Analyzing Oil Futures with a Dynamic Nelson-Siegel Model

NIELS STRANGE HANSEN & ASGER LUNDE

DEPARTMENT OF ECONOMICS AND BUSINESS, BUSINESS AND SOCIAL SCIENCES, AARHUS UNIVERSITY
AND
CENTER FOR RESEARCH IN ECONOMETRIC ANALYSIS OF TIME SERIES, CREATES

This version: September 1, 2012

ABSTRACT

In this paper we are interested in the term structure of futures contracts on oil. The objective is to specify a relatively parsimonious model which explains data well and performs well in out of sample forecasting. The dynamic Nelson-Siegel model is normally used to analyze and forecast interest rates of different maturities. The structure of oil futures resembles the structure of interest rates and this motivates the use of this model for our purposes. The Dynamic Nelson-Siegel model allows for a significant dimension reduction by introducing three factors. By performing a series of cross-section regressions we obtain time series for these factors and we focus on modeling their joint distribution. Using a copula decomposition we can set up a model for each factor individually along with a model for their dependence structure. When a reasonable model for the factors has been specified it can be used to forecast prices of futures contracts with different maturities. The outcome of this exercise is a model which describes the observed futures contracts well and forecasts better than conventional benchmarks. The model is flexible and is expected to work well for other energy based commodities.

1 Introduction

Following the financial crisis commodity markets have received a lot of attention. Commodity markets pose a lot of interesting challenges and insights for practitioners as well as people in academia. In this paper we set up a model which is very useful for analyzing and forecasting the term structure of energy based commodities. Data in these markets present the analyst with a lot of problems which we normally do not encounter in for example equity markets. The approach we undertake in this section is inspired by the approach to the modeling of the forward curve of interest rates in Diebold and Li (2006) and Noureldin (2011). The underlying assumption is that the term structure of prices of futures contracts can be explained by unobserved factors. These factors can be estimated and using the copula frame work we can model and forecast the factors individually without ignoring their dependence. With forecasts for the factors we can forecast the prices of futures contracts.

The rest of this paper is organized as follows. In section 2 we provide a thorough description of the data set analyzed in this paper. Section 3 contains a description of the model. In section 4 we present the results of the in sample analysis. Section 5 contains the results of the out of sample forecast analysis. Section 6 is devoted to the development of a trading strategy. Finally, some concluding remarks are presented in section 7.

2 Data

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

The data set we consider consists of daily closing prices of monthly futures contracts on oil. The contracts are on light sweet crude oil (WTI), more details can be found on the CME Group homepage[1]. At every day a number of futures with different time to maturity are traded. The maturities are approximately one month apart. Each contract expires on the third trading day prior to the 25th calendar day in the month before delivery. The first observation is from January 2nd 2005 and our estimation sample ends on July 1st 2011. We use the period until February 8th 2012 for forecast evaluation.

It is important to understand the nature of the data in order to understand the problems that we analyze in this paper. In order to grasp the structure and complexities of the data it is useful to consider an example of actual data. In figure 1 we have presented a small example of what data looks like. Dots are used to indicate that the data set extends in this direction. There are several important things to notice. First, note that on a given day only a limited number of contracts exists. Consider for example the first row of the table, on this day a contracts with 495 days to delivery does not exist.

---

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

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

Table 1

Such a contract exists on the next day, though. Secondly, the data contains a number of missing observations. On 06.09.2010 it was possible to trade in a contract with 495 days to maturity, but no one did. This means that we have two kinds of 'holes' in the data set. Contracts that do not exist and contracts which exist but are not traded. Both poses problems for a traditional time series analysis. Assume now that on 06.10.2010 we are interested in forecasting the price of a contract with 492 days to delivery which is traded on 06.11.2010. Futures contracts are characterized by their time to delivery so a good way of forecasting would be to consider a time series of prices of contracts with the same time to maturity. Such a time series are not observed, but it could be constructed by interpolation. Another possibility would be to consider the time series of prices for this particular contract. The drawbacks here are that the previous observations of prices are based on an other time to maturity. Furthermore, we have to decide what to do about missing values. If the contract was introduced to the market recently this approach might be infeasible due to the limited number of observations.

The data example is too limited to fully understand the data. Therefore we have sketched the structure of the data in figure 1. The figure illustrates how time to maturity, τ, for a number of different contracts changes over time. At every point in time we observe prices of a number of contracts with different maturities, these are represented by black dots. The gray dots in the top of the figure illustrate that a much higher number of contracts are traded on a given day. The line at T illustrates the end of our sample. This serves to highlight an important feature. namely that when we forecast, we know for certain the maturities of the contracts potentially traded the next day. These are represented by gray squares and it is the prices of these contracts which we are interested in forecasting.

In the top panel of figure 2 we present an example of how the observations from a single day, t, could look like. The black dots represent the prices observed in the market. And the line through the dots is the term structure of oil futures on this particular day. Our objective is to model and forecast this term structure. We do this in the following way. We analyze the observed data to determine which factors, X_t, determine the term structure at time t. We forecast these factors, $\hat{X}_{t+1}$, to construct a term structure for time $t + 1$ and because we know the maturities of the contracts traded at that time, we can forecast their prices, as illustrated in the bottom panel of figure 2. .

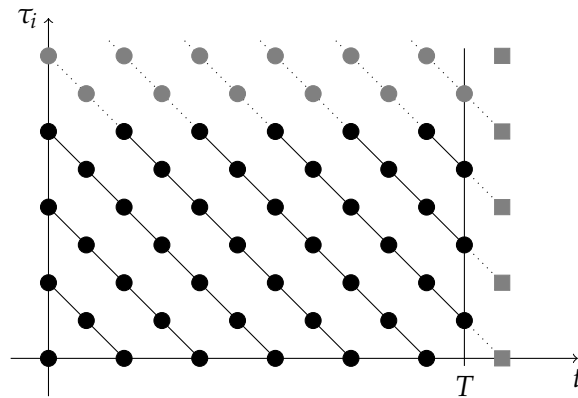The two figures presented here are simplifications and only used for illustrative

Figure 1

purposes. The actual data is far more extensive and not as nice and simple as these sketches. One thing that complicates matters significantly is that the number of different contracts which are traded differ from day to day. This is primarily a consequence of very limited trading in contracts with long maturities. These may simply not be traded every day. In figure 2 the prices are shown to lie on a nice smooth curve, but in the actual data this need not be the case. One of the questions we address in this analysis is whether a smooth curve is a reasonable approximation. Some days we might not have sufficient observations to fit a smooth curve to the data.

3   Model

We want to analyze the term structure of energy based commodity futures. The term structure of futures prices resembles the term structure of interest rates. Therefore we will use a term structure model known from the fixed income literature to analyze the commodity futures prices.

The model we use is the Dynamic Nelson-Siegel model. The Nelson-Siegel model was introduced to model yield curves by Nelson and Siegel (1987). A dynamic version of the model was introduced by Diebold and Li (2006). It states that at any time, t, the k different yields can be explained by three factors according to the following model
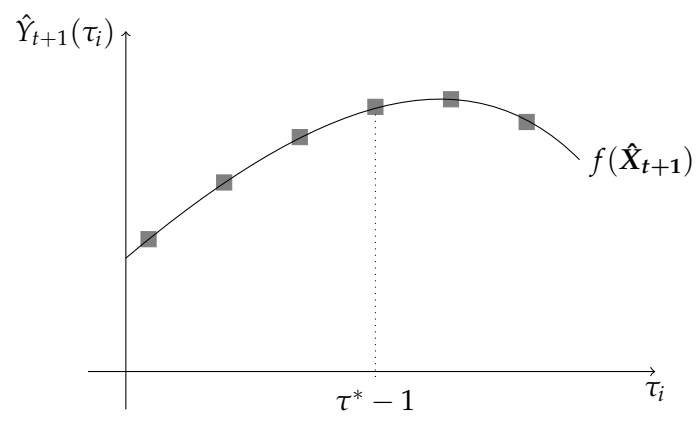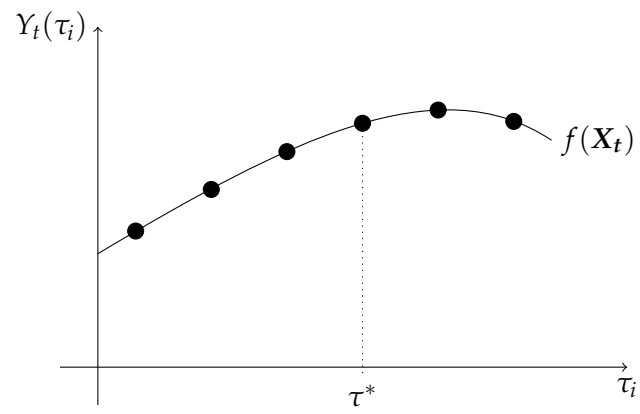
4

Figure 2

5

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

$$\underset{k \times 1}{\mathbf{Y}_t} = \underset{k \times 3}{\mathbf{Z}} \underset{3 \times 1}{\begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix}} + \underset{k \times 1}{\mathbf{e}_t}.$$

In this paper we consider futures prices. We use this model to describe the relationship between futures prices and their time to maturity. We extract the factors by a series of cross-section regressions.

We have time series for $x_{i,t}$ for $i = 1, 2, 3$, these are realizations of the random variables $\mathbf{X} = (X_1, X_2, X_3)'$ and we want to model their joint density, $f_{\mathbf{X}}(\mathbf{x})$. For this purpose we use the Copula-Margins decomposition as explained in Patton (2009). This allows us to express the joint density as

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^{3} f_{X_i}(x_i) c\left(F_{X_1}(x_1), F_{X_2}(x_2), F_{X_3}(x_3)\right) \\ &= \prod_{i=1}^{3} f_{X_i}(x_i) c\left(u_1, u_2, u_3\right), \end{aligned} \tag{1}$$

where $f_{X_i}(x_i)$ for $i = 1, 2, 3$ denote the pdfs for X_1, X_2 and X_3, respectively. Similarly, $F_{X_i}(x_i)$ denote the cdfs for the three factors. $c()$ denotes the copula density. u_i are realizations from uniform distributed random variables, U_1, U_2, U_3. Having established this, we can now turn to specifying the marginal distribution, $f_{X_i}(x_i)$ and the copula density.

3.1 Marginal Models

We must specify models for the conditional marginal distributions. To do this we first introduce the information set available at time t, $\mathcal{F}_t$. Each of the models will have the following structure

$$x_{i,t} = \mu_i(\mathbf{Z}_{i,t-1}) + \sigma_i(\mathbf{Z}_{i,t-1})\varepsilon_{i,t} \quad i = 1, 2, 3 \quad \text{and} \quad \mathbf{Z}_{i,t-1} \in \mathcal{F}_{t-1} \tag{2}$$
$$\varepsilon_{i,t}|\mathcal{F}_{t-1} \sim F_{X_i}(0, 1) \quad \forall t.$$

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

conditional on $Z_{i,t-1}$ must be independent from the distribution of x_i conditional on $Z_{j,t-1}$ for $j \neq i$ in the sense that

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

3.2 NIG-GARCH

I this analysis we use the NIG-GARCH class of models by Jensen and Lunde (2001) because of the great flexibility it offers. The Normal Inverse Gaussian (NIG) distribution allows us to model both skewness and excess kurtosis and it nests popular distributions like the normal and t-distributions as special cases. We assume that conditional on the partition $Z_{i,t-1}$, $X_{i,t}$ follows a NIG distribution. That is

$$X_{i,t}|Z_{i,t-1} \sim \text{NIG}\left(\bar{\alpha}, \bar{\beta}, m_t(Z_{i,t-1}), \sigma_t(Z_{i,t-1})\right)$$

Introducing the notation $\bar{\gamma} = \sqrt{\bar{\alpha}^2 - \bar{\beta}^2}$ this implies that

$$E[X_{i,t}] = \sigma_t \frac{\bar{\beta}}{\bar{\gamma}} \quad \text{and} \quad V[X_{i,t}] = \sigma_t^2 \frac{\bar{\alpha}^2}{\bar{\gamma}^3},$$

where the dependence on $Z_{i,t-1}$ of m_t and σ_t has been dropped to shorten notation. In order to formulate the model according to the structure in (2) we respecify the model slightly. We standardize ε_t in the following way

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

Now $E[\varepsilon_t] = 0$ and $V[\varepsilon_t] = 1$. We can now specify the model as in (2).

$$x_{i,t} = m_t + \frac{\sqrt{\bar{\gamma}}\bar{\beta}}{\bar{\alpha}}\sigma_t + \sigma_t\varepsilon_t. \tag{3}$$

Note, that the conditional mean of $X_{i,t}$ is the same as before our respecification. From (3) we can now see, that

$$X_{i,t}|Z_{i,t-1} \sim \text{NIG}\left(\bar{\alpha}, \bar{\beta}, m_t, \sigma_t \frac{\bar{\gamma}^{3/2}}{\bar{\alpha}}\right)$$

which is very convenient because it implies that

$$E[X_{i,t}|Z_{i,t-1}] = m_t + \sigma_t \frac{\sqrt{\bar{\gamma}}\bar{\beta}}{\bar{\alpha}} \quad \text{and} \quad V[X_{i,t}|Z_{i,t-1}] = \sigma_t^2,$$

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

In this paper we consider models for $\mu_i(Z_{i,t-1})$ where we include lags of x_i and allow for lags of the other factors in m_t, we also allow for GARCH in mean effects. For the conditional variance we consider a standard GARCH model for σ_t.

3.3 Misspecification tests

It is important to test for possible misspecification in the marginal models. The decomposition in (1) is only valid if the marginal models are well specified.

We carry out two tests for the presence of autocorrelation as suggested in Patton (2012). Both tests consider potential autocorrelation in the estimated probability integral transforms, $\hat{u}_{i,t}$. These are constructed based on the estimated parameters, $\hat{\boldsymbol{\psi}}_i$ as

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

$\hat{u}_i$ is sorted in ascending order and the Kolmogorov-Smirnov (KS) and the Cramer-von Mises (CvM) test statistics are calculated

$$KS = \max_t \left| \hat{u}_{i,t} - \frac{t}{T} \right|$$

$$CvM = \sum_{t=1}^{T} \left(\hat{u}_{i,t} - \frac{t}{T} \right)^2$$

In the case of NIG distributed error terms we have to estimate the parameters of the distribution. This means that we can not rely on the asymptotic distribution of these test statistics in our analysis. Instead we follow Patton (2012) and use a simulation based method to calculate the p-values of the tests. We simulate a sample of x based on the estimated parameters. Next, we estimate the model on the simulated data and compute the values of KS and CvM. We repeat this S times in order to simulate the distributions of the two test statistics.

3.4 Copula

The next step is to specify the copula density. For this purpose we consider the normal copula. The copula density function is given as

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

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

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

3.5 Estimation

After specifying the marginal models and the copula density we turn to parameter estimation. We collect all parameters in $\boldsymbol{\theta}$ the log likelihood can now be written as

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

8

where ψ_i contains the parameters in marginal model i. κ contains the parameters from the copula density and $u = (u_1, u_2, u_3)'$.

θ can now be estimated using (conditional) maximum likelihood. The maximization is carried out in two steps. First step is maximizing the likelihood functions associated with the marginal model, $l_{x_i,t}(\psi_i; x_i)$. In the second step we use the estimates, $\hat{\psi}_i$, to construct $\hat{u}_{i,t} = F_{X_i}(x_{i,t}; \hat{\psi}_i)$. We do not have a closed form expression for the cdf of the NIG distribution, so the probabilities are found using simulation.

Statistical inference in this case is based on a bootstrap. The conventional methods for statistical inference in two step estimation do not apply her because the analysis is carried out on estimated factors.

4   Empirical Analysis

The first step in the analysis is to estimate the factors. We do this in the following way. At each point in time we regress the observed futures prices on the matrix Z for a fixed value of λ. The parameter estimates are estimates of the factors, $(L_t, S_t, C_t)'$ at time t. A grid search is carried out to find the value of λ which minimizes the sum of the error terms over time,

$$\sum_{t=1}^{T} e_t' e_t.$$

The sum of the error terms in minimized for $\lambda = 0.006$. The estimated factors are presented in figure 3.

We analyze the time series properties of the factors to determine how the our model should be specified. We reject stationarity for L_t and S_t but not for C_t. In the following we carry out the analysis on the first differences of L_t and S_t along with the level of C_t. We introduce the following notation

$$x_t = \begin{pmatrix} \Delta L_t \\ \Delta S_t \\ C_t \end{pmatrix} = \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix}$$

For the marginal models we have chosen the NIG-GARCH(1,1) model. In each model we have included enough lags to remove the auto correlation. For x_1 and x_2 we have included two lags while six lags were needed for x_3. The test statistics and p-values for the test for auto correlation are presented in 2.

The p-values of the Goodness of fit tests are based on 1000 simulations. For p-values under 0.05 we reject that we have specified the distribution correctly. This is essential for the copula decomposition. We have chosen the normal copula to capture the dependence between the factors. The estimated parameters are presented in table 2. At this stage the table does not include p-values for the estimated parameters.

In figure 5 we have presented the estimated residual densities along with a NIG reference. The NIG distribution offers a reasonable approximation for the residual density.

9

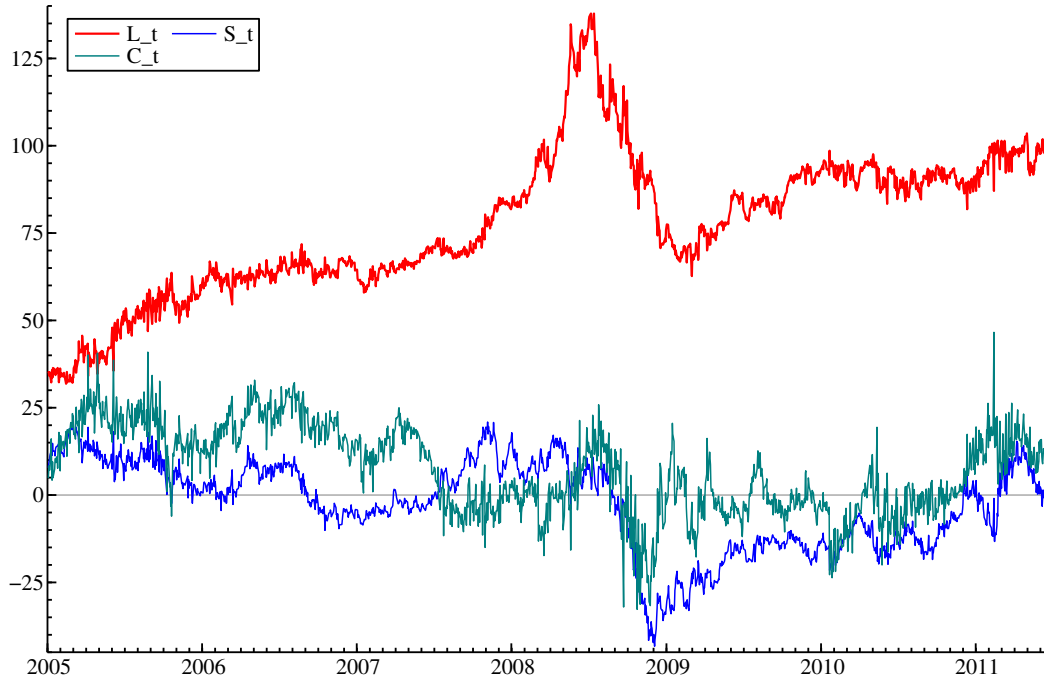|	$x_{1,t}$	$x_{2,t}$	$x_{3,t}$
	Conditional mean		
Constant	0.1802	0.094487	0.15246
	[ ]	[ ]	[ ]
$x_{i,t-1}$	-0.36331	-0.26337	0.45155
	[ ]	[ ]	[ ]
$x_{i,t-2}$	-0.11348	-0.091155	0.20709
	[ ]	[ ]	[ ]
$x_{i,t-3}$	-	-	0.097801
	-	-	[ ]
$x_{i,t-4}$	-	-	0.11072
	-	-	[ ]
$x_{i,t-5}$	-	-	0.030774
	-	-	[ ]
$x_{i,t-6}$	-	-	0.080040
	-	-	[ ]
	Conditional variance		
Uncond. variance ($\bar{\sigma}$)	5.0483	5.0757	23.560
	[ ]	[ ]	[ ]
ARCH term (α)	0.083791	0.057867	0.10711
	[ ]	[ ]	[ ]
GARCH term (β)	0.89895	0.93501	0.86245
	[ ]	[ ]	[ ]
	NIG distribution		
$\bar{\alpha}$	1.3737	2.4536	1.2873
	[ ]	[ ]	[ ]
$\bar{\beta}$	-0.067616	-0.070186	0.0072728
	[ ]	[ ]	[ ]
Log-Likelihood	-3406.14	-3381.72	-4678.82
	Misspecification tests (p-values)		
Breusch-Godfrey	16.861	13.123	12.722
	[0.15489]	[0.36004]	[0.38815]
Kolmogorov-Smirnov	0.016476	0.013097	0.016978
	[0.203]	[0.63]	[0.623]
CvM	0.035883	0.046861	0.053506
	[0.593]	[0.484]	[0.738]
	Copula		
	ρ_{12}	ρ_{13}	ρ_{23}
	-	-	-
	[ ]	[ ]	[ ]
Likelihood	-		

Table 2

Figure 3: The estimated factors

5 Forecasting

The forecasting exercise is quite simple. At every point in time we forecast the entire term structure of futures prices. We know which contracts may potentially be traded the next day, meaning that we know Z_{t+1} with certainty. For these contracts we construct directional forecasts. That is, we predict whether a contract will increase or decrease in the next day. The accuracy of the directional forecast is assessed in the following way. At every point in time, t and for each contract, j we construct the following variables.

$$d_{j,t} = sign\left(Y_t(\tau_j) - Y_{t-1}(\tau_{j+1})\right)$$
$$\hat{d}_{j,t} = sign\left(\hat{Y}_t(\tau_j) - Y_{t-1}(\tau_{j+1})\right).$$

Such that $d_{j,t}$ indicates the direction of the actual change and $\hat{d}_{j,t}$ indicates the direction predicted by our model. If the two are equal the forecast is successful and if they are not then the forecast is wrong.

The benchmark for forecasting is a simple autoregressive model fitted to the observed prices of the contract in question.

In this analysis we only consider one-step ahead forecasts. We consider the directional forecast and asses whether we are performing better for some maturities than for other. At every day in the out of sample period we perform the directional forecast. At

11
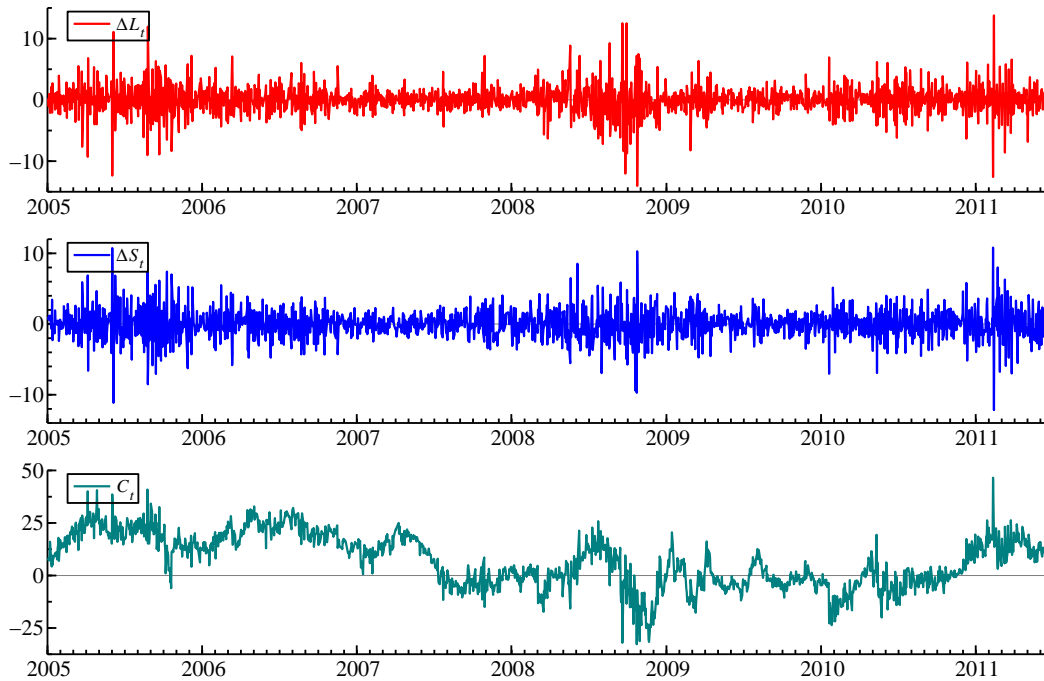
Figure 4: First differences of L_t and S_t along with level of C_t.

the end of the out of sample period we calculate the rate of success of our forecast for all possible maturities.

The forecasts are unfortunately not implemented at this moment.

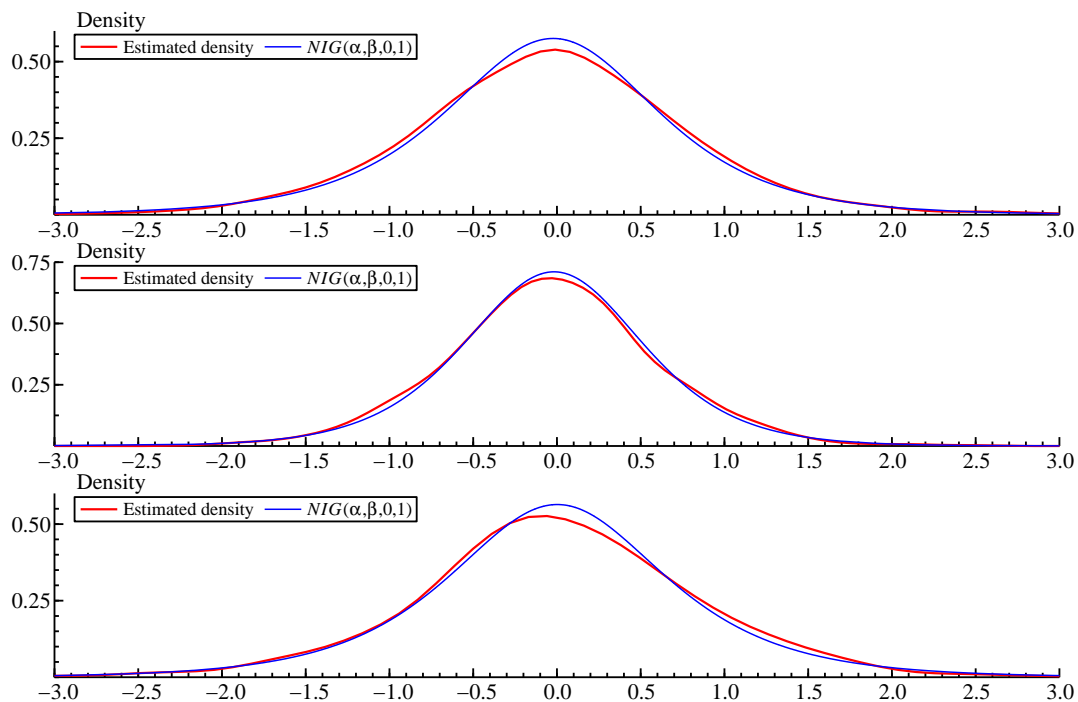# 6 Trading Strategy

To come...

7 Concluding Remarks

To come...

Figure 5

8 References

Diebold, F. X., Li, C., 2006. Forecasting the term structure of government bond yields. Journal of Econometrics, 337–364.

Jensen, M., Lunde, A., 2001. The nig-s & arch model: a fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. Econometrics Journal 4, 319–342.

Nelson, C. R., Siegel, A. F., 1987. Parsimonious modeling of yield curves. The Journal of Business 60 (4), 473–489.

Noureldin, D., 2011. Forecasting changes in the term structure of interest rates. Ph.D. thesis, Oxford University.

Patton, A., 2006. Modelling asymmetric exchange rate dependence. International Economic Review 47 (2).

Patton, A., 2012. Copula methods for forecasting. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting (Forthcoming).

Patton, A. J., 2009. Copula-based models for financial time series. In: Andersen, T. G., Davis, R. A., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer-Verlag, pp. 767–785.