Common Factors in Commodity Futures Curves*

Dennis Karstanje, Michel van der Wel, and Dick van Dijk†

December 8, 2017

Abstract

Literature studying comovement in commodity prices provides mixed evidence for whether commodity markets are segmented or driven by common factors. We provide a joint framework to study comovement across 24 of the most traded commodities over 20 years. The framework benefits from using the whole commodity futures curve, not just the nearest futures contract. Using all curve data we find that there is a strong comovement, though this is mostly due to sector commonalities. Furthermore, the importance of a common market factor is overestimated when using only the nearest futures data. Both these findings shed light on the contradicting conclusions in existing literature. Additionally we find that the importance of the common factors varies over time, with a stronger factor structure arising in recent years.

JEL classification: G12; G13.
Keywords: Commodity futures prices; comovement; term structure; dynamic Nelson Siegel model.

*Karstanje worked on this paper while being employed by Erasmus University Rotterdam and later Robeco Asset Management, Rotterdam. Van der Wel is from Erasmus University Rotterdam, CREATES, Tinbergen Institute and ERIM. Van Dijk is from Erasmus University Rotterdam, Tinbergen Institute and ERIM. The authors thank participants at the 51st Meeting of the EWGCFM, the MMF workshop on "Empirical Modelling of Financial Markets", the Workshop on "Factor Analysis in Economics and Finance" at the Canadian Economics Association Conference 2013, the SoFiE Annual Conference 2013, the SIRE Conference on "Finance and Commodities" 2013, the 19th International Conference on Computing in Economics and Finance, the SoFiE Large-Scale Factor Models in Finance Conference, the 2014 FMA European Conference, the CAMP workshop on Commodity Dynamics and Financialisation, the 2016 INFINITI conference on International Finance, and seminar participants at CREATES, IESEG Lille, Bristol University, and Warwick Business School. We are grateful to Marcin Jaskowski, John Hunter, Christian Stepanek and Ana-Maria Fuertes for their detailed comments. Michel van der Wel is grateful to Netherlands Organisation for Scientific Research (NWO) for a Veni grant; and acknowledges support from CREATES, funded by the Danish National Research Foundation. The computational tests of this research were performed on the Dutch National LISA cluster, and supported by NWO. We are responsible for all errors.

†Corresponding author. Tel.: +31 10 4081263. E-mail: djvandijk@ese.eur.nl. Other author's e-mail addresses: dennis@karstanje.com (Karstanje), vanderwel@ese.eur.nl (van der Wel). The views expressed in this paper are solely those of the authors and not necessarily shared by Robeco or its subsidiaries.

1 Introduction

Joint commodity price swings as witnessed during 2008-2010 raise the question whether commodities should be considered a heterogeneous or homogeneous asset class. Large fluctuations in commodity prices can have significant economic and political implications, which makes the question relevant not only to participants in commodity markets but also to governments and policy makers. Commonality across commodity prices can stem from various sources. Commodities can be economically related through production, substitution or complementary relationships (Casassus et al., 2013). Prices can also be intertwined due to links with the global economy (Aastveit et al., 2015; Kilian, 2009) or due to exchange rates (Chen et al., 2010). When investigating commodity price commonality and the driving forces behind their price movements, it is important to distinguish between commonality across all commodities (market-wide effects) and commonality within sectors (like energy, metals, grains, etc). Both types of commonality have different implications for market structure and investment decisions. We propose a novel approach that enables to make this distinction and furthermore benefits from using the whole commodity futures curve instead of only a single point on the curve.

The degree of commonality in commodity (futures) prices is subject to an ongoing debate. In a seminal paper, Pindyck and Rotemberg (1990) report that there is excess comovement among commodity prices after accounting for macroeconomic and market conditions. Other papers find weaker evidence or reject the excess comovement hypothesis when overcoming econometric issues or considering additional commodity-specific information (Deb et al., 1996; Ai et al., 2006). While these papers disagree whether it should be labeled "excess" comovement or not, they do agree on the point that commodity prices display commonality.[1] In contrast, recent asset pricing literature is not even unanimous on the existence of commonality as such. While Yang (2013), Szymanowska et al. (2014) and Bakshi et al. (2017) argue that there are common factors that can explain a large

[1] Deb et al. (1996) point out that, in general, a statement about "excess" is hard to investigate as it "is a relative concept which presupposes an underlying benchmark form model that defines what zero excess is".

part of the cross-sectional variation in returns, Daskalaki et al. (2014) and Christoffersen et al. (2017) find opposite results of no or a weak factor structure in returns.

We contribute to this debate by distinguishing between market-wide and sector-specific commonality and by taking the entire curve of commodity futures prices into account. The first aspect extends current literature, which typically enforces a factor structure that cannot discriminate between market-wide and sector-specific commonality. If the comovement is not equally strong across different commodities, a single factor extracted from all commodities may merely reflect sector-specific, rather than market-wide common dynamics. Disentangling sector-level comovement from market-wide comovement is not common in the commodities literature, with the exception of Delle Chiaie et al. (2017) and Ma et al. (2015). However, both do not make use of the entire futures curve. The second aspect, the use of full curve data, also extends existing research on common factors in commodity prices which typically uses only part of the available information. Baffes (2007), Delle Chiaie et al. (2017), Charlot et al. (2016) and West and Wong (2014) base their commonality findings on spot prices, while all commodities have an entire range of futures contracts available. Even though futures prices of the same commodity are related, we show it is important to include the full futures curve in commonality analyses. Intuitively, long-term contracts provide information about the equilibrium price level while the shape of the futures curve gives insight in short-term movements (Schwartz and Smith, 2000). Another reason to include more distant contracts is that over time these have become more traded and hence more liquid. Commodity trading strategies and index providers make more and more use of distant contracts (de Groot et al., 2014; Miffre, 2012).

To be able to study the price information of many commodity futures we introduce a novel factor approach. Diebold and Li (2006) reintroduce the Nelson and Siegel (1987) model to provide a parsimonious description of interest rates in terms of level, slope and curvature. We apply the model in the context of the term structure of commodity futures prices, due to the similarity with the term structure of interest rates. Grønborg and Lunde (2016) and Heidorn et al. (2015) also apply the Nelson and Siegel (1987) model to commodity futures curves. We develop a Nelson-Siegel type of model for commonality analysis. Diebold et al. (2008) include an additional model layer consisting of unobserved common components for the level, slope and curvature factors across yield curves. We adopt their approach and extend it to fit commodity futures data. We make use of two types of common components, i.e. a market-wide component that is common to all commodities, and sector-level components that are only common to commodities within the same sector. The nature of commodities motivates the inclusion of the sector-specific components. Besides the common components there are idiosyncratic components to allow for commodity-specific behavior. A related set-up is used in the business cycle literature, see, e.g., Kose et al. (2003) who use a decomposition into global, regional, and idiosyncratic components. Our modeling framework allows for studying the relative importance of the sector versus market-wide common components. We define both as commonality but a strong sector commonality points in the direction of commodities being a segmented asset class.

We apply our framework to the 24 most actively traded commodities over the sample period 1995-2012. We find that there are common factors in these commodity futures curves. The commonality is mostly driven by sector components (between 39% and 50% of the total price variation, compared to about 24% for the market-wide component), which implies that commodities are not a homogeneous asset class. We also study which (macroeconomic) variables, if any, drive the latent factors, and find that the market-wide level factor is related to hedging pressure, equity prices and exchange rates. When the net short positions of hedgers increase, the curves' price levels also increase (in line with Gorton et al., 2013). A rise in equity prices or a weakening of the US dollar also leads to an increase in the common level factor, conform, among others, Vansteenkiste (2009), Byrne et al. (2013), and Chen et al. (2010). Regarding the market-wide slope component, we find that it also increases when the net short positions of hedgers increase. This results in a more backwardated futures curve, which is consistent with the theory of normal backwardation (see, e.g., Bessembinder, 1992).[2] Furthermore, the slope of the futures

[2]Or, e.g., Carter et al. (1983); Chang (1985); de Roon et al. (2000).

curves is negatively related to commodity inventories. This relation between the slope of the futures curve (i.e. convenience yield) and commodity inventories is consistent with the theory of storage and findings of Gorton et al. (2013) and Deaton and Laroque (1992).

To investigate the impact of including the full curve of futures data, we re-estimate our model based on only the first nearby contract. We find a stronger factor structure where commonality is again mostly driven by sector-specific components. The increase in importance of common factors is however fully due to the market-wide component, which explains 33% more variation at the cost of idiosyncratic components. Hence, using only part of the available curve information does distort the commonality results. Finally, we investigate time variation in the degree of commonality. We find that only the commodity price level commonality varies over time. The market-wide level component becomes increasingly important, at the cost of idiosyncratic components. This increase in commonality is in line with the findings of Tang and Xiong (2012) that the financialization of the commodity markets increased correlations across commodities.

An alternative approach to model the commodity futures curve is to express it in terms of unobserved factors and derive futures prices under no-arbitrage conditions (see Gibson and Schwartz, 1990; Schwartz, 1997; Geman, 2005). This approach is more restrictive because it requires many assumptions on the market factors and it has difficulty incorporating seasonal patterns (West, 2012). Furthermore, the model complexity makes it difficult to jointly model multiple commodities. Cortazar et al. (2008), Ohana (2010) and Casassus et al. (2013) propose joint models for two commodities but these cannot easily be extended to larger dimensions.[3] As both approaches extract unobserved factors from the futures curve, it is possible that these factors are alike. Indeed, we show that for all individual commodities, our level and slope factors are similar to the corresponding unobserved spot price and convenience yield factors of the Schwartz (1997) three-factor model, while our curvature factor explains part of his residuals.

[3]Cortazar et al. (2008) propose a multi-commodity version of the Cortazar and Naranjo (2006) model for two oil related commodities that uses common and commodity-specific factors. Ohana (2010) captures the joint evolution of correlated futures curves by incorporating both the local and global dependence structures between slopes and levels. Casassus et al. (2013) jointly model the convenience yields of two commodities using a multi-commodity feedback affine model to match observed futures correlations.

The rest of this paper is organized as follows. In the next section we develop the methodology, followed by Section 3 where we present the data. Section 4 discusses the estimation results, while Section 5 provides detailed interpretations of these results. Section 6 concludes.

2 Methodology

In this section we develop our approach for decomposing the individual commodity futures curves into level, slope and curvature factors, and describe how commonality in these factors across commodities is modeled. The first subsection discusses our model. The second subsection discusses the estimation procedure.

2.1 Model

We model a collection of monthly futures prices for N different commodities. The futures price with maturity τ (measured in years) for commodity i at time t is denoted by $f_{i,t}(\tau)$. Our starting point is the dynamic version of the Nelson and Siegel (1987) model, as introduced in Diebold and Li (2006), to describe the futures curve of each individual commodity i, for $i = 1, 2, \ldots, N$, as

$$f_{i,t}(\tau) = l_{i,t} + s_{i,t}\left(\frac{1-\exp^{-\lambda_i \tau}}{\lambda_i \tau}\right) + c_{i,t}\left(\frac{1-\exp^{-\lambda_i \tau}}{\lambda_i \tau} - \exp^{-\lambda_i \tau}\right) + \nu_{i,t}(\tau), \qquad (1)$$

where $l_{i,t}$, $s_{i,t}$, $c_{i,t}$ are interpreted as time-varying unobserved factors, the decay parameter $\lambda_i > 0$ is assumed to be constant over time, and $\nu_{i,t}(\tau)$ is a disturbance term, where Σ_ν is the covariance matrix of $(\nu_{i,t}(\tau_1), \nu_{i,t}(\tau_2), \ldots, \nu_{i,t}(\tau_{J_i}))'$, with J_i the number of maturities for commodity i. The interpretation of the unobserved factors $l_{i,t}$, $s_{i,t}$, and $c_{i,t}$ is determined by their loadings. The loading on the first factor is a constant such that $l_{i,t}$ affects all futures prices in the same way irrespective of their maturity, hence it corresponds with the level factor. The loading on the second factor is a decreasing function of the contract maturity τ and $s_{i,t}$ can therefore be regarded as the slope of the futures curve. The loading on the third factor is a concave function of τ, which allows to fit humped-shaped futures curves, and $c_{i,t}$ can thus be interpreted as a curvature factor.[4]

We make two adjustments to the base model given in (1). First, we enhance the model to account for seasonality. Commodity futures curves display pronounced seasonal patterns due to seasonal supply and demand effects, for example related to crop cycles and weather conditions (see, e.g., Milonas, 1991). We account for seasonality by including a trigonometric function that depends on the expiry month $g_i(t,\tau)$ of the contract.[5] Trigonometric functions are often used to model seasonality, see for instance Sorensen (2002). We thus enhance the model in (1) with a seasonal term given by $\kappa_i \cos(\omega g_i(t,\tau) - \omega\theta_i)$, where the parameter κ_i determines the commodity-specific exposure to the seasonal, the constant ω determines the cycle length (fixed at a year, so $\omega = 2\pi/12$), and the parameter θ_i indicates the peak of the seasonal term (in months).

Second, for small values of λ_i, the loadings of the slope factor only slowly decline towards zero as the maturity increases. While this helps to fit futures curves that are very smooth, it becomes difficult to identify the level and slope factors when the term-structure dimension is limited. By re-centering the loadings of the slope and curvature factors to zero at the one year maturity, we make sure that the slope factor does not absorb movements of the curve's level. In this case, the level factor represents the price level of the one year futures contract, while the slope and curvature factor have a similar interpretation as before. With these two adjustments, the specification of the futures

[4] The loading for $s_{i,t}$ starts at 1 for $\tau \to 0$ and monotonically declines towards zero as the maturity τ increases. The loading for $c_{i,t}$ is equal to 0 for $\tau \to 0$ and $\tau \to \infty$, and reaches a maximum value for maturity τ^*, which depends on the value of λ_i.

[5] The mathematical expression for the expiration month is $g_i(t,\tau) = t + 12\tau - S\left\lfloor \frac{t+12\tau}{S} \right\rfloor$, with $S = 12$ the number of distinct "seasons", 12τ to scale the maturity in years to maturity in months, and the function $\lfloor x \rfloor$ returns the largest integer not greater than x. Therefore, $g_i(t,\tau)$ results in the integers $\{0, 1, \ldots, 11\}$. Our sample starts in January with $t = 0$ such that the integers $\{0, 1, \ldots, 11\}$ represent the expiry months January, February, ..., December.

price curve becomes

$$
\begin{aligned}
f_{i,t}(\tau) = & l_{i,t} + s_{i,t}\left[\left(\frac{1-\exp^{-\lambda_i\tau}}{\lambda_i\tau}\right) - \left(\frac{1-\exp^{-\lambda_i}}{\lambda_i}\right)\right] \\
& + c_{i,t}\left[\left(\frac{1-\exp^{-\lambda_i\tau}}{\lambda_i\tau} - \exp^{-\lambda_i\tau}\right) - \left(\frac{1-\exp^{-\lambda_i}}{\lambda_i} - \exp^{-\lambda_i}\right)\right] \\
& + \kappa_i \cos\left(\omega g_i(t,\tau) - \omega\theta_i\right) + \nu_{i,t}(\tau)\,. \qquad (2)
\end{aligned}
$$

The constants we subtract from the slope and curvature loadings are commodity specific (due to λ_i) but fixed over time. The constants only have an effect when λ_i becomes small as for large values of λ_i they are close to zero.

We investigate the comovement of commodity prices by linking the futures curves of the individual commodities. This link across commodities is accomplished by decomposing the level, slope, and curvature factors into latent market-wide, sector, and idiosyncratic components. We define the factor decompositions

$$
\begin{aligned}
l_{i,t} &= \alpha_i^L + \beta_i^L L_{market,t} + \gamma_i^L L_{sector,t} + L_{i,t}, \\
s_{i,t} &= \alpha_i^S + \beta_i^S S_{market,t} + \gamma_i^S S_{sector,t} + S_{i,t}, \qquad (3) \\
c_{i,t} &= \alpha_i^C + \beta_i^C C_{market,t} + \gamma_i^C C_{sector,t} + C_{i,t},
\end{aligned}
$$

where α_i^X for $X = \{L, S, C\}$ are constant terms, β_i^X are loadings on the latent market-wide component $X_{market,t}$, γ_i^X are loadings on sector components $X_{sector,t}$, and $X_{i,t}$ are idiosyncratic components. The (absolute) magnitude of the coefficients β_i^X and γ_i^X determines the degree of comovement across all commodities and across commodities within a specific sector, respectively.

We include sector components besides the market-wide components, because we may expect that commodities in the same sector are more closely related than commodities across different sectors. The market-wide, sector, and idiosyncratic components are assumed to have first-order autoregressive dynamics

$$\begin{pmatrix} \Delta L_{y,t} \\ S_{y,t} \\ C_{y,t} \end{pmatrix} = \begin{pmatrix} \phi^y_{11} & \phi^y_{12} & \phi^y_{13} \\ \phi^y_{21} & \phi^y_{22} & \phi^y_{23} \\ \phi^y_{31} & \phi^y_{32} & \phi^y_{33} \end{pmatrix} \begin{pmatrix} \Delta L_{y,t-1} \\ S_{y,t-1} \\ C_{y,t-1} \end{pmatrix} + \begin{pmatrix} \eta^L_{y,t} \\ \eta^S_{y,t} \\ \eta^C_{y,t} \end{pmatrix}, \tag{4}$$

where $y = \{market,\ sector,\ i\}$, and the disturbances $\eta_{y,t} = (\eta^L_{y,t}, \eta^S_{y,t}, \eta^C_{y,t})$ are normally distributed with covariance matrix Σ_{η_y}. The choice for VAR(1) dynamics for the components is in line with Diebold and Li (2006) and Diebold et al. (2008). One difference is that we assume the level factors to be non-stationary and model them as first differences. This assumption is in line with the theory of financial market efficiency that argues that prices are unpredictable (Fama, 1970). In contrast, the theory of storage argues that e.g. agricultural prices should be stationary (Deaton and Laroque, 1992). In our view the assumption is mild and of limited impact. We make it from a technical point of view to facilitate the estimation process and be able to make valid inference later on. Duffee (2011) makes a similar assumption for yield data, while Grønborg and Lunde (2016) do the same for oil commodity futures.

To facilitate tractable estimation in our applications and to let the covariation, if any, come from common factors, we do not allow for cross-correlation of the shocks $\nu_{i,t}(\tau)$ across maturities and commodities, and for the shocks $\eta_{y,t}$ across market-wide, sector, and idiosyncratic components. In other words, all the covariance disturbance matrices Σ_ν and Σ_{η_y} are diagonal. Also the autoregressive matrices in (4) are assumed to be diagonal.

As often with factor models, we need to make sure that our unobserved factors are uniquely identified. Here, we have two identification issues as neither the signs nor the scales of the market-wide and sector factors and their factor loadings are separately identified. We follow Sargent and Sims (1977) and Stock and Watson (1989) to identify the scales by assuming that each disturbance variance is equal to a constant, i.e. $\Sigma_{\eta_z}(j,j) = 0.01$ for $z = \{market,\ sector\}$ and $j = 1, 2, 3$.[6] We identify the factor signs by

[6]The value of 0.01 is in line with the estimated variance of the idiosyncratic factor' disturbances, as will become clear in the results section.

restricting one of the loadings for each of the market-wide and sector components to be positive.

2.2 Estimation

The model as given by (2)-(4) can be estimated either using a two-step approach (see, e.g., Diebold et al., 2008) or a one-step approach (see, e.g., Diebold et al., 2006). In the two-step approach one first extracts the latent factors $l_{i,t}$, $s_{i,t}$, and $c_{i,t}$ in (2) at each point in time for each commodity and in the second step decomposes the extracted factors into the market-wide, sector and idiosyncratic components. In the one-step approach one casts the complete model in a state space representation to allow for estimation of all parameters as well as the latent factors and their decomposition simultaneously, by means of the Kalman filter. We use the one-step approach as it has the advantage that it takes the estimation uncertainty in the extracted factors into account in their decomposition. Furthermore, it allows for using both time series and cross-sectional dimensions to accurately estimate the parameters. To initialize the one-step approach we use estimation results of smaller versions of our full model, e.g. a variant without market-wide or sector components.

The state space representation follows naturally from the model given by (2)-(4). The measurement equation (A.1) is given in Appendix A and based on (2) and (3). The individual latent level $l_{i,t}$, slope $s_{i,t}$, and curvature $c_{i,t}$ factors do not appear in the measurement equation, as we can link the observed futures prices $f_{i,t}(\tau)$ directly to the unobserved market-wide, sector and idiosyncratic components. The transition equations of the latent states are given by (4).[7] We collect all unknown coefficients in the parameter vector Ψ, i.e. the commodity-specific parameters in (2) and (3) $\left(\lambda_i, \kappa_i, \theta_i, \alpha_i^X, \beta_i^X, \gamma_i^X\right)$, and the diagonal elements of the VAR coefficient matrices Φ^y, and the variance matrices Σ_ν

[7] In our implementation, we consider a stacked form where the multivariate series is treated as univariate series, following Koopman and Durbin (2000). This is possible since we assume that there is no remaining cross-correlation across different commodities after accounting for the common factors. The univariate treatment gives not only computational gains but also allows the number of term-structure observations J_i to be time-varying and deals with an unbalanced panel. Note that for notational convenience we suppress the additional t subscript and simply maintain the notation of J_i, while in our implementation this does vary over time.

and Σ_{η_y}. Estimation of Ψ is based on the numerical maximization of the log-likelihood function that is constructed via the prediction error decomposition.

3 Data

We study futures curves for 24 commodities. We consider the period January 1995 to September 2012 and use all individual futures contracts that expire between January 1995 and December 2030.[8] Our commodity selection is based on the composition of the S&P Goldman Sachs Commodity Index (GSCI). The GSCI constituents can be split in five sectors: energy, metals, softs, grains, and meats. An overview of the data (as obtained from Thomson Reuters Datastream) is given in Table 1. All our analyses are done at the monthly frequency, and for this we use month-end log settlement prices. All prices are standardized since the pricing grid of the commodities is quite diverse.[9] To avoid liquidity issues we do not consider price information of (i) contracts with a monthly return that equals zero, and (ii) contracts in the expiration month. Furthermore, we filter out data errors by excluding contracts that have abnormal returns compared to adjacent contracts. These filters lead to exclusion of approximately 1.3% of the data.[10]

[insert Table 1]

Table 1 shows that the number of available contracts (term-structure observations) varies per commodity. This is caused by differences in (i) the number of distinct expiration months a year or (ii) the maximum time-to-maturity. Energy and industrial metal commodities have an expiring futures contract each month and also contracts with long dated maturities. In contrast, agricultural commodities have a small number of active futures because they have only five to eight distinct expiry months a year. Furthermore, the maximum maturity of these contracts is between one and two years, which is in line

[8] The start of the sample period is based on the availability of the metal commodities traded on LME.

[9] We standardize prices by setting the first nearby contract price in January 1995 of all commodities equal to 100. All other prices are adjusted such that (time series and term structure) returns remain unchanged.

[10] This percentage is a combination of expiring contracts (omits 0.8%), data errors identified due to abnormal returns (0.14%), and monthly returns equal to zero (0.4%).

with the length of the crop cycles and storability of these commodities. The variation in the number of contracts and the maximum time to maturity indicate that it is important to use a commodity-specific decay parameter λ_i in (2).

Besides differences in the number of available contracts across commodities, we notice the same within commodities over time. Especially the number of contracts for energy and industrial metal commodities greatly increases over our 17 years sample. Even though our estimation methods can deal with this increase of available contracts over time, our choice to use a fixed λ parameter limits the model flexibility. Instead of allowing λ to vary over time (see, e.g., Koopman, Mallee & Van der Wel, 2010), we choose to introduce a (commodity-specific) maturity bound to exclude long-dated contracts. By limiting the term-structure dimension variation within a commodity, we can keep using a fixed λ value. Furthermore, these long-dated contracts are possibly less liquid and hence have more noisy price information, which could otherwise affect our results. The introduced maturity bound excludes on average 10% of our observations.[11]

The summary statistics in Table 1 show that there are large return differences both across commodities and along their futures curves. The average annualized returns of the contracts range between -20.6% to 20.5% and are more extreme for the first nearby contract. The volatility of the returns confirms this as in almost all cases the 12 month contract returns are less volatile than the returns of the first nearby contract, also known as the Samuelson (1965) effect.

[insert Figure 1]

Figure 1 gives some insight in the data by showing the complete set of available futures prices for natural gas and coffee. For both commodities we observe that the shape of the futures curve varies substantially over time, with alternating periods of pronounced contango and backwardation especially for natural gas.[12] Both futures curves also clearly show the large increase in the general price level during the period 2005-2008. A notable

[11] Appendix B provides additional details on the introduced maturity bound.

[12] An upward sloping commodity futures curve is said to be in contango, while a downward sloping curve is in backwardation.

difference between these commodities is that the futures curve of natural gas displays a strong periodic pattern with spikes occurring for expiry months during the winter, while the curve of coffee does not show any signs of seasonality. Finally, Figure 1 illustrates that the number of available contracts varies over time. For both natural gas and coffee (and in fact also for most other commodities), contracts with longer maturities only have become available in the most recent years of our sample period.

4 Estimation results

We start by analyzing the individual commodities separately to decide on the exact model specification to be used for each commodity. Thereafter we incorporate the insights of the individual commodities' analysis to estimate the joint commodity model.

4.1 Individual commodities

In Section 2, we presented the model in general form, where all commodity curves are built up from a level, slope, and curvature factor combined with a seasonal term. However, not all commodity curves may show dynamics for which the flexibility of three factors is needed. Furthermore, not necessarily all commodities display periodic behavior. We therefore apply our Nelson-Siegel set-up to each individual commodity, i.e. we leave out the market-wide and sector components in (3) and based on the features of each specific commodity we decide on the number of factors to include and whether to include the seasonal term or not. We do this on an individual commodity basis as the choice for the number of factors is not likely to depend on possible cross-commodity relations.

Table 2 presents the final model choice and the estimated parameter values for each of the commodities. Column 3 shows the final model choice, which is based on several criteria.[13] First, we compare the results of three-factor models with and without seasonal

[13]Appendices C and D provide more detailed results that complement this section. Specifically, in Appendix C we report detailed results for individual models and compare our commonality results using the Nelson-Siegel set-up to results based on Principal Component Analysis, and in Appendix D we provide evidence for the appropriateness of the Nelson-Siegel model for fitting commodity curves.

term, to see if the exposure κ to the seasonal correction is significantly different from zero. The 13 commodities with the label "3FS" have highly significant κ parameters, ranging between 0.73 and 7.78. Most of these commodities have a clear crop cycle, i.e. a seasonal supply, while others, like natural gas, have well-known seasonal demand. We include three factors for all commodities with periodic behavior because even after the seasonal correction their curves display a large variety of shapes. Second, we decide whether or not to include a curvature factor for the remaining 11 commodities. Based on Akaike Information Criterion and Bayesian Information Criterion values, we should choose to include all three factors. However, low λ values for metal commodities lead to slope and curvature loadings that are highly correlated, which leads to identification problems. Hence, we decide to exclude the curvature factor for the metal commodities. The fourth column in Table 2 shows the in-sample fit of our models. With the exception of lean hogs all R^2 values are above 99.7%, which supports the use of the Nelson-Siegel framework for commodity futures prices.

[insert Table 2]

The remaining columns of Table 2 show the parameter estimates with corresponding standard errors in parentheses. The decay parameter λ determines the shape of the slope and curvature loadings. Large values of λ lead to quickly declining slope loadings and move the peak of the curvature loadings to the short-end of the curve. When $\lambda = 2$, the slope loading starts at 0.49, it declines to zero for $\tau = 1$ (by construction), and is equal to -0.19 and -0.33, for $\tau = 2$ and $\tau = 5$ years, respectively. The curvature loading peaks at 11 months maturity. By contrast, when lambda is equal to 0.01 the slope loading starts at 0.005, it gradually declines towards zero for $\tau = 1$, and is equal to -0.005 and -0.020, for $\tau = 2$ and $\tau = 5$ years, respectively. The curvature loadings are almost a mirror image of the slope loadings, which is why we exclude the third factor for the metal commodities. The maximum curvature loading (when lambda is equal to 0.01) is achieved for $\tau = 179.33$, i.e. 179 years and 4 months, which is way beyond the highest maturity that is included in our sample. The decay parameter λ varies substantially across commodities, ranging from 0.011 for gold to 5.2 for feeder cattle. A value of 0.01 is the lowest value we allow to prevent that the loadings of the level and slope factor become too similar. The variation in λ is both due to differences in curve shapes and maximum contract maturity. The effect of different futures' curves shapes on λ becomes clear when we compare results for commodities with a similar maximum contract maturity. Gasoil and soybeans both have futures up to 2.5 years until maturity, while their λ values differ greatly (1.688 versus 3.350, respectively). The seasonal correction terms are captured by the exposure κ and location θ. The θ estimates imply that gasoil, heating oil and natural gas contracts are most expensive between January and February, while most agricultural commodities are more expensive two months before their harvest. The last column shows the volatility of the errors. The volatilities of the errors are small, especially compared to the highly volatile observed futures prices.

4.2 Joint model for commodity curves

We return to the market-wide state space model given by (2)-(4) and incorporate the findings of the analysis discussed above. Specifically, when we estimate the full model, we use the choice of the number of factors and we fix the λ, κ, and θ parameters to their estimates based on commodity-specific data, as reported in Table 2. This reduces the computational burden as parts of the measurement equation are now constant. All other parameters are estimated using the Kalman filter and maximum likelihood.

The commonality across commodities is expressed by their loadings on market-wide and sector components, see also (3). Table 3 shows the estimated constant α, the loading on the market-wide component β, and the loading on the sector component γ for each commodity level, slope, and curvature factor. The α-parameters make sure that the idiosyncratic components have mean zero. All α level estimates are between 4 and 5, due to our standardization procedure (as $\log(100) \sim 4.61$). The α estimates corresponding to slope and curvature are almost all not significantly different from zero. Two noteworthy exceptions are gold and cocoa. The negative α parameter for the slope on gold is in line with expectations because its futures curve is often in contango.

All loadings on the market-wide level component (β's) are positive (or not significantly different from zero, in case of the negative coefficient of feeder cattle), which indicates that there exists a link between the levels of different commodity prices. The loadings on the market-wide slope component are positive for energy commodities, negative for most of the metal commodities, and close to zero for all other commodities. The loadings on the market-wide curvature component show substantial variation, ranging from -1.93 to 2.78. Most of them are not significantly different from zero. In general, sector loadings have the same sign within the corresponding sector, which implies commonality. The few slope and curvature loadings that have opposite signs are not significant. In line with previous results, all loadings on common components point in the direction of commonality.

The rightmost column of Table 3 shows the variance estimates of the measurement equation errors. We assume that these variances are commodity specific but within each commodity they are the same for all different contract maturities. We believe this assumption is appropriate because the factor structure can already account for volatility differences along the term-structure dimension due to the maturity dependent factor loadings. Almost all estimated variances are well below the variances of the factor disturbances.

[insert Table 3]

4.3 Factor dynamics

Before we answer the questions posed in the introduction, we turn to the dynamics of the factors extracted from the model. Table 4 shows the estimates related to the state equations in (4). The first three columns show the autoregressive coefficients. Focusing on the market-wide and sector components, all ϕ estimates are positive (except for one insignificant grains sector parameter). The slope factors are more persistent than the curvature components. The parameter estimates for the idiosyncratic factors are in line with the corresponding common factors. Most of the level ϕ coefficients are not significantly different from zero, which implies that levels evolve as pure random walk processes. The AR parameters of the slope factors range between 0.61 and 0.99. The only exception is heating oil with an AR(1) parameter of -0.33. The curvature factors show similar coefficients as those of the slope factors but are slightly less persistent.[14]

[insert Table 4]

[insert Figures 2, 3 and 4]

Figures 2, 3, and 4 show the unobserved level, slope, and curvature factors and their different components. In Subfigure A of each figure we show the market-wide component together with the five sector components. In Subfigures B-F we show the sector components together with the corresponding idiosyncratic commodity components. The market-wide level factor shows an increase in 2006-2007 and 2011, in line with patterns observed in all major commodity indices. The energy and metal sectors show similar behavior although the factors peak at different points in time, in line with subindices like GSCI Energy, Precious Metals and Industrial Metals. The other subfigures show that the scale of the idiosyncratic level factors varies across commodities but is in line with the size of the estimated disturbance variances. The slope factors in Figure 3 show quite some variability. They are both negative and positive, which indicates that the futures curves interchange between contango and backwardation. The scale of the metals' slope factors is large compared to the other commodities due to the λ-dependent loading correction we have introduced in (2). Last, the curvature factors in Figure 4 show more mean-reversion compared to the slope factors.

5 Interpretation

The parameter estimates and factor dynamics are only part of the story. This section interprets the estimation results and answers the questions we posed in the introduction. Most importantly, in Section 5.1 we discuss our findings regarding commonality. In

[14]The variances of the factor disturbances are in the last three columns in Table 4, but harder to interpret. The variances of the common factor disturbances are fixed for identification purposes. The estimates of the idiosyncratic disturbance variance cannot be compared across commodities or factors because the magnitude of the factor loadings is not the same. This can be seen in Appendix A Equation (A.1) where all idiosyncratic states are premultiplied by the matrix A, which contains the commodity-specific factor loadings. For metal commodities the λ_i parameter is small, which results in close to zero loadings on the slope factor for most maturities and hence seemingly large error variances.

the subsequent sections we turn to the importance of using data of the full curve, time variation in commonality, and the economic interpretation of the unobserved factors.

5.1 Importance of common factors

The often significant factor loading estimates reported in Section 4.2 already hint at commonality across commodities. The model however contains a large number of parameters, so it is tough to draw firm conclusions. To study the existence of commonality and its sources we decompose the variation in commodity level, slope, and curvature factors into parts driven by the market-wide, sector, and idiosyncratic components. As mentioned in Kose et al. (2003) and Diebold et al. (2008), the market-wide, sector, and commodity-specific components may be correlated as they are extracted from a finite sample. Hence we orthogonalize the extracted components using a Cholesky decomposition to ensure that they add up.[15] Then, we can use (3) to write

$$\begin{aligned} \text{var}\left(\Delta l_{i,t}\right) &= \left(\beta_i^L\right)^2 \text{var}\left(\Delta L_{market-wide,t}\right) + \left(\gamma_i^L\right)^2 \text{var}\left(\Delta L_{sector,t}\right) + \text{var}\left(\Delta L_{i,t}\right), \\ \text{var}\left(s_{i,t}\right) &= \left(\beta_i^S\right)^2 \text{var}\left(S_{market-wide,t}\right) + \left(\gamma_i^S\right)^2 \text{var}\left(S_{sector,t}\right) + \text{var}\left(S_{i,t}\right), \\ \text{var}\left(c_{i,t}\right) &= \left(\beta_i^C\right)^2 \text{var}\left(C_{market-wide,t}\right) + \left(\gamma_i^C\right)^2 \text{var}\left(C_{sector,t}\right) + \text{var}\left(C_{i,t}\right). \end{aligned} \tag{5}$$

For the level factors we decompose the variances of the first differenced series as the variance of a non-stationary series is undefined.

The fractions of explained variance per component are shown in Table 5. The market-wide level component explains on average 23.3% of the variance of the commodity level factors, while the sector component explains 38.6%. However, the differences across commodities are large. For example, in the case of feeder cattle the market-wide component explains close to nothing of its level variation while for corn the market-wide component explains 72.1% of its variation. Overall, the market-wide component explains quite some variation of the softs and grains levels. For many of the other commodities the sector

[15] We put the market-wide component first, followed by the sector component, and last the idiosyncratic component.

component is more important, i.e. explains more price variation, than the market-wide component. The energy sector component explains around 65% of all energy commodities (except natural gas) compared to only 12% for the market-wide component. Similar observations can be made for the metal commodities (precious metals versus industrial metals) and the grain commodities (wheat versus corn and soybeans). These results clearly show that there is not a single common market-wide factor that affects all commodity price levels.

[insert Table 5]

The variance decomposition results for the slope factor are again diverse but confirm the importance of sector components. The market-wide, sector, and idiosyncratic components explain on average 24.9%, 48.6%, and 26.5%, respectively, of the commodities' slope factors. Notably, the market-wide component explains 70% or more of the energy commodities slope variation, with the exception of natural gas. Also the precious metals gold and silver are driven by a common market component for more than 50%. All other slope factors are mostly explained by a sector component, except for cocoa, cotton, sugar and soybeans. For the curvature factor, the market-wide component explains on average 24.0% of the commodities' curvature factors, although there is no clear pattern in these results. The sector-specific component explains 50% on average and is more important than the market-wide component for all sectors.

The variance decomposition results indicate that 61.9% of the level variation, 73.5% of the slope variation and 74.2% of curvature variation is explained by common factors. However, in all three cases the market-wide component explains only a quarter of the variation while commonality within sectors is two times more important.

5.2 Importance of curve data

In our analysis of commonality in commodity prices we use extensive term-structure information. An obvious advantage of this additional data is that it allows for investigating commonality in curve shapes, besides the often investigated commonality in levels. We argue that the inclusion of more distant futures data not only gains additional insights but is also important for the common level component analysis.

To quantify potential differences due to the inclusion of term structure information, we redo some of our analyses with only first nearby contract information. This allows us to compare the market-wide level factors that we obtained using our full model and dataset with the market-wide level factor that we obtain from a restricted version of our model and a limited dataset. We construct a one-factor model based on only first nearby contracts by restricting (2)-(4) to

$$f_{i,t} = l_{i,t} + \nu_{i,t} = \alpha_i^L + \beta_i^L L_{market,t} + \gamma_i^L L_{sector,t} + L_{i,t} + \nu_{i,t}, \tag{6}$$

where

$$\Delta L_{x,t} = \phi^x \Delta L_{x,t-1} + \eta_{x,t}^L. \tag{7}$$

We estimate this model using the same Kalman filter methodology but use only data on first nearby contracts.

Figure 5 shows a comparison of the extracted market-wide factors. The results from both methodologies are in line but there are some notable differences. First, the factor from the full model is pretty constant during the periods 2004-2006 and 2009-2010, while the factor from the one-factor model rises sharply. However, after both subperiods, the full model factor quickly picks up and rises to the same level as the one-factor model. The differences in behavior are also reflected by the correlation of 0.53 between the first differenced series of both factors. This implies that there is a large difference between quantifying the commodity price levels based on only the first nearby contract data versus the levels based on the full commodities curve.

[insert Figure 5]

Of most interest for our research question is the amount of commonality. The extracted market-wide factors are just one side of the story, and for the full picture the interplay of loadings and common components is what matters. When we redo the variance decompositions we find that the market-wide component explains a larger part of the variance in the one-factor model, 30.3% versus 23.3% for the full model. The sector components are also more important, with 40.0%, which is again higher than the market-wide component. The finding that commonality seems to be larger when only the front contracts are used, was not expected as front contracts are more volatile than contracts further down the curve (see Table 1 and Samuelson 1965). A possible explanation is the effect of commodity indices. These indices (or funds that track them) are mostly invested in front contracts, which could lead to increased commonality (Tang and Xiong, 2012). Taken together, when using only first nearby contract there is an overestimation of commonality in commodity prices.

5.3 Time variation

In the period of the so-called financialization of commodities markets, the entrance of financial investors (around 2004-2005) has changed the commodities market dynamics (Cheng and Xiong, 2014). All our analyses so far are based on the full sample period from 1995 to 2012. It is interesting to see if the amount of commonality varies over time during this period.

To investigate variation over time, we re-estimate the model attaching weights to the likelihood contributions of different observations in such a way that we emphasize specific data periods. The likelihood contribution of the observation at time $t_c + k$ is given weight $\delta^{|k|}$, for $k = \ldots, -2, -1, 0, 1, 2, \ldots$, with $0 < \delta < 1$, resulting in estimates centered at $t = t_c$. Repeating this for $t_c = 1, 2, \ldots, T$ yields a sequence of smoothly time-varying parameter estimates. Although by design, this weighting will not produce any abrupt change, it nevertheless provides information about the presence, or otherwise, of temporal variation. We favor this approach to the use of subsamples because our data covers a limited time period and we do not want to impose breakdates ourselves. We use $\delta = 0.99$ to ensure that each estimate reflects information in a sample of reasonable effective size.

[insert Figure 6]

Of particular interest is the relative importance over time of the common, market-wide and sector components. Figure 6 shows results of time-varying variance decompositions, which are at each point in time based on estimation results of the above described methodology. Both the lines corresponding to sector and idiosyncratic explained variation are based on averages across all sectors or commodities, respectively. Subfigure A shows that the market-wide component in the level factors becomes increasingly important over time. In 1995 this component explained just 14% of total variance, while this percentage steadily increases to 32% in 2008 after which it stays constant. Most of the increase happened at the expense of the idiosyncratic components. Hence, the percentage of explained variance by common factors increased over time. Still, most of this variance is explained by sector components and not the market-wide component. Subfigure B shows the results for the slope factors. All three lines are pretty constant over time. Subfigure C shows slightly more variation over time for the curvature components. From 2007 onward, the sector components explain a larger part of total curvature variation, while the market-wide component explains less variation. The total amount of curvature variation explained by common components remains quite stable over time.

Our findings show that only the level factors show an increase in commonality over time. These findings are in line with Tang and Xiong (2012). In all cases our main finding that sector components are more important than the market-wide component does not change.

5.4 Economic interpretation of unobserved states

We have established that there are common factors in commodities futures curves and these factors explain a substantial part of the variation in level, slope and curvature factors that drive observed futures prices. As these common components are unobserved, it is not straightforward which economic mechanism is underlying these components and, thus, the comovement. In this section we link the unobserved common factors to observed macroeconomic variables.

Existing literature provides a range of variables that could be related to the level, slope or curvature factors. The theory of normal backwardation (Keynes, 1930) argues that commodity producers and inventory holders hedge their risk by shorting futures. To induce risk-averse speculators into taking the opposite long positions, current futures prices are set at a discount (i.e. are "backwardated") to expected future spot prices at maturity.[16] Therefore the ratio of hedgers versus speculators could be related to the shape of the futures curve and hence to the factors. Alternatively, the theory of storage (Kaldor, 1939; Working, 1949) argues that convenience yield, basis, and inventories are closely related. In our set-up we capture the convenience yield in the slope factors.[17] Besides commodity-specific variables, we also use macroeconomic variables to interpret the factors.

We collect a large database of macroeconomic and commodity-specific variables. We consider the same set of 108 macroeconomic variables as in Stock & Watson (2012). Following Stock & Watson (2012) we transform the variables to ensure stationarity and assign them to 12 categories: GDP components, industrial production, employment, unemployment rate, housing, business inventories, wages, interest rates, money, exchange rates, stock prices, and consumer expectations.[18] For each category we apply a Principal Components Analysis (PCA) to summarize the variables within that group and proceed with the principal component that explains most of the variation in each group. In this way we reduce the dimension to 12 series that all correspond to a particular macroeconomic category. Besides the Stock & Watson (2012) dataset, we add a "financial conditions"

[16]Fernandez-Perez et al. (2017) document predictive power for commodity portfolios that take into account shape phases of the curve.

[17]Appendix E provides evidence of this empirical link. In this appendix, we compare the factors from our model to that of the Schwartz (1997) three-factor model. We find that for all individual commodities, our level and slope factors are similar to the corresponding unobserved spot price and convenience yield factors. However, the third factor in the Schwartz (1997) model is the interest rate and this is not the same as our curvature factors. Moreover, our curvature factor explains part of the residuals of the Schwartz (1997) three-factor model. This is another motivation for using the Nelson-Siegel approach over the Schwartz (1997) approach, on top of the complexity of extending the structural approach to a multi-curve setting.

[18]We leave out the Prices category due to endogeneity issues.

group to capture investor expectations and market conditions. This category consists of the Aruoba, Diebold & Scotti (2009) (ADS) business conditions index, the Baker and Wurgler (2006) sentiment index (Lutzenberger, 2014), and the Baltic dry shipping index (Bakshi et al., 2011).[19] Even though our financial conditions variable is based on macroeconomic series that are already included in other categories, it is not highly correlated with these other explanatory variables. We collect commodity inventory and hedging pressure data following the methodology of Gorton et al. (2013). Finally, we add volatility of the Commodity Research Bureau (CRB) spot market price index (Pindyck, 2004) as possible candidate. The in total 209 collected individual series result in 17 stationary variables. A complete overview of the individual series, their sources, categories and transformations is given in Appendix F.

[insert Table 6]

To determine which variables are most related to each common component we use a multivariate regression. Table 6 shows the variables that have a statistically significant coefficient. Up to one third of the variation of the differenced level components can be explained by the explanatory variables. These percentages are higher for the slope and curvature components.[20]

Focusing on Panel A, the results for the common market-wide components are in line with expectations. Changes in the level component, i.e. returns, are related to returns in foreign exchange and equity markets. As all commodity futures contracts we examine are denominated in dollars, a stronger dollar leads to lower commodity prices, which is reflected by the negative coefficient. The positive relation between equity and commodity prices is surprising as commodities are often used for diversification purposes (see, e.g., Erb and Harvey, 2006; Gorton and Rouwenhorst, 2006). However it is in line with more recent

[19] The ADS index is designed to track real business conditions at high frequency. Baker and Wurgler (2006) define sentiment as investor propensity to speculate. The Baltic Dry Index is an indicator of transportation costs for raw materials shipped by sea.

[20] We investigate the time-variation in the explanatory power of these variables using the methodology described in Section 5.3 (Results are available upon request.) The most interesting weighted OLS results show that Equity and Exchange rates explain ΔLevel better at the end of the sample. Furthermore, Hedging pressure data has a downward trend in terms of explanatory power for the market-wide slope and curvature components.

findings of Singleton (2014) and Tang and Xiong (2012). The (almost significant) positive relation with hedging pressure is consistent with the theory of normal backwardation Keynes (1930) and the results in Gorton, Hayashi & Rouwenhorst (2013); Hamilton & Wu (2014).[21]

Business inventories (new orders growth), housing (construction growth) and financial conditions are all considered leading economic indicators. Their positive relation with the market-wide slope factor implies a shift to more backwardated commodities curves when economic future perspectives are positive.[22] The same holds for hedging pressure. When it goes up (more net short positions by hedgers), the slope component goes up, which results in a backwardated futures curve. The relations between the market-wide slope and industrial production growth and changes in employment are negative, hence the commodity curves are more backwardated at times when the economy slows down. This is not in line with expectations, e.g. Fama and French (1988) find that when metal inventories are high, convenience yields are lower and the curve will be in contango.

The curvature component is related to business inventories, hedging pressure and interest rates. An increase in the curvature component leads to an increase in the price of mid-term contracts, while the contracts with very short or very long time to maturities are less affected. This seems to be coinciding with higher Treasury yields and lower hedging pressures.

Panels B to F show for each sector to which variables the components are linked. For the commodity-specific variables we use only data of the commodities included in the sector of interest. In general, all level components are related to exchange rates, equity and hedging pressure. The grains level component is an exception as it is hardly explained by any of the explanatory variables, shown by the R^2 of 7.1%. The sector slope components relate positively to hedging pressure, just as the market-wide slope component. Only for the metals sector the coefficient for hedging pressure is negative, which

[21]Note that Gorton et al. (2013) Table XI reports significant negative slope coefficients. However, they define hedging pressure as net long positions of hedgers while we look at net short positions.

[22]Recall that the loadings on the slope factors are convex, hence a positive (negative) slope factor implies a backwardated (contangoed) futures curve.

is opposite to what we would expect. It is also interesting to see that two sector slope components have a negative coefficient for commodity inventories. This is in line with the theory of storage as shown by, e.g., Gorton et al. (2013); Geman and Nhuyen (2005). Last, the sector curvature components show similar results as the market-wide curvature component. Business inventories, hedging pressure and yield curve variables have significant explanatory power. Furthermore, industrial production is negatively related to most sector curvature components, while employment variables are positively related.[23]

Concluding, the unobserved common components relate to observed macroeconomic and commodity-specific variables. In general, changes in the level components are related to equity, exchange rates and hedging pressure, which is in line with results of existing literature. Also the slope components show significant coefficients for hedging pressure, housing, and commodity inventories that are in line with expectations. Last, the curvature factor, which to the best of our knowledge has not yet been investigated in existing research, is positively related to interest rates and business inventories and negatively to industrial production.

6 Conclusion

We investigate comovement across commodities by examining the commonality in the price levels and shapes of their futures curves. To process all this information we use an enhanced version of the Nelson and Siegel (1987) model and extend the framework of Diebold et al. (2008) to extract the factors that drive the individual commodity futures curves. Comovement across commodities is investigated by decomposing each individual factor in a market-wide, sector, and idiosyncratic component.

Our commonality investigation differs from existing research by making use of the entire futures curve and accounting for the possibility of sector-specific commonality, next to a market-wide component. Using a monthly dataset of 24 commodities that are part of the S&P Goldman Sachs Commodity Index (GSCI) we show that there is comovement

[23]Hammoudeh et al. (2015) also provide evidence of a monetary policy and macroeconomic link with commodity prices, and highlight a heterogeneous relationship across sectors.

across commodity futures curves. Common components explain between 62% and 74% of the variation of individual commodities, but in contrast to existing research we show that this is mostly due to commonality within sectors. The exclusion of curve data leads to an overestimation of the importance of a common market-wide factor. We conclude that the commodities in our sample should still be considered as a heterogeneous asset class.

Additionally, we investigate if there is time variation in the importance of common components. For the shape related factors we find almost no variation over time. This is in contrast to the level factors, where the market-wide component explains more variation in more recent times. The percentage of explained variation starts at 14% in 1995 and increases to 32% in 2012. Still, the sector component remains more important, as the increase of importance of the market-wide component comes at the cost of the individual components. The unobserved common components relate to macroeconomic and commodity-specific variables in ways consistent with existing literature. The level components relate to equity markets, exchange rates and hedging pressure. The slope components are linked to hedging pressure (theory of normal backwardation) and commodity inventories (theory of storage). Last, the newly introduced curvature components relate to the yield curve, business inventories and industrial production.

The presented framework provides a way to include more futures data to investigate commonality across commodities. Using this framework we show that it is important to include the term-structure dimension and account for sector-specific effects in the analysis of comovement, as it alters the findings on the extent of comovement.

References

Aastveit, K. A., H. C. Bjørnland, and L. A. Thorsrud, 2015: What drives oil prices? Emerging versus developed economies. *Journal of Applied Econometrics*, **30**(7), 1013–1028.

Ai, C., A. Chatrath, and F. Song, 2006: On the comovement of commodity prices. *American Journal of Agricultural Economics*, **88**(3), 574–588.

Aruoba, S., F. Diebold, and C. Scotti, 2009: Real-time measurement of business conditions. *Journal of Business and Economic Statistics*, **27**(4), 417–427.

Baffes, J., 2007: Oil spills on other commodities. *Resources Policy*, **32**(3), 126 – 134.

Baker, M. and J. Wurgler, 2006: Investor sentiment and the cross-section of stock returns. *Journal of Finance*, **61**(4), 1645–1680.

Bakshi, G., X. Gao, and A. G. Rossi, 2017: Understanding the sources of risk underlying the cross-section of commodity returns. *Management Science (forthcoming)*.

Bakshi, G., G. Panayotov, and G. Skoulakis, 2011: The Baltic dry index as a predictor of global stock returns, commodity returns, and global economic activity. Working paper.

Bessembinder, H., 1992: Systematic risk, hedging pressure, and risk premiums in futures markets. *The Review of Financial Studies*, **5**(4), 637–667.

Byrne, J. P., G. Fazio, and N. Fiess, 2013: Primary commodity prices: Co-movements, common factors and fundamentals. *Journal of Development Economics*, **101**, 16–26.

Carter, C. A., G. C. Rausser, and A. Schmitz, 1983: Efficient asset portfolios and the theory of normal backwardation. *Journal of Political Economy*, **91**(2), 319–331.

Casassus, J., P. Liu, and K. Tang, 2013: Economic linkages, relative scarcity, and commodity futures returns. *Review of Financial Studies*, **26**(5), 1324–1362.

Chang, E. C., 1985: Returns to speculators and the theory of normal backwardation. *Journal of Finance*, **40**(1), 193–208.

Charlot, P., O. Darne, and Z. Moussa, 2016: Commodity returns co-movement: Fundamentals or "style" effect? *Journal of International Money and Finance*, **68**, 130–160.

Chen, Y.-C., K. S. Rogoff, and B. Rossi, 2010: Can exchange rates forecast commodity prices? *Quarterly Journal of Economics*, **125**(3), 1145–1194.

Cheng, I.-H. and W. Xiong, 2014: The financialization of commodities markets. *Annual Review of Financial Economics*, **6**(1), 1–23.

Christoffersen, P., A. Lunde, and K. V. Olesen, 2017: Factor structure in commodity futures return and volatility. Rotman school of management working paper no. 2495779.

Cortazar, G., C. Milla, and F. Severino, 2008: A multicommodity model of futures prices: Using futures prices of one commodity to estimate the stochastic process of another. *Journal of Futures Markets*, **28**(6), 537–560.

Cortazar, G. and L. Naranjo, 2006: An N-factor Gaussian model of oil futures. *The Journal of Futures Markets*, **26**(3), 243–268.

Daskalaki, C., A. Kostakis, and G. S. Skiadopoulos, 2014: Are there common factors in individual commodity futures returns? *Journal of Banking and Finance*, **40**(3), 346–363.

de Groot, W., D. Karstanje, and W. Zhou, 2014: Exploiting commodity momentum along the futures curves. *Journal of Banking and Finance*, **48**, 79–93.

de Roon, F. A., T. E. Nijman, and C. Veld, 2000: Hedging pressure effects in futures markets. *Journal of Finance*, **55**(3), 1437–1456.

Deaton, A. and G. Laroque, 1992: On the behaviour of commodity prices. *Review of Economic Studies*, **59**(1), 1–23.

Deb, P., P. Trivedi, and P. Varangis, 1996: The excess comovement of commodity prices reconsidered. *Journal of Applied Econometrics*, **11**(3), 275–291.

Delle Chiaie, S., L. Ferrara, and D. Giannone, 2017: Common factors of commodity prices. Banque de france working paper no. 645.

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

Diebold, F. X., C. Li, and V. Z. Yue, 2008: Global yield curve dynamics and interactions: A dynamic Nelson-Siegel approach. *Journal of Econometrics*, **146**(2), 351–363.

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

Duffee, G., 2011: Forecasting with the term structure: The role of no- arbitrage restrictions. working paper.

Erb, C. and C. Harvey, 2006: The tactical and strategic value of commodity futures. *Financial Analysts Journal*, **62**(2), 69–97.

Fama, E., 1970: Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, **25**(2), 383–417.

Fama, E. and K. French, 1988: Business cycles and the behavior of metals prices. *Journal of Finance*, **43**(5), 1075–1093.

Fernandez-Perez, A., A.-M. Fuertes, and J. Miffre, 2017: Commodity markets, long-run predictability, and intertemporal pricing. *Review of Finance*, **21**(3), 1159–1181.

Geman, H., 2005: *Commodities and Commodity Derivatives: Modeling and Pricing for Agriculturals, Metals, and Energy.* Wiley Finance, London, UK.

Geman, H. and V.-N. Nhuyen, 2005: Soybean inventory and forward curve dynamics. *Management Science*, **51**(7), 1076–1091.

Gibson, R. and E. S. Schwartz, 1990: Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, **45**(3), 959–976.

Gorton, G., F. Hayashi, and K. G. Rouwenhorst, 2013: The fundamentals of commodity futures returns. *Review of Finance*, **17**(1), 1–71.

Gorton, G. and K. G. Rouwenhorst, 2006: Facts and fantasies about commodity futures. *Financial Analysts Journal*, **62**(2), 47–68.

Grønborg, N. S. and A. Lunde, 2016: Analyzing oil futures with a dynamic Nelson-Siegel model. *Journal of Futures Markets*, **36**(2), 153–173.

Hamilton, J. D. and J. C. Wu, 2014: Risk premia in crude oil futures. *Journal of International Money and Finance*, **42**(4), 9–37.

Hammoudeh, S., D. K. Nguyen, and R. M. Sousa, 2015: US monetary policy and sectoral commodity prices. *Journal of International Money and Finance*, **57**, 61–85.

Heidorn, T., F. Mokinski, C. Rühl, and C. Schmaltz, 2015: The impact of fundamental and financial traders on the term structure of oil. *Energy Economics*, **48**(3), 276–287.

Kaldor, N., 1939: Speculation and economic stability. *Review of Economic Studies*, **7**(1), 1–27.

Keynes, J., 1930: *A Treatise on Money*, volume 2. Macmillan, London.

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), 1053–1069.

Koopman, S. and J. Durbin, 2000: Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis*, **21**(3), 281–296.

Koopman, S. J., M. I. P. Mallee, and M. Van der Wel, 2010: Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters. *Journal of Business and Economic Statistics*, **28**(3), 329–343.

Kose, M. A., C. Otrok, and C. H. Whiteman, 2003: International business cycles: World, region, and country-specific factors. *American Economic Review*, **93**(4), 1216–1239.

Lutzenberger, F., 2014: The predictability of aggregate returns on commodity futures. *Review of Financial Economics*, **23**(3), 120–130.

Ma, J., A. Vivian, and M. E. Wohar, 2015: What drives commodity returns? Market, sector or idiosyncratic factors? Working paper.

Miffre, J., 2012: Comparing first, second and third generation commodity indices. Working paper.

Milonas, N. T., 1991: Measuring seasonalities in commodity markets and the half-month effect. *Journal of Futures Markets*, **11**(3), 331–345.

Nelson, C. and A. Siegel, 1987: Parsimonious modeling of yield curves. *Journal of Business*, **60**(4), 473–489.

Ohana, S., 2010: Modeling global and local dependence in a pair of commodity forward curves with an application to the US natural gas and heating oil markets. *Energy Economics*, **32**(2), 373–388.

Pindyck, R. and J. Rotemberg, 1990: The excess co-movement of commodity prices. *Economic Journal*, **100**, 1173–1189.

Pindyck, R. S., 2004: Volatility and commodity price dynamics. *Journal of Futures Markets*, **24**(11), 1029–1047.

Samuelson, P., 1965: Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, **2**, 41–49.

Sargent, T. J. and C. A. Sims, 1977: Business cycle modeling without pretending to have too much a priori economic theory. Working paper.

Schwartz, E. S., 1997: The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, **52**(3), 923–973.

Schwartz, E. S. and J. E. Smith, 2000: Short-term variations and long-term dynamics in commodity prices. *Management Science*, **46**(7), 893–911.

Singleton, K. J., 2014: Investor flows and the 2008 boom/bust in oil prices. *Management Science*, **60**(2), 300–318.

Sorensen, C., 2002: Modeling seasonality in agricultural commodity futures. *Journal of Futures Markets*, **22**(5), 393–426.

Stock, J. H. and M. W. Watson, 1989: New indexes of coincident and leading economic indicators. In *NBER Macroeconomics Annual 1989, Volume 4*. MIT Press, 351–409.

Stock, J. H. and M. W. Watson, 2012: Generalized shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics*, **30**(4), 481–493.

Szymanowska, M., F. de Roon, T. Nijman, and R. van den Goorbergh, 2014: An anatomy of commodity futures risk premia. *Journal of Finance*, **69**(1), 453–482.

Tang, K. and W. Xiong, 2012: Index investment and financialization of commodities. *Financial Analysts Journal*, **68**(6), 54–74.

Vansteenkiste, I., 2009: How important are common factors in driving non-fuel commodity prices? A dynamic factor analysis. Ecb working paper no. 1072, European Central Bank.

West, J., 2012: Long-dated agricultural futures price estimates using the seasonal Nelson-Siegel model. *International Journal of Business and Management*, **7**(3), 78–93.

West, K. D. and K.-F. Wong, 2014: A factor model for co-movements of commodity prices. *Journal of International Money and Finance*, **42**(4), 289–309.

Working, H., 1949: The theory of price storage. *American Economic Review*, **39**(6), 1254–1262.

Yang, F., 2013: Investment shocks and the commodity basis spread. *Journal of Financial Economics*, **110**(1), 164–184.

Table 1 Commodity data overview

The table presents an overview of the 24 commodity futures series that are all present in the S&P Goldman Sachs Commodity Index (GSCI). The sector classification is the same as GSCI. We consider the period January 1995 to September 2012. We show the number of cross-sectional contracts that are available in the first and last year of our dataset. Furthermore we show the average annualized return and its volatility of both the first maturing futures contract and the futures contract with 12 months to maturity.

		# contracts		1st nearby contract		12M contract	
Sector	Commodity	Begin	End	$\bar{r}$	$\sigma(r)$	$\bar{r}$	$\sigma(r)$
Energy	Brent crude oil	12	56	12.9%	32.3%	13.4%	19.3%
	WTI crude oil	26	67	9.0%	32.8%	7.8%	17.1%
	Gasoil	14	30	12.8%	32.8%	7.3%	20.7%
	Heating oil	17	19	10.7%	34.8%	6.6%	25.2%
	Natural gas	20	83	−20.6%	53.1%	−1.4%	17.4%
	Gasoline	10	24	18.8%	39.2%	7.6%	23.3%
Metals	Gold	19	19	5.9%	16.4%	4.8%	15.9%
	Silver	19	19	8.4%	30.6%	8.5%	29.8%
	Aluminium	11	60	−4.9%	19.9%	7.8%	17.1%
	Copper	11	60	7.5%	27.6%	20.5%	31.4%
	Lead	11	29	4.7%	30.2%	10.8%	35.1%
	Nickel	11	30	4.2%	36.4%	10.3%	33.7%
	Zinc	11	30	−2.2%	27.6%	14.2%	29.7%
Softs	Cocoa	8	10	0.6%	31.4%	−4.6%	24.9%
	Coffee	7	10	−5.3%	36.7%	−5.6%	28.9%
	Cotton	9	11	−6.9%	29.9%	−2.0%	18.1%
	Sugar	7	8	0.5%	38.2%	8.6%	18.6%
Grains	Corn	8	14	−4.4%	29.0%	6.6%	12.6%
	Soybeans	11	17	7.3%	26.6%	3.1%	17.8%
	Chicago wheat	7	10	−10.2%	30.3%	4.4%	17.5%
	Kansas wheat	5	9	−0.2%	29.3%	1.3%	21.5%
Meats	Feeder cattle	8	8	1.5%	14.5%	4.2%	10.3%
	Lean hogs	8	9	−7.2%	28.0%	7.4%	14.1%
	Live cattle	7	7	0.5%	14.9%	3.8%	8.5%

Table 2 Individual commodity state space results

The table presents the estimation results and the fit of the individual commodity state space models. We show the final specification we use for each individual commodity, whereby 2F stands for a model with only a level and slope factor, 3F also includes a curvature factor, and 3FS adds a seasonal correction term. The model fit is shown in terms of R^2. The estimated parameter values for the decay parameter λ, the exposure κ to the seasonal term, and the most expensive contract expiry month θ, where $\theta = 0$ corresponds to January. The last column presents the volatility of errors $\sigma(\nu)$. Standard errors of all estimates are provided between brackets. For models that do not contain a seasonal component, the κ and θ parameters are irrelevant, which is represented by a horizontal dash.

Sector	Commodity	Model	R^2	λ		κ		θ		$\sigma(\nu)$
Energy	Brent crude oil	3F	99.99%	1.144	(0.01)	-	-	-	-	0.65%
	WTI crude oil	3F	99.99%	1.272	(0.01)	-	-	-	-	0.79%
	Gasoil	3FS	99.98%	1.688	(0.04)	0.92	(0.02)	0.4	(0.00)	1.03%
	Heating oil	3FS	99.97%	3.600	(0.03)	2.80	(0.01)	0.8	(0.00)	1.21%
	Natural gas	3FS	99.40%	1.137	(0.02)	6.38	(0.01)	0.9	(0.00)	3.56%
	Gasoline	3FS	99.94%	3.285	(0.04)	4.78	(0.01)	6.2	(0.00)	1.52%
Metals	Gold	2F	99.99%	0.011	(6.36)	-	-	-	-	0.44%
	Silver	2F	99.99%	0.026	(1.79)	-	-	-	-	0.54%
	Aluminium	2F	99.91%	0.187	(0.04)	-	-	-	-	0.77%
	Copper	2F	99.97%	0.111	(0.10)	-	-	-	-	1.13%
	Lead	2F	99.99%	0.324	(0.13)	-	-	-	-	0.57%
	Nickel	2F	99.98%	0.095	(0.28)	-	-	-	-	0.91%
	Zinc	2F	99.97%	0.059	(0.19)	-	-	-	-	0.69%
Softs	Cocoa	3F	99.97%	1.413	(0.03)	-	-	-	-	0.58%
	Coffee	3F	99.98%	1.468	(0.03)	-	-	-	-	0.57%
	Cotton	3FS	99.74%	3.482	(0.03)	0.91	(0.06)	5.9	(0.01)	1.21%
	Sugar	3FS	99.91%	3.272	(0.04)	1.13	(0.05)	2.5	(0.01)	1.32%
Grains	Corn	3FS	99.81%	2.743	(0.04)	1.80	(0.03)	5.7	(0.00)	1.56%
	Soybeans	3FS	99.89%	3.350	(0.03)	1.30	(0.03)	5.7	(0.00)	1.19%
	Chicago wheat	3FS	99.84%	1.496	(0.17)	1.53	(0.04)	2.0	(0.00)	1.51%
	Kansas wheat	3FS	99.86%	2.461	(0.08)	1.44	(0.05)	2.2	(0.01)	1.35%
Meats	Feeder cattle	3FS	99.91%	5.221	(0.07)	0.73	(0.07)	10.0	(0.01)	0.71%
	Lean hogs	3FS	97.80%	4.109	(0.07)	7.78	(0.02)	6.1	(0.00)	3.09%
	Live cattle	3FS	99.60%	4.487	(0.10)	2.32	(0.03)	1.9	(0.00)	1.29%

Table 3 Joint model factor loadings

The table presents the estimated loadings on various components in our joint model. Each level, slope, and curvature factor is decomposed in a constant part α, a market-wide part with loading β, a sector part with loading γ, and a commodity-specific part. Standard errors of all estimates are provided between brackets. The commodity-specific variance estimates, $\sigma^2_{\nu_i}$, are multiplied by 1,000 for readability reasons. Note that we do not include a curvature factor for the metal commodities, which is represented by a horizontal dash.

	Level						Slope						Curvature						Variance	
Commodity	α		β		γ		α		β		γ		α		β		γ		$\sigma^2_{\nu_i}$	
Brent crude oil	4.59	(0.06)	0.81	(0.09)	1.79	(0.09)	0.10	(0.08)	2.99	(0.03)	−0.90	(0.43)	0.06	(0.07)	0.11	(4.23)	3.88	(0.20)	0.01	(0.55)
WTI crude oil	4.58	(0.06)	0.78	(0.15)	1.79	(0.10)	0.08	(0.08)	2.92	(0.18)	−0.70	(0.43)	−0.02	(0.06)	0.36	(0.79)	3.26	(0.23)	0.02	(0.34)
Gasoil	4.70	(0.06)	0.78	(0.15)	1.82	(0.09)	0.03	(0.08)	2.83	(0.17)	−0.10	(0.42)	0.05	(0.04)	1.32	(0.53)	1.81	(0.29)	0.06	(0.35)
Heating oil	4.69	(0.06)	0.75	(0.15)	1.78	(0.09)	0.02	(0.07)	2.44	(0.21)	1.25	(0.38)	0.07	(0.05)	1.34	(0.54)	−0.54	(0.34)	0.12	(0.24)
Natural gas	4.84	(0.07)	0.46	(0.16)	0.81	(0.09)	−0.06	(0.11)	1.51	(0.39)	1.43	(0.15)	0.19	(0.13)	0.83	(0.86)	0.54	(0.64)	1.15	(0.04)
Gasoline	4.57	(0.06)	0.79	(0.15)	1.79	(0.10)	0.08	(0.06)	2.10	(0.18)	0.51	(0.36)	0.05	(0.03)	−0.95	(0.60)	0.12	(0.40)	0.17	(0.23)
Gold	4.66	(0.05)	0.50	(0.10)	0.40	(0.10)	−6.78	(1.67)	−3.09	(1.34)	−2.51	(1.61)	-	-	-	-	-	-	0.02	(0.59)
Silver	4.71	(0.08)	1.12	(0.17)	0.90	(0.16)	−2.12	(1.51)	−1.60	(0.83)	1.30	(0.98)	-	-	-	-	-	-	0.02	(0.53)
Aluminium	4.60	(0.05)	0.49	(0.11)	1.06	(0.04)	−0.15	(0.19)	0.30	(0.33)	3.26	(0.05)	-	-	-	-	-	-	0.03	(0.29)
Copper	4.50	(0.07)	0.78	(0.16)	1.81	(0.13)	0.60	(0.40)	0.47	(0.50)	3.77	(0.57)	-	-	-	-	-	-	0.09	(0.17)
Lead	4.63	(0.08)	0.60	(0.17)	1.41	(0.15)	0.05	(0.20)	−0.86	(0.53)	3.08	(0.55)	-	-	-	-	-	-	0.02	(0.65)
Nickel	4.62	(0.09)	0.69	(0.21)	1.66	(0.19)	0.65	(0.55)	2.26	(1.18)	8.71	(1.22)	-	-	-	-	-	-	0.04	(0.36)
Zinc	4.65	(0.07)	0.65	(0.16)	1.67	(0.13)	−0.34	(0.86)	−0.34	(1.34)	13.57	(1.49)	-	-	-	-	-	-	0.04	(0.35)
Cocoa	4.67	(0.08)	0.78	(0.16)	0.37	(0.25)	−0.20	(0.08)	−0.17	(0.12)	0.17	(0.13)	−0.17	(0.03)	−0.17	(0.33)	−0.29	(0.37)	0.02	(1.07)
Coffee	4.64	(0.08)	0.85	(0.18)	0.70	(0.21)	−0.19	(0.11)	−0.46	(0.25)	3.43	(0.03)	−0.11	(0.06)	−0.58	(0.53)	0.83	(0.56)	0.01	(1.34)
Cotton	4.42	(0.06)	1.02	(0.13)	0.54	(0.25)	−0.06	(0.13)	−0.10	(0.21)	−0.31	(0.22)	−0.10	(0.08)	−0.67	(0.58)	4.29	(4.54)	0.17	(0.31)
Sugar	4.50	(0.07)	0.61	(0.15)	0.36	(0.21)	0.05	(0.07)	−0.13	(0.27)	0.20	(0.27)	0.06	(0.08)	−1.54	(0.64)	0.23	(0.79)	0.17	(0.37)
Corn	4.70	(0.06)	1.65	(0.11)	0.21	(0.26)	−0.10	(0.08)	0.04	(0.21)	1.60	(0.07)	0.06	(0.05)	−0.57	(0.72)	2.79	(0.08)	0.26	(0.24)
Soybeans	4.67	(0.06)	1.48	(0.11)	0.14	(0.12)	0.03	(0.06)	0.05	(0.16)	0.90	(0.17)	0.05	(0.05)	−1.93	(0.66)	2.49	(0.56)	0.15	(0.29)
Chicago wheat	4.53	(0.06)	1.34	(0.13)	1.42	(0.10)	−0.07	(0.12)	0.08	(0.27)	3.72	(0.22)	0.16	(0.05)	2.08	(0.71)	3.04	(0.52)	0.28	(0.27)
Kansas wheat	4.52	(0.06)	1.32	(0.13)	1.42	(0.10)	−0.05	(0.09)	0.10	(0.21)	2.61	(0.18)	−0.00	(0.05)	2.78	(0.80)	3.73	(0.58)	0.23	(0.32)
Feeder cattle	4.58	(0.03)	−0.02	(0.07)	0.86	(0.03)	−0.01	(0.02)	−0.02	(0.11)	1.26	(0.09)	0.02	(0.02)	0.93	(0.33)	0.87	(0.17)	0.03	(0.93)
Lean hogs	4.71	(0.04)	0.43	(0.10)	0.34	(0.09)	−0.04	(0.07)	0.69	(0.33)	0.86	(0.38)	0.14	(0.11)	0.55	(1.28)	−1.68	(1.17)	1.09	(0.13)
Live cattle	4.48	(0.02)	0.34	(0.05)	0.61	(0.04)	−0.03	(0.03)	−0.06	(0.15)	1.37	(0.25)	−0.03	(0.03)	1.03	(0.48)	3.91	(0.39)	0.17	(0.39)

Table 4 Joint model factor dynamics

The table presents the dynamics of the unobserved states by showing the parameter estimates of the state equations. The autoregressive parameters are the diagonal elements of the Φ-matrices. All level factors are modeled in first differences due to their non-stationary behavior. The disturbance variances correspond to elements in Σ_{η_y}, which are multiplied by 1,000 for readability reasons. Note that the disturbance variances of the market-wide and sector components are fixed for identifications purposes, and we do not include a curvature factor for the metal commodities, which is represented by a horizontal dash.

		Autoregressive parameters (ϕ)						Disturbance variances					
Sector	Factor	ΔLevel		Slope		Curv.		ΔLevel		Slope		Curv.	
	Market-wide	0.11	(0.05)	0.93	(0.46)	0.78	(0.26)	1.00	-	1.00	-	1.00	-
	Energy	0.12	(0.04)	0.79	(0.22)	0.88	(0.29)	1.00	-	1.00	-	1.00	-
	Metals	0.12	(0.05)	0.96	(0.94)	-	-	1.00	-	1.00	-	-	-
	Softs	0.45	(0.28)	0.93	(0.53)	0.87	(0.63)	1.00	-	1.00	-	1.00	-
	Grains	−0.12	(0.06)	0.94	(0.54)	0.88	(0.35)	1.00	-	1.00	-	1.00	-
	Meats	0.05	(0.05)	0.88	(0.38)	0.72	(0.18)	1.00	-	1.00	-	1.00	-
Energy	Brent crude oil	−0.07	(0.09)	0.92	(0.75)	0.83	(3.20)	0.11	(1.95)	0.20	(6.64)	0.03	(66.53)
	WTI crude oil	0.11	(0.07)	0.82	(0.32)	0.83	(0.23)	0.16	(1.39)	0.72	(0.97)	4.81	(0.23)
	Gasoil	−0.60	(0.15)	0.77	(0.20)	0.64	(0.14)	0.16	(1.46)	1.62	(0.50)	13.23	(0.19)
	Heating oil	−0.53	(0.16)	−0.33	(0.37)	0.83	(0.24)	0.08	(2.72)	0.28	(6.88)	12.52	(0.21)
	Natural gas	−0.03	(0.05)	0.91	(0.40)	0.88	(0.32)	3.92	(0.22)	19.60	(0.12)	53.52	(0.11)
	Gasoline	0.06	(0.06)	0.61	(0.12)	0.60	(0.11)	0.35	(0.88)	5.82	(0.21)	34.36	(0.12)
Metals	Gold	−0.09	(0.05)	0.98	(2.10)	-	-	1.73	(0.33)	299.85	(86.52)	-	-
	Silver	−0.24	(0.06)	0.99	(2.91)	-	-	4.96	(0.20)	133.71	(0.09)	-	-
	Aluminium	0.01	(0.06)	0.85	(0.40)	-	-	0.98	(0.52)	11.67	(0.21)	-	-
	Copper	0.09	(0.08)	0.97	(1.05)	-	-	1.10	(0.72)	34.54	(0.12)	-	-
	Lead	0.04	(0.05)	0.82	(0.23)	-	-	3.31	(0.26)	51.64	(0.09)	-	-
	Nickel	−0.02	(0.05)	0.86	(0.32)	-	-	5.23	(0.20)	205.89	(0.20)	-	-
	Zinc	−0.07	(0.07)	0.91	(0.53)	-	-	1.67	(0.45)	200.08	(0.25)	-	-
Softs	Cocoa	−0.14	(0.06)	0.96	(0.81)	0.80	(0.21)	5.05	(0.21)	2.80	(0.28)	8.37	(0.21)
	Coffee	−0.14	(0.07)	0.96	(0.76)	0.82	(0.22)	5.70	(0.23)	0.03	(31.22)	20.69	(0.15)
	Cotton	−0.12	(0.09)	0.96	(0.81)	0.82	(0.72)	2.17	(0.48)	7.85	(0.18)	6.87	(9.71)
	Sugar	0.09	(0.05)	0.88	(0.30)	0.85	(0.26)	3.77	(0.24)	13.97	(0.13)	28.52	(0.14)
Grains	Corn	0.04	(0.09)	0.92	(0.47)	0.76	(0.19)	0.66	(1.26)	5.82	(0.22)	16.20	(0.21)
	Soybeans	−0.19	(0.08)	0.92	(0.47)	0.81	(0.32)	1.14	(0.60)	4.40	(0.23)	7.71	(0.79)
	Chicago wheat	−0.25	(0.68)	0.89	(4.13)	0.88	(0.53)	0.01	(88.34)	0.03	(88.67)	2.79	(1.54)
	Kansas wheat	0.20	(0.09)	0.87	(0.32)	0.92	(1.63)	0.15	(2.77)	1.60	(0.61)	0.73	(12.67)
Meats	Feeder cattle	0.28	(0.21)	0.81	(0.43)	0.75	(0.18)	0.04	(24.97)	0.53	(5.93)	6.37	(0.28)
	Lean hogs	0.12	(0.05)	0.88	(0.32)	0.80	(0.22)	1.26	(0.47)	16.42	(0.15)	104.90	(0.13)
	Live cattle	0.29	(0.16)	0.82	(0.28)	0.69	(1.63)	0.02	(29.99)	1.94	(1.08)	0.18	(89.28)

Table 5 Variance decompositions

The table presents the percentage of explained variation of the level, slope and curvature factors by the market-wide, sector, and idiosyncratic components. As the level factors are non-stationary, their analysis is done on first differences.

	ΔLevel			Slope			Curvature		
Commodity	Market	Sector	Idio.	Market	Sector	Idio.	Market	Sector	Idio.
Brent crude oil	13.5%	65.8%	20.7%	83.2%	7.5%	9.3%	0.1%	93.7%	6.2%
WTI crude oil	12.6%	66.6%	20.8%	85.1%	4.9%	10.0%	1.1%	90.4%	8.5%
Gasoil	12.3%	67.3%	20.5%	88.8%	0.1%	11.1%	28.8%	54.6%	16.6%
Heating oil	11.9%	66.9%	21.3%	69.8%	18.4%	11.7%	58.0%	9.5%	32.5%
Natural gas	11.2%	35.1%	53.7%	43.0%	38.4%	18.6%	34.5%	15.0%	50.5%
Gasoline	12.9%	66.3%	20.8%	77.8%	4.5%	17.7%	47.1%	0.8%	52.1%
Gold	17.9%	11.6%	70.5%	56.8%	37.4%	5.8%	-	-	-
Silver	40.6%	26.7%	32.7%	50.1%	32.7%	17.2%	-	-	-
Aluminium	10.2%	47.9%	42.0%	0.8%	90.7%	8.5%	-	-	-
Copper	12.3%	67.1%	20.6%	1.5%	92.2%	6.3%	-	-	-
Lead	10.6%	59.5%	29.8%	6.7%	84.4%	8.9%	-	-	-
Nickel	11.4%	64.9%	23.7%	6.3%	92.5%	1.2%	-	-	-
Zinc	10.0%	66.3%	23.7%	0.1%	99.4%	0.5%	-	-	-
Cocoa	34.6%	8.0%	57.4%	2.7%	2.8%	94.4%	2.6%	7.5%	89.9%
Coffee	32.7%	22.0%	45.4%	1.6%	90.7%	7.7%	16.8%	34.2%	49.1%
Cotton	44.3%	12.6%	43.1%	1.0%	9.0%	90.0%	2.3%	92.7%	5.0%
Sugar	24.6%	8.7%	66.7%	1.5%	3.8%	94.7%	69.2%	1.6%	29.3%
Corn	72.1%	1.2%	26.7%	0.0%	71.6%	28.4%	3.6%	85.4%	11.0%
Soybeans	68.2%	0.6%	31.2%	0.1%	44.6%	55.3%	34.1%	56.8%	9.1%
Chicago wheat	37.2%	41.9%	20.9%	0.0%	93.2%	6.8%	29.7%	63.4%	6.9%
Kansas wheat	36.3%	42.7%	21.0%	0.1%	87.0%	12.9%	34.2%	61.4%	4.4%
Feeder cattle	0.0%	42.4%	57.6%	0.0%	61.2%	38.8%	32.9%	28.8%	38.3%
Lean hogs	14.3%	8.7%	76.9%	21.5%	33.4%	45.1%	7.3%	68.6%	24.1%
Live cattle	7.7%	24.6%	67.7%	0.1%	65.3%	34.6%	6.1%	88.1%	5.8%
Average	23.3%	38.6%	38.0%	24.9%	48.6%	26.5%	24.0%	50.2%	25.8%

Table 6 Interpretation common components

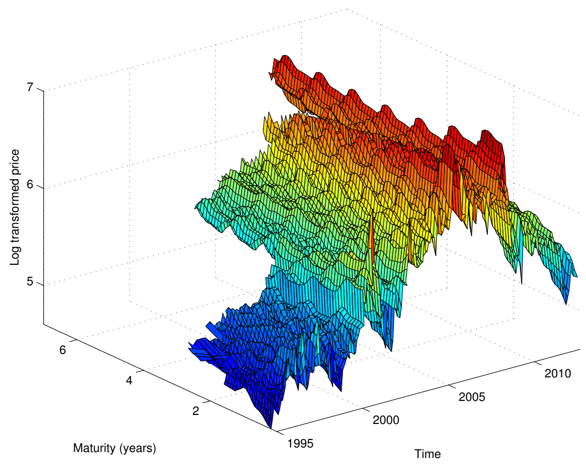
The table presents observed macroeconomic and commodity-specific variables that are related to our unobserved states. Each component is regressed on a set of 17 explanatory variables. The table presents the statistically significant coefficients with their corresponding *t*-stat (based on Newey-West corrected standard errors). The macroeconomic variables are taken from Stock and Watson (2012), while the commodity-specific variables are collected as described by Gorton et al. (2013). Full details on the data series are given in Appendix F.

ΔLevel		Slope		Curvature	
Variable	*t*-stat	Variable	*t*-stat	Variable	*t*-stat
Panel A Market-wide components					
$R^2 = 16.5\%$		$R^2 = 60.1\%$		$R^2 = 22.6\%$	
Equity	3.29	Business inventories	4.43	Business inventories	2.53
Exchange rates	−2.48	Employment	−2.74	Hedging pressure	−3.39
Hedging pressure	1.93	Financial conditions	8.29	Interest rates	3.76
		Hedging pressure	10.04		
		Housing	8.54		
		Industrial production	−2.78		
Panel B Energy sector components					
$R^2 = 31.0\%$		$R^2 = 24.3\%$		$R^2 = 31.1\%$	
Employment	−2.49	Business inventories	−4.63	Business inventories	3.74
Equity	2.09	Commodity inventories	−3.02	Financial conditions	4.90
Exchange rates	−2.05	Financial conditions	−5.10	Hedging pressure	−4.75
Financial conditions	2.80			Industrial production	−3.55
Hedging pressure	6.92			Interest rates	1.99
Housing	3.69			Unemployment	−2.23
Interest rates	2.34				
Panel C Metals sector components					
$R^2 = 29.9\%$		$R^2 = 49.1\%$			
Employment	−3.64	Business inventories	7.93		
Equity	4.72	Commodity inventories	−2.09		
Exchange rates	−3.36	Hedging pressure	−7.53		
Industrial production	2.40	Housing	8.41		
Interest rates	2.43				
Wages	−2.68				
Panel D Softs sector components					
$R^2 = 36.9\%$		$R^2 = 47.0\%$		$R^2 = 41.7\%$	
Employment	−4.00	Employment	6.84	Business inventories	2.53
Equity	2.00	Hedging pressure	5.10	Commodity volatility	−2.06
Hedging pressure	8.02	Housing	−4.83	Employment	4.99
Housing	3.09			Housing	−5.12
				Industrial production	−2.28
Panel E Grains sector components					
$R^2 = 7.1\%$		$R^2 = 56.1\%$		$R^2 = 60.8\%$	
		Business inventories	−2.90	Employment	4.36
		Commodity volatility	−2.67	Financial conditions	−5.20
		Employment	2.23	Hedging pressure	13.26
		Financial conditions	−7.82	Industrial production	−2.32
		Hedging pressure	10.62		
		Housing	3.74		
		Interest rates	−3.20		
Panel F Meats sector components					
$R^2 = 19.0\%$		$R^2 = 39.8\%$		$R^2 = 10.9\%$	
Hedging pressure	4.82	Business inventories	4.66	Hedging pressure	2.14
		Employment	−4.44		
		Hedging pressure	3.27		
		Housing	8.43		
		Wages	−2.88		

Figure 1 Commodity futures curves

These figures gives insight in the data by showing the complete set of available futures prices for natural gas and coffee. The figures show the commodity futures curves at each month in time.

(a) Natural gas



(b) Coffee

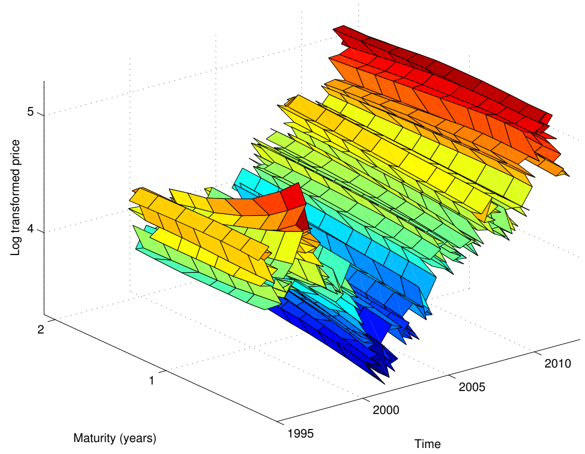
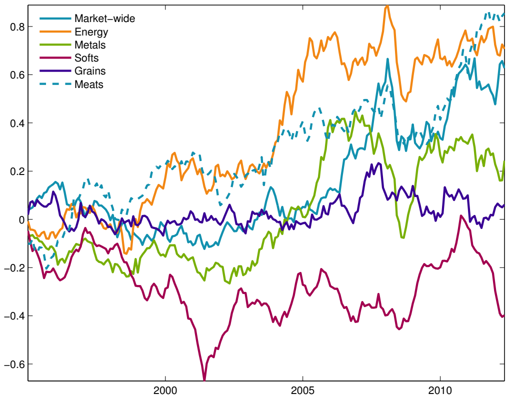


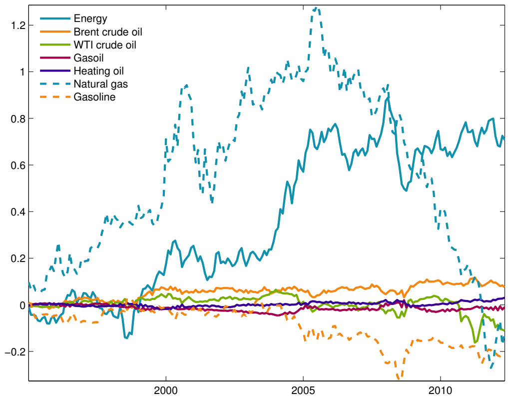
Figure 2 Joint model - extracted commodity level factors

These figures show the extracted level factors of our joint model estimated using all 24 commodities. Subfigure A shows the market-wide and sector components. Subfigures B-F show the estimated idiosyncratic level factors for the commodities of a specific sector.

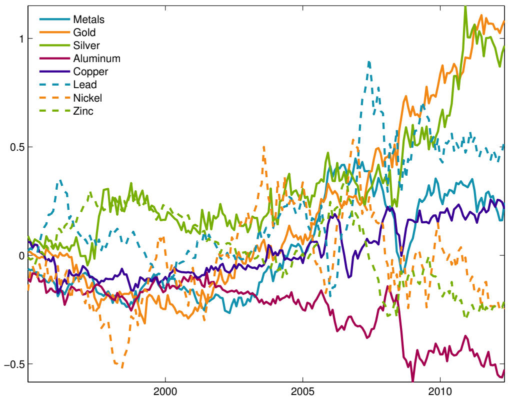
(a) Level - market-wide



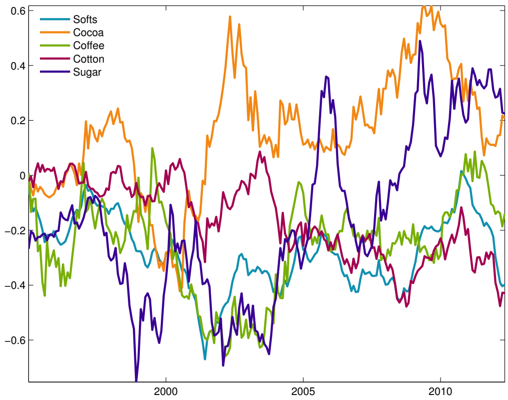
(b) Level - energy



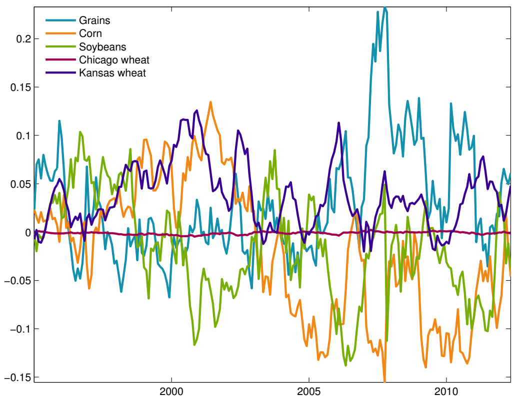
(c) Level - metals



(d) Level - softs



(e) Level - grains



(f) Level - meats

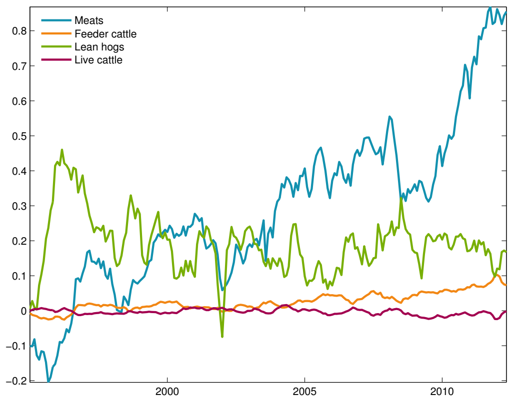
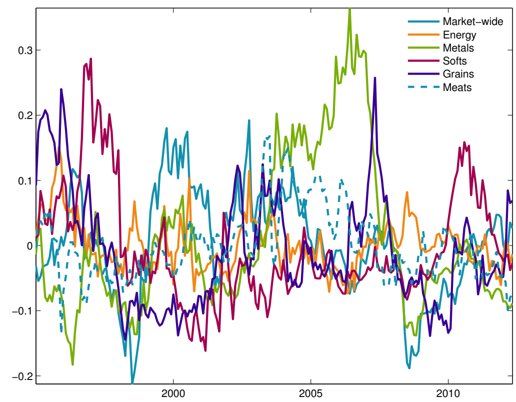


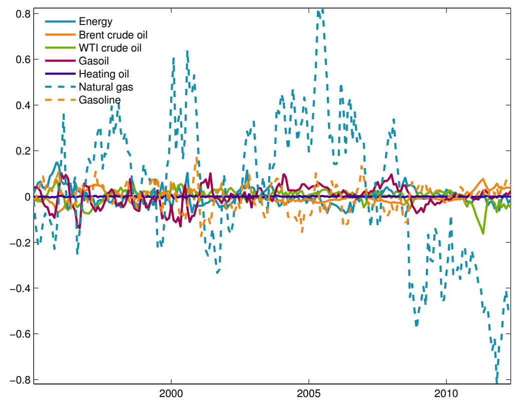
Figure 3 Joint model - extracted commodity slope factors

These figures show the extracted slope factors of our joint model estimated using all 24 commodities. Subfigure A shows the market-wide and sector components. Subfigures B-F show the estimated idiosyncratic slope factors for the commodities of a specific sector.

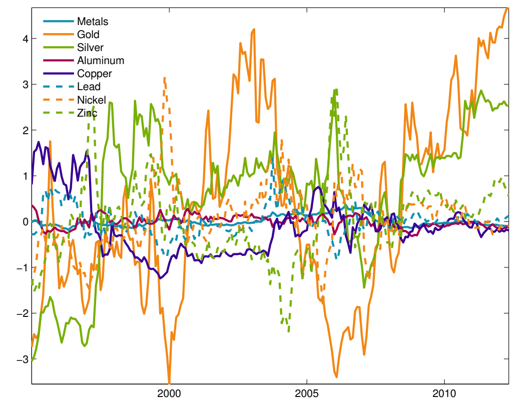
(a) Slope - market-wide



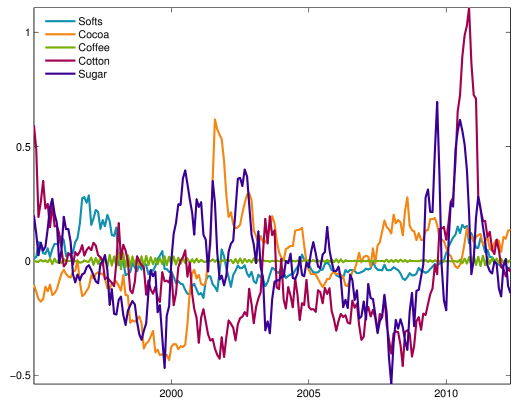
(b) Slope - energy



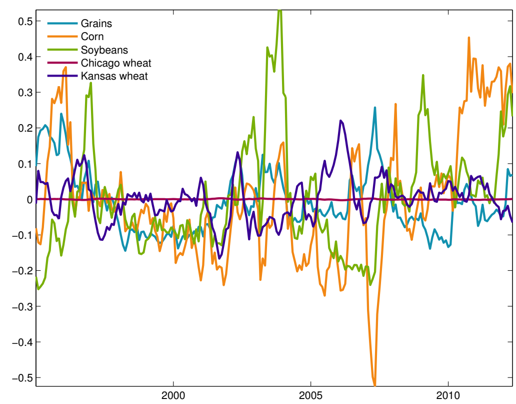
(c) Slope - metals



(d) Slope - softs



(e) Slope - grains



(f) Slope - meats

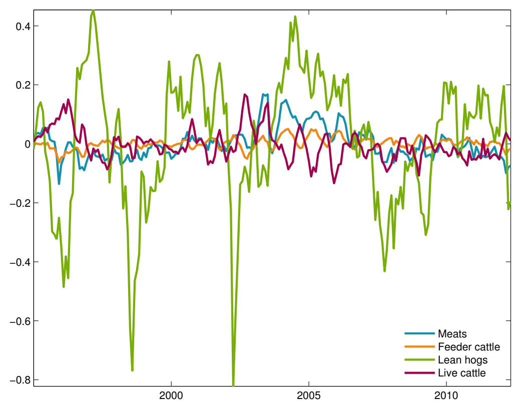
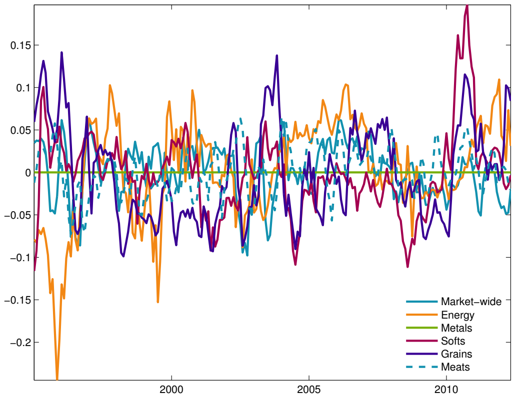


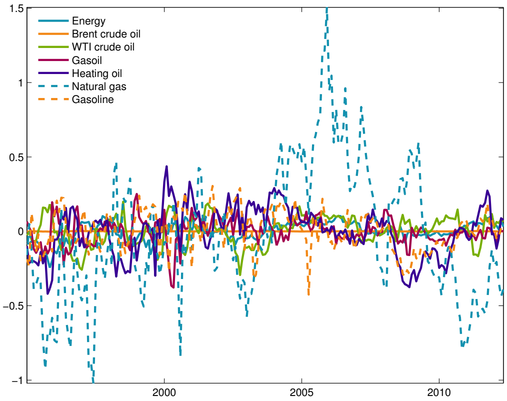
Figure 4 Joint model - extracted commodity curvature factors

These figures show the extracted curvature factors of our joint model estimated using all 24 commodities. Subfigure A shows the market-wide and sector components. Subfigures B-E show the estimated idiosyncratic curvature factors for the commodities of a specific sector. Note that the metal commodities subfigure is missing because we use a two-factor model for the metal commodities.

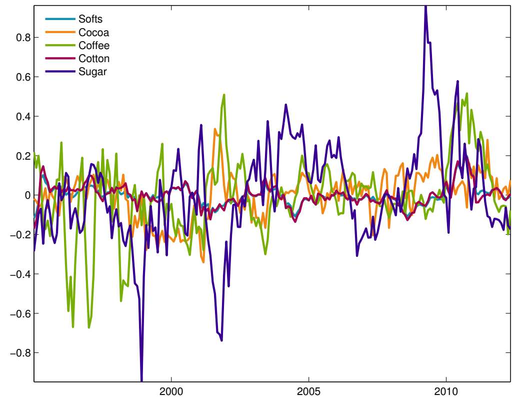
(a) Curvature - market-wide



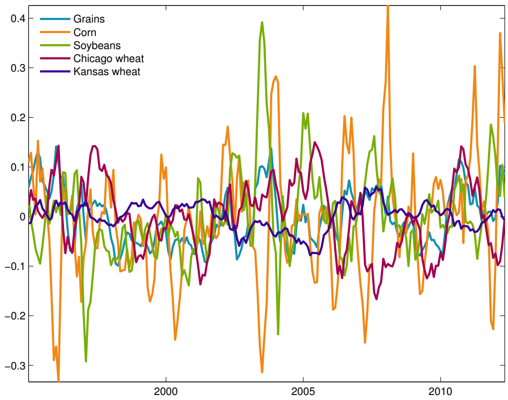
(b) Curvature - energy



(c) Curvature - softs



(d) Curvature - grains



(e) Curvature - meats

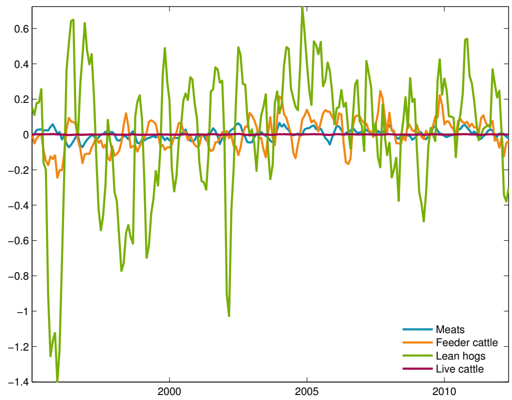


Figure 5 Comparison market-wide level components

Each line shows the market-wide level component of a particular model. The blue line corresponds to the full joint model with slope and curvature factors and curve information. The orange line is a restricted version of the full model with only level factors and first nearby contract information.

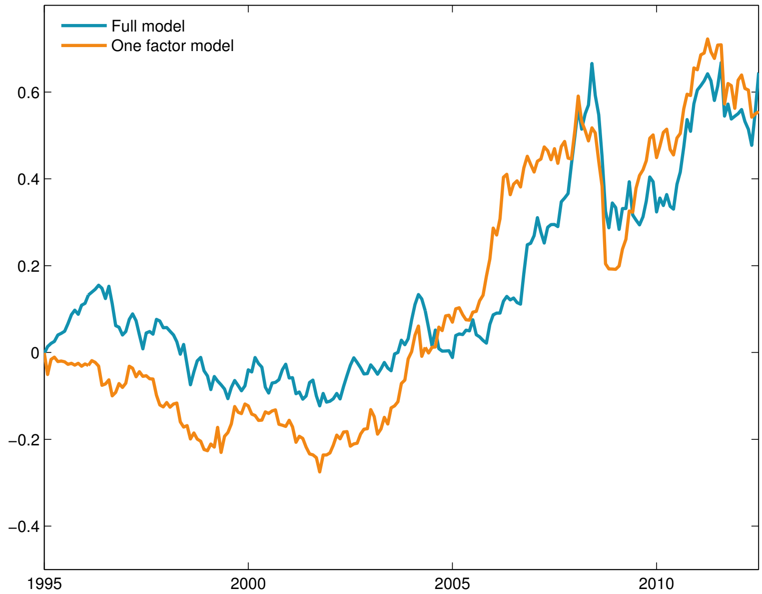
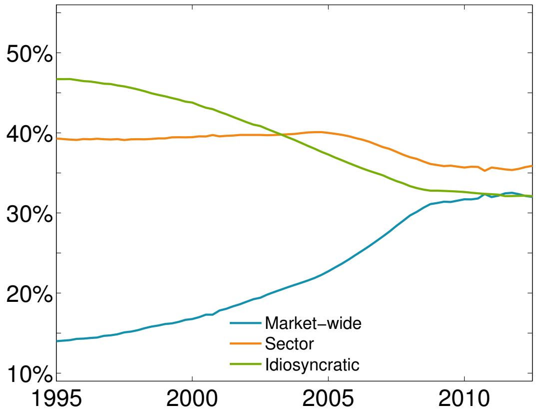


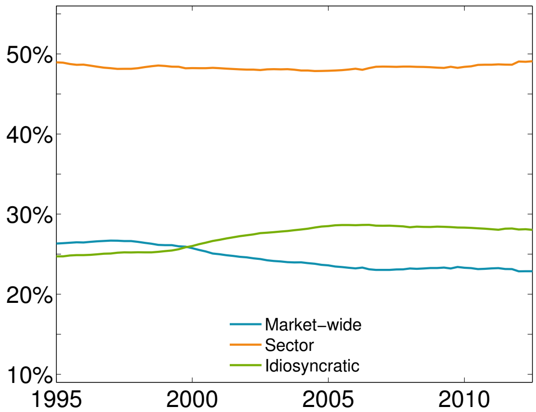
Figure 6 Moving window variance decompositions

These figures show the percentage of explained variation per component. Time-variation is introduced by estimating the model using a rolling window and applying a variance decomposition analysis at each point in time. The time-varying specification is estimated by attaching lower weights to more distant observations. The observation at time $\Theta + k$ is given weight $\delta^{|k|}$, for $k = \ldots, -2, -1, 0, 1, 2, \ldots$, with $0 < \delta < 1$, resulting in estimates centered at $t = \Theta$.

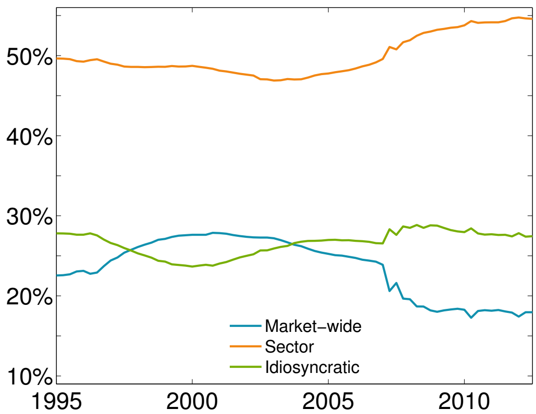
(a) ΔLevel



(b) Slope



(c) Curvature



Common Factors in Commodity Futures Curves

Web Appendix

Dennis Karstanje, Michel van der Wel, and Dick van Dijk*

December 8, 2017

Contents

*Corresponding author. Tel.: +31 10 4081263. E-mail: djvandijk@ese.eur.nl. Other author's e-mail addresses: dennis@karstanje.com (Karstanje), vanderwel@ese.eur.nl (van der Wel).

Appendix

A State space representation

The state space representation follows naturally from the model given by (2)-(4). The measurement equation in (A.1) is a combination of (2) and (3). Note that the individual latent level $l_{i,t}$, slope $s_{i,t}$, and curvature $c_{i,t}$ factors do not appear in the measurement equation, as we can link the observed futures prices $f_{i,t}(\tau)$ directly to the unobserved market-wide, sector and idiosyncratic components.

$$\begin{pmatrix} f_{1,t}(\tau_1) \\ f_{1,t}(\tau_2) \\ \vdots \\ f_{1,t}(\tau_{J_1}) \\ \vdots \\ f_{N,t}(\tau_{J_N}) \end{pmatrix} = A \begin{pmatrix} \alpha_1^L \\ \alpha_1^S \\ \alpha_1^C \\ \vdots \\ \alpha_N^C \end{pmatrix} + B \begin{pmatrix} L_{market,t} \\ S_{market,t} \\ C_{market,t} \end{pmatrix} + C \begin{pmatrix} L_{Energy,t} \\ S_{Energy,t} \\ C_{Energy,t} \\ \vdots \\ L_{Meats,t} \\ S_{Meats,t} \\ C_{Meats,t} \end{pmatrix} + A \begin{pmatrix} L_{1,t} \\ S_{1,t} \\ C_{1,t} \\ \vdots \\ C_{N,t} \end{pmatrix} + D \begin{pmatrix} \kappa_1 \\ \kappa_2 \\ \vdots \\ \kappa_N \end{pmatrix} + \begin{pmatrix} \nu_{1,t}(\tau_1) \\ \nu_{1,t}(\tau_2) \\ \vdots \\ \nu_{1,t}(\tau_{J_1}) \\ \vdots \\ \nu_{N,t}(\tau_{J_N}) \end{pmatrix} \tag{A.1}$$

where J_i is the number of available contracts of commodity i. As discussed in Section 2 J_i varies over time, yet for readability reasons we keep writing J_i instead of J_{i_t}.

$$A = \begin{pmatrix} 1 & \left(\frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1}\right) & \left(\frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} - e^{-\lambda_1\tau_1}\right) & 0 & \cdots & \cdots & 0 \\ 1 & \left(\frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2}\right) & \left(\frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} - e^{-\lambda_1\tau_2}\right) & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & \left(\frac{1-e^{-\lambda_N\tau_{J_N}}}{\lambda_N\tau_{J_N}}\right) & \left(\frac{1-e^{-\lambda_N\tau_{J_N}}}{\lambda_N\tau_{J_N}} - e^{-\lambda_N\tau_{J_N}}\right) \end{pmatrix},$$

$$B = \begin{pmatrix} \beta_1^l & \beta_1^s\left(\frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1}\right) & \beta_1^c\left(\frac{1-e^{-\lambda_1\tau_1}}{\lambda_1\tau_1} - e^{-\lambda_1\tau_1}\right) \\ \beta_1^l & \beta_1^s\left(\frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2}\right) & \beta_1^c\left(\frac{1-e^{-\lambda_1\tau_2}}{\lambda_1\tau_2} - e^{-\lambda_1\tau_2}\right) \\ \vdots & \vdots & \vdots \\ \beta_N^l & \beta_N^s\left(\frac{1-e^{-\lambda_N\tau_{J_N}}}{\lambda_N\tau_{J_N}}\right) & \beta_N^c\left(\frac{1-e^{-\lambda_N\tau_{J_N}}}{\lambda_N\tau_{J_N}} - e^{-\lambda_N\tau_{J_N}}\right) \end{pmatrix},$$

$$C = \begin{pmatrix} \gamma_1^l & \gamma_1^s \left(\frac{1-e^{-\lambda_1 \tau_1}}{\lambda_1 \tau_1}\right) & \gamma_1^c \left(\frac{1-e^{-\lambda_1 \tau_1}}{\lambda_1 \tau_1} - e^{-\lambda_1 \tau_1}\right) & 0 & \cdots & \cdots & 0 \\ \gamma_1^l & \gamma_1^s \left(\frac{1-e^{-\lambda_1 \tau_2}}{\lambda_1 \tau_2}\right) & \gamma_1^c \left(\frac{1-e^{-\lambda_1 \tau_2}}{\lambda_1 \tau_2} - e^{-\lambda_1 \tau_2}\right) & 0 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \gamma_N^l & \gamma_N^s \left(\frac{1-e^{-\lambda_N \tau_{J_N}}}{\lambda_N \tau_{J_N}}\right) & \gamma_N^c \left(\frac{1-e^{-\lambda_N \tau_{J_N}}}{\lambda_N \tau_{J_N}} - e^{-\lambda_N \tau_{J_N}}\right) \end{pmatrix},$$

$$D = \begin{pmatrix} \cos\left(\omega g_1\left(t, \tau_1\right) - \omega\theta_1\right) & 0 & \cdots & 0 \\ \cos\left(\omega g_1\left(t, \tau_2\right) - \omega\theta_1\right) & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \cos\left(\omega g_1\left(t, \tau_{J_1}\right) - \omega\theta_1\right) & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cos\left(\omega g_N\left(t, \tau_{J_N}\right) - \omega\theta_N\right) \end{pmatrix}.$$

The transition equations of the latent states are given by (4). The market-wide, sector, and idiosyncratic components are assumed to have first-order autoregressive dynamics. For completeness, we also present them here:

$$\begin{pmatrix} \Delta L_{y,t} \\ S_{y,t} \\ C_{y,t} \end{pmatrix} = \begin{pmatrix} \phi_{11}^y & \phi_{12}^y & \phi_{13}^y \\ \phi_{21}^y & \phi_{22}^y & \phi_{23}^y \\ \phi_{31}^y & \phi_{32}^y & \phi_{33}^y \end{pmatrix} \begin{pmatrix} \Delta L_{y,t-1} \\ S_{y,t-1} \\ C_{y,t-1} \end{pmatrix} + \begin{pmatrix} \eta_{y,t}^L \\ \eta_{y,t}^S \\ \eta_{y,t}^C \end{pmatrix}, \tag{A.2}$$

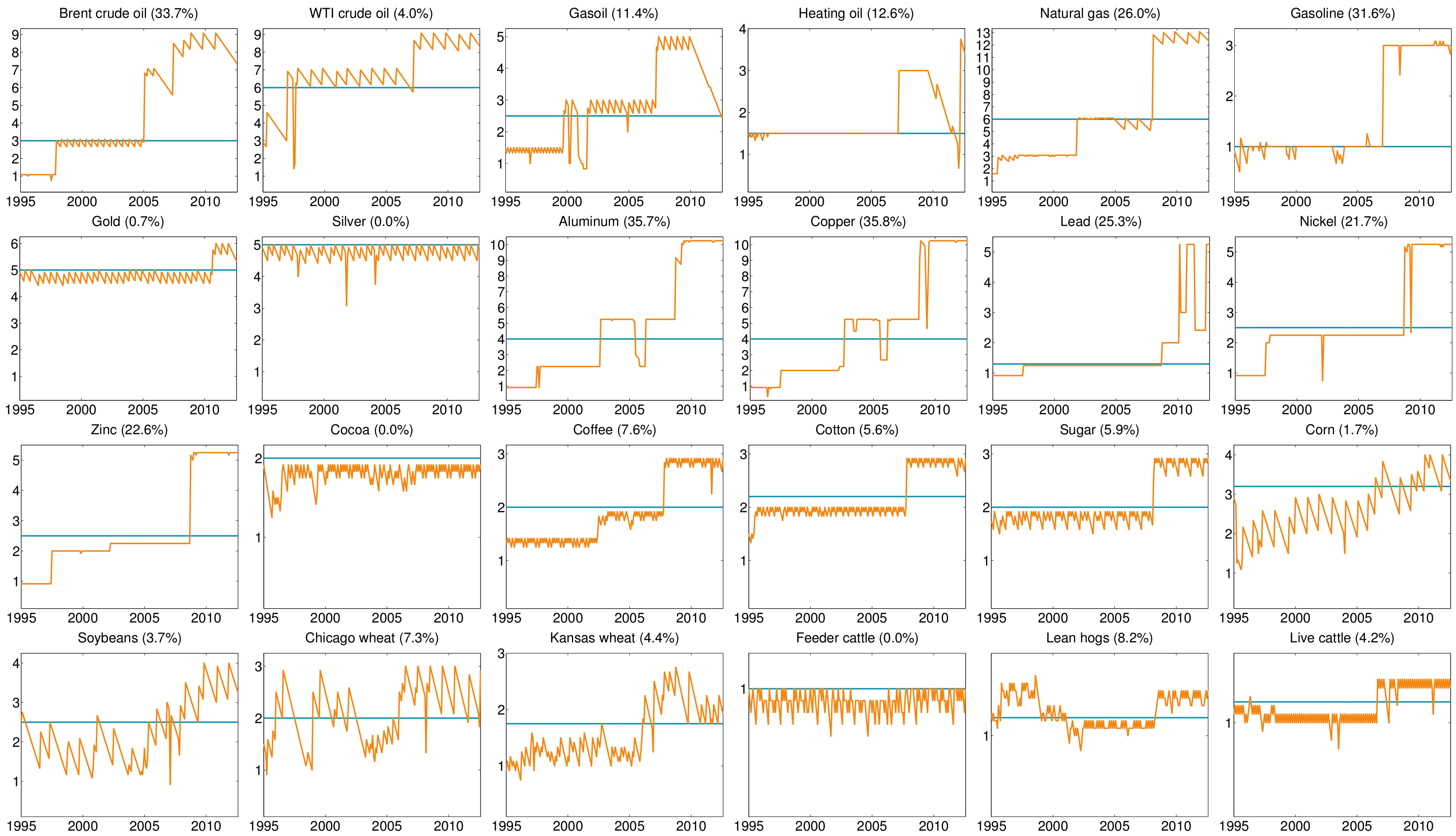
where $y = \{$*market, sector, idiosyncratic*$\}$, and the disturbances $\eta_{y,t} = (\eta_{y,t}^L, \eta_{y,t}^S, \eta_{y,t}^C)$ are normally distributed with covariance matrix Σ_{η_y}.

B Maturity bound

We introduce a maturity bound to exclude long-dated contracts. As discussed in Section 3, by limiting the term-structure dimension variation within a commodity we can keep using a fixed λ value. Furthermore, these long-dated contracts are possibly less liquid and hence have more noisy price information, which could otherwise affect our results. The introduced maturity bound excludes on average 10% of our observations (all at high maturities). Figure B.1 gives more details on individual commodities. Note that the longest available maturity varies substantially over time.

Figure B.1 Maturity bounds

These figures show for each commodity the maturity bound (blue lines) and the longest available maturity (orange lines, in years).



C Additional individual commodity results

Detailed individual model results

Table C.1 provides the estimation results for all individual model specifications. Based on these results we decide on the final model specification for every commodity. The final model choice is based on several criteria. First, we compare the results of three-factor models with and without seasonal term, to see if the exposure κ to the seasonal correction is significantly different from zero. There are 13 commodities in Table C.1 that have highly significant κ parameters, ranging between 0.73 and 7.78. There is a big gap between the 11 commodities which have a t-stat below 5, and the remaining 13 commodities which have a t-stat above 400. We do not include a seasonal correction when κ is below 0.1. Second, for the non-seasonal commodities, we need to decide if we include a third curvature factor or not. Non-reported AIC or BIC values indicate that we should always choose for the larger models. However, the λ values for some metal commodities in Table C.1 are below 0.5 which leads to slope and curvature loadings that are close to opposite with a correlation below -0.80. Therefore we decide to exclude the curvature factor for all metal commodities.

Figures C.1, C.2, and C.3 show the extracted level, slope, and curvature factors per commodity sector. In general, we see a similar pattern in all level factors. Until 2004 they are relatively constant, then they increase until they peak in 2008, whereafter they again remain constant. The level factors within the energy, metals and grains sectors seem to comove the most. The slope factors in Figure C.2 show some peaks and troughs. Especially for the energy commodities we see a sharp decline in 2008 and a gradual increase thereafter. As the Nelson-Siegel loading on the slope factor in (2) is a decreasing function of maturity, a negative factor estimate signifies an upward sloping (i.e. contangoed) futures curve. This implies that in 2008 all the backwardated energy futures curves quickly went into contango, and only gradually returned back to being backwardated. Last, the curvature factors in Figure C.3 show again some degree of comovement. Note that the metal commodities are missing, as we find that two factors are enough to capture their curve dynamics. Of the four sectors the energy commodities have the strongest comoving curvature factors. Based on the plots of the individual factors, there seems to be commonality across commodities. The advantage of our framework is that we can easily accommodate this.

Table C.1 All individual model results

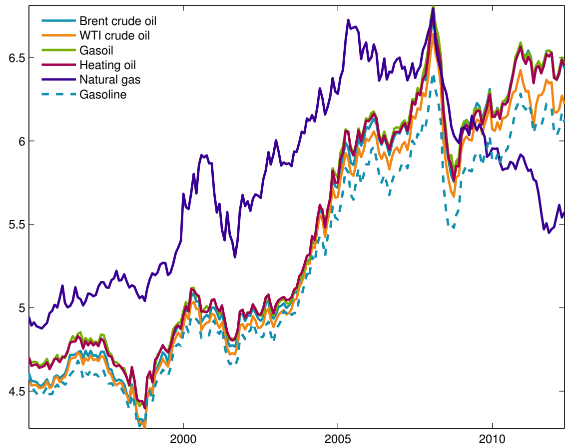
The table presents the estimation results and the fit of the individual commodity state space models. We show the results of three different specification applied to all individual commodities. The two-factor (2F) model contains only a level and slope factor; the three-factor (3F) model includes a curvature factor; and the 3FS model adds a seasonal correction term. For each model we present the estimated parameter values for the decay parameter λ, and (if relevant) the exposure κ to the seasonal term and the most expensive contract expiry month θ, where $\theta = 0$ corresponds to January. The last column presents the volatility of errors $\sigma(\nu)$. Standard errors of all estimates are provided between brackets.

		two-factor (2F)				three-factor (3F)				three-factor and seasonal (3FS)							
Sector	Commodity	λ		$\sigma(\nu)$	R^2	λ		$\sigma(\nu)$	R^2	λ		κ		θ		$\sigma(\nu)$	R^2
Energy	Brent crude oil	0.884	(0.01)	1.28%	99.96%	1.144	(0.01)	0.65%	99.99%	1.143	(0.01)	0.04	(0.99)	1.1	(0.02)	0.65%	99.99%
	WTI crude oil	1.262	(0.01)	1.49%	99.95%	1.272	(0.01)	0.79%	99.99%	1.273	(0.01)	0.03	(1.24)	0.1	(0.03)	0.79%	99.99%
	Gasoil	1.644	(0.02)	1.41%	99.96%	1.991	(0.04)	1.17%	99.97%	1.688	(0.04)	0.92	(0.02)	0.4	(0.00)	1.03%	99.98%
	Heating oil	1.956	(0.04)	2.40%	99.88%	3.392	(0.06)	2.17%	99.90%	3.600	(0.03)	2.80	(0.01)	0.8	(0.00)	1.21%	99.97%
	Natural gas	0.822	(0.03)	6.39%	98.07%	1.171	(0.03)	5.72%	98.45%	1.137	(0.02)	6.38	(0.01)	0.9	(0.00)	3.56%	99.40%
	Gasoline	2.298	(0.07)	3.59%	99.67%	4.346	(0.04)	3.11%	99.75%	3.285	(0.04)	4.78	(0.01)	6.2	(0.00)	1.52%	99.94%
Metals	Gold	0.011	(6.36)	0.44%	99.99%	0.129	(0.08)	0.19%	100.00%	0.129	(0.08)	0.00	(22.65)	0.1	(0.79)	0.19%	100.00%
	Silver	0.026	(1.79)	0.54%	99.99%	0.278	(0.05)	0.35%	100.00%	0.278	(0.05)	0.02	(2.08)	2.6	(0.04)	0.35%	100.00%
	Aluminium	0.187	(0.04)	0.77%	99.91%	0.574	(0.01)	0.32%	99.98%	0.574	(0.01)	0.01	(2.81)	9.6	(0.03)	0.32%	99.98%
	Copper	0.111	(0.10)	1.13%	99.97%	0.489	(0.01)	0.41%	100.00%	0.489	(0.01)	0.02	(1.35)	7.3	(0.02)	0.41%	100.00%
	Lead	0.324	(0.13)	0.57%	99.99%	1.771	(0.02)	0.36%	100.00%	1.767	(0.02)	0.04	(0.72)	9.2	(0.02)	0.36%	100.00%
	Nickel	0.095	(0.28)	0.91%	99.98%	1.250	(0.01)	0.71%	99.99%	1.250	(0.01)	0.00	(30.31)	4.5	(0.17)	0.71%	99.99%
	Zinc	0.059	(0.19)	0.69%	99.97%	1.132	(0.01)	0.32%	99.99%	1.131	(0.01)	0.03	(0.79)	9.6	(0.02)	0.32%	99.99%
Softs	Cocoa	0.360	(0.10)	0.74%	99.96%	1.413	(0.03)	0.58%	99.97%	1.415	(0.03)	0.08	(0.71)	8.8	(0.02)	0.58%	99.97%
	Coffee	0.954	(0.03)	0.80%	99.96%	1.468	(0.03)	0.57%	99.98%	1.468	(0.03)	0.00	(56.48)	0.1	(1.61)	0.57%	99.98%
	Cotton	1.484	(0.03)	1.76%	99.45%	3.566	(0.03)	1.37%	99.67%	3.482	(0.03)	0.91	(0.06)	5.9	(0.01)	1.21%	99.74%
	Sugar	1.610	(0.05)	2.13%	99.76%	3.585	(0.04)	1.53%	99.88%	3.272	(0.04)	1.13	(0.05)	2.5	(0.01)	1.32%	99.91%
Grains	Corn	1.603	(0.03)	2.37%	99.57%	0.696	(0.18)	2.10%	99.67%	2.743	(0.04)	1.80	(0.03)	5.7	(0.00)	1.56%	99.81%
	Soybeans	1.284	(0.05)	1.82%	99.73%	3.536	(0.03)	1.46%	99.83%	3.350	(0.03)	1.30	(0.03)	5.7	(0.00)	1.19%	99.89%
	Chicago wheat	1.345	(0.05)	2.09%	99.70%	1.143	(0.15)	1.86%	99.76%	1.496	(0.17)	1.53	(0.04)	2.0	(0.00)	1.51%	99.84%
	Kansas wheat	1.027	(0.11)	1.94%	99.72%	3.318	(0.07)	1.64%	99.80%	2.461	(0.08)	1.44	(0.05)	2.2	(0.01)	1.35%	99.86%
Meats	Feeder cattle	2.532	(0.07)	0.96%	99.83%	4.522	(0.05)	0.78%	99.89%	5.221	(0.07)	0.73	(0.07)	10.0	(0.01)	0.71%	99.91%
	Lean hogs	1.701	(0.18)	6.66%	89.78%	4.022	(0.04)	5.12%	93.95%	4.109	(0.07)	7.78	(0.02)	6.1	(0.00)	3.09%	97.80%
	Live cattle	3.200	(0.10)	2.19%	98.85%	4.099	(0.05)	1.77%	99.25%	4.487	(0.10)	2.32	(0.03)	1.9	(0.00)	1.29%	99.60%

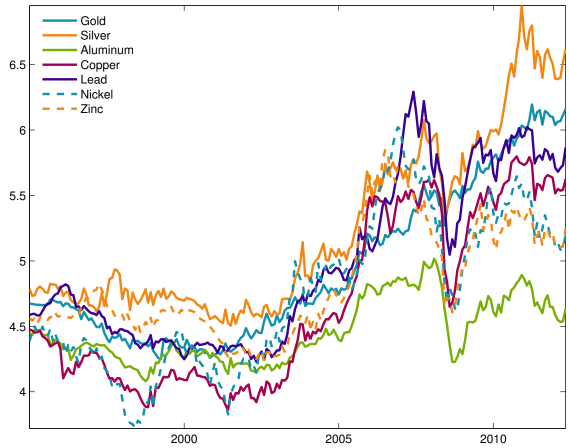
Figure C.1 Individual models - extracted commodity level factors

These figures show the extracted level factors based on individual models applied to all 24 commodities. Each subfigure shows the estimated level factors for the commodities of a specific sector.

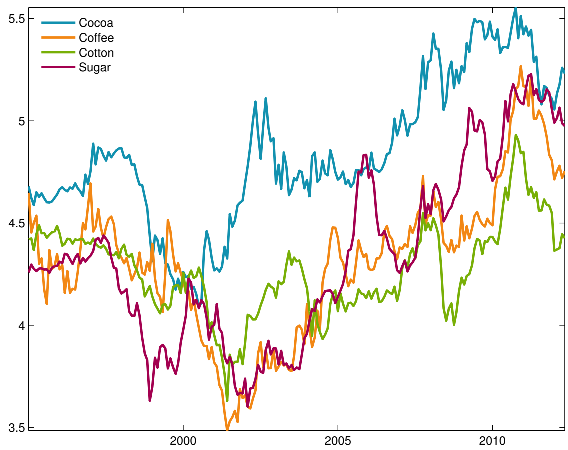
(a) Level - energy



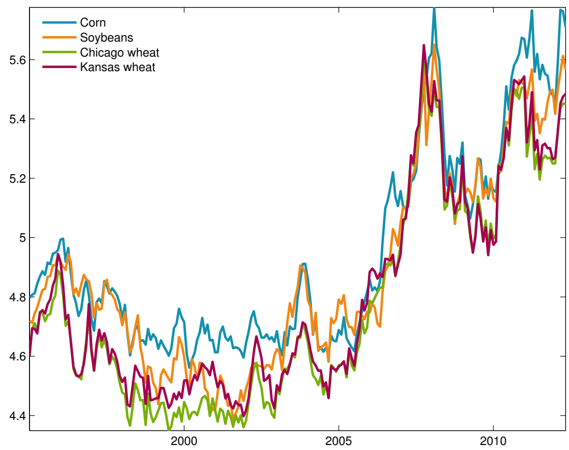
(b) Level - metals



(c) Level - softs



(d) Level - grains



(e) Level - meats

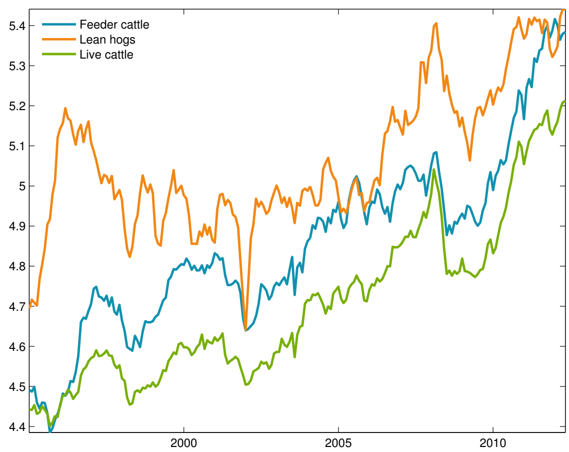
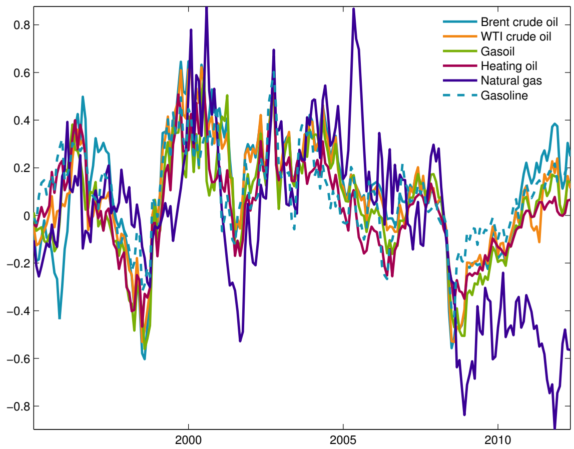


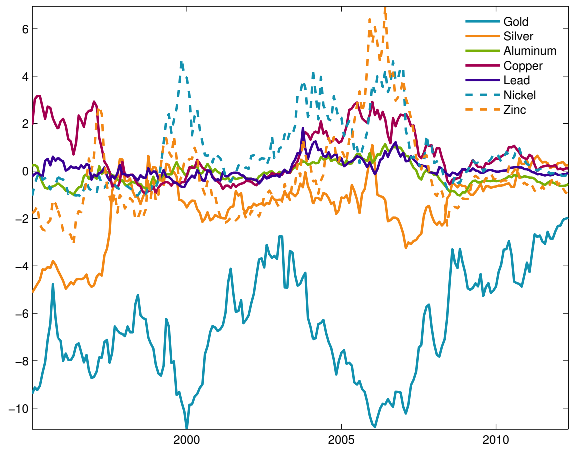
Figure C.2 Individual models - extracted commodity slope factors

These figures show the extracted slope factors based on individual models applied to all 24 commodities. Each subfigure shows the estimated slope factors for the commodities of a specific sector.

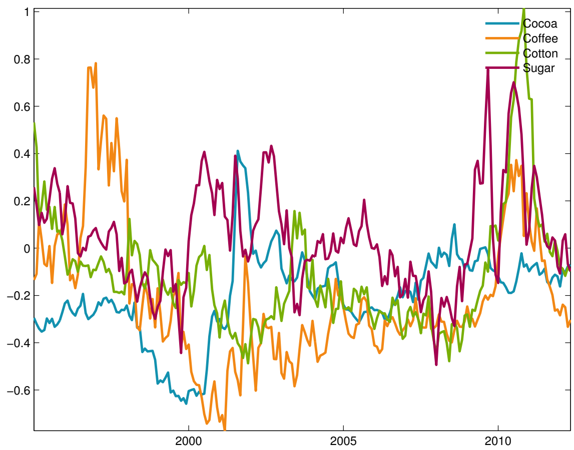
(a) Slope - energy



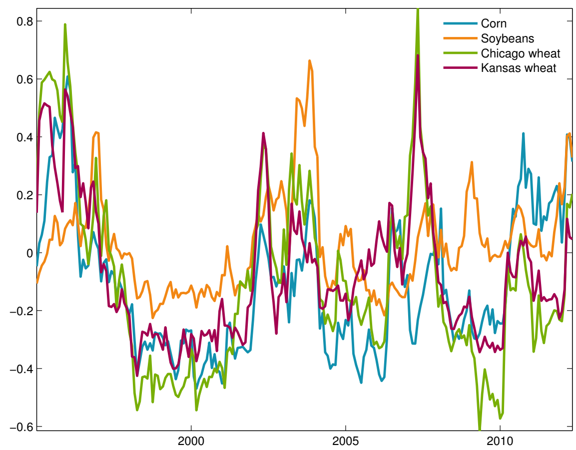
(b) Slope - metals



(c) Slope - softs



(d) Slope - grains



(e) Slope - meats

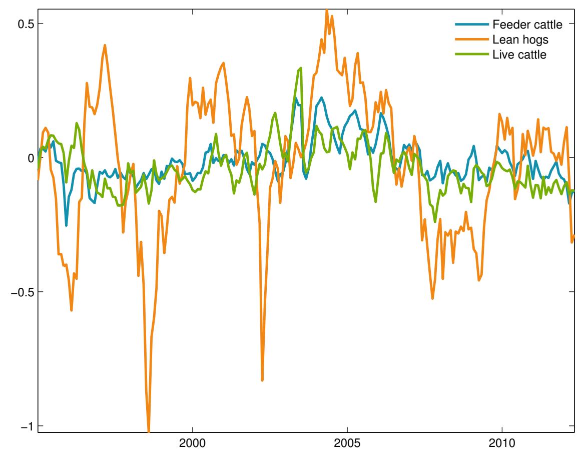
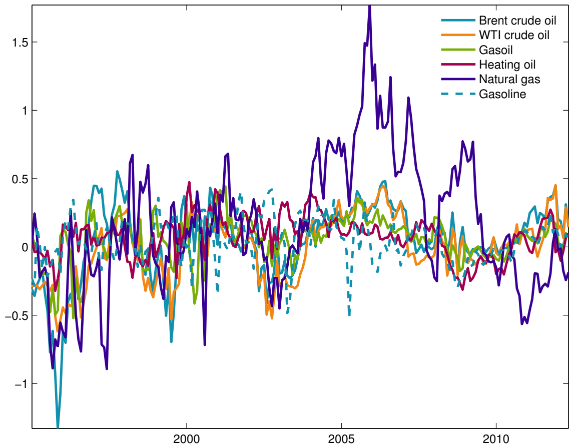


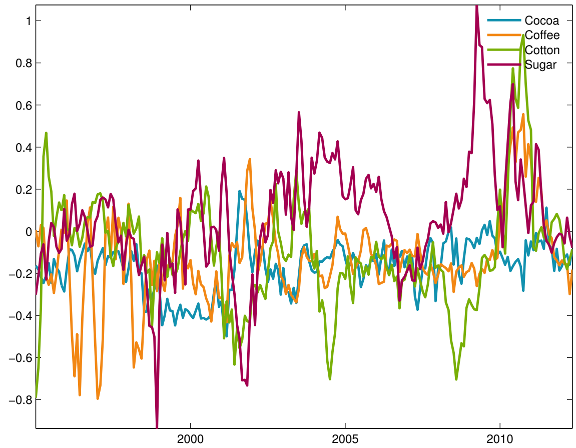
Figure C.3 Individual models - extracted commodity curvature factors

These figures show the extracted curvature factors based on individual models applied to all 24 commodities. Each subfigure shows the estimated curvature factors for the commodities of a specific sector.

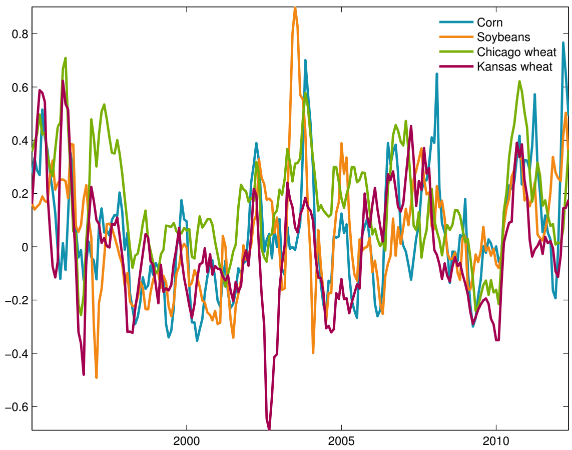
(a) Curvature - energy



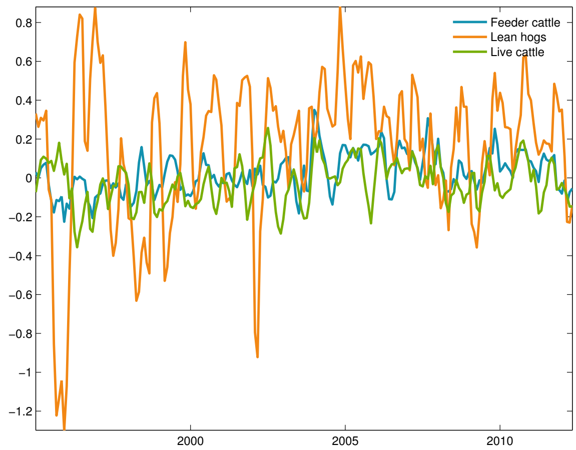
(b) Curvature - softs



(c) Curvature - grains



(d) Curvature - meats



Preliminary commonality results

A preliminary check for commonality is obtained by comparing the unobserved level, slope and curvature factors from the individual models. Using PCA, we check to what extent the factors can be explained by the first principal component. The market-wide level component is approximated by the first principal component when PCA is applied to all 24 extracted level factors. For each commodity we can compute the fraction of "individual" variance that is explained by this first principal component. We then take the average over all commodities in a particular sector. The same analysis is also applied on all slope factors and on all curvature factors. If the first principal explains a large part of the factors variations, it is an indication of commonality.

Panel A in Table C.2 shows that there seems to be a market-wide component that drives the level factors as the first principal component explains 78.7% of the variation in individual level factors. Especially, the energy, metals, and grains level factors comove with the market-wide level component, as more than 80% of their variation is explained. For the other two commodity sectors we observe that 56.1% and 77.3% of their variation is explained by the market-wide level component. Investigating the market-wide slope component shows that there is less comovement on average indicated by the explained variation of 26.6%. The decomposition in sectors shows that the market-wide slope component still explains half of the variation in the energy slope factors, while it hardly explains variation for the softs and grains sectors. The market-wide curvature component shows similar results as the market-wide slope component. In general, the market-wide curvature component explains 19.9% of the variation in the individual commodity curvature factors. For energy the percentage of explained variation is much higher (38.2%), while for the softs sector it explains none of the variation at all. In Panel B of the table, the analysis is repeated only at a sector-level. At this disaggregated level, the explained variation is due to both the common market-wide and sector-specific components. Compared to Panel A, particularly the slope and curvature commonality are greatly increased. This implies that for these sectors the sector-specific component is important in explaining comovement.

Table C.2 PCA commonality results

The table presents the percentage of explained variation of the extracted level, slope and curvature factors for the first three principal components. Panel A shows how much variation is explained by the market-wide level, slope and curvature components. In Panel B the percentages refer to the explained variation when PCA is applied to a particular sector, i.e. this variation can be due to both the market-wide and sector-specific component. The differences in percentages explained variation between Panel A and B, indicate the comovement due to the sector-specific component.

	level factor			slope factor			curvature factor		
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
Panel A Market-wide component									
1) all commodities	78.7%	9.8%	3.5%	26.5%	17.0%	14.9%	19.9%	16.7%	9.4%
2) energy	80.9%	16.3%	0.3%	58.6%	2.4%	22.8%	38.2%	5.9%	12.4%
3) metals	86.5%	1.4%	6.7%	23.9%	12.5%	25.8%	-	-	-
4) softs	56.1%	22.4%	2.7%	6.5%	13.2%	1.4%	0.0%	0.0%	0.0%
5) grains	85.5%	7.0%	1.1%	1.9%	61.5%	4.0%	0.1%	1.1%	7.9%
6) meats	77.3%	3.4%	6.7%	27.9%	2.6%	6.1%	6.5%	11.5%	12.1%
Panel B Market-wide and Sector component									
1) energy	93.5%	6.4%	0.0%	79.1%	11.7%	4.9%	46.0%	21.4%	12.4%
2) metals	90.3%	5.8%	1.7%	49.9%	24.4%	12.2%	-	-	-
3) softs	77.3%	12.8%	6.9%	43.4%	26.3%	20.3%	33.6%	28.3%	24.1%
4) grains	97.8%	1.3%	0.8%	69.2%	19.3%	9.4%	60.4%	20.1%	11.0%
5) meats	90.4%	9.0%	0.6%	62.6%	26.2%	11.2%	47.6%	35.3%	17.1%

D Appropriateness of Nelson-Siegel model for commodities data

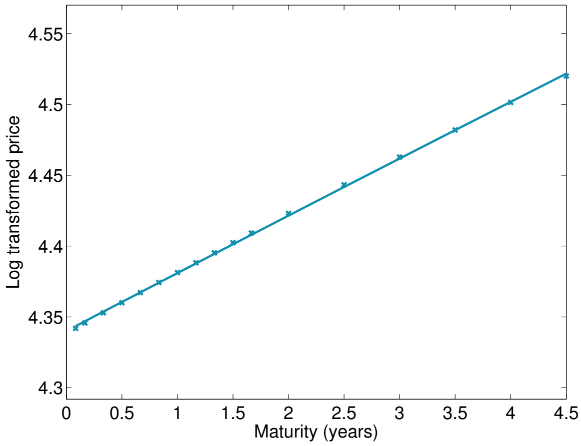
Figure D.1 shows representative examples of the model fit for three of the 24 commodities. Subfigures D.1.a and D.1.b correspond to the gold futures curve, Subfigures D.1.c and D.1.d correspond to the curve of WTI crude oil, while Subfigures D.1.e and D.1.f correspond to natural gas. The crosses represent the observed price data while the lines correspond to the fitted values of our models. The presented figures are snapshots at one particular point in time. For each of these three commodities we use a different version of our model. The futures curves of gold are almost straight lines. Hence, we can easily fit the prices with only a level and slope factor and we do not need a curvature factor or a seasonal term. The futures curves of WTI crude oil do not show seasonal patterns but do have a curved shape. In a static case, it would be possible to fit this curve with just a level and a slope factor. However, curves change over time and a two-factor model is not flexible enough to cope with these changes. The natural gas futures curve displays a pronounced seasonal pattern. Therefore, we use all three factors (level, slope and curvature) plus a seasonal term. In general our fitted values are close to the real prices, with some exceptions at the short end of the curve. The inclusion of the seasonal term seems an appropriate solution to model the periodic behavior.

To provide additional evidence that the Nelson and Siegel (1987) model is suited for commodity futures, we apply PCA to raw price data. In order to apply PCA, we need a balanced data panel. Therefore we exclude contract data if they have missing price data for more than 10% of the time periods. Then, for this selection, we exclude months where one of the contracts has missing price data. In the end we are left with a balanced sample with no missing observations. Figure D.2 shows the loadings of the first three principal components (based on the covariance matrix). For most commodities these loadings resemble the level, slope and curvature factor loadings. The only exceptions are some commodities with a pronounced seasonal pattern, namely heating oil, natural gas, gasoline, cotton and lean hogs. Due to the Samuelson (1965) effect, results based on the correlation matrix are even more similar to level, slope and curvature loadings.

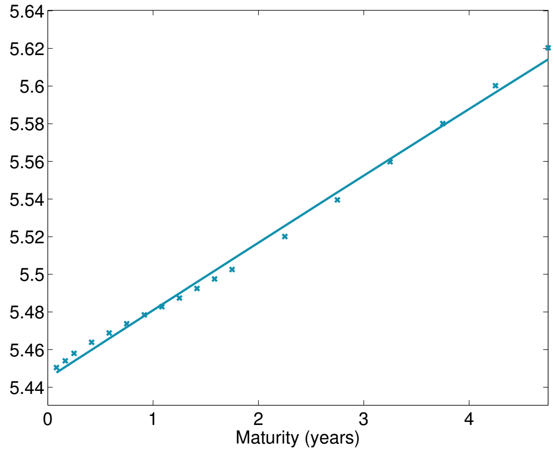
Figure D.1 Model fit commodity futures curves

These figures show example of the fit of our individual models. The crosses represent the observed price data while the lines correspond to the fitted values of our models. The raw prices are first standardized and thereafter we apply a log-transformation. The horizontal axis shows the time to maturity (τ) in years.

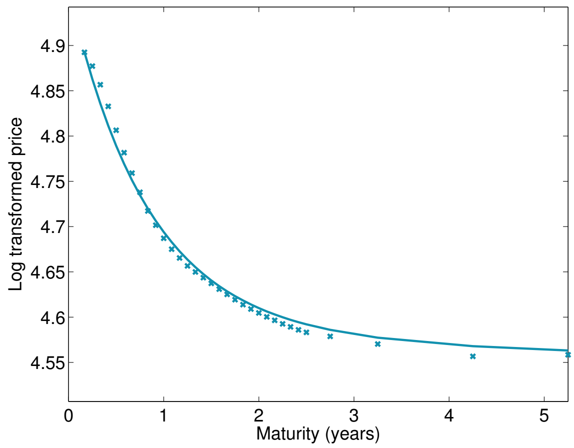
(a) Gold January 2000



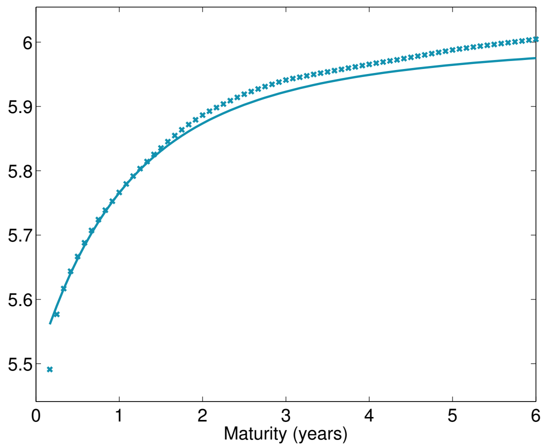
(b) Gold November 2008



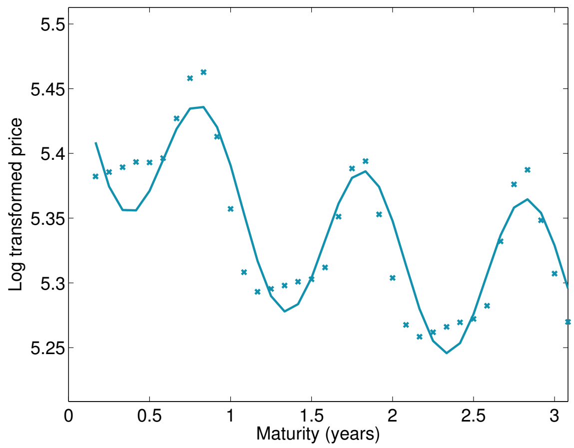
(c) WTI crude oil January 2000



(d) WTI crude oil November 2008



(e) Natural gas January 2000



(f) Natural gas November 2008

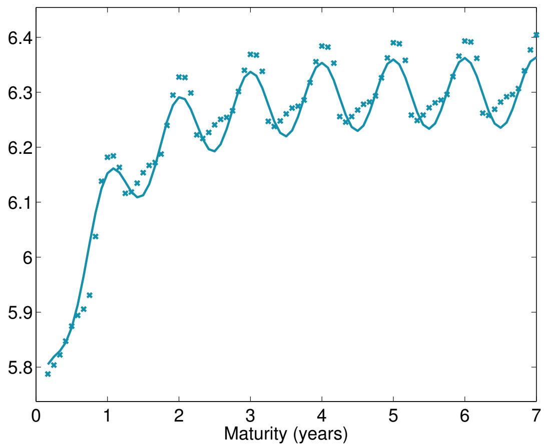
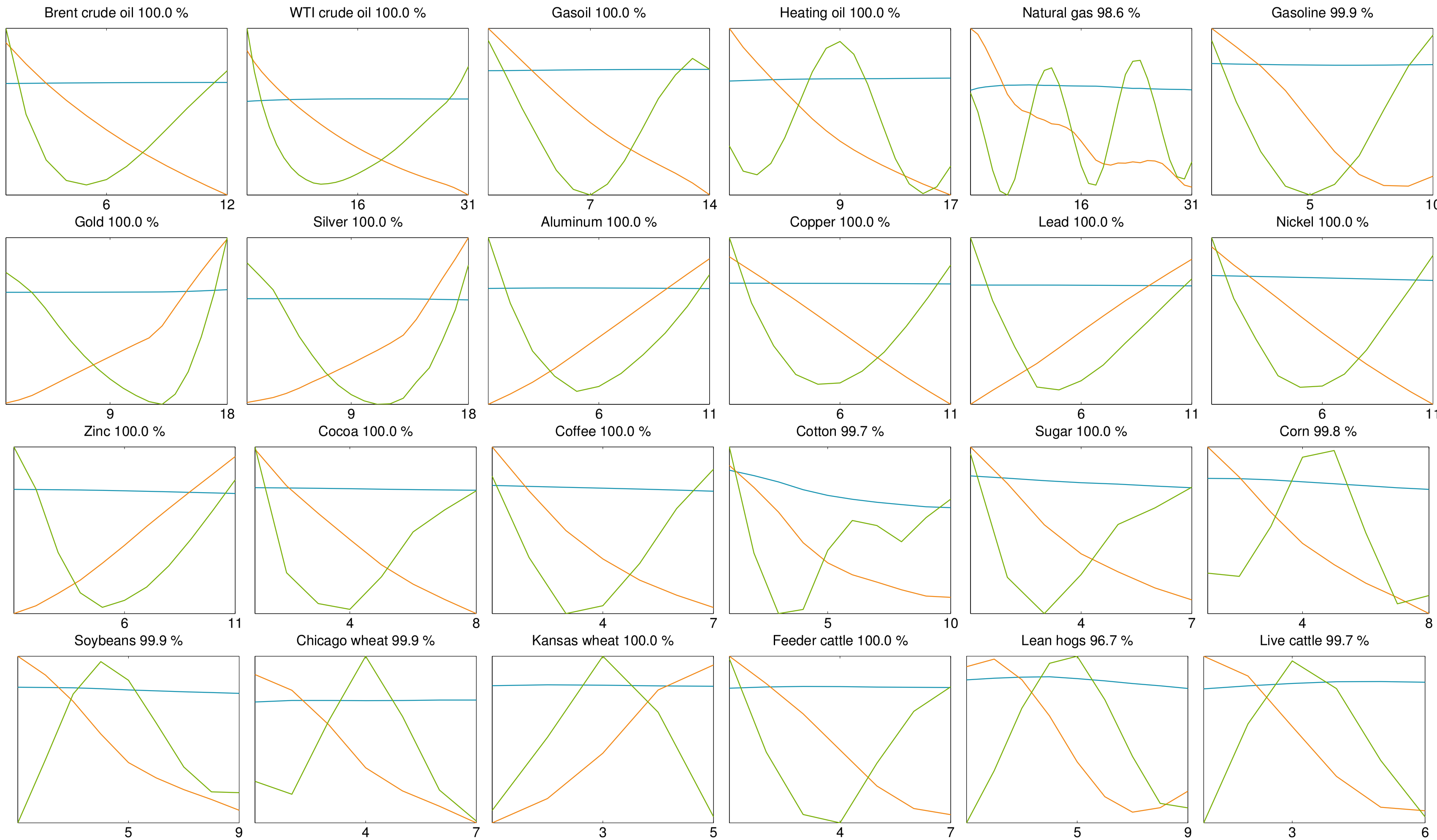


Figure D.2 Principal components

These figures show for each commodity the first three principal component loadings extracted from the covariance matrix of raw prices (where blue is the first, orange the second, and green the third principal component). Contracts are included when they have valid observations for at least 90% of the months in our sample.



E Comparison with the Schwartz (1997) three-factor model

Existing models like in Gibson and Schwartz (1990), Schwartz (1997) and Schwartz and Smith (2000) all begin by assuming a functional form for a set of underlying state variables. The futures curve can be derived from these state variables under no arbitrage conditions. Even though our framework is different, both approaches assume that commodity prices are driven by unobserved factors. Hence it is possible that the extracted unobserved factors from both methods are similar, e.g., through factor rotation. In this section we compare the results of our Nelson-Siegel type model with the three-factor model described in Schwartz (1997).

Extending the model of Gibson and Schwartz (1990), Schwartz (1997) assumes that commodity prices are driven by three stochastic factors namely the commodity spot price, the convenience yield and the interest rate.[1] Variations on this approach are given by many subsequent papers on commodity prices (see among others Schwartz and Smith, 2000; Casassus and Collin-Dufresne, 2005). Both the log spot price and the convenience yield are assumed to be mean reverting. Also the instantaneous interest rate is assumed to follow a mean reverting process as in Vasicek (1977).

Schwartz (1997) estimates a simplified version of his three-factor model by assuming that the interest rate process is independent of both commodity processes. He first estimates the interest rate parameters and then plugs these into the model.[2] The loadings on the unobserved log spot price and the instantaneous convenience yield show great resemblance with the loadings on our level and slope factors.

Figure E.1 shows the unobserved level factor of our model and the unobserved spot price of the three-factor Schwartz (1997) model, while Figure E.2 compares our slope factor with the unobserved convenience yield. Both models are estimated using the same dataset. The similarities are very clear in both figures. The level factor and the spot price both show, in general, an increasing trend with a pronounced dip around the recent financial crisis. The slope and convenience yield factors show more peaks and troughs, which implies upward and downward sloping futures curves. The resemblance of all lines is confirmed by the pair-wise correlations. The average (median) correlation is 0.67 (0.79) between the level factor and the spot price, and 0.76 (0.86) between the slope factor and

[1] Brennan (1991) defines the convenience yield as "the flow of services which accrues to the owner of a physical inventory but not to the owner of a contract for future delivery"

[2] For the complete model specification we refer to page 933 in Schwartz (1997). The interest rate process is estimated separately from the spot price and convenience yield processes, and is based on an observed 3-month Treasury Bill series. Essentially, this simplified version contains only two unobserved states, while still allowing for a time-varying interest rate. When estimating his model we follow the same estimation methodology.

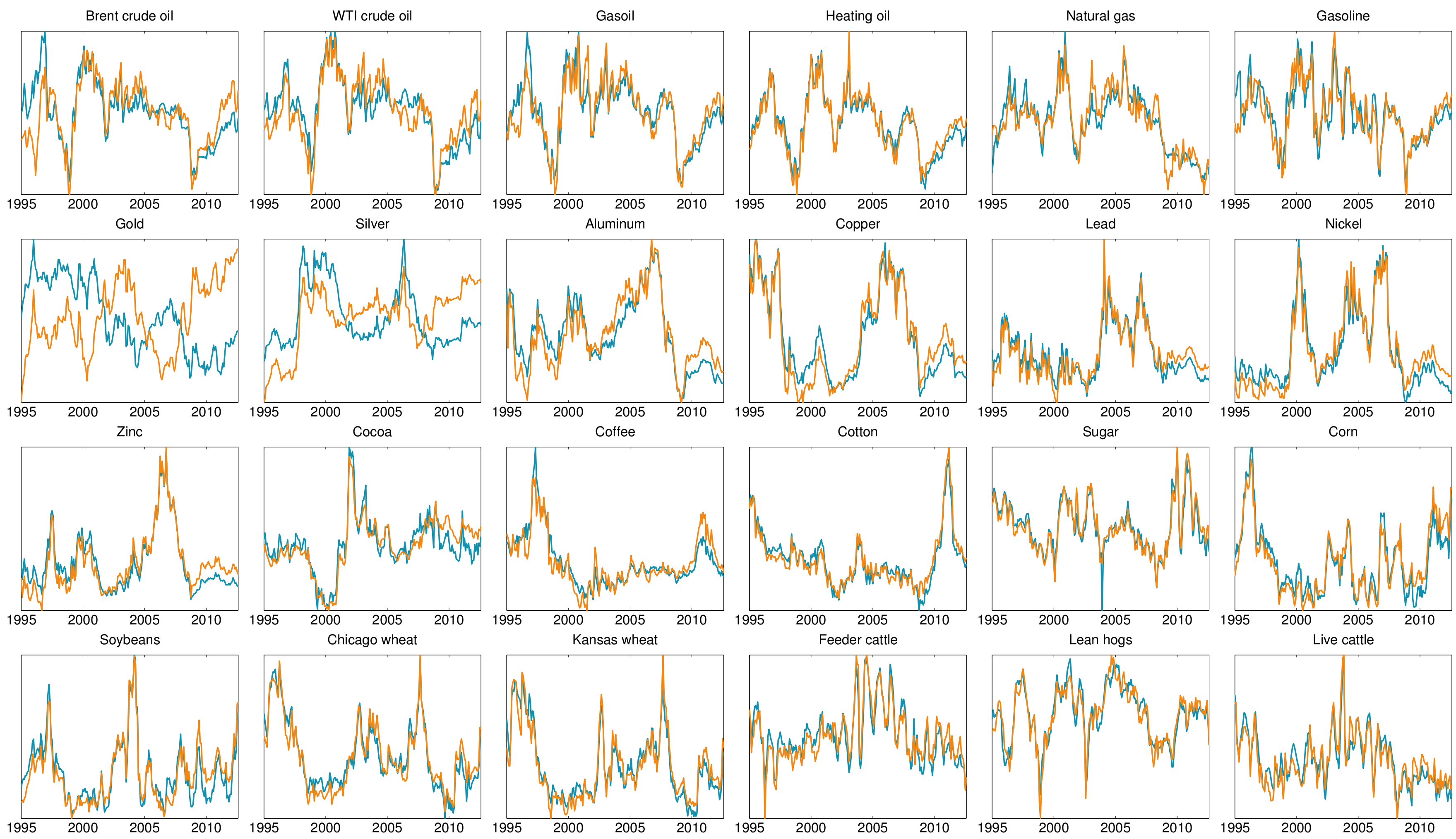
the convenience yield, respectively. Concluding, our statistical factors level and slope are strongly related to the spot price and convenience yield.

Figure E.1 Comparison level factor and spot price

These figures show the unobserved level factor of our Nelson-Siegel type models and the unobserved spot price series of the three-factor Schwartz (1997) model. The blue line is the unobserved spot price series and the orange line is our level factor.

Figure E.2 Comparison slope factor and convenience yield

These figures show the unobserved slope factor of our Nelson-Siegel type model and the unobserved convenience yield series of the three-factor Schwartz (1997) model. The blue line is the unobserved convenience yield and the orange line is our slope factor.



Although our level and slope factors show great resemblance with Schwartz' factors, we have an additional third factor, namely curvature. This gives us additional flexibility to better fit the observed futures prices. In terms of R^2 we increase the model fit by 5.6%. When we examine the residuals of the Schwartz (1997) model, we find that for most commodities there is a strong common factor present. These common factors have on average a correlation of 0.20 with our corresponding curvature factors.

Table E.1 presents results on the differences in model fit between our Nelson-Siegel framework and the Schwartz (1997) model. In Section 4.1 Table 2 we present the fit of our individual models. In Table E.1 we present these same numbers together with the R^2 of the corresponding Schwartz (1997) model. For all commodities our model has a better fit. The minimum fit increase is 1.6%, the average increase is 5.6% and the maximum increase is 29.2%.

Table E.1 Model fit comparison

This table compares the differences in fit $\left(R^2\right)$ between the Schwartz (1997) model and our Nelson-Siegel type model. The difference is computed as $(R^2_{\text{Nelson-Siegel}} - R^2_{\text{Schwartz}})/R^2_{\text{Nelson-Siegel}}$.

Commodity	Schwartz	Nelson-Siegel	diff.	Commodity	Schwartz	Nelson-Siegel	diff
Brent crude oil	97.06%	99.99%	3.0%	Cocoa	93.22%	99.97%	7.2%
WTI crude oil	97.28%	99.99%	2.8%	Coffee	94.76%	99.98%	5.5%
Gasoil	98.40%	99.98%	1.6%	Cotton	88.66%	99.74%	12.5%
Heating oil	98.41%	99.97%	1.6%	Sugar	96.51%	99.91%	3.5%
Natural gas	92.14%	99.40%	7.9%				
Gasoline	97.43%	99.94%	2.6%	Corn	95.72%	99.81%	4.3%
				Soybeans	95.32%	99.89%	4.8%
Gold	96.73%	99.99%	3.4%	Chicago wheat	95.47%	99.84%	4.6%
Silver	96.13%	99.99%	4.0%	Kansas wheat	95.15%	99.86%	5.0%
Aluminium	90.29%	99.91%	10.7%				
Copper	96.99%	99.97%	3.1%	Feeder cattle	97.51%	99.91%	2.5%
Lead	98.05%	99.99%	2.0%	Lean hogs	75.68%	97.80%	29.2%
Nickel	96.45%	99.98%	3.7%	Live cattle	95.77%	99.60%	4.0%
Zinc	94.80%	99.97%	5.5%				

F Macroeconomic data

Table F.1 lists all macroeconomic and commodity-specific data series. For each series we provide their name, code, transformation, category, source, and description. An overview of the categories is shown in Table F.2 while the transformation codes are explained in Table F.3.

Table F.1 Data series

The table presents all macroeconomic and commodity specific data series (in line with Stock and Watson, 2012, and Gorton, Hayashi & Rouwenhorst, 2013). For each series we provide their name, code, transformation, category, source, and description. An overview of the categories is shown in Table F.2 while the transformation codes are explained in Table F.3. The codes correspond to the database identifiers of the source.
Used abbreviations: St. Louis, Federal Reserve Economic Data (FRED); Commodity Futures Trading Commission (CFTC); Department of Energy (DOE); Intercontinental Exchange (ICE); U.S. Department of Agriculture (USDA).
[a] Author's website: `http://people.stern.nyu.edu/jwurgler/data/Investor_Sentiment_Data_v23_POST.xlsx`
[b] We follow the details given in Appendix B of Gorton, Hayashi & Rouwenhorst (2013)

Name	Code	T	Cat	Source	Description
Cons-Dur	DNDGRG3M086SBEA	5	1	FRED	Personal consumption expenditures: Nondurable goods, Price index (2009=100), SA
Cons-NonDur	DPCERA3M086SBEA	5	1	FRED	Real personal consumption expenditures, Quantity index (2009=100), SA
Cons-Serv	DSERRG3M086SBEA	5	1	FRED	Personal consumption expenditures: Services, Price index (2009=100), SA
Exports	USEXNGS.B	5	1	Datastream	Real exports
Imports	USIMNGS.B	5	1	Datastream	Real imports
IP: cons dble	IPDCONGD	5	2	FRED	Industrial Production: Durable Consumer Goods Index (2007=100), SA
IP: cons nondble	IPNCONGD	5	2	FRED	Industrial Production: Nondurable Consumer Goods Index (2007=100), SA
IP: bus eqpt	IPBUSEQ	5	2	FRED	Industrial Production: Business Equipment Index (2007=100), SA
IP: dble mats	IPDMAT	5	2	FRED	Industrial Production: Durable Materials Index (2007=100), SA
IP: nondble mats	IPNMAT	5	2	FRED	Industrial Production: nondurable Materials Index (2007=100), SA
IP: mfg	IPMANSICS	5	2	FRED	Industrial Production: Manufacturing (SIC) Index (2007=100), SA
IP: fuels	IPUTIL	5	2	FRED	Industrial Production: Electric and Gas Utilities Index (2007=100), SA
NAPM prodn	NAPMPI	1	2	FRED	ISM Manufacturing: Production Index, SA
Capacity Util	TCU	1	2	FRED	Capacity Utilization: Total Industry % of Capacity, SA
Emp: mining	CES1021000001	5	3	FRED	All Employees: Mining and Logging: Mining, Thous. of Persons, SA
Emp: const	USCONS	5	3	FRED	All Employees: Construction, Thous. of Persons, SA
Emp: dble gds	DMANEMP	5	3	FRED	All Employees: Durable goods, Thous. of Persons, SA
Emp: nondbles	NDMANEMP	5	3	FRED	All Employees: Nondurable goods, Thous. of Persons, SA
Emp: services	SRVPRD	5	3	FRED	All Employees: Service-Providing Industries, Thous. of Persons, SA
Emp: TTU	USTPU	5	3	FRED	All Employees: Trade, Transportation & Utilities, Thous. of Persons, SA
Emp: wholesale	USWTRADE	5	3	FRED	All Employees: Wholesale Trade, Thous. of Persons, SA
Emp: retail	USTRADE	5	3	FRED	All Employees: Retail Trade, Thous. of Persons, SA

Continued on next page

Name	Code	T	Cat	Source	Description
Emp: FIRE	USFIRE	5	3	FRED	All Employees: Financial Activities, Thous. of Persons, SA
Emp: Govt	USGOVT	5	3	FRED	All Employees: Government, Thous. of Persons, SA
Emp. Hours	AWHI	5	3	FRED	Aggr. Wkly Hours: Prod. and Nonsuperv. Employ.: Total Private Industries (2002=100), SA
Avg hrs	CES0600000007	1	3	FRED	Avg. Wkly Hours of Prod. and Nonsuperv. Employ.: Goods-Producing Hours, SA
Overtime: mfg	AWOTMAN	2	3	FRED	Avg. Wkly Overtime Hours of Prod. and Nonsuperv. Employees: Manufacturing Hours, SA
U: all	UNRATE	2	4	FRED	Unemployment rate: all workers, 16 years and over, Percentage, SA
U: mean duration	UEMPMEAN	2	4	FRED	Average (Mean) Duration of Unemployment, Weeks, SA
U: < 5 wks	UEMPLT5	5	4	FRED	Number of Civilians Unemployed - Less Than 5 Weeks, Thous. of Persons, SA
U: 5-14 wks	UEMP5TO14	5	4	FRED	Number of Civilians Unemployed for 5 to 14 Weeks, Thous. of Persons, SA
U: 15+ wks	UEMP15OV	5	4	FRED	Number of Civilians Unemployed for 15 Weeks and Over, Thous. of Persons, SA
U: 15-26 wks	UEMP15T26	5	4	FRED	Number of Civilians Unemployed for 15 to 26 Weeks, Thous. of Persons, SA
U: 27+ wks	UEMP27OV	5	4	FRED	Number of Civilians Unemployed for 27 Weeks and Over, Thous. of Persons, SA
HStarts: NE	HOUSTNE	4	5	FRED	Housing Starts in Northeast Census Region, Thous. of Units, SAAR
HStarts: MW	HOUSTMW	4	5	FRED	Housing Starts in Midwest Census Region, Thous. of Units, SAAR
HStarts: S	HOUSTS	4	5	FRED	Housing Starts in South Census Region, Thous. of Units, SAAR
HStarts: W	HOUSTW	4	5	FRED	Housing Starts in West Census Region, Thous. of Units, SAAR
PMI	NAPM	1	6	FRED	ISM Manufacturing: PMI Composite Index, SA
NAPM new orders	NAPMNOI	1	6	FRED	ISM Manufacturing: New Orders Index, SA
NAPM vendor del	NAPMSDI	1	6	FRED	ISM Manufacturing: Supplier Deliveries Index, SA
NAPM Invent	NAPMII	1	6	FRED	ISM Manufacturing: Inventories Index, NSA
Orders (ConsGoods)	ACOGNO	5	6	FRED	Manufacturers New Orders for Cons. Goods Indus., Mil. of $, SA
Orders (NDCapGoods)	ANDENO	5	6	FRED	Manufacturers New Orders for Capital Goods: Nondef. Capital Goods Indus., Mil. of $, SA
CPI-core	CPIULFSL	6	7	FRED	Consumer Price Index for All Urban Cons.: All Items Less Food, Index (1982-84=100), SA
PCED	PCEPI	6	7	FRED	Personal Consumption Expenditures, Price Index (2009=100), SA
AHE: const	CES2000000008	5	8	FRED	Avg. Hourly Earnings of Prod. and Nonsuperv. Employees: Construction, $ per Hour, SA
AHE: mfg	CES3000000008	5	8	FRED	Avg. Hourly Earnings of Prod. and Nonsuperv. Employees: Manufacturing, $ per Hour, SA
FedFunds	FEDFUNDS	2	9	FRED	Effective Federal Funds Rate, % per annum, NSA
3mo T-bill	TB3MS	2	9	FRED	3-Month Treasury Bill: Secondary Market Rate, % per annum, NSA
3mo T-bill	TB6MS	2	9	FRED	6-Month Treasury Bill: Secondary Market Rate, % per annum, NSA
M1	M1SL	6	10	FRED	M1 Money Stock, Bil. of $, SA
M2	M2SL	6	10	FRED	M2 Money Stock, Bil. of $, SA
MB	AMBSL	6	10	FRED	St. Louis Adjusted Monetary Base, Bil. of $, SA
Reserves tot.	TOTRESNS	6	10	FRED	Total Reserves of Depository Institutions, Bil. of $, NSA
BUSLOANS	BUSLOANS	6	10	FRED	Commercial and Industrial Loans, All Commercial Banks, Bil. of $, SA
Cons credit	NONREVSL	6	10	FRED	Total Nonrevolving Credit Owned and Securitized, Outstanding, Bil. of $, SA
Ex rate: avg	TWEXMMTH	5	11	FRED	Trade Weighted U.S. Dollar Index: Major Currencies, Index (Mar 1973=100), NSA
Ex rate: Switz	EXSZUS	5	11	FRED	Switzerland / U.S. Foreign Exchange Rate Swiss Francs to 1 U.S., $, NSA
Ex rate: Japan	EXJPUS	5	11	FRED	Japan / U.S. Foreign Exchange Rate Japanese Yen to 1 U.S., $, NSA
Ex rate: UK	EXUSUK	5	11	FRED	U.S. / U.K. Foreign Exchange Rate U.S. $ to 1 British Pound, £, NSA
Ex rate: Canada	EXCAUS	5	11	FRED	Canada / U.S. Foreign Exchange Rate Canadian $ to 1 U.S., $, NSA

Continued on next page

Name	Code	T	Cat	Source	Description
S&P 500	SP500	5	12	FRED	S&P 500, Index, NSA
DJIA	USSHRPRCF	5	12	Datastream	Dow Jones Industrial Average
Consumer expect	UMCSENT	2	13	FRED	University of Michigan: Consumer Sentiment, Index (1966Q1=100), NSA
ADS	ADS	1	14	FRB of Phil.	Aruoba Diebold Scotti financial conditions index
Sentiment	Sentiment	1	14	Author's website[a]	Baker Wurgler paper
Baltic Dry index	BALTICF	1	14	Datastream	Dry bulk shipping price
Comm. vol.	CRBSPOT	1	15	Datastream	Volatility of Commodity Research Bureau (CRB) spot market price index
HP WTI	hedging WTI	1	16A	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP gasoline	hedging GL	1	16A	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP heating oil	hedging HO	1	16A	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP natural gas	hedging NG	1	16A	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP copper	hedging CP	1	16B	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP cocoa	hedging CC	1	16C	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP coffee	hedging CF	1	16C	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP sugar	hedging SG	1	16C	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP cotton	hedging CT	1	16C	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP soybeans	hedging S	1	16D	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP wheat	hedging W	1	16D	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP corn	hedging C	1	16D	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP feeder cattle	hedging FC	1	16E	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP live cattle	hedging LC	1	16E	CFTC	Hedging pressure computed from Commitment of Traders Reports
HP lean hogs	hedging LH	1	16E	CFTC	Hedging pressure computed from Commitment of Traders Reports
inventory crude oil	inventory WTI	5	17A	DOE	U.S. ending stocks excluding SPR of crude oil, thousands of barrels[b]
inventory gasoline	inventory GL	5	17A	DOE	U.S. motor gasoline ending stocks, thousands of barrels[b]
inventory heating oil	inventory HO	5	17A	DOE	U.S. total stocks of distillate fuel oil[b]
inventory natural gas	inventory NG	5	17A	DOE	U.S. total natural gas in underground storage (working gas), millions of cubic feet[b]
inventory gold	COMXGOLD Index	5	17B	Bloomberg	Comex warehouse stocks[b]
inventory silver	COMXSILV Index	5	17B	Bloomberg	Comex warehouse stocks[b]
inventory copper	LSCA Index	5	17B	Bloomberg	LME warehouse stocks[b]
inventory aluminium	LSAH Index	5	17B	Bloomberg	LME warehouse stocks[b]
inventory lead	LSPB Index	5	17B	Bloomberg	LME warehouse stocks[b]
inventory nickel	LSNI Index	5	17B	Bloomberg	LME warehouse stocks[b]
inventory zinc	LSZS Index	5	17B	Bloomberg	LME warehouse stocks[b]
inventory cocoa	inventory CC	5	17C	ICE	ICE wareh. stocks (ports of New York, Delaware River, Hampton Roads, Albany, Baltimore)[b]
inventory coffee	inventory CF	5	17C	ICE	ICE wareh. stocks (ports of New York, New Orleans, Houston, Miami, Antwerp, Hamburg, Barcelona)[b]
inventory cotton	inventory CT	5	17C	ICE	Hist. Certif. Stock Report (ports of Dallas, Galveston, Greenville, Houston, Memphis, New Orleans)[b]
inventory sugar	inventory SG	5	17C	USDA	U.S. sugar stocks held by primary distributors[b]
inventory corn	inventory C	5	17D	USDA	Stocks of Grain at Selected Terminals and Elevator Sites, thousands of bushels[b]
inventory soybeans	inventory S	5	17D	USDA	Stocks of Grain at Selected Terminals and Elevator Sites, Thousands of Bushels[b]
inventory wheat	inventory W	5	17D	USDA	Stocks of Grain at Selected Terminals and Elevator Sites, Thousands of Bushels[b]
inventory feeder cattle	inventory FC	5	17E	USDA	United States Cattle Placed on Feed in 7 States[b]

Continued on next page

Name	Code	T	Cat	Source	Description
inventory live cattle	inventory LC	5	17E	USDA	Frozen beef stocks in cold storage in the U.S.[b]
inventory lean hogs	inventory LH	5	17E	USDA	Frozen pork stocks in cold storage in the U.S.[b]

Table F.2 Data categories

The table presents the data category codes (following Stock and Watson, 2012).

Category code	Category name
1	GDP components
2	Industrial production
3	Employment
4	Unemployment rate
5	Housing
6	Business inventories
7	Prices
8	Wages
9	Interest rates
10	Money
11	Exchange rates
12	Stock prices
13	Consumer expectations
14	Financial conditions
15	Commodity volatility
16	Hedging pressure
17	Commodity inventories

Table F.3 Data transformations

The table presents the data transformation codes (following Stock and Watson, 2012). Z_t denotes the raw series and X_t the transformed series used to compute the principal components.

Transformation code	X_t
1	Z_t
2	$Z_t - Z_{t-1}$
3	$(Z_t - Z_{t-1}) - (Z_{t-1} - Z_{t-2})$
4	$\ln(Z_t)$
5	$\ln(Z_t/Z_{t-1})$
6	$\ln(Z_t/Z_{t-1}) - \ln(Z_{t-1}/Z_{t-2})$

References

Brennan, M., 1991: The price of convenience and the valuation of commodity contingent claims. In *Stochastic Models and Option Values: Applications to Resources, Environment and Investment Problems*, Lund, D. and Oksendal, B., editors. North-Holland, New York.

Casassus, J. and P. Collin-Dufresne, 2005: Stochastic convenience yield implied from commodity futures and interest rates. *Journal of Finance*, **60**(5), 2283–2331.

Gibson, R. and E. S. Schwartz, 1990: Stochastic convenience yield and the pricing of oil contingent claims. *Journal of Finance*, **45**(3), 959–976.

Gorton, G., F. Hayashi, and K. G. Rouwenhorst, 2013: The fundamentals of commodity futures returns. *Review of Finance*, **17**(1), 1–71.

Nelson, C. and A. Siegel, 1987: Parsimonious modeling of yield curves. *Journal of Business*, **60**(4), 473–489.

Samuelson, P., 1965: Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, **2**, 41–49.

Schwartz, E. S., 1997: The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, **52**(3), 923–973.

Schwartz, E. S. and J. E. Smith, 2000: Short-term variations and long-term dynamics in commodity prices. *Management Science*, **46**(7), 893–911.

Stock, J. H. and M. W. Watson, 2012: Generalized shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics*, **30**(4), 481–493.

Vasicek, O., 1977: An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**(2), 177–188.