identify the latent Nelson-Siegel factors. This happens due to the critical deficit of price observations for the contracts with long maturities if the data from just one trading day is used. In case we obtain a "suspicious" outlying volatility estimate, we use the data from the two consecutive trading days for likelihood maximization. It helps to smooth out very noisy estimates coming from the days when the intra-daily observation panel is especially scarse.

correlation matrix R and only estimate volatilities h_t of the latent factors.

This issue points out an interesting trade-off between the use of the one- and three-component dynamic specifications of the time-varying covariance matrix Ω_t. On one hand, the three-component dynamics provides more flexibility to the term structure volatility, on the other hand the one-component specification suggests a better precision of the realized measures. Consequently, this trade-off makes it ex-ante unclear which type of approach is able to better model and predict the density of the futures prices and we relegate this question to our empirical evaluation.

5 Evaluation Strategy

In this section, we outline the out-of-sample evaluation of our modeling framework. At first, we briefly describe the set of alternative (benchmark) model specifications. The corresponding state-space model formulations are given in Appendix B. Then we explain our strategy for the out-of-sample density forecast evaluation.

5.1 Pool of Considered Models

In our empirical evaluation we consider several alternative models along with our baseline framework. The goal is to investigate whether time-varying volatility and realized measures suggested in our main framework provide improvements in the out-of-sample performance compared to the alternative models with no such features. For better comparison, all the alternative models have the DNS factor dynamics in the conditional mean, have autocorrelated idiosyncratic components, and the random walk assumption is maintained for the Nelson-Siegel factor dynamics. In other words, the part described by equations (2)-(6) is common for all considered models and the differences come in the specification of the factor covariance matrix Ω_t. Particularly, we examine alternative specifications *a)* with constant volatility and *b)* with time-varying volatility, but with no use of realized measures.

DNS with constant variance (DNS-CV). This model is the standard state-space DNS framework from Diebold et al. (2006) where all variance parameters are assumed to be constant. The model is characterized by (2)-(6) with the constant covariance matrix Ω in (6). The DNS-CV represents a natural constant volatility benchmark in our analysis.

DNS with GARCH effects (DNS-GARCH). In this class of models we add time-varying volatility components to the factor covariance matrix Ω_t. We suggest two specifications with one and three dynamic volatility components, so they possess the same parametrization of Ω_t as proposed in our baseline framework. Similarly, these time-varying components are supposed to have GARCH-type dynamics, but they are not related to any kind of realized volatility measures. We follow the methodology from Harvey et al. (1992) by incorporating GARCH effects into a linear Gaussian state-space model. Therefore, instead of realized measures we use the conditional expectations of the factor innovation shocks in order to specify the updating terms in the GARCH equations for dynamic volatilities.[26] Since these conditional expectations appear explicitly in the forward Kalman recursions, the models can be estimated directly by means of the Kalman filter. We consider them in order to assess the gains from introducing time-varying volatility in the Nelson-Siegel factor dynamics. At the same time, we rule out the effect from the realized measures which is presented in our baseline models.

DNS with GARCH effects and realized measures (DNS-RGARCH). This class of models is our baseline framework and is described in details in Section 2. We consider specifications of Ω_t with the one and three dynamic volatility components exploiting the information from realized measures of volatility. Therefore, in addition to the time-varying volatility aspect, these models also incorporate the effect from the use of volatility measures based on high frequency data. We examine the baseline models with and without leverage effects specified in (12) and (19).

[26] The use of conditional expectations of factor volatility shocks implies that the model should be seen as an approximation to a standard GARCH model.

Model	Dynamic covariance components	Realized measures	Leverage effect
CV	0	No	No
G-1	1	No	No
G-3	3	No	No
RG-1	1	Yes	No
RG-1-lev	1	Yes	Yes
RG-3	3	Yes	No
RG-3-lev	3	Yes	Yes

Table 4: Set of competing models used in the paper. All models are the DNS factor models with the random walk dynamics of the Nelson-Siegel factors and the autocorrelated idiosyncratic components. In all models correlations in the factor dynamics are constant.

Table 4 provides the complete list of the models used in our evaluation study.

5.2 Out-of-sample Evaluation

Distinct specifications of the conditional second moment in the models naturally lead to differences in the model-implied volatility dynamics and, as a consequence, in the model-implied degree of uncertainty around the expected futures prices. In our evaluation study we aim to investigate the out-of-sample performance of these volatility specifications. At first, we simply compare variance forecasts of futures prices produced by our models using a realized variance benchmark. Then, we focus on forecasting uncertainty of futures prices. For this, we employ a standard density forecasting approach. One key difference between these two schemes is that in case of variance prediction we use a noisy proxy of an unobservable volatility, whereas the evaluation of density forecasting is based on perfectly observable futures prices.

Given the fact that we estimate our models with futures prices at a daily frequency, we focus on short-term forecasting. We consider 8 forecast horizons: 1, 2, 3, 4, 5, 10, 15, and 22 business days ahead. Note that for commodity futures the maturities of traded contracts τ change from day to day, so we can not observe the same set of contracts in the course of the forecasting period. Thus, we partition the traded contracts into several maturity ranges (e.g. 1-3 months, 1-2 years, etc.) and consider an averaged performance across the traded contracts from a given maturity range.

The forecasting period lasts from January 2, 2014 until January 16, 2015 and contains 255 trading days in total.[27] The oil prices demonstrate very diverse dynamic regimes throughout the forecasting period (see Figure 6). In particular, it spans both the relatively calm period in the first half of 2014 and the sharp downfall of oil prices that occurred during the second half of 2014.

As a variance forecasting criterion we use R^2 from the Mincer-Zarnowitz regressions. Namely, for each model $m \in \mathcal{M}$ and for each maturity range $\mathcal{T}_\tau \in \mathcal{T}$ we run the following pooling regression out-of-sample for a given forecast horizon h

$$RV^{ss}_{t+h}(\tau_{t+k,i}) = \beta_0 + \beta_1 \hat{V}_{t+k}(\tau_{t+k,i}; m) + \varepsilon_{t,i}, \qquad \text{for } i:\ \tau_{t+k,i} \in \mathcal{T}_\tau \tag{27}$$

where $\hat{V}_{t+k}(\tau; m)$ stands for the k-step ahead daily log-return variance forecast generated by model m conditional on the information available up to day t. Variable $RV^{ss}_{t+k}(\tau_{t+k,i})$ is a sub-sampled realized variance estimate which is defined in (24) and serves as a proxy for the latent variance.

This approach is especially suitable for our analysis for several reasons. Firstly, R^2 computed from the Mincer-Zarnowitz regressions with non-transformed variables provide a robust ranking of models in case a noisy variance

[27] In our evaluation analysis we delete the days from the out-of-sample period where log predictive likelihoods get extremely large negative values for all of considered models (14 trading days in total).

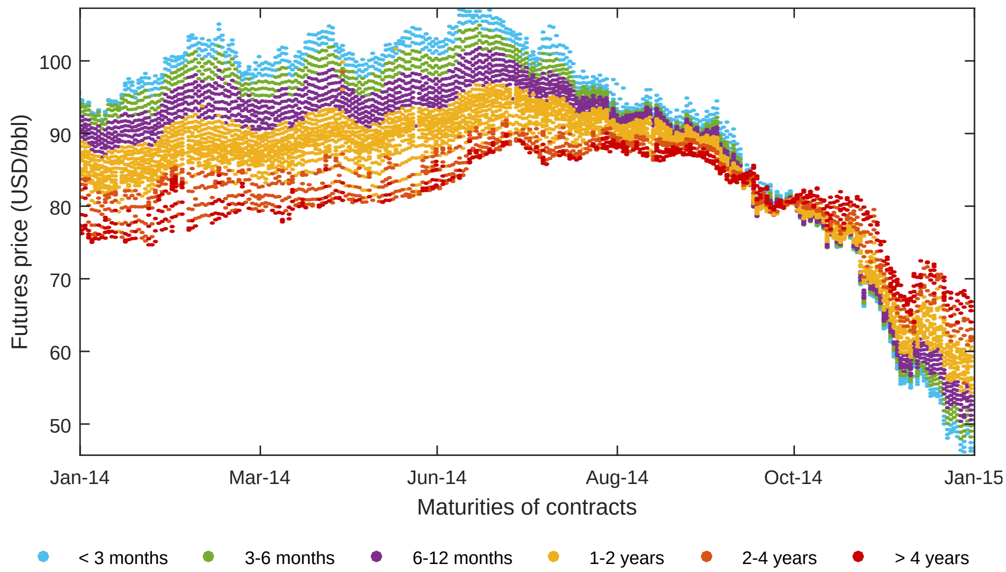


Figure 6: Futures prices during the forecasting period (January 2, 2014 - January 16, 2015)

proxy is used (see Hansen and Lunde (2006a)). Furthermore, regression in (27) accounts for the possible systematic bias between the levels of realized and model-implied variances (e.g., due to overnight variation that does not contribute to realized measures).

We use the log predictive likelihood as a criterion for the density forecast evaluation. The comparison of scoring rules across the pool of competing models to assess their relative ability of producing density forecasts has a long tradition in empirical studies (see e.g. Diebold and Lopez (1996), Lopez (2001), Geweke and Amisano (2010)). The log predictive likelihood of a log futures price $y_{t+k}(\tau)$ reads

$$L^m_{t+k,\tau} = \log \hat{p}_{t+k,\tau}\big(y_{t+k}(\tau)|\mathcal{F}_t, m\big) \tag{28}$$

where h denotes a forecast horizon and $\hat{p}_{t+k,\tau}\big(\cdot|\mathcal{F}_t, m\big)$ is a forecast of the probability density function of $y_{t+k}(\tau)$ that is generated by model m conditional on the information $\mathcal{F}_t$ available up to day t. As it is evident from (28), the log predictive likelihood gets a higher value if the corresponding forecasted density assigns a higher probability to the actual future realization of the predicted variable ($y_{t+k}(\tau)$, in our case). Intuitively, the model that systematically exhibits higher predictive likelihoods has a better performance.

Definitely, a Gaussian distribution, that is used in our model, may inadequately describe certain aspects of daily futures returns. Nontheless, we treat the density forecasting approach as a valid tool for evaluation of the predicted uncertainty around point forecasts. Thus, the Normal density in our analisys serves as a weight function that turns the predicted variance into a loss value conditional on an inaccuracy of a point forecast. Thus, if a price misprediction is small enough, a model that has predicted a lower uncertainty is credited better. Conversely, if a price misprediction is large enough, such a model is penalized more heavily.

In order to forecast the variance and density of futures prices from a given model we refer to simulation techniques (apart from the case of the constant volatility DNS model where the predictions can be derived analytically). More precisely, to produce k-step ahead variance and density forecasts we simulate S sequences of the variance components h_t and the factors f_t of length k, which are initialized at day t given information $\mathcal{F}_t$. While simulating, we use a stochastic law of motion defined by model m with the pre-estimated parameters. To obtain the variance forecast, $\hat{V}_{t+k}(\tau; m)$, we take the median terminal value of S simulated variance components h_t and use (7) and (8) in order to convert it into a variance prediction. Using the terminal values of S simulated factor sequences, we obtain S

simulated prices $\{\hat{y}^s_{t+k}(\tau)\}_{s=1}^S$ using (2). Then, we apply a kernel density estimator to get $\hat{p}_{t+k,\tau}(\cdot|\mathcal{F}_t, m)$ and evaluate the log predictive likelihood from (28) at the observable futures price realization $y_{t+k}(\tau)$. In our analysis, we take $S = 10,000$ simulations and use the MATLAB built-in function *ksdensity* with Normal kernel smoother and the corresponding optimal bandwidth.

In order to statistically assess the differences in density forecasting performance between the competing models, we construct the Model Confidence Set (MCS) of Hansen et al. (2011) that provides a framework for *multiple* comparison of the competing models. MCS consists of the models that exhibit significantly superior performance with respect to a given loss function and is formulated as follows

$$MCS_{k,\tau} = \{m \in \mathcal{M} : \ E(L^{m^*}_{t+k,\tau}) \geq E(L^m_{t+k,\tau}), \ \forall m \in \mathcal{M}\} \tag{29}$$

We use Kevin Sheppard's MFE Toolbox[28] for MATLAB in order to obtain the corresponding p-values for MCS of 99, 95, and 90% coverage, while doing so we set the number of stationary bootstrap replications at 10,000 and the average window length at 12.

6 Empirical Results

6.1 Parameter estimates and factors

Table 5 presents the parameter estimates for the seven different models. Significance at a 5% level is indicated in bold face.

The parameter estimates are obtained by numerically maximizing the log-likelihood (23) using data from the estimation period (January 3, 2005 to December 31, 2013). To start the estimation with appropriate values of the three Nelson-Siegel factors and the volatility components we make use of the 100 days initialization period (August 9, 2004 to December 30, 2004).[29]

The parameters $\lambda, \sigma^2_w, \beta, \rho_{ls}, \rho_{lc}, \rho_{sc}$ are common for all models and their estimates tend to be relatively close across the models. Hansen and Lunde (2013) fix $\lambda = 0.005$, which is slightly smaller than the estimates found in Table 5. The parameter β is significant and relatively high for all models, which reflects the importance of incorporating autocorrelation in the idiosyncratic component especially when applying the model to daily data. The most pronounced correlation parameter is the large negative correlation between the level and the curvature. Hence, an increase in the overall level of futures prices tends to be offset by a decrease in the prices of the medium maturity contracts.

The parameter estimates controlling the volatility dynamics in the standard GARCH models are not directly comparable to their realized counterparts. The volatility dynamics in the GARCH-1 and GARCH-3 models are linear while the models based on realized measures have a log-linear volatility specification.

If we consider the realized models, we immediately see that the realized measure of the RGARCH-1(-lev) model is less noisy than the realized measures of the RGARCH-3(-lev) model by comparing the σ^2_u parameter estimates. The parameter γ_{1l} governing the impact on the volatility from the realized measure is also substantially larger in the RGARCH-1(-lev) model compared to the corresponding parameters in the RGARCH-3(-lev) model.

In the RGARCH-1-lev model we find a negative δ_1, which is the common finding in finance literature. Here, a negative return in the shortest futures often leads to an increase in the volatility of all three factors. In the RGARCH-3-lev model the sign of the δ_1 parameters is not unambiguous and for the level factor we observe an inverse leverage effect $\delta_{1l} > 0$. All δ_1 parameter estimates are, however, insignificant in this model. The δ_2

[28] https://www.kevinsheppard.com/MFE_Toolbox

[29] In the initialization period the filter is started by assuming the factors are multivariate normal with a mean equal to their empirical counterparts (see Diebold and Li (2006)) and a high variance equal to the identity matrix. The volatility components are initialized at their unconditional means.

parameter is positive in the RGARCH-1-lev, which indicates that large standardized returns of either sign in the shortest futures often lead to an increase in the general volatility level. For RGARCH-3-lev we find an even more pronounced effect for all factors.

Figure 7 shows the filtered factors and the annualized volatility series of the different models. Here we only present the filtered factors for the RGARCH-3 model including its 99% confidence interval. The filtered factors from the other models are visually identical. In the volatility plots we leave out the volatility series of two models including leverage since they are visually identical to their non-leverage counterparts.

By comparing the level factor with the data (Figure 1) we may see that the factor tracks the overall tendency of the futures prices. The slope factor is slightly positive (corresponding to a decreasing futures curve) in the beginning of the estimation period and rapidly turns negative to reach its lowest values during the financial crisis. After the financial crisis the slope factor gradually returns to positive values. The curvature is positive in nearly 75% of the days in the estimation period. A hump-shaped or inverted hump-shaped futures curve may appear if the curvature is relatively large compared to the slope. Most of the hump-shaped futures curves are observed in late 2005 to 2006 and again in 2012. A few inverted hump-shaped futures curves, with a negative curvature and a relatively small slope, are observed in late 2007.

The time-varying volatility series of the three factors all share the same pattern. All factors are most volatile during the financial crisis. The curvature is by far the most volatile factor followed by the slope. The level is generally the factor with the smallest volatility. It is evident that the volatility series of the RGARCH-3 models are less "rough" than the series of the competing models. This is consistent with the relatively low estimates of the γ_1 parameters.

6.2 Partial Likelihood Comparison

To evaluate the relative fit of the estimated models we compare the log-likelihood values both in- and out-of-sample. Since the realized models (RG1, RG1-lev, RG3, RG3-lev) contain an additional contribution to the likelihood function that stems from the fit of realized measures within the measurement equations, we compare only those contributions which all of the models share. Namely, our partial likelihoods contain only the contribution from the fit of daily futures prices that is given by the first term in (23) and in (B.1).

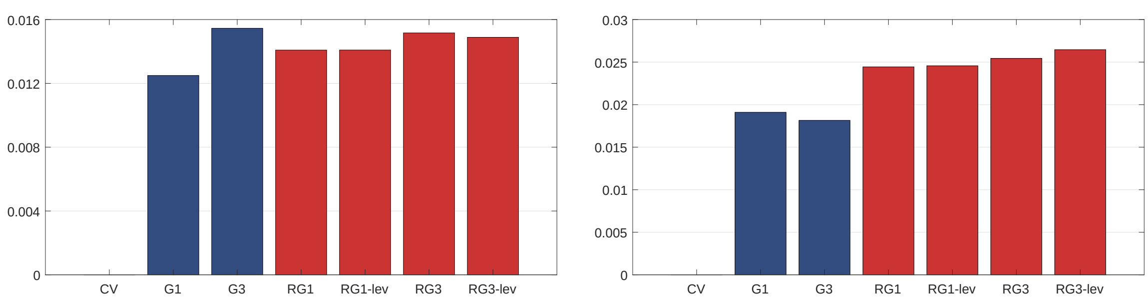
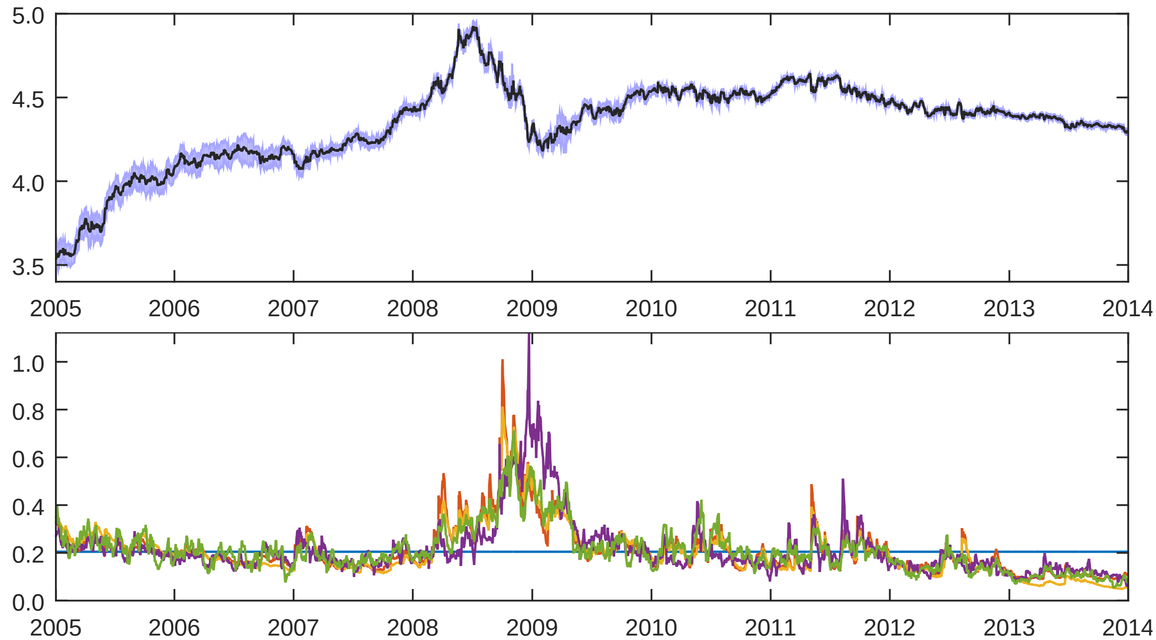


Figure 8: In-sample (left) and out-of-sample (right) partial likelihood comparison

Figure 8 provides a graphical illustration of the in-sample and out-of-sample log-likelihood comparisons. The average normalized daily log-likelihood value of the estimated model with constant volatility (CV) is taken as a benchmark.[30] The performance of other models is provided in terms of log-likelihood increments relative to the benchmark. As we can see from both plots, models with time-varying volatility strongly outperform the constant volatility benchmark both in- and out-of-sample. Whereas both realized and non-realized models perform similarly in-sample, an advantage of realized models out-of-sample is clearly pronounced.

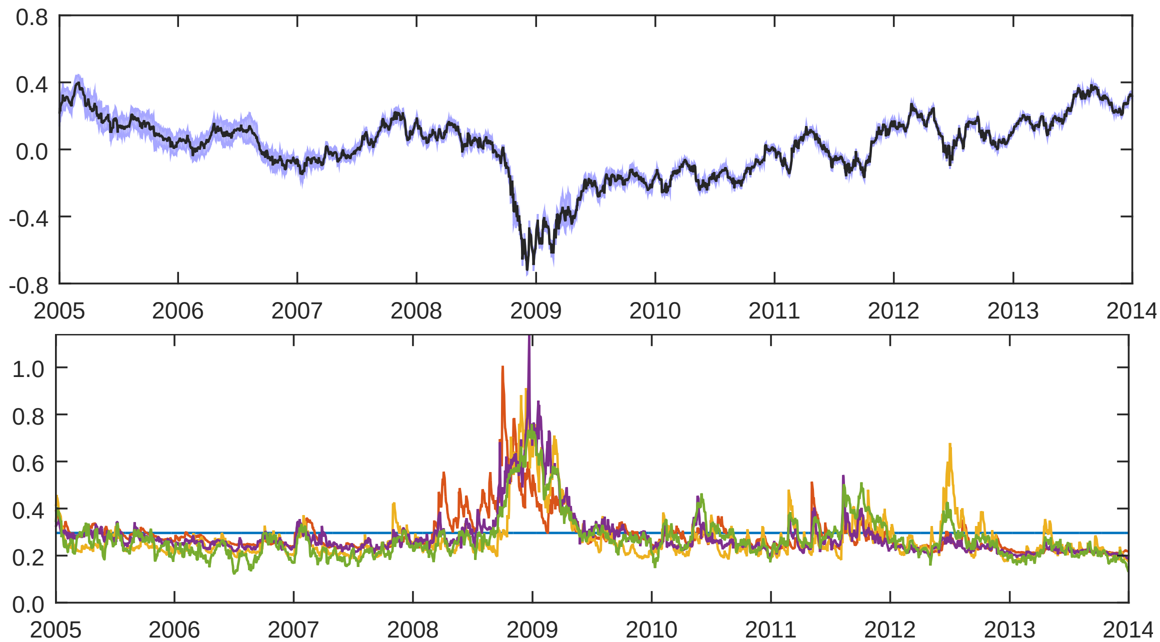
[30] More precisely, we first compute the corresponding log-likelihood values for all the trading days in a given period, then normalize them by the number of contracts traded at these day and then compute an average normalized daily log-likelihood.

Level



5.0
4.5
4.0
3.5
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
1.0
0.8
0.6
0.4
0.2
0.0
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014

Slope



0.8
0.4
0.0
-0.4
-0.8
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
1.0
0.8
0.6
0.4
0.2
0.0
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014

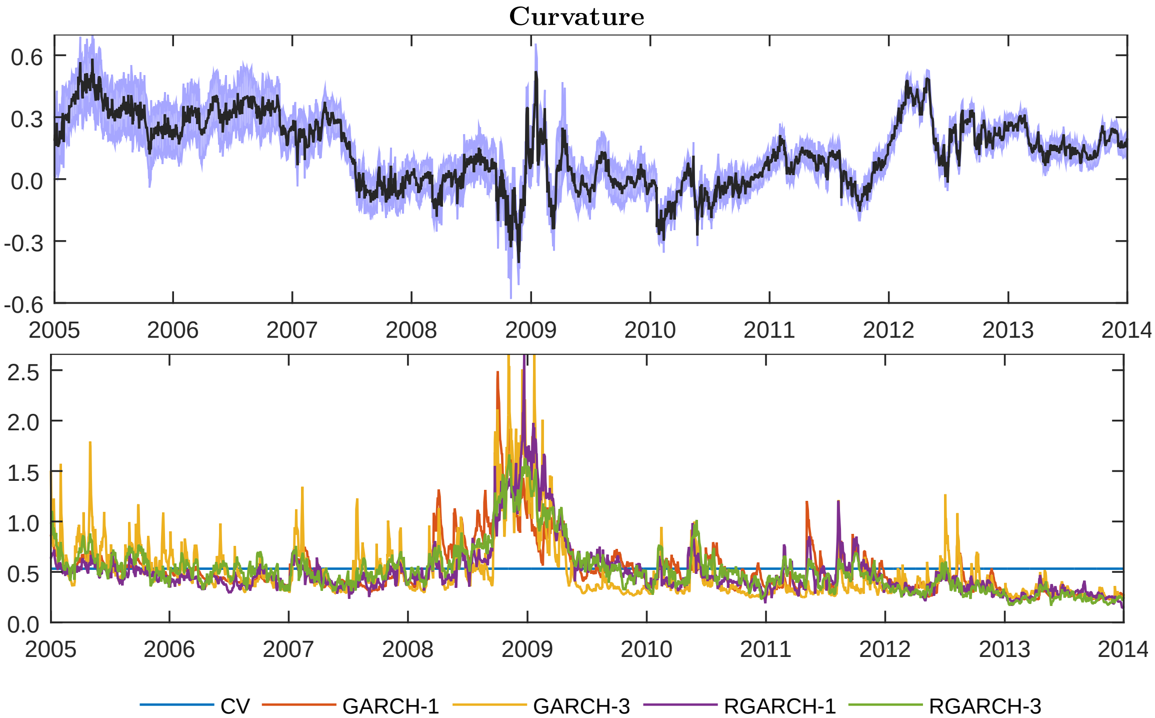


Figure 7: Filtered factors with 99% confidence intervals and the annualized filtered conditional factor volatilities.

6.3 Forecasting Results

We provide R^2 from the Mincer-Zarnowitz regressions in Tables 6-7. The results suggest that both in case of the one- and three-component modifications, GARCH models augmented with the realized information generally demonstrate higher R^2 than its non-realized counterparts. This is especially pronounced for the models with the single dynamic variance component. The forecast comparison of one- and three-component specifications reveals a systematic advantage of the latter. Such an advantage is especially evident for non-realized models. In case of realized models the situation is a bit more intricate. Namely, the benefits of the three-component specification appear to be significant only for longer than 3 days ahead forecast horizons. In particular, this may be a consequence of the relatively more noisy realized measures used in the three-component covariance specification. Thus, an advantage of the richer three-component volatility structure is deteriorated for 1-2 days ahead forecasts, where the role of the precision of realized measures is more salient.

The Mincer-Zarnowitz R^2 become noticeably smaller with time to maturity. This observation can be explained not only by the unability of the Nelson-Siegel models to properly capture the variance of the long-term futures prices. Another reason is the presence of noise in RV estimates which is likely to be higher for the futures with longer tenors. This fact is a direct consequence of the evidence described in Section 4, that the average number of intra-daily transactions per trading day for long-term contracts is low. It leads to a higher dispersion of the RV estimator which makes the variance proxy more noisy and, as a consequence, negatively affects the fit of the Mincer-Zarnowitz regressions.

Tables 8-9 contain the results of our density forecasting analysis. In addition, we provide density forecasting results based on the 6 month sub-period (March, 2014 - August, 2014) where the oil prices exhibit relatively stable dynamics with no long negative or positive trends (see Figure 6). The results are presented in Tables C.3-C.4 in Appendix C.

The uncertainty forecasting reveals several important implications. First of all, it is evident that models with time-varying volatility strongly outperform the constant volatility benchmark.[31] This finding is quite natural in the context of variance prediction. Furthermore, it agrees with the corresponding results on the term structure density forecasting of interest rates (see Carriero et al. (2014), Shin and Zhong (2015)). As far as the forecasting horizon increases, the difference between constant and time-varying volatility models reduces. This fact is also not surprising since in the absence of long-memory effects the GARCH volatility components converge to their unconditional means relatively fast. We also note that such difference is more pronounced for the futures with short and medium time to maturity, whereas for the long contracts (more than 4 years until expiration) an advantage of time-varying volatility is significant only for the very short-term forecasting horizons.

Secondly, we find that the use of realized volatility measures provides not large, but systematic gains. Since the volatility is persistent, realized measures make an appreciable difference only when the latent volatility state is rapidly changing. Whereas the standard models can take a while to adjust the volatility process, the use of realized measures can speed the adjustment up. By comparing DNS-GARCH-1 and DNS-GARCH-3 models to DNS-RGARCH-1(-lev) and DNS-RGARCH-3(-lev) respectively, we may see that the realized models demonstrate higher R^2 and predictive likelihoods at almost all the considered horizons for the short- and mid-term contracts. With respect to density forecasting, this finding is qualitatively similar to the results of Shin and Zhong (2015) who documented significant gains from the use of realized measures in the bond yield density forecasting within a DNS framework. However, in our case these gains are quantitatively much smaller.[32] The "realized" models always appear in the MCS (for both full and calm out-of-sample periods), which suggest that the gains from realized measures are systematically significant, though they are not so large. The relative performance of the models with and without realized measures is getting less discernible as the forecasting horizon increases. This is natural since the realized information about the current volatility state is more relevant for the short-term volatility (and, hence, density) predictions rather than for the far ahead forecasts.

Third, the comparison of models with the one and three dynamic volatility components (DNS-(R)GARCH-1(-lev) to DNS-(R)GARCH-3(-lev)) reveals a significant advantage of the three component specifications in the ability to produce density forecasts for the short and mid term contracts. This is especially evident from the results based on the calm forecasting sub-period where for the contracts below 2 years until maturity only 3-component models appear in MCS at all considered horizons. It may imply that the use of three dynamic volatility components allows to better control the distribution of the futures term structure than the use of just one dynamic volatility component. This result conforms with the study of Creal and Wu (2014) who found that one volatility component is not sufficient to properly capture the distribution dynamics of the interest rate term structure. At the same time, we may see that for the long-term contracts ($>$ 4 years until maturity) the one component models provide superior predictive likelihoods at the horizons above 1 week ahead. The possible explanation of this effect is the following. The volatility of the level factor implied by the one-component models is systematically more conservative (higher) than the level volatility in the three factor models. Since the conditional variance of long term futures is mostly explained by the volatility of the level factor, one-component models provide more conservative density forecasts for the long-end contracts. Thus, the large mispredictions of the long futures prices at turbulent periods lead to less severe losses in the corresponding predictive likelihoods for the models with the one volatility component.[33]

[31] We note that the constant volatility DNS model produces pretty conservative volatility and density forecasts. This is because the estimation sample includes the exceedingly volatile Great Recession episode which raises the estimated constant variance parameters.

[32] For instance, realized models with the one dynamic volatility component, DNS-RGARCH-1(lev), provide about 8% gain in predictive likelihood compared to its non-realized counterpart DNS-GARCH-1 for the 1-day ahead horizon for the short and mid term futures. The corresponding gain for the 3-component DNS-RGARCH-3(lev) is much smaller and is about 2.5%. We note, however, that the latter result can be explained by relatively noisy realized measures used in the 3-component model specifications. The relatively small benefits from using realized measures in forecasting the oil futures densities resemble the findings of Lunde and Olesen (2014) from the electricity markets. In particular, they demonstrated that the use of realized volatility measures does not give any significant improvement in the density predictions of the nearby NOMXC forward returns in the context of the univariate Realized GARCH framework.

[33] As we can see from the results, for the long-term contracts the gains from the time-varying volatility (and the realized measures) are small and insignificant for the long enough forecasting horizons (which is especially evident from the results based on the full out-of-sample period). One possible reason is an inability of the DNS models to appropriately capture the variance dynamics of the long-term

Finally, we note that the leverage effect does not seem to significantly improve the forecasting performance. We observe only slight gains from the introduction of the leverage effect presumably for the short-term contracts and short forecasting horizons.

7 Conclusion

In this paper, we analyze the role of time-varying volatility in modeling commodity futures prices. Our approach is based on a dynamic version of the classical Nelson-Siegel model. The model effectively reduces the dimensionality by modeling the whole term structure as a function of three factors: level, slope, and curvature. The dynamic factor structure allows for direct forecasting of the whole term structure of futures prices. We improve the forecasting properties of the model by introducing time-varying volatility directly into the conditional variance of the three unobserved factors. When modeling the volatility we follow the Realized GARCH framework by utilizing precise volatility information extracted from high-frequency data. The realized measures provide a better signal about the latent volatility than the daily returns used in conventional GARCH models. The volatility specification is therefore better suited to periods where the volatility changes rapidly. This makes the model very attractive for forecasting future volatility and in turn providing superior density forecasts.

Empirically we apply the model to data on light crude oil futures. By an extensive out-of-sample forecasting exercise we examine to what extent time-varying volatility improves the variance and density predictability and whether we gain from the use of realized measures. We examine the forecasting performance across different forecast horizons and different parts of the term structure futures curve. We compare the performance of variance forecasts using R^2 from the Mincer-Zarnowitz regressions and use the log predictive likelihood as a criterion for the density forecast evaluation.

We find that the models with time-varying volatility strongly outperform the constant volatility benchmark. The benefit from time-varying volatility is generally larger for short horizon forecasts and for futures contracts with short/medium time to maturity. The use of realized measures provides small but systematic gains relative to standard GARCH volatility. When comparing the one- and three factor specifications of the volatility we find that the three component specification generally produces better density forecasts except for long maturity contracts at long forecast horizons. Finally, we found that leverage effects do not seem to improve forecasting significantly.

futures prices. When the time-varying volatility is introduced into the three Nelson-Siegel factors, the variance of the term structure long-end is regulated by the variance of the level factor solely, whereas the volatilities of the short and mid futures are additionally explained by the variances of the slope and curvature factors respectively. Since the fraction of the futures with long maturities is relatively small in our panel, the volatility of the level factor may be not very well adjusted to the volatility dynamics of the long term futures prices. Therefore, an introduction of an additional volatility factor associated with the long-term futures prices could lead to a potential improvement in the term structure modeling.

Table 5: Maximum likelihood parameter estimates

		CV	GARCH-1	GARCH-3	RGARCH-1	RGARCH-1-lev	RGARCH-3	RGARCH-3-lev
λ	10^3	**5.800**	**5.734**	**5.829**	**5.774**	**5.772**	**5.770**	**5.803**
σ_w^2	10^5	**4.493**	**4.466**	**4.471**	**4.472**	**4.472**	**4.467**	**4.455**
β		**0.633**	**0.634**	**0.632**	**0.632**	**0.633**	**0.633**	**0.631**
ρ_{ls}		**0.030**	**−0.092**	−0.047	**−0.113**	**−0.111**	**−0.081**	**−0.097**
ρ_{lc}		**−0.504**	**−0.535**	**−0.505**	**−0.531**	**−0.533**	**−0.563**	**−0.537**
ρ_{sc}		**−0.139**	−0.030	−0.047	−0.005	−0.005	−0.011	−0.003
σ_l^2	10^3	**0.168**						
σ_s^2	10^3	**0.351**						
σ_c^2	10^3	1.132						
a_s	10^4		**1.458**		1.274	1.140		
b_s			**0.957**		**1.006**	**1.094**		
a_c	10^7		88.809		0.044	0.000		
b_c			**6.078**		**5.578**	**5.571**		
γ_{0l}			**0.000**	**0.000**	−0.023	−0.062	**0.147**	0.127
γ_{0s}				**0.000**			−0.068	0.009
γ_{0c}				**0.000**			−0.033	−0.042
γ_{1l}			**0.146**	**0.093**	**0.431**	**0.436**	**0.135**	**0.143**
γ_{1s}				**0.119**			**0.138**	**0.156**
γ_{1c}				**0.347**			**0.155**	**0.120**
γ_{2l}			**0.845**	**0.907**	**0.607**	**0.597**	**0.860**	**0.849**
γ_{2s}				**0.850**			**0.838**	**0.828**
γ_{2c}				**0.629**			**0.822**	**0.859**
ξ_l					−0.382	−0.414	**−2.166**	**−2.117**
ξ_s							−0.537	**−1.107**
ξ_c							−0.631	−0.478
ϕ_l					**0.864**	**0.861**	**0.913**	**0.916**
ϕ_s							**1.047**	**0.981**
ϕ_c							**1.026**	**1.055**
$\sigma_{u,l}^2$					**0.154**	**0.138**	**1.278**	**1.085**
$\sigma_{u,s}^2$							**0.632**	**0.538**
$\sigma_{u,c}^2$							**1.082**	**0.921**
δ_{1l}						**−0.049**		0.040
δ_{1s}								−0.041
δ_{1c}								−0.030
δ_{2l}						**0.079**		**0.232**
δ_{2s}								**0.190**
δ_{2c}								**0.191**
$\log \mathcal{L}$		160,551.5	161,130.2	161,255.1	160,101.6	160,230.1	151,578.7	152,128.9
$\sqrt{250E[V_l]}$		0.205	0.235	0.195	0.223	0.221	0.224	0.220
$\sqrt{250E[V_s]}$		0.296	0.299	0.273	0.286	0.286	0.279	0.286
$\sqrt{250E[V_c]}$		0.532	0.582	0.559	0.526	0.522	0.549	0.487

Maximum likelihood parameter estimates for the seven different models. Significance at a 5% level is indicated in bold face. For visualization some rows have been scaled by the number given in the second column.

	Specification of factor variance dynamics within DNS model						
Maturity Category	CV	G-1	G-3	RG-1	RG-1-lev	RG-3	RG-3-lev
			1 day-ahead				
< 3 months	0.001	0.490	0.679	0.736	**0.738**	0.701	0.697
3-6 months	0.000	0.450	0.661	0.699	**0.702**	0.686	0.681
6-12 months	0.002	0.248	0.360	0.401	**0.402**	0.398	0.394
1-2 years	0.004	0.076	0.094	0.102	**0.102**	0.088	0.086
2-4 years	0.000	0.051	**0.082**	0.076	0.076	0.071	0.072
> 4 years	0.031	0.094	0.106	0.154	0.155	**0.169**	0.165
			2 days-ahead				
< 3 months	0.001	0.458	0.600	0.649	0.650	0.660	**0.660**
3-6 months	0.000	0.426	0.586	0.610	0.613	**0.641**	0.640
6-12 months	0.002	0.235	0.317	0.365	0.367	**0.368**	0.365
1-2 years	0.004	0.077	0.083	0.100	**0.101**	0.087	0.085
2-4 years	0.000	0.048	0.064	0.076	**0.076**	0.071	0.072
> 4 years	0.031	0.094	0.106	0.150	0.152	**0.155**	0.153
			3 days-ahead				
< 3 months	0.001	0.459	0.557	0.584	0.586	0.618	**0.628**
3-6 months	0.000	0.428	0.548	0.542	0.544	0.608	**0.613**
6-12 months	0.002	0.235	0.296	0.315	0.317	0.343	**0.346**
1-2 years	0.004	0.076	0.084	0.087	**0.087**	0.081	0.081
2-4 years	0.000	0.053	0.068	0.069	0.070	0.071	**0.073**
> 4 years	0.031	0.098	0.110	0.121	0.122	0.147	**0.147**
			4 days-ahead				
< 3 months	0.001	0.469	0.538	0.570	0.571	0.644	**0.652**
3-6 months	0.000	0.439	0.538	0.530	0.533	0.627	**0.632**
6-12 months	0.002	0.258	0.298	0.295	0.297	0.348	**0.352**
1-2 years	0.004	0.079	0.079	0.086	**0.086**	0.082	0.082
2-4 years	0.000	0.065	0.076	0.068	0.069	0.083	**0.084**
> 4 years	0.031	0.085	0.102	0.107	0.108	**0.142**	0.142

Table 6: R^2 from Micer-Zarnowitz regressions. Bold values indicate the highest R^2 across competing models for a given maturity range and a forecast horizon.

Maturity Category	Specification of factor variance dynamics within DNS model						
	CV	G-1	G-3	RG-1	RG-1-lev	RG-3	RG-3-lev
			1 week-ahead				
< 3 months	0.001	0.516	0.576	0.557	0.561	0.658	**0.667**
3-6 months	0.000	0.485	0.577	0.532	0.537	0.643	**0.648**
6-12 months	0.002	0.286	0.340	0.294	0.297	0.365	**0.368**
1-2 years	0.004	0.083	**0.088**	0.083	0.083	0.083	0.084
2-4 years	0.000	0.076	**0.084**	0.058	0.059	0.077	0.078
> 4 years	0.031	0.088	0.102	0.109	0.111	0.147	**0.147**
			2 weeks-ahead				
< 3 months	0.001	0.469	0.607	0.544	0.549	0.661	**0.664**
3-6 months	0.000	0.456	0.575	0.547	0.554	0.635	**0.635**
6-12 months	0.002	0.249	0.309	0.313	0.317	0.354	**0.355**
1-2 years	0.004	0.067	0.077	0.074	0.074	0.081	**0.081**
2-4 years	0.000	0.063	0.075	0.051	0.052	**0.078**	0.078
> 4 years	0.031	0.095	0.105	0.129	0.131	**0.146**	0.143
			3 weeks-ahead				
< 3 months	0.001	0.470	**0.680**	0.636	0.637	0.671	0.672
3-6 months	0.000	0.466	**0.641**	0.603	0.608	0.633	0.631
6-12 months	0.002	0.219	0.365	0.339	0.342	**0.367**	0.366
1-2 years	0.004	0.052	**0.102**	0.075	0.076	0.082	0.082
2-4 years	0.000	0.080	0.085	**0.097**	0.095	0.086	0.086
> 4 years	0.031	0.109	0.114	0.124	0.127	**0.131**	0.129
			1 month-ahead				
< 3 months	0.001	0.257	0.523	0.514	0.506	0.567	**0.578**
3-6 months	0.000	0.277	0.526	0.536	0.531	0.556	**0.561**
6-12 months	0.002	0.112	0.280	0.304	0.299	0.318	**0.319**
1-2 years	0.004	0.032	0.073	0.068	0.068	0.073	**0.073**
2-4 years	0.000	0.047	0.059	0.073	**0.073**	0.071	0.072
> 4 years	0.031	0.053	0.101	**0.123**	0.123	0.118	0.116

Table 7: R^2 from Micer-Zarnowitz regressions. Bold values indicate the highest R^2 across competing models for a given maturity range and a forecast horizon.

Maturity Category	CV	G-1	G-3	RG-1	RG-1-lev	RG-3	RG-3-lev
	Specification of factor variance dynamics within DNS model						
			1 day-ahead				
< 3 months	*2.656*	0.174	0.248	0.242	0.249	0.270	0.271
3-6 months	*2.784*	0.146	0.243	0.237	0.244	0.268	0.272
6-12 months	*2.921*	0.168	0.257	0.253	0.259	0.285	0.289
1-2 years	*3.094*	0.148	0.286	0.274	0.277	0.308	0.306
2-4 years	*3.126*	0.246	0.326	0.322	0.317	0.344	0.326
> 4 years	*3.210*	0.239	0.247	0.241	0.240	0.272	0.274
			2 days-ahead				
< 3 months	*2.368*	0.200	0.249	0.228	0.237	0.265	0.263
3-6 months	*2.500*	0.191	0.249	0.230	0.237	0.271	0.270
6-12 months	*2.633*	0.203	0.271	0.254	0.260	0.300	0.301
1-2 years	*2.787*	0.238	0.308	0.281	0.285	0.327	0.327
2-4 years	*2.861*	0.261	0.333	0.301	0.300	0.324	0.337
> 4 years	*2.884*	0.246	0.223	0.245	0.246	0.273	0.275
			3 days-ahead				
< 3 months	*2.176*	0.207	0.255	0.233	0.240	0.272	0.269
3-6 months	*2.305*	0.187	0.250	0.233	0.239	0.275	0.273
6-12 months	*2.435*	0.196	0.279	0.261	0.266	0.310	0.309
1-2 years	*2.585*	0.203	0.302	0.291	0.293	0.333	0.339
2-4 years	*2.679*	0.245	0.307	0.296	0.295	0.331	0.332
> 4 years	*2.696*	0.248	0.190	0.249	0.247	0.270	0.273
			4 days-ahead				
< 3 months	*2.033*	0.200	0.246	0.223	0.226	0.259	0.256
3-6 months	*2.157*	0.187	0.243	0.224	0.229	0.269	0.266
6-12 months	*2.286*	0.188	0.267	0.251	0.254	0.299	0.299
1-2 years	*2.443*	0.211	0.301	0.284	0.286	0.326	0.328
2-4 years	*2.528*	0.249	0.304	0.299	0.298	0.322	0.322
> 4 years	*2.565*	0.243	0.137	0.239	0.239	0.247	0.250
MCS coverage			99%		95%		90%

Table 8: Comparison of predictive log-likelihoods across the pool of competing models. In the first column, the mean predictive log-likelihoods for the constant volatility DNS model are reported. In the remaining columns, we provide changes in mean predictive log-likelihoods relative to the values reported in the first column. The background color indicates that the model is in MCS with a coverage 99%, 95%, and 90% depending on the color density.

	Specification of factor variance dynamics within DNS model						
Maturity Category	CV	G-1	G-3	RG-1	RG-1-lev	RG-3	RG-3-lev
				1 week-ahead			
< 3 months	*1.916*	0.192	0.240	0.219	0.221	0.254	0.253
3-6 months	*2.039*	0.175	0.234	0.215	0.220	0.254	0.256
6-12 months	*2.167*	0.179	0.254	0.239	0.243	0.282	0.284
1-2 years	*2.321*	0.201	0.286	0.275	0.280	0.311	0.311
2-4 years	*2.404*	0.212	0.272	0.294	0.294	0.294	0.295
> 4 years	*2.456*	0.228	0.092	0.235	0.235	0.230	0.231
				2 weeks-ahead			
< 3 months	*1.501*	0.166	0.222	0.217	0.219	0.264	0.261
3-6 months	*1.627*	0.147	0.209	0.212	0.212	0.249	0.250
6-12 months	*1.763*	0.138	0.210	0.222	0.224	0.244	0.249
1-2 years	*1.939*	0.158	0.218	0.249	0.249	0.240	0.243
2-4 years	*2.052*	0.167	0.147	0.250	0.249	0.191	0.202
> 4 years	*2.154*	0.197	-0.104	0.184	0.187	0.146	0.158
				3 weeks-ahead			
< 3 months	*1.256*	0.123	0.235	0.240	0.238	0.294	0.295
3-6 months	*1.382*	0.094	0.217	0.227	0.229	0.270	0.270
6-12 months	*1.524*	0.075	0.198	0.226	0.231	0.244	0.243
1-2 years	*1.710*	0.088	0.181	0.239	0.239	0.202	0.203
2-4 years	*1.835*	0.037	0.063	0.209	0.208	0.111	0.115
> 4 years	*1.945*	0.138	-0.382	0.138	0.142	0.091	0.102
				1 month-ahead			
< 3 months	*0.925*	-0.216	0.232	0.199	0.214	0.272	0.282
3-6 months	*1.027*	-0.399	0.209	0.183	0.208	0.234	0.242
6-12 months	*1.172*	-0.434	0.180	0.195	0.211	0.199	0.203
1-2 years	*1.371*	-0.276	0.107	0.201	0.206	0.105	0.129
2-4 years	*1.490*	-0.415	-0.238	0.159	0.170	-0.073	-0.000
> 4 years	*1.650*	-0.003	-1.543	0.090	0.106	-0.072	-0.038
MCS coverage			99%		95%		90%

Table 9: Comparison of predictive log-likelihoods across the pool of competing models. In the first column, the mean predictive log-likelihoods for the constant volatility DNS model are reported. In the remaining columns, we provide changes in mean predictive log-likelihoods relative to the values reported in the first column. The background color indicates that the model is in MCS with a coverage 99%, 95%, and 90% depending on the color density.

References

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. 39(4):885–905.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, 160(1):220–234.

Bandi, F. M. and Russell, J. R. (2008). Microstructure Noise, Realized Variance, and Optimal Sampling. *Review of Economic Studies*, 75:339–69.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise. *Econometrica*, 76(6):1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: trades and quotes. *Econometrics Journal*, 12(3):C1–C32.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17(5):457–477.

Basak, S. and Pavlova, A. (2015). A Model of Financialization of Commodities. CEPR Discussion Papers 10651, C.E.P.R. Discussion Papers.

Bhardwaj, G., Gorton, G., and Rouwenhorst, G. (2015). Facts and Fantasies about Commodity Futures Ten Years Later. NBER Working Papers 21243, National Bureau of Economic Research, Inc.

Bianchi, F., Mumtaz, H., and Surico, P. (2009). The great moderation of the term structure of UK interest rates. *Journal of Monetary Economics*, 56(6):856–871.

Bollerslev, T. (1990). Modelling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. *The Review of Economics and Statistics*, 72(3):498–505.

Brennan, M. J. and Schwartz, E. S. (1985). Evaluating Natural Resource Investments. *The Journal of Business*, 58(2):135–57.

Buyuksahin, B. and Robe, M. A. (2014). Speculators, commodities and cross-market linkages. *Journal of International Money and Finance*, 42(C):38–70.

Carriero, A., Clark, T., and Marcellino, M. (2014). No Arbitrage Priors, Drifting Volatilities, and the Term Structure of Interest Rates. CEPR Discussion Papers 9848, C.E.P.R. Discussion Papers.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947.

Christoffersen, P., Lunde, A., and Olesen, K. V. (2014). Factor Structure in Commodity Futures Return and Volatility. CREATES Research Papers 2014-31, School of Economics and Management, University of Aarhus.

Clark, T. E. (2011). Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility. *Journal of Business & Economic Statistics*, 29(3):327–341.

Collier, P. and Goderis, B. (2012). Commodity prices and growth: An empirical investigation. *European Economic Review*, 56(6):1241–1260.

Corsi, F., Peluso, S., and Audrino, F. (2015). Missing in Asynchronicity: A Kalman-EM Approach for Multivariate Realized Covariance Estimation. *Journal of Applied Econometrics*, 30(3):377–97.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, 8(2):pp. 93–115.

Creal, D., Koopman, S. J., and Lucas, A. (2008). A General Framework for Observation Driven Time-Varying Parameter Models. Tinbergen Institute Discussion Papers 08-108/4, Tinbergen Institute.

Creal, D. D. and Wu, J. C. (2014). Monetary Policy Uncertainty and Economic Fluctuations. NBER Working Papers 20594, National Bureau of Economic Research, Inc.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364.

Diebold, F. X. and Lopez, J. A. (1996). Forecast evaluation and combination. In Maddala, G. and Rao, C., editors, *Handbook of Statistics*, volume Volume 14: Statistical Methods in Finance, pages 241–268. Amsterdam: North-Holland.

Diebold, F. X., Rudebusch, G. D., and Boragan Aruoba, S. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131(1-2):309–338.

Dobrev, D. and Szerszen, P. J. (2010). The information content of high-frequency data for estimating equity return models and forecasting risk. International Finance Discussion Papers 1005, Board of Governors of the Federal Reserve System (U.S.).

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods: Second Edition*. Number 9780199641178 in OUP Catalogue. Oxford University Press.

Engle, R. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics*, 20(3):339–50.

Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1-2):3–27.

Engle, R. F., Ng, V. K., and Rothschild, M. (1990). Asset pricing with a factor-arch covariance structure : Empirical estimates for treasury bills. *Journal of Econometrics*, 45(1-2):213–237.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366):291–298.

Geman, H. (2005). Commodities and commodity derivatives : modeling and pricing for agriculturals, metals and energy. Economics Papers from University Paris Dauphine 123456789/607, Paris Dauphine University.

Geweke, J. and Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26(2):216–230.

Golosnoy, V., Gribisch, B., and Liesenfeld, R. (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics*, 167(1):211–223.

Gorton, G. and Rouwenhorst, K. G. (2006). Facts and Fantasies about Commodity Futures. *Financial Analysts Journal*, 62(2):47–68.

Hall, S. G. and Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23(1):1–13.

Hansen, N., Lunde, A., Olesen, K., and Elst, H. (2014). Realizing Commodity Correlations and the Market Beta. Creates research papers, School of Economics and Management, University of Aarhus.

Hansen, N. S. and Lunde, A. (2013). Analyzing Oil Futures with a Dynamic Nelson-Siegel Model. CREATES Research Papers 2013-36, School of Economics and Management, University of Aarhus.

Hansen, P. R. and Huang, Z. (2016). Exponential garch modeling with realized measures of volatility. *Journal of Business & Economic Statistics*, 34(2):269–287.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Hansen, P. R. and Lunde, A. (2005). A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data. *Journal of Financial Econometrics*, 3(4):525–554.

Hansen, P. R. and Lunde, A. (2006a). Consistent ranking of volatility models. *Journal of Econometrics*, 131(1-2):97–121.

Hansen, P. R. and Lunde, A. (2006b). Realized Variance and Market Microstructure Noise. *Journal of Business & Economic Statistics*, 24:127–161.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The Model Confidence Set. *Econometrica*, 79(2):453–497.

Harvey, A., Ruiz, E., and Sentana, E. (1992). Unobserved component time series models with Arch disturbances. *Journal of Econometrics*, 52(1-2):129–157.

Harvey, A. C. (1991). *Forecasting. Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hautsch, N. and Ou, Y. (2012). Analyzing interest rate risk: Stochastic volatility in the term structure of government bond yields. *Journal of Banking & Finance*, 36(11):2988–3007.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and their Applications*, 119(7):2249–2276.

Karstanje, D., van der Wel, M., and van Dijk, D. (2015). Common Factors in Commodity Futures Curves. Technical report.

Kilian, L. (2008). The Economic Effects of Energy Price Shocks. *Journal of Economic Literature*, 46(4):871–909.

Kilian, L. (2009). Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. *American Economic Review*, 99(3):1053–69.

Koopman, S. J., Lucas, A., and Scharth, M. (2015). Predicting time-varying parameters with parameter-driven and observation-driven models. *Review of Economics and Statistics*, Forthcoming.

Koopman, S. J., Mallee, M. I. P., and Van der Wel, M. (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.

Koopman, S. J. and Scharth, M. (2012). The Analysis of Stochastic Volatility in the Presence of Daily Realized Measures. *Journal of Financial Econometrics*, 11(1):76–115.

Liu, L. Y., Patton, A. J., and Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187(1):293–311.

Lopez, J. A. (2001). Evaluating the Predictive Accuracy of Volatility Models. *Journal of Forecasting*, 20(2):87–109.

Lunde, A. and Olesen, K. V. (2014). Modeling and Forecasting the Distribution of Energy Forward Returns - Evidence from the Nordic Power Exchange. CREATES Research Papers 2013-19, School of Economics and Management, University of Aarhus.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious Modeling of Yield Curves. *The Journal of Business*, 60(4):473–89.

Ravazzolo, F. and Lombardi, M. J. (2012). Oil price density forecasts: Exploring the linkages with stock markets. Working Papers 0008, Centre for Applied Macro- and Petroleum economics (CAMP), BI Norwegian Business School.

Samuelson, P. (1965). Proof that properly anticipated prices fluctates randomly. *Industrial management review*, 6(Spring):13–31.

Schwartz, E. and Smith, J. E. (2000). Short-Term Variations and Long-Term Dynamics in Commodity Prices. *Management Science*, 46(7):893–911.

Schwartz, E. S. (1997). The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging. *Journal of Finance*, 52(3):923–73.

Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25(2):197–231.

Shephard, N. and Xiu, D. (2012). Econometric analysis of multivariate realised QML: efficient positive semi-definite estimators of the covariation of equity prices. Economics Series Working Papers 604, University of Oxford, Department of Economics.

Shin, M. and Zhong, M. (2015). Does Realized Volatility Help Bond Yield Density Prediction? Finance and economics discussion series 2015-115, Washington: Board of Governors of the Federal Reserve System.

Stock, J. and Watson, M. (2011). Dynamic factor models. *The Oxford Handbook of Economic Forecasting*.

Takahashi, M., Omori, Y., and Watanabe, T. (2009). Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics & Data Analysis*, 53(6):2404–2426.

Wasserfallen, W. and Zimmermann, H. (1985). The behavior of intra-daily exchange rates. *Journal of Banking & Finance*, 9(1):55–72.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50(1):1–25.

Zhang, L., Mykland, P. A., and Ait-Sahalia, Y. (2005). A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data. *Journal of the American Statistical Association*, 100:1394–1411.

A Realized Variances for a Panel of Contracts

In this appendix, we investigate the co-dynamics of realized variances computed for contracts with different maturities. The purpose of this inspection is to validate the assumption that the variance of the oil futures term structure can be adequately captured by means of just few dynamic volatility components.

We consider the period between January 3, 2012 and January 16, 2015, which amounts to 785 trading days. This period is chosen due to availability of the large cross-section of traded contracts, so we are able to consider a broad spectrum of maturities. We partition this spectrum into 10 maturity categories: up to 2 month, 2-4, 4-6, 6-9, 9-12, 12-18, 18-24, 24-36, 36-48 and more than 48 months until expiration. For each trading day from the considered sample period and for each maturity category we compute a subsampled realized variance as in (24) using high frequency transaction data. If at a given day t we observe several traded contracts that belong to the same maturity category, we take an average of the corresponding realized variances. As a result, we have a balanced panel of realized variances that spans 10 maturity ranges and 785 trading days (see Figure A.1).

The matrix of correlations between the log realized variance series of different maturity ranges is provided in Table A.1. We can see that the realized variances are highly correlated across all maturity categories.

$$\begin{array}{r} < 2 \\ 2-4 \\ 4-6 \\ 6-9 \\ 9-12 \\ 12-18 \\ 18-24 \\ 24-36 \\ 36-48 \\ > 48 \end{array} \begin{bmatrix} 1.000 & 0.990 & 0.961 & 0.885 & 0.788 & 0.766 & 0.697 & 0.740 & 0.703 & 0.541 \\ & 1.000 & 0.970 & 0.892 & 0.799 & 0.775 & 0.702 & 0.749 & 0.709 & 0.551 \\ & & 1.000 & 0.900 & 0.796 & 0.758 & 0.688 & 0.725 & 0.692 & 0.515 \\ & & & 1.000 & 0.832 & 0.726 & 0.634 & 0.686 & 0.657 & 0.453 \\ & & & & 1.000 & 0.789 & 0.647 & 0.657 & 0.671 & 0.492 \\ & & & & & 1.000 & 0.753 & 0.696 & 0.678 & 0.553 \\ & & & & & & 1.000 & 0.643 & 0.588 & 0.463 \\ & & & & & & & 1.000 & 0.669 & 0.508 \\ & & & & & & & & 1.000 & 0.653 \\ & & & & & & & & & 1.000 \end{bmatrix}$$

Table A.1: Correlation matrix of log realized variances computed for distinct maturity categories (January 3, 2012 - January 16, 2015). The maturity range is specified in the left column.

These sample correlations, however, are silent about the co-dynamic properties of the realized variance series as well as about the presence of some common factors in such dynamics. To gauge the possible co-variation we implement a principal component analysis. The results are provided in Table A.2.

PC	Variation	Total
1	0.690	0.690
2	0.137	0.826
3	0.048	0.874
4	0.038	0.913
5	0.033	0.945
6	0.027	0.972
7	0.019	0.991

Table A.2: Principal component analysis of the log realized variance panel. Seven main principal components appear in the first column. In the second column there is a percentage of the variation explained by the corresponding principal components. In the third column we report the cumulative fraction of the explained variation.

As it can be seen, the first principal component explains roughly two third of the total variation in the joint dynamics of the log realized variance series, whereas the first 4 explain more than 90% of the total variation.

On one hand, the principal component analysis indicates that the volatility panel reveals a significant degree of comovement. To a certain extent this finding supports the idea that even only one time-varying component can describe an appreciable amount of the common realized variance dynamics. Note that the realized variance of the shortest contract that is used as a volatility state variable in our one-component dynamic volatility specification has a correlation about 0.9 with the first principal component.

On the other hand, if just a one component is used, a sufficient amount of realized variance dynamics (about 30%) is left unexplained. This motivates to make use of several dynamic components in order to better capture the conditional second moment of the whole term structure distribution.

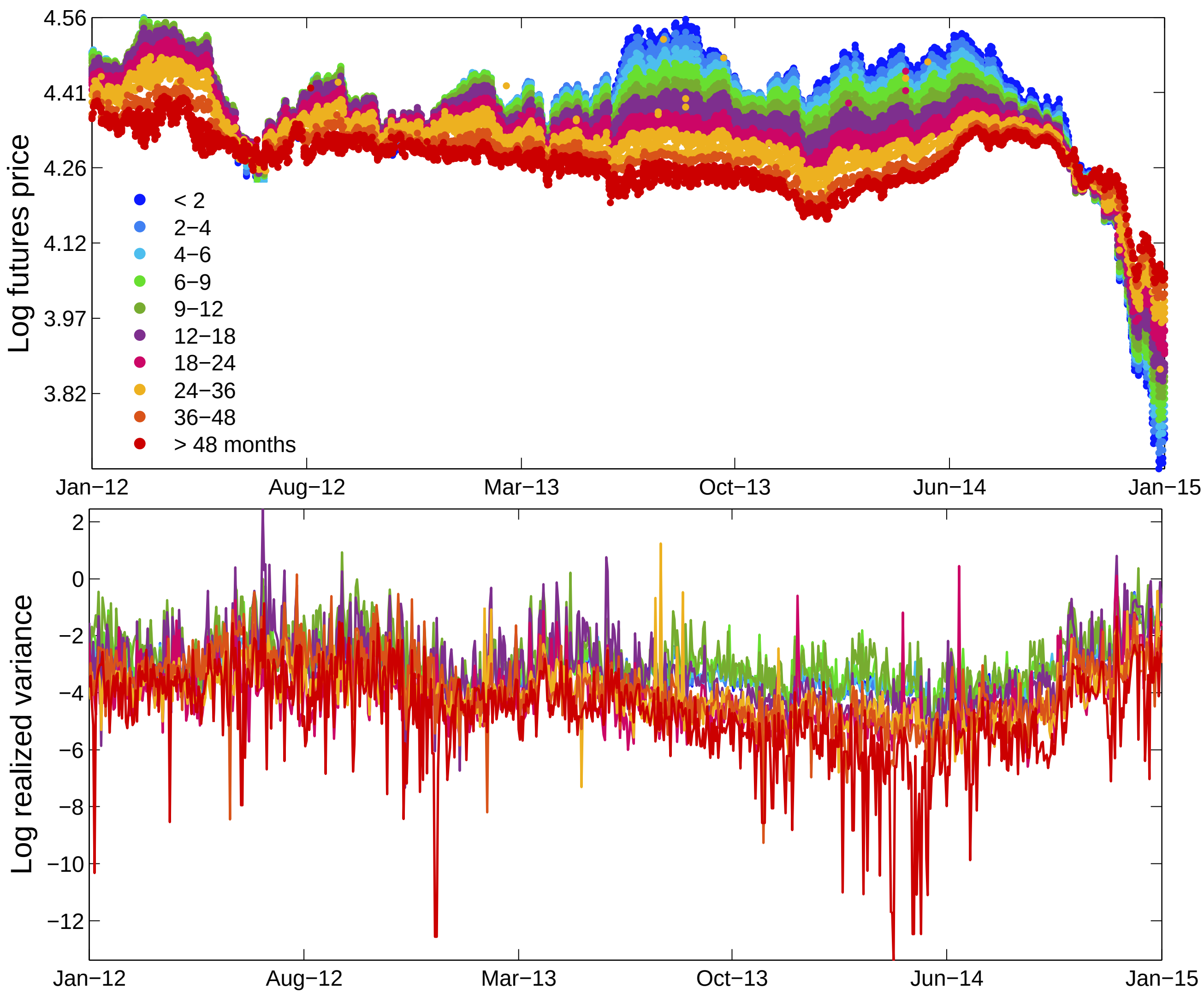


Figure A.1: Dynamics of the log prices (top panel) and the corresponding log annualized realized variances (bottom panel) for oil futures (January 3, 2012 - January 16, 2015).

B State space formulations

In the technical appendix we use the following notation inspired by Durbin and Koopman (2012).

The standard linear Gaussian state space system is given by

Observation equation

$$x_t = d_t + Z_t \alpha_t + \nu_t \qquad \nu_t \sim \mathcal{N}(0, V_t)$$

State equation

$$\alpha_{t+1} = T_t \alpha_t + c_t + W_{t+1} q_{t+1} \qquad q_{t+1} \sim N(0, Q_{t+1})$$

Where the Kalman filter is defined by the following recursions:

Kalman filter

$$e_t = x_t - Z_t a_t - d_t \qquad F_t = Z_t P_t Z_t^{'} + V_t$$

Updating step

$$a_{t|t} = a_t + P_t Z_t^{'} F_t^{-1} e_t \qquad P_{t|t} = P_t - P_t Z_t^{'} F_t^{-1} Z_t P_t$$

Prediction step

$$a_{t+1} = T_t a_{t|t} + c_t \qquad P_{t+1} = T_t P_{t|t} T_t^{'} + W_{t+1} Q_{t+1} W_{t+1}^{'}$$

where e_t denotes the one-step ahead forecast error. $F_t = \mathrm{Var}(e_t | \mathcal{F}_{t-1})$ is the conditional variance of the forecast errors. $a_{t|t} = \mathrm{E}(\alpha_t | \mathcal{F}_t)$ is the filtered states and $P_{t|t} = \mathrm{Var}(\alpha_t | \mathcal{F}_t)$ denotes the time t conditional variance of the states. $a_{t+1} = \mathrm{E}(\alpha_{t+1} | \mathcal{F}_t)$ is the one-step ahead predicted states and $P_{t+1} = \mathrm{Var}(\alpha_{t+1} | \mathcal{F}_t)$.

The log-likelihood function of a standard linear Gaussian state space system is given by

$$\log p(x_1, ..., x_T) = \sum_{t=1}^{T} \log p(x_t | \mathcal{F}_{t-1})$$

where

$$\log p(x_t | \mathcal{F}_{t-1}) = -\frac{n_t}{2} \log(2\pi) - \frac{1}{2}\left(\log |F_t| + e_t^{'} F_t^{-1} e_t\right) \tag{B.1}$$

The next sections contain the specific state space formulations of each model.

B.1 Constant volatility

The constant volatility dynamic Nelson-Siegel model can be formulated directly in the standard linear Gaussian state space formulation. Matching the notation of the standard linear Gaussian state space system with the notation of the constant volatility dynamic Nelson-Siegel model we have that:

State space notation		Dynamic NS notation
x_t	$=$	y_t
Z_t	$=$	$(\Lambda_t f_t, -\beta C_t \Lambda_{t-1})$
α_t	$=$	$\begin{pmatrix} f_t \\ f_{t-1} \end{pmatrix}$
d_t	$=$	$\beta C_t y_{t-1}$
ν_t	$=$	w_t
T_t	$=$	$\begin{pmatrix} I_{3\times3} & 0_{3\times3} \\ I_{3\times3} & 0_{3\times3} \end{pmatrix}$
W_t	$=$	$\begin{pmatrix} I_{3\times3} \\ 0_{3\times3} \end{pmatrix}$
q_{t+1}	$=$	η_{t+1}
Q_{t+1}	$=$	$\begin{pmatrix} \sigma_l^2 & \rho_{ls}\sigma_l\sigma_s & \rho_{lc}\sigma_l\sigma_c \\ \rho_{ls}\sigma_l\sigma_s & \sigma_s^2 & \rho_{sc}\sigma_s\sigma_c \\ \rho_{lc}\sigma_l\sigma_c & \rho_{sc}\sigma_s\sigma_c & \sigma_c^2 \end{pmatrix}$
c_t	$=$	$0_{3\times1}$

B.2 GARCH-1

In the GARCH-1 model we follow Harvey et al. (1992).

State space notation		Dynamic NS notation
x_t	$=$	y_t
Z_t	$=$	$(\Lambda_t f_t, -\beta C_t \Lambda_{t-1}, 0_{n_t \times 1})$
α_t	$=$	$\begin{pmatrix} f_t \\ f_{t-1} \\ \eta_{t,l} \end{pmatrix}$
d_t	$=$	$\beta C_t y_{t-1}$
ν_t	$=$	w_t
T_t	$=$	$\begin{pmatrix} I_{3\times3} & 0_{3\times3} & 0_{3\times1} \\ I_{3\times3} & 0_{3\times3} & 0_{3\times1} \\ 0_{1\times3} & 0_{1\times3} & 0_{1\times1} \end{pmatrix}$
W_t	$=$	$\begin{pmatrix} I_{3\times3} \\ 0_{3\times3} \\ (1,0,0) \end{pmatrix}$
q_{t+1}	$=$	η_{t+1}
Q_{t+1}	$=$	$H_{t+1} R H_{t+1}$
c_t	$=$	$0_{3\times1}$

where

$$
\begin{aligned}
H_{t+1} &= \begin{pmatrix} \sqrt{h_{t+1,l}} & 0 & 0 \\ 0 & \sqrt{h_{t+1,s}} & 0 \\ 0 & 0 & \sqrt{h_{t+1,c}} \end{pmatrix} \\
R &= \begin{pmatrix} 1 & \rho_{ls} & \rho_{lc} \\ \rho_{ls} & 1 & \rho_{sc} \\ \rho_{lc} & \rho_{sc} & 1 \end{pmatrix} \\
h_{t+1} &= \gamma_{0l} + \gamma_{1l}\left(a^2_{t|t,7} + P_{t|t\,7,7}\right) + \gamma_{2l} h_t
\end{aligned}
$$

Here $a_{t|t,7}$ denotes the 7th element in $a^2_{t|t}$ and $P_{t|t\,7,7}$ is the $(7,7)$ element in $P_{t|t}$. Since $\eta^2_{t,l}$ is not directly observable we use $\mathrm{E}\left(\eta^2_{t,l}\,|\mathcal{F}_t\right) = a^2_{t|t,7} + P_{t|t\,7,7}$ as the "signal" in the volatility process. Hence, the model should be seen as an approximation to a "standard GARCH" model.

B.3 GARCH-3

State space notation		Dynamic NS notation
x_t	$=$	y_t
Z_t	$=$	$(\Lambda_t f_t, -\beta C_t \Lambda_{t-1}, 0_{n_t \times 3})$
α_t	$=$	$\begin{pmatrix} f_t \\ f_{t-1} \\ \eta_t \end{pmatrix}$
d_t	$=$	$\beta C_t y_{t-1}$
ν_t	$=$	w_t
T_t	$=$	$\begin{pmatrix} I_{3\times 3} & 0_{3\times 3} & 0_{3\times 3} \\ I_{3\times 3} & 0_{3\times 3} & 0_{3\times 3} \\ 0_{3\times 3} & 0_{3\times 3} & 0_{3\times 3} \end{pmatrix}$
W_t	$=$	$\begin{pmatrix} I_{3\times 3} \\ 0_{3\times 3} \\ I_{3\times 3} \end{pmatrix}$
q_{t+1}	$=$	η_{t+1}
Q_{t+1}	$=$	$H_{t+1} R H_{t+1}$
c_t	$=$	$0_{3\times 1}$

where

$$\begin{aligned}
H_{t+1} &= \begin{pmatrix} \sqrt{h_{t+1,l}} & 0 & 0 \\ 0 & \sqrt{h_{t+1,s}} & 0 \\ 0 & 0 & \sqrt{h_{t+1,c}} \end{pmatrix} \\
R &= \begin{pmatrix} 1 & \rho_{ls} & \rho_{lc} \\ \rho_{ls} & 1 & \rho_{sc} \\ \rho_{lc} & \rho_{sc} & 1 \end{pmatrix} \\
h_{t+1,l} &= \gamma_{0l} + \gamma_{1l}\left(a_{t|t,7}^2 + P_{t|t\,7,7}\right) + \gamma_{2l} h_{t,l} \\
h_{t+1,s} &= \gamma_{0s} + \gamma_{1s}\left(a_{t|t,8}^2 + P_{t|t\,8,8}\right) + \gamma_{2s} h_{t,s} \\
h_{t+1,c} &= \gamma_{0c} + \gamma_{1c}\left(a_{t|t,9}^2 + P_{t|t\,9,9}\right) + \gamma_{2c} h_{t,c}
\end{aligned}$$

Here $a_{t|t,i}$ denotes the i'th element in $a_{t|t}$ and $P_{t|t\,i,j}$ is the (i,j) element in $P_{t|t}$. Since $\eta_{t,k}^2$ is not directly observable we use

$$\mathrm{E}\left(\eta_{t,k}^2 \,|\mathcal{F}_t\right) = \begin{cases} a_{t|t,7}^2 + P_{t|t\,7,7} & \text{for } k = l \\ a_{t|t,8}^2 + P_{t|t\,9,9} & \text{for } k = s \\ a_{t|t,8}^2 + P_{t|t\,9,9} & \text{for } k = c \end{cases}$$

as the "signal" in the volatility process. Hence, the model should be seen as an approximation to a "standard multivariate GARCH" model.

B.4 RGARCH-1-lev

State space notation		Dynamic NS notation
x_t	$=$	y_t
Z_t	$=$	$(\Lambda_t f_t, -\beta C_t \Lambda_{t-1})$
α_t	$=$	$\begin{pmatrix} f_t \\ f_{t-1} \end{pmatrix}$
d_t	$=$	$\beta C_t y_{t-1}$
ν_t	$=$	w_t
T_t	$=$	$\begin{pmatrix} I_{3\times3} & 0_{3\times3} \\ I_{3\times3} & 0_{3\times3} \end{pmatrix}$
W_t	$=$	$\begin{pmatrix} I_{3\times3} \\ 0_{3\times3} \end{pmatrix}$
q_{t+1}	$=$	η_{t+1}
Q_{t+1}	$=$	$H_{t+1} R H_{t+1}$
c_t	$=$	$0_{3\times1}$

Hence, the standard space formulation is equivalent to the constant volatility model except for the fact that Q_{t+1} is time-varying and driven by realized measures. Here we have that:

$$H_{t+1} = \begin{pmatrix} \sqrt{\exp(h_t)} & 0 & 0 \\ 0 & \sqrt{a_s + b_s \exp(h_t)} & 0 \\ 0 & 0 & \sqrt{a_c + b_c \exp(h_t)} \end{pmatrix}$$

with

$$\begin{aligned} h_t &= \gamma_0 + \gamma_1 s_{t-1} + \gamma_2 h_{t-1} \\ s_t &= \begin{cases} \xi + \phi h_t + \delta(z_t) + u_t & \text{with leverage} \\ \xi + \phi h_t + u_t & \text{without leverage} \end{cases} \end{aligned}$$

In the Realized models we need to account for s_t being an observable. Hence, we need to add

$$\log p(s_t | y_t, \mathcal{F}_{t-1}; \theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\left(\log(\sigma_u^2) + \frac{u_t^2}{\sigma_u^2}\right)$$

to the NS log-likelihood contribution (B.1).

B.5 RGARCH-3-lev

State space notation		Dynamic NS notation
x_t	$=$	y_t
Z_t	$=$	$(\Lambda_t f_t, -\beta C_t \Lambda_{t-1})$
α_t	$=$	$\begin{pmatrix} f_t \\ f_{t-1} \end{pmatrix}$
d_t	$=$	$\beta C_t y_{t-1}$
ν_t	$=$	w_t
T_t	$=$	$\begin{pmatrix} I_{3\times 3} & 0_{3\times 3} \\ I_{3\times 3} & 0_{3\times 3} \end{pmatrix}$
W_t	$=$	$\begin{pmatrix} I_{3\times 3} \\ 0_{3\times 3} \end{pmatrix}$
q_{t+1}	$=$	η_{t+1}
Q_{t+1}	$=$	$H_{t+1} R H_{t+1}$
c_t	$=$	$0_{3\times 1}$

Hence, the standard space formulation is equivalent to the constant volatility model except for the fact that Q_{t+1} is time-varying and driven by realized measures. Here we have that:

$$H_{t+1} = \begin{pmatrix} \sqrt{\exp(h_{t,l})} & 0 & 0 \\ 0 & \sqrt{\exp(h_{t,s})} & 0 \\ 0 & 0 & \sqrt{\exp(h_{t,c})} \end{pmatrix}$$

with

$$\begin{aligned} h_{t,j} &= \gamma_{0,j} + \gamma_{1,j} s_{t-1,j} + \gamma_{2,j} h_{t-1,j} \\ s_{t,j} &= \begin{cases} \xi_j + \phi_j h_{t,j} + \delta_j(z_{t,j}) + u_{t,j} & \text{with leverage} \\ \xi_j + \phi_j h_{t,j} + u_{t,j} & \text{without leverage} \end{cases} \\ & \text{for } j = l, s, c \end{aligned}$$

In the Realized models we need to account for s_t being an observable. Hence, we need to add

$$\begin{aligned} \log p(s_t \,|y_t, \mathcal{F}_{t-1}; \theta) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\left(\log\left(\sigma_{u,l}^2\right) + \frac{u_{t,l}^2}{\sigma_{u,l}^2}\right) \\ &\quad -\frac{1}{2}\log(2\pi) - \frac{1}{2}\left(\log\left(\sigma_{u,s}^2\right) + \frac{u_{t,s}^2}{\sigma_{u,s}^2}\right) \\ &\quad -\frac{1}{2}\log(2\pi) - \frac{1}{2}\left(\log\left(\sigma_{u,c}^2\right) + \frac{u_{t,c}^2}{\sigma_{u,c}^2}\right) \end{aligned}$$

to the NS log-likelihood contribution (B.1).

C Additional Results on Density Forecasting

Forecasting sub-period with relatively mild price dynamics (March 3, 2014 - August 29, 2014)

	Specification of factor variance dynamics within DNS model						
Maturity Category	CV	G-1	G-3	RG-1	RG-1-lev	RG-3	RG-3-lev
			1 day-ahead				
< 3 months	*2.805*	0.290	0.368	0.333	0.345	0.386	0.391
3-6 months	*2.947*	0.286	0.354	0.327	0.338	0.379	0.382
6-12 months	*3.077*	0.309	0.373	0.352	0.360	0.406	0.408
1-2 years	*3.212*	0.349	0.413	0.389	0.394	0.443	0.442
2-4 years	*3.305*	0.360	0.445	0.395	0.396	0.435	0.433
> 4 years	*3.287*	0.300	0.325	0.321	0.320	0.331	0.334
			2 days-ahead				
< 3 months	*2.469*	0.305	0.398	0.352	0.368	0.417	0.422
3-6 months	*2.622*	0.305	0.386	0.351	0.363	0.416	0.418
6-12 months	*2.756*	0.339	0.415	0.388	0.398	0.456	0.459
1-2 years	*2.890*	0.382	0.457	0.426	0.431	0.494	0.493
2-4 years	*2.996*	0.412	0.513	0.448	0.448	0.507	0.504
> 4 years	*2.963*	0.307	0.302	0.319	0.321	0.330	0.336
			3 days-ahead				
< 3 months	*2.282*	0.313	0.408	0.362	0.373	0.434	0.439
3-6 months	*2.431*	0.316	0.393	0.360	0.372	0.425	0.429
6-12 months	*2.567*	0.352	0.427	0.401	0.411	0.475	0.476
1-2 years	*2.707*	0.401	0.479	0.445	0.452	0.522	0.521
2-4 years	*2.815*	0.435	0.543	0.470	0.470	0.539	0.536
> 4 years	*2.781*	0.313	0.282	0.317	0.315	0.332	0.334
			4 days-ahead				
< 3 months	*2.140*	0.309	0.399	0.357	0.369	0.422	0.423
3-6 months	*2.288*	0.316	0.391	0.359	0.371	0.423	0.426
6-12 months	*2.427*	0.351	0.427	0.402	0.410	0.475	0.475
1-2 years	*2.571*	0.410	0.492	0.457	0.460	0.533	0.532
2-4 years	*2.684*	0.448	0.561	0.482	0.482	0.554	0.553
> 4 years	*2.655*	0.326	0.255	0.316	0.318	0.324	0.331
MCS coverage			99%		95%		90%

Table C.3: Comparison of predictive log-likelihoods across the pool of competing models. In the first column, the mean predictive log-likelihoods for the constant volatility DNS model are reported. In the remaining columns, we provide changes in mean predictive log-likelihoods relative to the values reported in the first column. The background color indicates that the model is in MCS with a coverage 99%, 95%, and 90% depending on the color density.

Forecasting sub-period with relatively mild price dynamics (March 3, 2014 - August 29, 2014)

	Specification of factor variance dynamics within DNS model						
Maturity Category	CV	G-1	G-3	RG-1	RG-1-lev	RG-3	RG-3-lev
			1 week-ahead				
< 3 months	*2.022*	0.307	0.391	0.356	0.362	0.413	0.415
3-6 months	*2.172*	0.313	0.383	0.356	0.363	0.411	0.413
6-12 months	*2.317*	0.351	0.422	0.400	0.407	0.467	0.466
1-2 years	*2.464*	0.413	0.494	0.459	0.463	0.535	0.533
2-4 years	*2.580*	0.458	0.574	0.492	0.492	0.566	0.563
> 4 years	*2.557*	0.332	0.240	0.312	0.316	0.332	0.335
			2 weeks-ahead				
< 3 months	*1.680*	0.295	0.364	0.344	0.351	0.393	0.391
3-6 months	*1.829*	0.301	0.357	0.348	0.349	0.388	0.384
6-12 months	*1.976*	0.343	0.398	0.395	0.399	0.444	0.442
1-2 years	*2.133*	0.414	0.480	0.468	0.469	0.521	0.518
2-4 years	*2.254*	0.453	0.546	0.486	0.483	0.531	0.529
> 4 years	*2.230*	0.312	0.068	0.232	0.242	0.239	0.257
			3 weeks-ahead				
< 3 months	*1.490*	0.299	0.371	0.346	0.347	0.405	0.405
3-6 months	*1.639*	0.302	0.362	0.347	0.347	0.400	0.395
6-12 months	*1.786*	0.338	0.391	0.388	0.389	0.444	0.437
1-2 years	*1.941*	0.402	0.464	0.457	0.451	0.507	0.499
2-4 years	*2.052*	0.422	0.497	0.432	0.425	0.471	0.465
> 4 years	*2.017*	0.276	-0.197	0.157	0.171	0.160	0.178
			1 month-ahead				
< 3 months	*1.302*	0.294	0.380	0.342	0.343	0.405	0.401
3-6 months	*1.448*	0.290	0.358	0.335	0.335	0.391	0.385
6-12 months	*1.602*	0.330	0.387	0.381	0.378	0.443	0.436
1-2 years	*1.753*	0.385	0.443	0.433	0.425	0.487	0.479
2-4 years	*1.857*	0.380	0.438	0.380	0.376	0.420	0.415
> 4 years	*1.814*	0.221	-0.477	0.083	0.112	0.088	0.116
MCS coverage			99%		95%		90%

Table C.4: Comparison of predictive log-likelihoods across the pool of competing models. In the first column, the mean predictive log-likelihoods for the constant volatility DNS model are reported. In the remaining columns, we provide changes in mean predictive log-likelihoods relative to the values reported in the first column. The background color indicates that the model is in MCS with a coverage 99%, 95%, and 90% depending on the color density.