EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Anastasia Allayioti, Fabrizio Venditti

The role of comovement and time-varying dynamics in forecasting commodity prices

No 2901

Abstract

Commodity prices co-move, but the strength of this co-movement changes over time due to structural factors, like changing energy intensity in production and consumption as well as changing composition of underlying shocks. This paper explores whether econometric models that exploit this co-movement and account for parameter instability provide more accurate point and density forecasts of ten major commodity indices viz-a-viz constant-coefficient models. Improvements in point forecast accuracy are small, with predictability varying substantially across forecast horizons and commodity indices, but they are large and significant in terms of density forecasting. An economic evaluation reveals that allowing for parameter time variation and commonalities leads to higher portfolios returns, and to higher utility values for investors.

JEL classification: C32 , C52, C53, C55, E37

Keywords: Commodities; Commonalities; Instabilities; Density forecasting; Economic evaluation

Non-technical summary

Commodity prices have been a recurrent source of concern for policymakers, given their effects on terms of trade, inflation, inflation expectations and disposable income. Forecasting commodity prices, in particular, plays a key role in inflation projections in central banks, which typically condition their inflation forecasts on a given path of oil and other commodity prices. The predictability of commodity prices therefore remains an active and policy-relevant area of applied research.

A striking feature of commodity prices, and one that might prove useful in forecasting, is the extent of their co-movement. Significant co-movement is found not only for the prices of closely related commodities but also for those that exhibit near-zero cross-price elasticities of demand and supply. The strength of this co-movement, however, varies significantly over time, owing to both structural as well as cyclical factors. Structural factors include the changing intensity of commodity use (both in production and consumption) as well as the rising importance of commodities in financial markets. Cyclical factors relate to the nature of the structural shocks affecting prices. Whereas demand shocks are a common source of fluctuations across prices, supply shocks are commodity specific, and they weaken co-movement across commodity prices. Motivated by these two stylized features (co-movement and instability), this paper investigates the merits of constructing forecasts for key commodity prices from models that exploit co-movement in large data sets and that can deal with structural breaks. Among others, we consider large dynamic factor models with time varying parameters (TVP) - including specifications that impose the presence of specific blocks on the factor structure of commodity prices - as well as their constant-coefficient counterparts. Evaluation is based on statistical criteria but also on economic grounds, in particular on the gains accrued to an investor who uses the model-implied forecasts for portfolio allocation.

In terms of forecast accuracy, we find that the largest improvements come from modeling time variation (especially with respect to the second-order moments) rather than from the way in which the factor structure of commodity prices is modeled. This is particularly evident when we evaluate the ability of competing specifications in generating accurate and well-calibrated predictive forecast densities. The economic evaluation of the competing forecasting models, on the other hand, reveals an economic value of allowing for both parameter time variation and commonalities.

1 Introduction

Commodity prices have been a recurrent source of concern for policymakers, given their effects on terms of trade, inflation , inflation expectations and disposable income. Forecasting commodity prices is also an important block of inflation projections in central banks. Also, commodities are now widely traded in derivative markets and form an important part of investors portfolios. In spite of numerous empirical studies exploring whether commodity prices can be predicted, and, if so, by which variables, forecasting commodity prices with reasonable accuracy remains a challenging endeavor. Future prices, the most natural candidate, have proven to be inadequate predictors of future spot prices (Hong and Yogo, 2012).

A striking feature of commodity prices, hence one that might prove useful in a forecasting context, is the extent of their co-movement. Significant co-movement is found not only for the prices of closely related commodities (think, for instance, of cocoa and sugar) but also for those of unrelated commodities i.e. those for which the cross-price elasticities of demand and supply are close to zero, according to the definition by Pindyck and Rotemberg (1990). Recent work aiming at explaining the mechanism behind the high-degree of co-movement since the turn of the twenty-first century includes, among others, Krugman (2008) who argues that the increase in oil prices was responsible for the subsequent increase in food prices. More recently, Alquist et al. (2014) show that the vast majority of historical commodity price movements can be attributed to a general-equilibrium response to aggregate non-commodity shocks rather than specific shocks to commodity markets. Building on this literature, Delle Chiaie et al. (2022), henceforth DFG (2022), extract a common factor across a large panel of commodity prices and validate its use as an indicator of the global business cycle.

Yet, the strength of the co-movement across commodity prices varies significantly over time, as we document in Section 2. This is due to a host of factors, both structural as well as cyclical. Structural factors include the changing intensity of commodity use (both in production and consumption) as well as the rising importance of commodities in financial markets. Cyclical factors relate to the nature of the structural shocks affecting prices. Indeed, whereas demand shocks are a common source of fluctuations across prices, supply shocks are commodity specific, and their occurrence weakens co-movement across commodity prices. In a recent contribution, for instance, Peersman et al. (2021) document sizeable spillovers across commodity markets, but find that the size of such spillovers (and therefore the intensity of price co-movement) changes over time.

Motivated by these two stylized features (co-movement and instability), this paper studies the performance of a specific class of econometric models, namely factor models with time-varying parameters, in forecasting commodity prices and indices using 68 spot commodity prices. Among the competing forecast specifications, we evaluate a dynamic factor model (DFM) with 10 factors, a hierarchical-DFM with the same number of factors that imposes a blocks structure, in the spirit of DFG (2022), and their respective time-varying counterparts. Our work assesses whether the accuracy of point and density forecasts for the price of a given commodity could benefit from (i) imposing a block structure in cross-correlations and (ii) dealing with structural breaks. The analysis is based on monthly data and evaluates the forecasting performance for the period 2001:M1-2021M3.

The first result of our analysis is that the largest forecast improvements come from modeling time variation, while the precise way in which co-movement is modeled is of lesser relevance. This is particularly evident when we evaluate the ability of competing specifications in generating accurate and well-calibrated predictive forecast densities. Second, we document that allowing for time-varying second-order moments substantially enhances the predictability of commodity prices. Third, we find that predictability is heterogeneous across forecast horizons and for different types of commodity prices. Finally, using a simple mean-variance portfolio allocation framework we perform an economic evaluation of the competing forecasting models. This analysis reveals an economic value of allowing for both parameter time variation and comovement.

Our paper is connected to three strands of the literature. The first is the stream of papers that explore the power of different predictors to forecast commodity prices. Chen et al. (2010) and Pesenti and Groen (2011) look at the relationship between exchange rates and commodity prices. Chen et al. (2010) find evidence that exchange rates predict commodity prices both in-sample and out-of-sample, while Pesenti and Groen (2011) document much weaker support for commodity exchange rates. Gargano and Timmermann (2014) focus on a set of different predictors taken from the stock return predictability literature, in addition to macroeconomic predictors such as inflation, money supply growth, growth in industrial production, and the unemployment rate, along with exchange rates for commodity currencies, and indicators of global economic activity. In an out-of-sample point forecasting exercise, DFG (2022) find that a DFM with constant parameters has adequate point forecasting properties at short horizons. We include their model in our empirical analysis and explore how it compares to other constant and time-varying specifications in terms of point and density forecast accuracy, as well as in terms of accrued economic gains. Our results illustrate how introducing time-varying dynamics enhances both statistical and economic gains.

We also contribute to the literature exploiting empirical factor models to summarize a wide range of information contained in commodity prices. For example, Byrne et al. (2013) investigate the relationship between commodity prices and macroeconomic determinants using a Factor Augmented Vector Auto Regression (FAVAR) approach. Moreover, Chen et al. (2014) show that the bulk of movements of 51 tradeable commodities is mostly due to the first common component. The main drawback with extracting a single common component from commodities is that the estimated factor can be difficult to interpret and may not fully account for heterogeneity (Moench et al., 2013; Byrne et al., 2019). Recent research has attempted to address these concerns. For example, Yin and Han (2015) use a multilevel factor model to decompose commodity returns into global, sectoral and idiosyncratic components.

Our paper is also related to the research on the role of structural instability in commodity markets. The majority of existing work on the determinants of commodity prices relies upon a time-invariant methodology (e.g. Poncela et al., 2014). There are, however, compelling reasons to believe that changes occur in the composition of underlying structural shocks over time (e.g. Kilian, 2009) and in their transmission mechanism (e.g. Riggi and Venditti, 2015; Blanchard and Gali, 2007).

Our study differs from past empirical work in many aspects. First, we evaluate the forecasting performance of various models instead of restricting our attention to a single model. The model space includes specifications that account for the factor structure of commodities, potential heterogeneity and the presence of structural breaks. This extensive comparison allows us to disentangle the importance of each one of these characteristics in improving forecast accuracy. To our knowledge, commodity heterogeneity and time-variation have not been simultaneously considered in the forecasting literature within a unified framework.[1] A second contribution of this paper is that we evaluate not only point, but also density forecasts. Third, we look beyond traditional statistical performance and perform an economic evaluation of the various forecast models by analysing whether an investor conditioning on the models' forecasts would be able

[1]For instance, DFG (2022) restrict their attention to a constant-parameter factor model that accounts for commodity heterogeneity but does not allow for the presence of instabilities. Conversely, Chen et al. (2010) account for parameter instability but only explore the predictive content of exchange rates for commodity prices.

to obtain tangible economic gains. To this purpose, we test whether an investor might rationally use the predicted return (and its estimated variance) for portfolio optimization. Our economic evaluation exercise is similar in spirit to existing studies forecasting stock returns (e.g. Rapach et al., 2010) and exchange rate fluctuations (e.g. Abbate and Marcellino, 2014).

The rest of the paper is organized as follows. Section 2 introduces the data and presents some stylized facts that motivate our analysis. Section 3 describes the theoretical framework of the constant and time-varying specifications under examination. Section 4 discusses estimation algorithms. Section 5 provides details about the nature of the forecasting setting. Section 6 presents the results of the analysis. Section 7 concludes.

2 Data and motivating evidence

The dataset consists of 68 monthly spot commodity prices in nominal U.S. dollars (see Table A2 in Appendix A).[2] The source of our data is the International Monetary Fund (IMF) primary commodity price database and covers the period January 1992 to March 2021. The IMF also constructs various price indices and sub-indices as the weighted average of the 68 individual commodity prices. The weights of the indices correspond to the global import share over a 3-year period (2014-2016).[3]

We forecast ten core indices that reflect the different categories of individual prices described in Table 1: one overall index for the entire set of commodities *All Commodity* that is constructed as the weighted average of the 68 individual commodity prices and 9 sub-indices that further decompose the individual prices into categories.[4] Two core blocks are represented by, first, the *Non-Fuel* index, which consists of 59 non-fuel prices and, second, the *Fuel* index which consists of 9 fuel prices. Subsequently, these two main blocks are further decomposed. The *Non-Fuel* index consists of the *Agriculture* index with 42 prices, that is further decomposed into the *Agriculture Raw Materials* index (9 prices), the *Food* index (29 prices) and the *Beverages* index (4 prices), the *Fertilizers* index with 3 prices and the *Metals* index (14 prices). [5] Similarly, the *Fuel* index, which consists of 9 fuel prices, is decomposed into the *Oil* index (3 prices) and the *Coal, Natural Gas & Propane* (*CNP*) index (4 prices).

[2]Note that the prices are period averages and are not seasonally adjusted.
[3]For some indices, data was not available starting from January 1992 so we re-constructed these using the IMF methodology.
[4]For the aggregate commodity index, *All Commodity* and *All* will be used interchangeably.
[5]To keep our hierarchical factor model parsimonious, we combine the *Agriculture* and *Fertilizers* indices into one.

Table 1: Description of the adopted block structure. Values in parenthesis reflect trade weights (%)

Global	*Blocks*	*Subblock*	*Groups*	*N*
All Commodities (100)				(68)
	Non-Energy (59.1)			(59)
		Agriculture (34.5)		(42)
			Agricultural Raw Materials (4.3)	(9)
			Beverages (2.3)	(4)
			Food (27.8)	(29)
		Fertilizers (1.9)		(3)
		Metals (22.7)		(14)
	Energy (40.9)			(9)
		Coal (3.0)		(2)
		Crude Oil (28.6)		(3)
		Natural Gas (7.8)		(3)
		Propane (1.5)		(1)

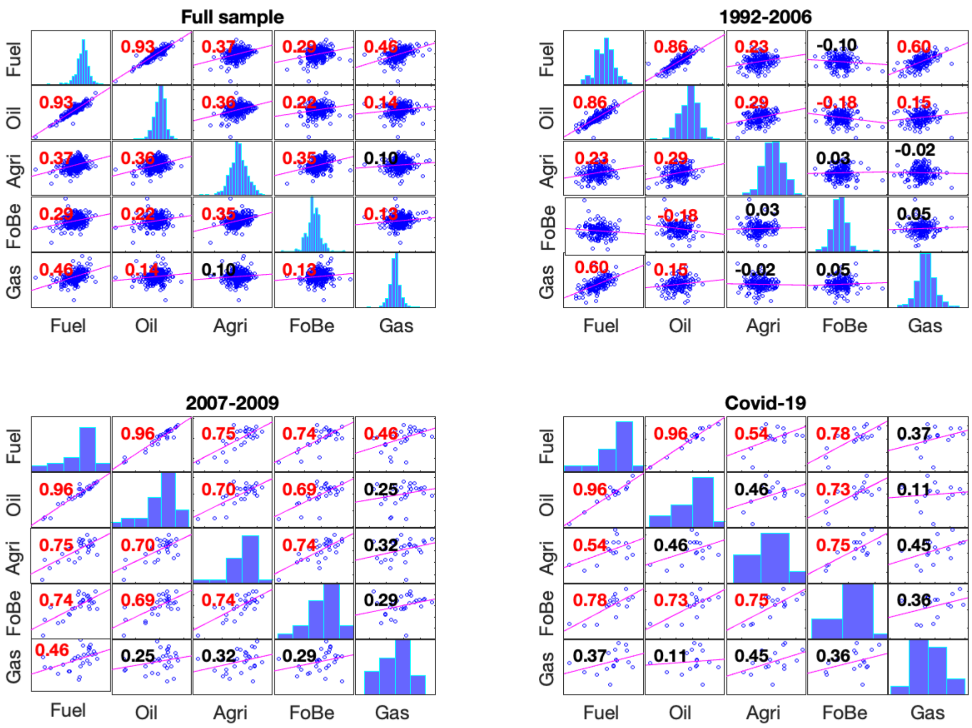
To motivate our analysis, we start by looking at the cross-correlations and volatility of selected core commodity indices over the period January 1992 to March 2021. Figure 1 presents a diagram of the (unconditional) cross-correlation coefficients between five core indices and across different sub-samples. Off-diagonal entries illustrate a scatterplot of the relevant variables with a least squares reference line whose slope equals the displayed correlation coefficient. The coefficients in red indicate that the null hypothesis of zero correlation cannot be rejected. Each diagonal subplot depicts the distribution of each commodity index (returns) as a histogram. The left-hand side graph in the first row (*Full sample*) depicts correlations over the full sample, while the right-hand side graph runs from 1992 up until the beginning of the 2007/8 financial turmoil. Moving to the second row, we present cross-correlations over December 2007 to June 2009 (*Figure 1 - LHS*), and December 2019 to the end of sample (*Figure 1 - RHS*). The former sample corresponds to the eighteenth-month recession that characterized the Global Financial Crisis as identified by the NBER, and the latter sample signals the Covid-19 pandemic.

It is evident that the years prior to the 2007/08 turmoil exhibit weaker evidence of co-movement. For example, while between 1992 and mid-2000s the lowest correlation (in absolute value) is as low as (approximately) 1%, during the Global Financial Crisis years the lowest correlation was approximately 25%. Also, the sign of certain correlation coefficients changes going from the pre-GFC to the post-GFC years. During the first subsample we osberve (statistically *insignificant*) negative correlation for specific commodity-pairs, e.g. Food & Beverages vs. Fuel. In contrast, the Great Recession years are characterised by positive and statistically significant correlations. A closer look at the correlation coefficients between December 2007 and June 2009 emphasizes the increased degree of co-movement during high-uncertainty periods. Similar conclusions can be reached when examining the first half of the pandemic crisis.[6] Overall, these findings are consistent with previously documented results in the literature showing that, for macroeconomic and financial variables, downturn periods are characterized by increased co-movement (D'Agostino and Giannone, 2012).[7]

[6]The plots in the second row illustrate relationships based on a relatively small sample size. Employing the Kendall correlation which offers a non-parametric alternative robust to smaller sample sizes results in qualitatively similar results.

[7]The question of what drives commodities co-movement and whether these commonalities are originating from short-run or long-run forces is beyond the scope of our paper. Notable (co-movement) drivers include the interest rate (Byrne et al., 2013), US dollar exchange rate (Vansteenkiste, 2009), fluctuations in global economic activity that are more relevant on the medium and long run (Delle Chiaie et al., 2022; Alquist et al., 2020; Baumeister et al., 2020) and uncertainty that has been shown to exhibit an important role in determining short-run commonalities (Poncela et al., 2014). In a recent application, Casoli and Lucchetti (2022) estimate factor models by taking cointegration into account and find that commodity prices move together mostly due to long-term common forces.

Figure 1: Cross-correlations between main commodity indices over the full sample and three subsamples



Notes: The chart presents correlation coefficients over the full sample, the first half between *1992 - 2006*, the highly volatile period between *2007-2009* and the Covid-19 years for 5 main indices: Energy (*Fuel*), Oil (*Oil*), Agricultural (*Agri*), Food & Beverages (*FoBe*), and Natural Gas (*Gas*).

Figure 2: Commodity prices volatility measured as the square monthly percentage changes in the commodity reported in the title

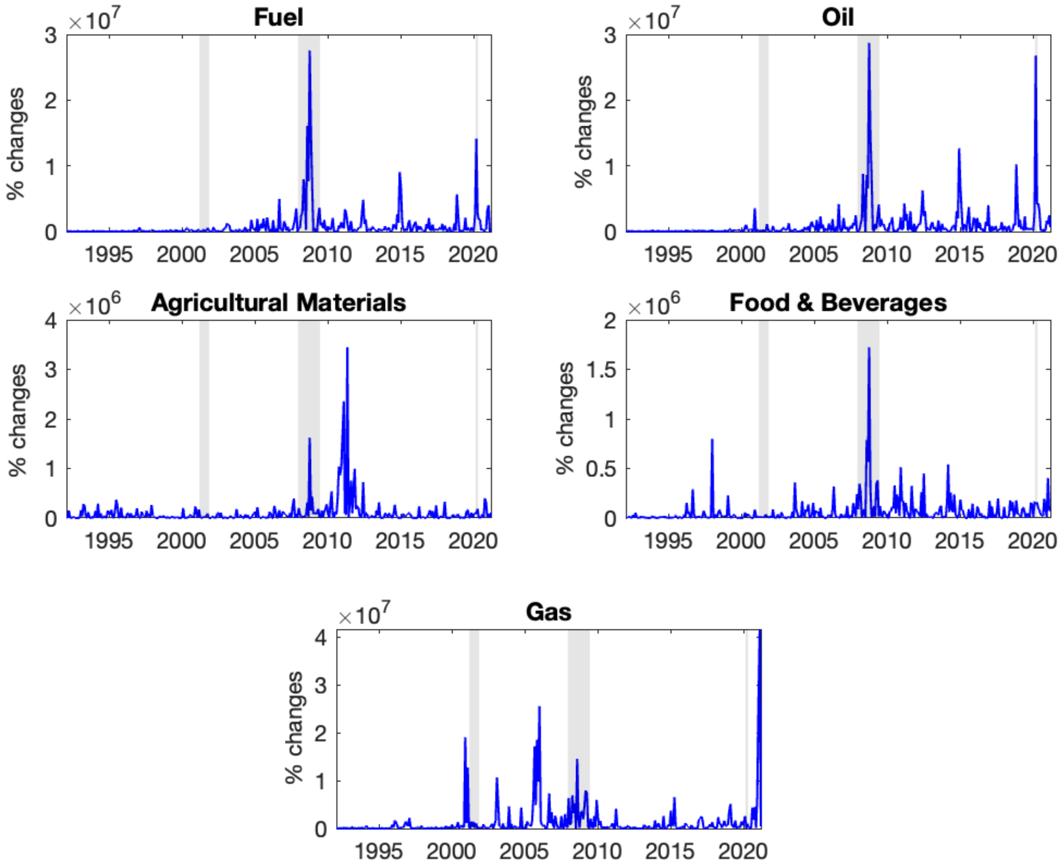


Figure 2 provides some preliminary evidence on the presence of time-varying volatilities (as measured by squared monthly percentage changes). Volatility was mostly low in the early 1990s but it spiked dramatically during the 2007/08 to fall again thereafter. More recently, the Covid-19 shock affected prominently energy prices as containment measures led to a collapse of traveling and transportation.[8]

In conclusion, summary statistics show evidence of time-variation in the degree of commonality and in the volatility of commodity prices. In the remaining of this paper, we explore whether accounting for these two features can improve point and density forecasting accuracy and yield meaningful economic gains.

3 Dynamic Factor Models (DFM): a suite of specifications

Our empirical investigation makes use of a suite of models that can be nested in the more general specification of a dynamic factor model with time-varying parameters (TVP-DFM). Table 2 illustrates the full set of competing models and their respective estimation methodologies.[9]

Let $x_{it} = (x_{1t}, ..., x_{nt})'$ be an $n-$ dimensional vector of commodity returns expressed as the log change in month t of a given commodity price $i = 1, ..., n$, that follows a dynamic factor model of the form

$$x_{it} = \lambda_{it} f_t + \epsilon_{it} \tag{1}$$

$$f_t = B_t f_{t-1} + \eta_t \tag{2}$$

where f_t is the $k \times 1$ vector of factors, λ_{it} is the $n \times k$ factor loadings vector, B_t is a $k \times k$ matrix of VAR(1) coefficients and ϵ_{it} and η_t are disturbance terms. It is further assumed that $\epsilon_t \sim N(0, V_t)$ and $\eta_t \sim N(0, Q_t)$ where Q_t and V_t are the $n \times n$ and $k \times k$ diagonal covariance matrices respectively. Note that the ϵ_{it} are uncorrelated with both f_t or η_t at all leads and lags.

We let λ_t and β_t evolve as driftless random walks:

$$\lambda_t = \lambda_{t-1} + u_t \quad u_t \sim N(0, R_t) \tag{3}$$

$$\beta_t = \beta_{t-1} + v_t \quad v_t \sim N(0, W_t) \tag{4}$$

[8]Excess supply turned out to be so extreme and storing capacity so tight, that a given WTI crude oil futures contract fell to a negative value on April 20, 2020, to -$38.

[9]We have also evaluated the statistical and economic gains of a large-dimensional non-parametric VAR proposed by Kapetanios et al. (2019). This specification does not impose any type of factor structure but, instead, makes use of the information content of the complete set of 68 commodities. Results are available upon request.

Table 2: Prediction models

	Specification	Estimation
Constant Parameters models	DFM **dfm** DFM with block structure **dfmBlocks**	QML
Time-Varying Parameters models	TVP-DFM **tvp** TVP-DFM with block structure **tvpBlocks**	Forgetting Factors

The law of motion of the covariances R_t and W_t is described in Appendix A.2. We follow a standard assumption in assuming that u_t and v_t are *iid* errors, uncorrelated with each other as well as with ϵ_t and η_t at all leads and lags.[10] The specification hereby adopted is largely inspired by Del Negro and Otrok (2008), who explore the empirical performance of a TVP-DFM in explaining international business cycles. Variants of this model have been used to study the properties of global inflation (Mumtaz and Surico, 2012) and the interaction between the macroeconomy and financial markets (Bianchi et al., 2009).

[10]See, for instance, Koop and Korobilis (2010).

We further impose a block structure that reflects the composition of the IMF commodity price database and follows DFG (2022) in spirit.[11] In particular, we extract one global factor (*All Commodities*), two block factors (*Non-Fuel* and *Fuel*), four sub-block factors (*Agriculture & Fertilizers*, *Metals*, *Crude Oil* and *Coal, Natural Gas and Propane*) and three group factors (*Agricultural Raw Materials*, *Food*, and *Beverage*).[12] Table 1 provides an overview of the imposed block structure. Appendix A.2 offers details on the adopted factor specification and the corresponding block structure.

The setup of DFM with structural instabilities given in Equations 1 and 2 nests the typical time-invariant parameter DFM, when we restrict variation in all parameters. The time-invariant parameter DFM is derived by replacing the time varying-specifications in equations 3 and 4 with

$$R_t = W_t = 0 \tag{5}$$

and setting the time-varying covariances V_t and Q_t to their constant values

$$V_t = V \quad \text{and} \quad Q_t = Q \tag{6}$$

One additional nested model that imposes the assumption of time-invariant parameters and similarly allows for the existence of group-specific comovement can be derived in a straightforward manner. This constant-coefficient specification has been recently adopted by DFG (2022) in order to study comovement in international commodity prices.

4 Estimation algorithms

In this section we introduce the estimation methodologies for both the time-varying and constant-parameter specifications.

4.1 Time-varying models

Time variation is modelled through a forgetting factor framework, introduced in the macroeconomic forecasting literature by Koop and Korobilis (2013). This model can be considered as an approximation of the Bayesian VAR with time-varying parameters and stochastic volatility developed by Cogley and Sargent (2005) and Primiceri (2005). Instead of using a sampler to

[11]Note that our block structure differs from DFG (2022) in the number and composition of the distinct factors.

[12]Using the trade weights published in the IMF website, and with the purpose of maintaining a parsimonious factor structure, we merge specific factors into one. In particular, we construct one common factor for Coal, Natural Gas and Propane, and another one for Agriculture and Fertilizers.

draw the covariance matrices of the time-varying coefficients and volatility parameters, the proposed framework estimates them as a weighted average of previous estimates and of the current forecast error variance. It is a more parsimonious set up that considerably speeds up computational time, and enables the analysis with a larger number of variables. See Appendix A.3 for estimation details.

4.2 Time-invariant models

For the estimation of the constant-coefficient specifications we adopt an extension of the two-step estimation algorithm for constant-parameter dynamic factor models of Doz et al. (2011). In the first step, the algorithm is initialized by computing principal components, and the model parameters are estimated by an OLS regression. The next step updates the factors estimates by using the Kalman smoother as in Doz et al. (2011). The additional, and final, step we take is to re-estimate the factor loadings based on the updated Kalman filter estimates of the previous step.

5 The design of the forecasting exercise

We estimate the models over the period 1992M1-2000:M12 and evaluate their forecasting performance for the period 2001:M1 through 2021M3-h for $h = 1, 3, 6, 12$ months ahead. In addition to the assessment of the competing models over the full sample, we present an evaluation based on distinct subsamples. The first subsample forecasts are for the period 2001M1 + h to 2007M1, while the second subsample begins just before the 2007/08 financial turmoil and extends to the end of the sample.

The benchmark model against which we evaluate the predictive ability of the competing specifications, unless otherwise stated, is the random-walk (no-drift) model. For both constant- and time-varying specifications, we obtain the forecasts of the 68 individual commodity prices iteratively using the Kalman filter formulation.

$$\hat{x}_{t+h|t} \equiv E[x_{t+h}|I_t] = \Lambda\hat{E}[f_{t+h}|I_t] = \Lambda\hat{f}_{t+h|t} = \Lambda B^h f_{t|t} \tag{7}$$

$$\hat{x}_{t+h|t} \equiv E[x_{t+h}|I_t] = \hat{E}[\lambda_{t|t} f_{t+h}|I_t] = \lambda_{t|t}\hat{f}_{t+h|t} = \lambda_{t|t}\beta_{t+h} f_{t|t} \tag{8}$$

The respective point and density forecasts of the ten major commodity indices are, subsequently, calculated as the weighted averages of the individual commodity price forecasts using the import weights published in the IMF primary commodity prices database.

For generating samples from the predictive density of the time-varying factor specifications (eq. 8) we follow (Koop and Korobilis, 2013, henceforth KK, 2013). We derive the predictive densities from the constant-parameter specifications (eq. 7) using bootstrap methods. The focus of the forecasting literature has been steadily moving from point forecasts to density forecasts that incorporate the uncertainty about the future evolution of the variables of interest; see, among others, Jore et al. (2010) and Clark (2011). A multivariate forecast density for a given horizon can be obtained assuming Gaussian forecast errors and known lag order and model parameters. However, documented departures from these assumptions pose a serious concern when forecasting with multivariate models, calling into question widely-used techniques for constructing predictive densities. Bootstrap procedures provide a flexible methodology for constructing forecast densities that incorporate parameter uncertainty and can, furthermore, relax assumptions on the error distributions for a more general treatment. [13]

The steps for obtaining a bootstrap forecast density and associated statistics for the 68 commodities in vector $x_{it}, i = 1,,,.68$ are as follows:

1. For $t = 1, ..., T$, estimate the factor ($\tilde{f_t}$) and associated factor loadings ($\tilde{\Lambda}$) and obtain the residuals: $\tilde{\epsilon_{it}} = x_{it} - \tilde{\Lambda}\tilde{f_t}$;
2. Generate $x^*_{it} = \tilde{\Lambda}\tilde{f_t} + \epsilon^*_{it}$ where ϵ^*_{it} is a resampled bootstrap version of the residuals $\tilde{\epsilon_{it}}$;
3. Using the new bootstrap set x^*_{it}, estimate the bootstrap factors f^*_t;
4. Estimate the corresponding bootstrap parameters, e.g. Λ^*, B^*_1 etc.;
5. The (bootstrap) forecast is given by Eq., (7), replacing the respective quantities with their bootstrap counterparts. The forecast density $F(x_{it+h})$ is given by the density function of the bootstrap forecasts $x_{it+h,b}, b = 1, ..., B$ where B is the desired number of replications.

5.1 Forecast Evaluation Metrics

We perform a recursive out-of-sample forecasting exercise to evaluate the performance of the competing forecasting methods in terms of both point and density h-step-ahead iterated forecasts.

[13] Focusing on the constant-coefficient factor models (eq. 7), adopting a bootstrapping algorithm allows us to capture uncertainty resulting from the estimation of the factors.[14] The methodology is inspired by recent work on prediction intervals for factor models. For example, Gonçalves and Perron (2014) propose a general residual-based bootstrap method for inference in a factor-augmented regression, while Gonçalves et al. (2017) extend this bootstrapping methodology to prediction intervals. Our algorithm, instead, can be applied to dynamic factor models of the form described above and is based on a residual-based bootstrap scheme.

An extensive literature on statistical evaluation methods proposes various loss functions and test statistics in order to assess statistical significance. We employ the Root Mean Squared Forecast Error (RMSFE) criterion to explore whether it is possible to beat the random walk forecast for a given forecast horizon. For each competing model, we compute the ratio of the RMSFE of the individual specification relative to the RMSFE of the driftless random walk.

To evaluate the statistical significance of differences in forecast accuracy across the competing specifications, we consider tests of equal predictive accuracy (Diebold and Mariano, 1995; West, 1996). The tests compare the forecasts from the factor models against the benchmark. Additionally, we incorporate the finite sample correction proposed by Harvey et al. (1997).[15]

We evaluate the entire predictive density and compare the cumulative differences in log-predictive likelihoods of the competing models. Predictive density accuracy is evaluated on the basis of the logarithmic score, see Mitchell and Hall (2005) and Amisano and Giacomini (2007). See Appendix A.4 for additional details.

We investigate the correct specification of predictive densities using different tests. The tests we consider include tests of uniformity, accounting for the possibility of serial correlation. Focusing on tests based on the Probability Integral Transform (PIT), we consider the histogram-based evaluation technique employed by Diebold et al. (1997) and Diebold et al. (1999). As was highlighted in these papers, a correctly conditionally calibrated density produces PITs that are uniformly distributed. For the one-step ahead forecasts, we evaluate the calibration of the densities by looking at the results of Berkowitz (2001)'s LR test. For horizons beyond one step ahead, we report the test of Knüppel (2015), which is robust to the presence of serial correlation of the PITs.

6 Empirical results

6.1 Point forecasts

Tables 3 and 4 present RMSFE ratios for the monthly forecast horizons 1, 3, 6 and 12 months ahead of the competing models relative to the benchmark random walk model over the subsamples 1992 - 2007 and 2008 - 2021. Entries with values less than one indicate that the forecast model

[15] It is important to note that the literature examining testing procedures shows that a number of issues arise when testing for significant differences in forecasting performance, including the size of the in-sample relative to the size of the out-of-sample period, and the type of estimation window used; see Corradi and Swanson (2006) for a useful survey of some of these issues. In light of such evidence, the adopted test procedure is only used to provide a rough guide for assessing statistical significance.

under examination delivers accuracy gains relative to the benchmark. [16]

Looking at the overall forecasting performance of each model across the 10 indices, our findings highlight the difficulties in finding a model that performs uniformly better than the others across forecast periods, horizons and commodity indices. Focusing on the subsample coinciding with the pre-GFC years (Table 3) , the competing factor models consistently outperform the random walk for specific indices belonging to the *NonFuel* block, namely *Agriculture and Fertilisers, Metals* and *Agriculture Raw Materials*. At the shortest horizon ($h = 1$) evidence of predictability is weak, but as the forecast horizon increases, forecasting performance becomes stronger. At $h = 3$, the time-invariant factor specification (*dfm*) improves upon the random walk benchmark by 10% for the *Non-Fuel* index, while accounting for time-variation and commodity-specific comovement appears to be particularly beneficial for the *Metals* index. Focusing on the six-month ahead point forecasts, we observe that, for all but two indices (*Oil* and *CoalGas*), introducing a factor structure is successful at outperforming the random walk benchmark. For example, looking at the *All Commodities* index, the *tvpBlocks* specification outperforms the competing factor models and the random walk benchmark, enhancing accuracy by more than 10%. At the longest horizon of one-year ahead, significant gains in point forecast accuracy are documented across all ten commodity indices. On average, the forecasting model that captures time-evolving dynamics and commodity-specific heterogeneity improves upon both its constant-coefficient variant and the model that does not impose a specific block structure.

Results for the subsample from 2008 onward (Table 4) appear to reinstate the empirical difficulty of finding a single modelling framework that outperforms its competitors at all horizons. At the short and medium-term horizons (1-, 3-, and 6- months ahead), most models deliver improvements over their pre-GFC performance.[17] Nevertheless, we do note that at the longest horizon of 12-months ahead, evidence of a superior post-Crisis performance is weaker. For example, focusing on the *Fuel* index, we find statistically significant improvements over the random walk model for the one-month and three-months forecast horizons. In contrast, all competing specifications fail to outperform the benchmark model at the one-year ahead

[16] One concern related to macroeconomic and financial forecasting studies is the effect of data mining on the size of tests of predictability. Data mining occurs when a researcher searches over alternative forecast specifications, but only reports results for the model with the highest predictive content, causing the size of the test of predictability to be inflated, and thereby resulting in spurious rejections of the no-predictability null (Inoue and Kilian, 2005). As a first attempt in tackling issues related to collective data mining (Denton, 1985), we report results for high levels of significance which should provide some confidence in the results.

[17] This finding is in line with prior literature documenting an enhanced point forecast performance of factor models during the Great Recession, e.g. DFG (2022).

horizon. A similar pattern is observed for the majority of commodity indices. Results for the *Metals* and *Agriculture and Fertilisers* indices paint a different picture, with all factor specifications outperforming the benchmark random walk model at short-, medium-, and long-term horizons. For instance, the evaluation exercise for the *Metals* index points to a particularly enhanced performance of all models relative to the random walk, with the time-varying specifications outperforming their constant-coefficient counterparts. At the longer horizons, no model exhibits a consistently improved performance over the benchmark random walk. Nonetheless, a comparison between the time-invariant forecasting models and their time-varying counterparts suggests that introducing time-evolving dynamics to constant-coefficient factor models (with or without a hierarchical structure) lowers forecast errors for $h = 12$. Turning to the effect of imposing the block structure described in Table 1, we focus on the relationship between *dfm* and *dfmBlocks*. Looking at the one-month ahead point forecasts, introducing a hierarchical structure is beneficial for eight out of ten commodity indices. A similar performance is documented at three- and six-months ahead. Improvements appear to diminish at the longest horizon of one-year ahead, with the block structure model delivering marginal gains.

Table 3: RMSFE ratios of the competing models relative to a random walk for various forecast horizons over the first sub-sample period **1992M1 to 2007M12**. Underlined values denote significantly different forecast errors according to a Diebold − Mariano test, modified using the small sample correction of Harvey et al. (1998).

Model↓|Horizon→	h=1	h=3	h=6	h=12	||	h=1	h=3	h=6	h=12
		All					**Fuel**		
dfm	1.09	1.03	0.93	0.84		1.12	1.08	1.00	0.90
dfmBlocks	1.01	0.95	0.88	0.83		1.02	0.99	0.94	0.89
tvp	1.06	1.04	0.98	0.87		1.11	1.10	1.03	0.94
tvpBlocks	0.99	0.95	0.86	0.81		1.04	1.05	0.96	0.88
		NonFuel					**Agr/Fuel**		
dfm	0.99	0.90	0.89	0.82		0.97	0.95	0.94	0.94
dfmBlocks	1.02	0.93	0.92	0.83		0.98	0.95	0.94	0.94
tvp	1.04	0.96	0.93	0.85		0.97	0.95	0.95	0.93
tvpBlocks	1.03	0.94	0.91	0.81		0.97	0.95	0.93	0.93
		Metals					**Oil**		
dfm	0.98	0.95	0.94	0.93		1.16	1.17	1.11	0.94
dfmBlocks	0.97	0.95	0.94	0.92		1.08	1.08	1.05	0.92
tvp	0.97	0.96	0.95	0.93		1.19	1.19	1.16	0.99
tvpBlocks	0.97	0.94	0.93	0.91		1.11	1.13	1.08	0.90
		CGP					**Food**		
dfm	1.09	1.09	1.07	0.98		1.07	1.01	0.95	0.90
dfmBlocks	1.07	1.06	1.04	0.98		1.02	1.00	0.96	0.92
tvp	1.11	1.10	1.07	1.04		1.09	1.12	1.04	1.01
tvpBlocks	1.09	1.11	1.05	0.96		1.03	1.01	0.96	0.91
		Beverages					**AgriRaw**		
dfm	1.09	1.07	1.05	0.96		0.96	0.97	0.98	0.91
dfmBlocks	1.03	1.05	1.06	0.97		0.96	0.95	0.96	0.91
tvp	1.06	1.04	0.97	0.93		0.98	0.99	0.96	0.92
tvpBlocks	1.03	1.04	1.01	0.94		0.95	0.97	0.96	0.91

Table 4: RMSFE ratios of the competing models relative to a random walk for various forecast horizons over the first sub-sample period **2008M1 to 2021M3**. Underlined values denote significantly different forecast errors according to a Diebold – Mariano test, modified using the small sample correction of Harvey et al. (1998).

Model↓|Horizon→	h=1	h=3	h=6	h=12	||	h=1	h=3	h=6	h=12
		All					**Fuel**		
dfm	0.95	0.98	1.04	1.14		0.96	1.00	0.99	1.13
dfmBlocks	0.91	0.97	1.03	1.13		0.92	0.98	1.04	1.13
tvp	0.94	0.99	1.02	1.08		0.95	1.00	1.03	1.09
tvpBlocks	0.93	0.98	1.01	1.05		0.96	1.01	1.03	1.08
		NonFuel					**Agr/Fer**		
dfm	0.95	0.98	1.02	1.11		0.98	0.95	0.95	0.95
dfmBlocks	0.92	0.96	1.02	1.10		0.99	0.95	0.93	0.94
tvp	0.94	0.98	0.99	1.07		0.99	0.97	0.95	0.95
tvpBlocks	0.92	0.95	0.99	1.07		0.99	0.96	0.94	0.94
		Metals					**Oil**		
dfm	0.98	0.93	0.93	0.92		1.05	1.07	1.11	1.19
dfmBlocks	0.97	0.93	0.91	0.91		1.03	1.04	1.09	1.18
tvp	0.98	0.95	0.90	0.89		1.02	1.05	1.08	1.15
tvpBlocks	0.97	0.93	0.90	0.90		1.00	1.05	1.06	1.13
		CGP					**Food**		
dfm	0.97	0.94	0.99	1.12		0.97	0.98	1.04	1.11
dfmBlocks	0.94	0.97	1.01	1.11		0.95	0.97	1.03	1.10
tvp	0.97	1.00	1.00	1.07		0.96	0.98	1.01	1.10
tvpBlocks	0.96	0.98	0.99	1.05		0.94	0.98	1.01	1.08
		Beverages					**AgriRaw**		
dfm	0.99	1.05	1.02	1.11		0.94	0.95	1.01	1.08
dfmBlocks	0.98	1.03	1.02	1.09		0.97	0.97	1.01	1.06
tvp	0.99	1.05	1.02	1.09		0.92	0.94	1.00	1.05
tvpBlocks	0.98	1.03	1.03	1.07		0.97	0.96	1.00	1.05

6.2 Density forecasts

In this section, we present an assessment of commodity price density forecasts. First, we investigate the importance of imposing a factor block structure. Figure 3 presents the results of the analysis. It shows the cumulative differences in log-predictive likelihoods between the model that adopts a block structure (*dfmBlocks*) and the baseline constant-coefficient factor model (*dfm*). Increasing and positive differentials indicate a better performance of the *dfmBlocks* specification.

For the overall commodity index, a strong superiority of the *dfmBlocks* model emerges at almost all horizons. This result, at the aggregate index level, is driven by the energy block, that is oil (panel b) and coal, gas and propane (panel g), as well as by metals (panel e). For the non-energy block as a whole (panel c), results are instead in favor of the *dfm*. This is mostly due to the better performance the *dfm* specification for the food block (panel h), which accounts for almost half of the non-energy block. This first comparison shows that, for some specific commodities, exploiting within block information may improve density forecast accuracy. Among the total 40 examined cases (10 indices, 4 horizons), a superior performance of the block structure model is indeed documented for 27 cases (67%).

We test whether the differences observed in Figure 3 are statistically significant, using the Amisano and Giacomini (2007) test, see Table 5. Positive (negative) values with asterisk denote forecast horizons at which the benchmark constant-factor model performs significantly better (worse) than the model indicated in the row header.[18] Focusing on the comparison with the *dfmBlocks* specification, the test confirms that imposing a hierarchical structure yields a significantly lower density forecast error for 7 out of the 10 indices under investigation.

Next, we compare the density forecasting performance of the factor model that allows for time-varying slope and volatility parameters (*tvp*) and its respective constant-coefficient variant (*dfm*). This exercise tells us whether time variation improves forecast accuracy. [19] Figure 4 plots the cumulative log-likelihood score at one-month to one-year forecast horizons, where the benchmark is again the constant-coefficient factor model *dfm*.[20] As evident from increasing and positive log-scores, the model that allows for time-variation in both slope and volatility emerges uniformly as the best model for all commodities at all horizons. At the shortest horizon of one-month ahead, the time-varying parameters specification consistently outperforms its time-invariant competitor for nine out of ten currency pairs. The largest gains are documented for the Non-fuel block of commodities, with the bulk of forecast improvements driven by the Agriculture & Fertilisers index. The test statistics of the equal predictive ability test corresponding to this comparison can be found in the middle panel of Table 5.[21] The statistics are all negative and significantly different from zero, confirming the benefits of introducing time-varying dynamics in the context of density forecasting. The results obtained with the model allowing for time-varying parameters *and* a block structure (*tvpBlocks*) are very similar to the ones obtained on the basis of the *tvp* model. We do not report the log-scores for brevity, but comparing the middle to the bottom panels of Table 5 provides a clear indication that adding a block structure to the *tvp* model does not lead to any further material improvements in density forecast accuracy.

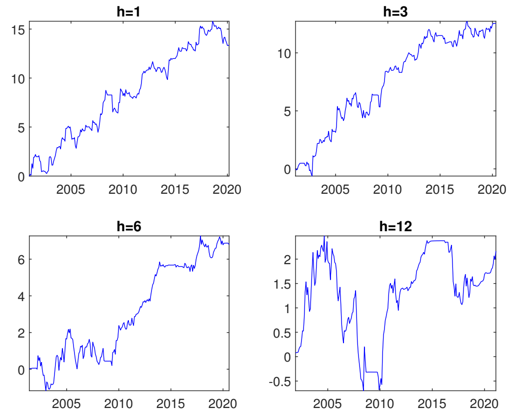
[18]For the purpose of evaluating the importance of adopting a block structure for density forecasting, we set the constant factor model *dfm* as the benchmark model.

[19]Density forecast comparisons of selected models and the random walk benchmark can be found in Appendix A.5.1 (Figures A.1).

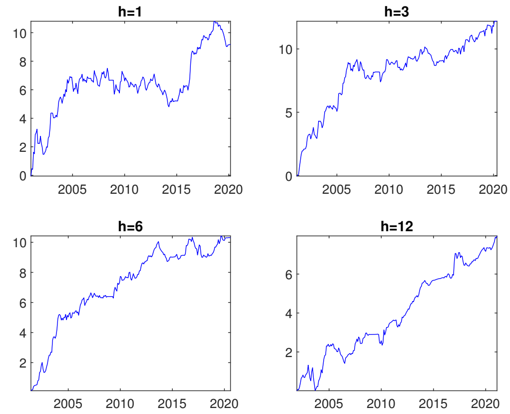
[20]Figure A.2 in Appendix A.5.1 presents a comparison between the tvpBlocks and the dfmBlocks specifications, also aiming at illustrating the benefits of introducing time-variation in terms of density forecast accuracy.

[21]Additional comparisons for this test can be found in Appendix A.5.1 - Table A3.

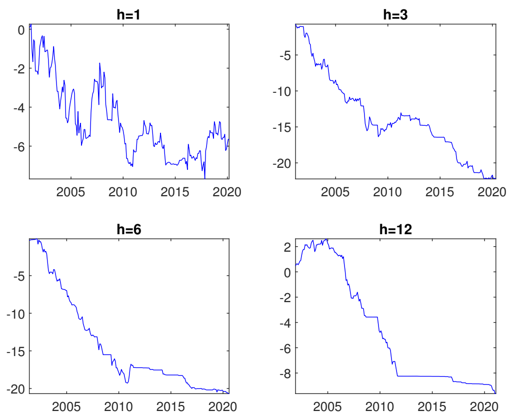
Figure 3: Cumulative differences in log-predictive likelihood of the constant-parameters dynamic factor model with block structure (`dfmBlocks`) relative to the constant-parameters dynamic factor model (`dfm`). Increases in the statistic denote dates in which `dfmBlocks` outperforms the alternative.



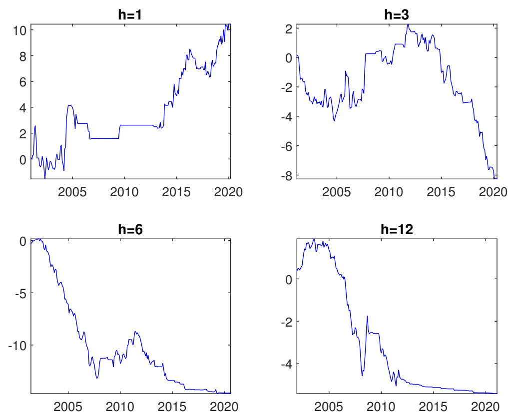
(a) All Commodity



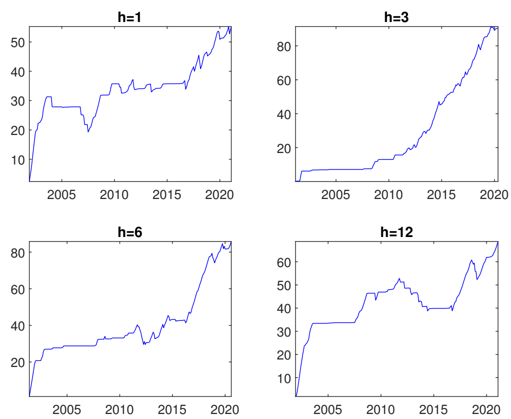
(b) Fuel



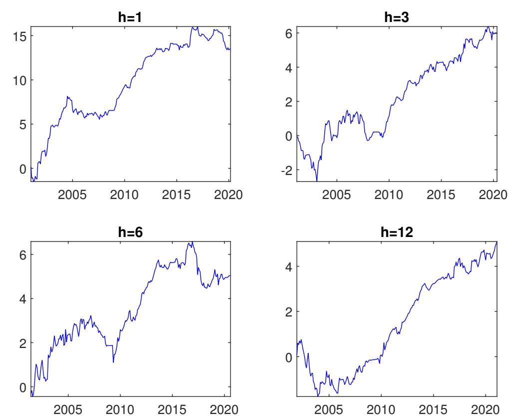
(c) Non-Fuel



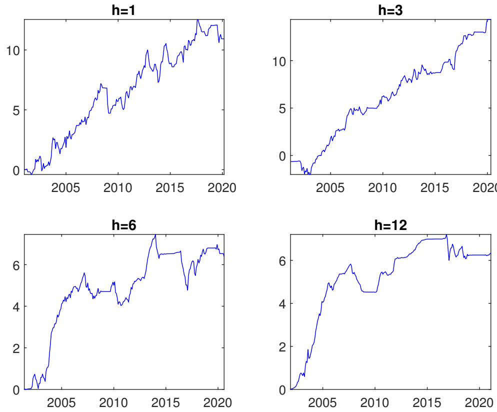
(d) Agriculture & Fertilisers



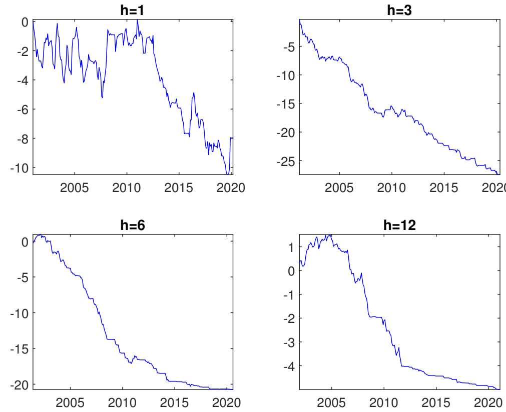
(e) Metals



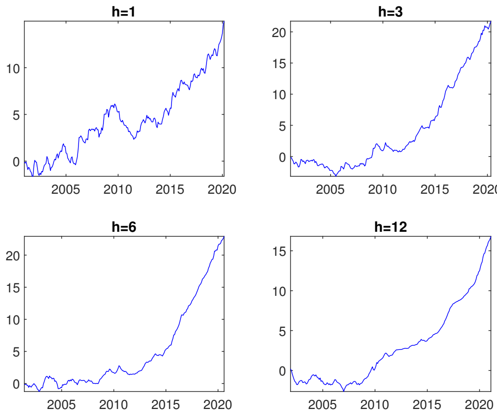
(f) Oil



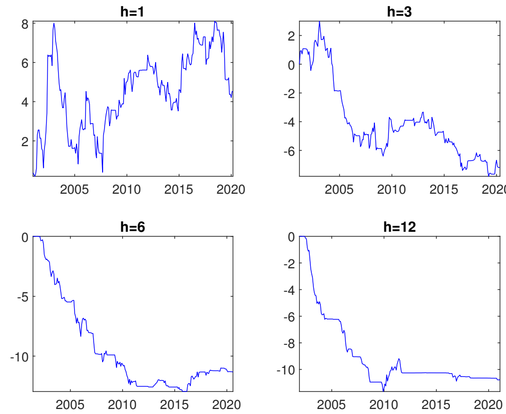
(g) Coal, Gas & Propane



(h) Food

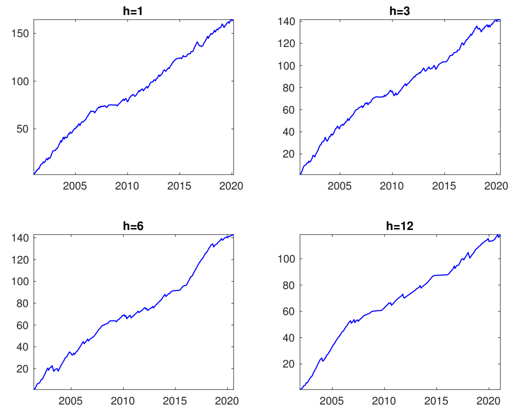


(i) Beverages

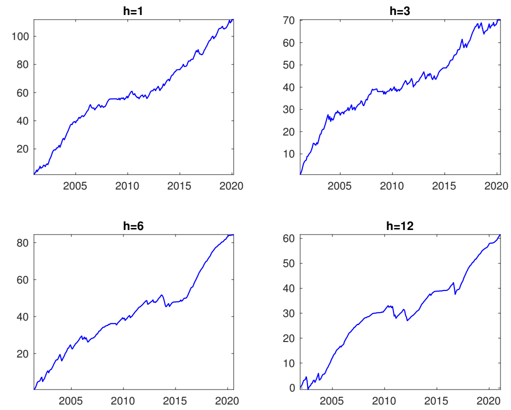


(j) Agriculture Raw Materials

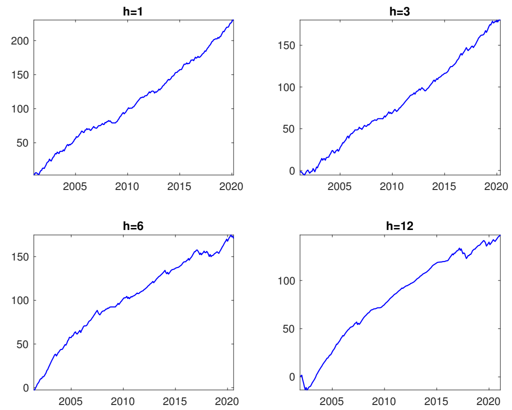
Figure 4: Cumulative differences in log-predictive likelihood of the time-varying dynamic factor model (tvp) relative to the constant-coefficient dynamic factor model (dfm). Increases in the statistic denote dates in which tvp outperforms the alternative.



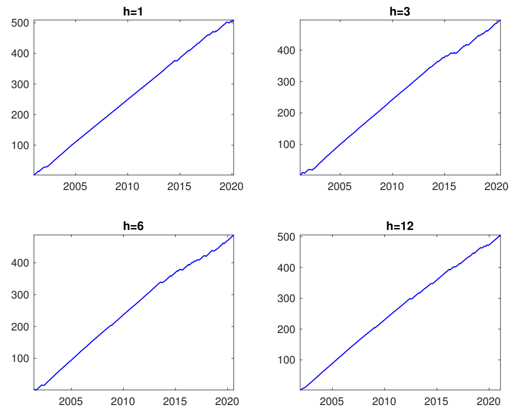
(a) All Commodity



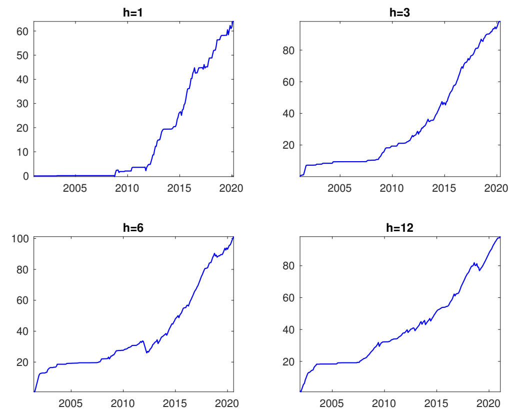
(b) Fuel



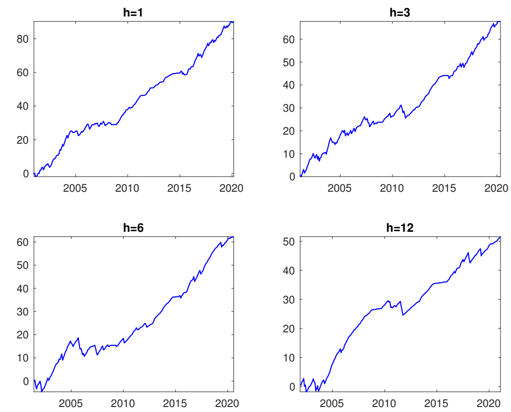
(c) Non-Fuel



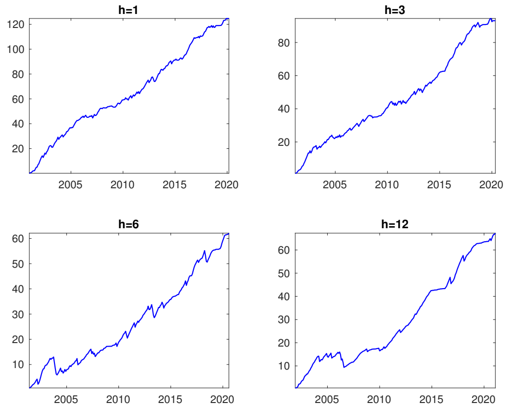
(d) Agriculture & Fertilizers



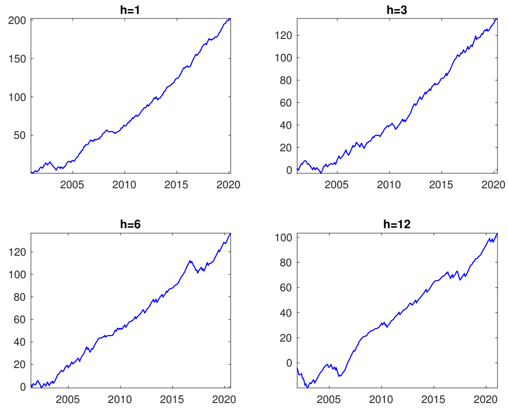
(e) Metals



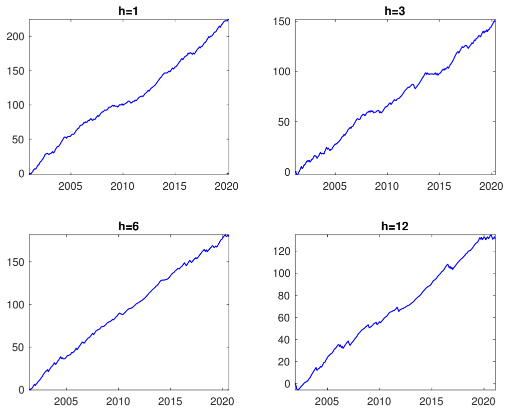
(f) Oil



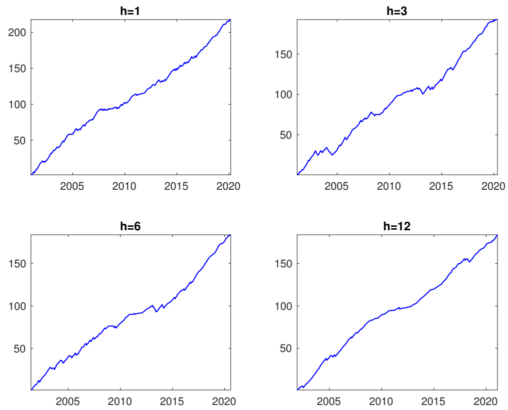
(g) Coal, Gas & Propane



(h) Food



(i) Beverages



(j) Agriculture Raw Materials

Table 5: Equal predictive ability test.

	All	Fuel	NF	AF	Metals	Oil	CGP	Food	Beve	AR
dfmBlocks										
h=1	-2.17*	-1.91*	0.77	-1.54*	-5.23*	-3.58*	-2.84*	1.00	-2.73*	-0.56
h=3	-2.39*	-3.03*	3.70*	1.41	-7.57*	-1.80*	-3.43*	5.25*	-5.04*	1.34
h=6	-1.80	-3.60*	4.32*	3.26*	-5.60*	-1.70*	-3.98*	4.85*	-6.96*	2.01*
h=12	-0.79	-3.46*	2.25*	1.98*	-5.03**	-2.08*	-4.92*	2.26*	-6.73*	2.16*
tvp										
$h = 1$	-9.94*	-6.34*	-11.16*	-8.47*	-6.25*	-7.67*	-8.44*	-10.15*	-12.17*	-10.93*
$h = 3$	-8.66*	-4.75*	-11.10*	-8.19*	-7.78*	-4.68*	-7.35*	-7.78*	-8.21*	-7.27*
$h = 6$	-6.28*	-5.28*	-6.83*	-9.87*	-4.54*	-4.83*	-4.65*	-7.42*	-12.50*	-8.60*
$h = 12$	-5.84*	-5.80*	-4.26*	-7.67*	-4.45*	-3.63*	-4.74*	-2.97*	-7.74*	-9.00*
tvpBlocks										
$h = 1$	-10.45*	-6.86*	-11.38*	-8.24*	-5.58*	-7.83*	-7.81*	-10.28*	-12.50*	-11.16*
$h = 3$	-9.63*	-5.29*	-10.20*	-7.47*	-6.91*	-5.03*	-7.48*	-8.37*	-9.89*	-7.80*
$h = 6$	-9.88*	-5.83*	-6.34*	-8.80*	-4.61*	-6.05*	-4.97*	-8.36*	-13.54*	-8.65*
$h = 12$	-9.87*	-5.58*	-5.09*	-7.00*	-4.38*	-4.09*	-5.06*	-4.54*	-7.22*	-9.18*

[1] The table reports the t-statistics for the null hypothesis that the model under investigation has the same predictive ability as the dynamic factor model with constant coefficients. Positive values with one asterisk (*) denote higher log-predictive likelihood at 5% significance level.

[2] NF correspond to the Non-Fuel index, AF to the Acriculture & Fertilizers, and AR to the Agriculture Raw Materials.

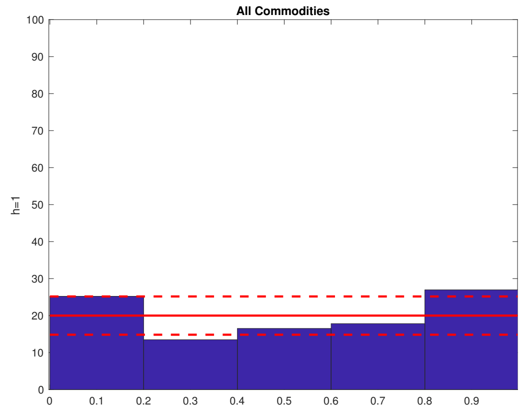
Finally, we test uniformity for both short (one-month) and long (one-year) horizon forecasts. Figure 5 reports results based on the Diebold et al. (1999) test for one-month ahead forecasts of the All Commodities index.[22] We focus on the distribution functions of the PITs for four core models, namely *dfm*, *dfmBlocks*, *tvp* and *tvpBlocks*. It is evident that models with constant coefficients (*dfm* & *dfmBlocks*) tend to produce densities that are somewhat U-shaped and in which some realisations fall outside the bounds implied by i.i.d uniform PITs. The two models with time-varying slope and volatility, instead, produce well-calibrated one-month ahead density forecasts. Figure 6 shows the relevant results for one-year ahead forecasts. Here the difference is all the more striking, and the role of time-varying parameters in delivering well-calibrated forecasts emerges more clearly. As a last exercise, we report results based on the inverse normal of the PIT. Table 6 shows the results for Berkowitz (2001) test at the one-month ahead horizon. According to Berkowitz (2001), this is a test of joint uniformity and (lack of) serial correlation. It is more powerful than some of the alternatives tests. For the longer-term horizons we employ the Knüppel (2015) test which is robust to the presence of serial correlation of the PITs. The results

[22] Relevant figures for the remaining commodity indices can be found in Appendix A.5.2 and A.5.3 for the one-month and one-year horizon, respectively.

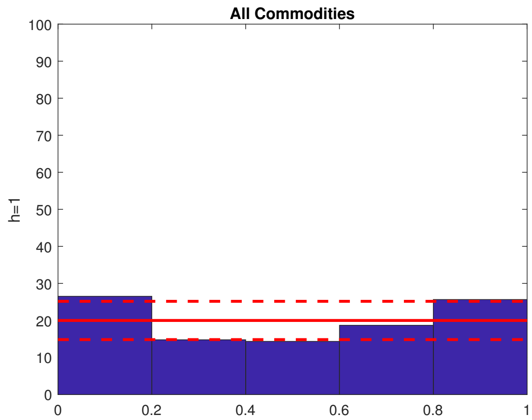
confirm that the density forecasts of constant-parameters models are poorly calibrated, both at short and long-horizons, with the exception of Fuel and Agricultural Raw Materials.[23] For time-varying models, density forecasts are better calibrated and rejections of the null hypothesis of uniformity are sporadic.

[23]Relevant figures for the remaining commodity indices can be found in Appendix A.5.2 and A.5.3 for the one-month and one-year horizon, respectively.

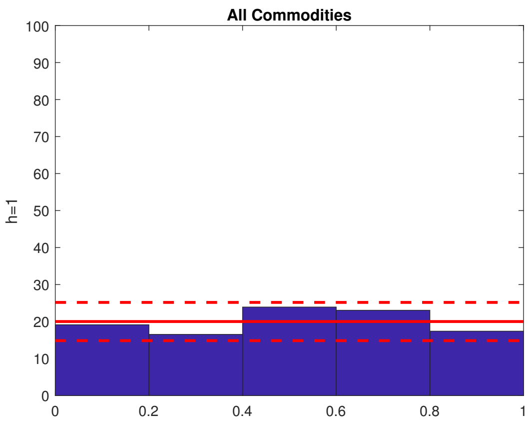
Figure 5: Probability density functions of the PITs for four core models for one-month ahead forecasts of the *All Commodity* index. The dashed lines are 95% confidence intervals, constructed using a normal approximation to a binomial distribution, as per Diebold et al. (1998).



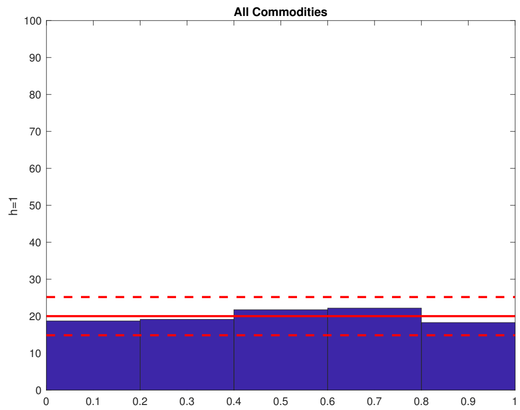
(a) *dfm*



(b) *dfmBlocks*

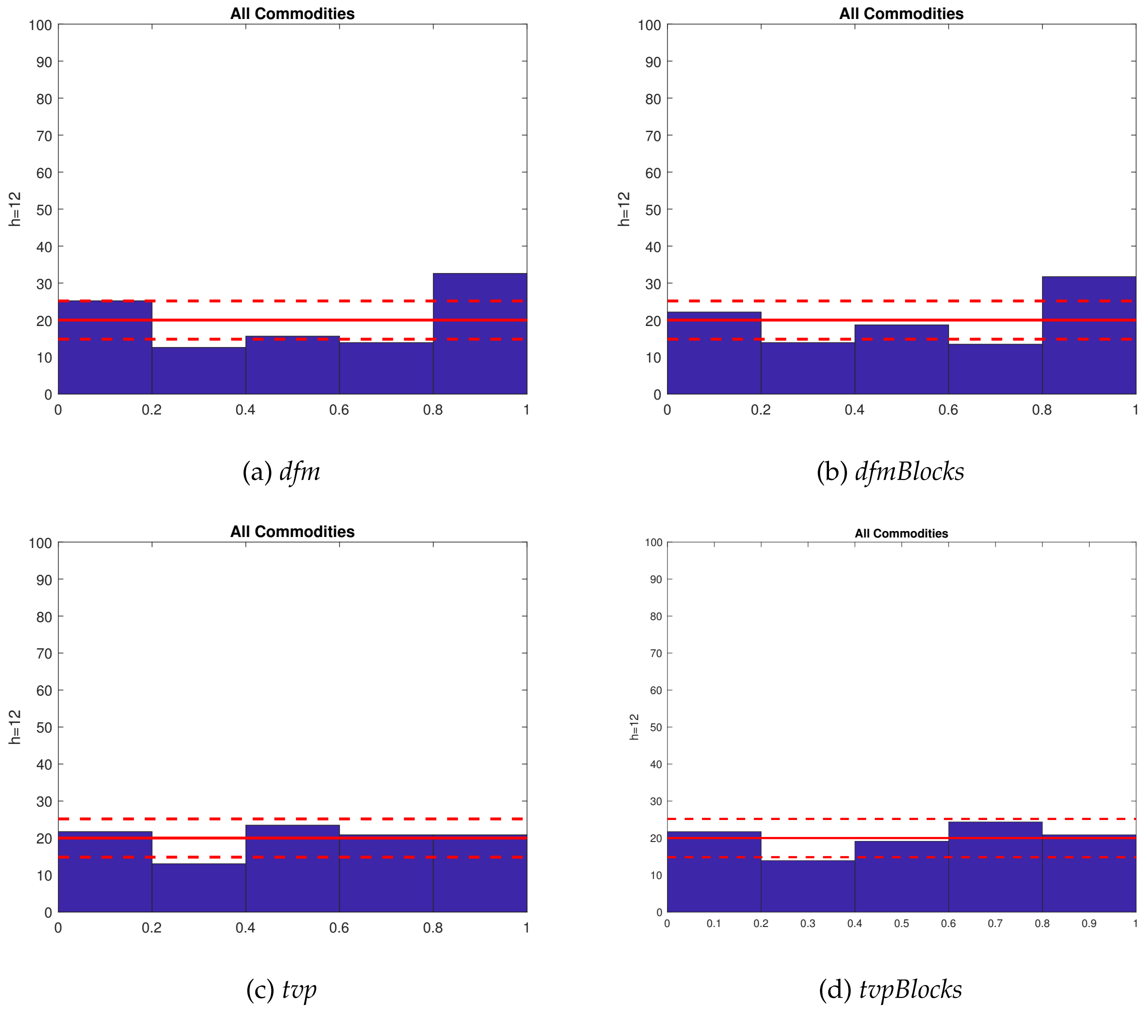


(c) *tvp*



(d) *tvpBlocks*

Figure 6: Probability density functions of the PITs for four core models for one-year ahead forecasts of the *All Commodity* index. The dashed lines are 95% confidence intervals, constructed using a normal approximation to a binomial distribution, as per Diebold et al. (1998).



(a) *dfm*

(b) *dfmBlocks*

(c) *tvp*

(d) *tvpBlocks*

Table 6: Berkowitz (2001) Likelihood Ratio and Knüppel (2015) tests.

	All	Fuel	NonFuel	AgrFer	Metals	Oil	CGP	Food	Beve	AgrRaw
dfm										
$h = 1$	0.05	0.79	0.00*	0.03*	0.07	0.00*	0.09	0.00*	0.00*	0.00*
$h = 12$	0.07	0.08	0.20	0.15	0.05	0.00*	0.03*	0.09	0.08	0.16
dfmBlocks										
$h = 1$	0.00*	0.06	0.00*	0.00*	0.01*	0.00*	0.00*	0.02*	0.37	0.00*
$h = 12$	0.04*	0.09	0.12	0.34	0.08	0.09	0.01*	0.80	0.07	0.03*
tvp										
$h = 1$	0.25	0.10	0.08	0.05	0.05	0.01*	0.09	0.04*	0.23	0.11
$h = 12$	0.28	0.21	0.16	0.24	0.24	0.24	0.08	0.59	0.04*	0.23
tvpBlocks										
$h = 1$	0.27	0.86	0.57	0.05	0.48	0.50	0.52	0.12	0.04*	0.56
$h = 12$	0.52	0.52	0.36	0.29	0.21	0.49	0.13	0.65	0.40	0.14

[1] Notes: Results for $h = 1$ are based on the p.value of the test proposed by Berkowitz (2001).Results for $h = 12$ are based on the p.value of Knuppel's (2015) test. * marks rejection at the 5% significance level.

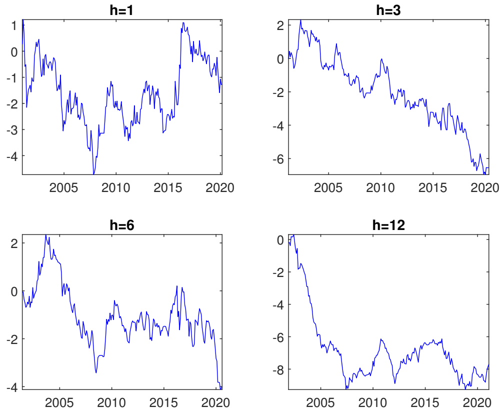
Discussion: Decomposing parameter time variation

Our empirical analysis shows that introducing time-variation in slope and volatility improves forecast accuracy for most commodity indies. We explore here which source of parameter time variation matters the most: whether it is in the parameters of the slope coefficients, or in the volatility of the innovations.

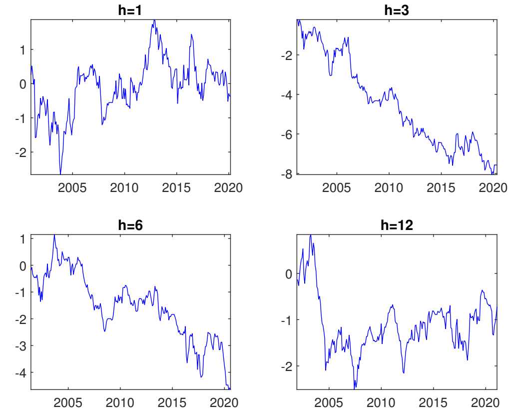
We start by comparing the performance, in terms of density forecast accuracy, of the specification that accounts for time-variation in both the slopes and volatility with that of a dynamic factor model with time-varying volatiliy but constant slopes. The results are shown in Figure 7. For most commodities, the cumulative differences in log-predictive likelihoods are negative, and are either constant or (especially at longer horizons) falling. This implies that switching-off time-variation in the slope coefficients actually yields an improvement in density forecast accuracy.

The second exercise consists of comparing the model with time-variation in both the slope and volatility with that of a model that has constant volatility but time-varying slope (Figure 8). In this case, for almost all the 10 commodities, we observe a steady increase in cumulative predictive likelihood differences. These positive, rising lines indicate that switching-off time-variation in volatility parameters, while retaining time-evolving dynamics for slope coefficients, *reduces* forecast accuracy.

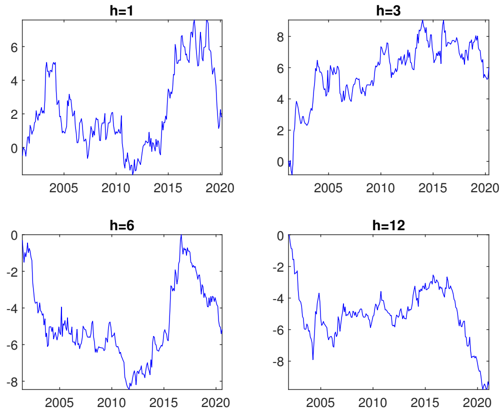
Figure 7: Cumulative differences in log-predictive likelihood of the model allowing for both time-varying slope and volatility (*tvp*) relative to the model that only allows for time-varying volatility. Increases in the statistic denote dates in which *tvp* outperforms the alternative.



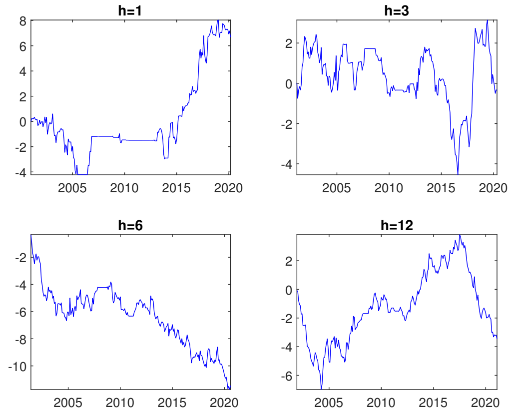
(a) All Commodity



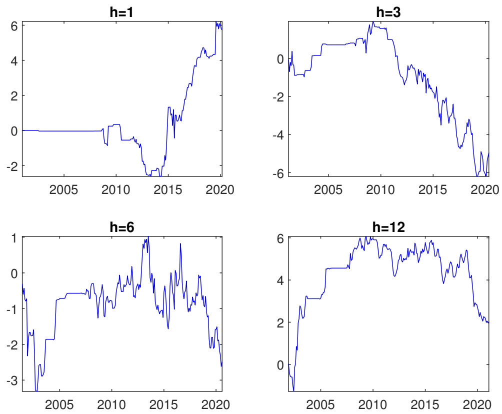
(b) Fuel



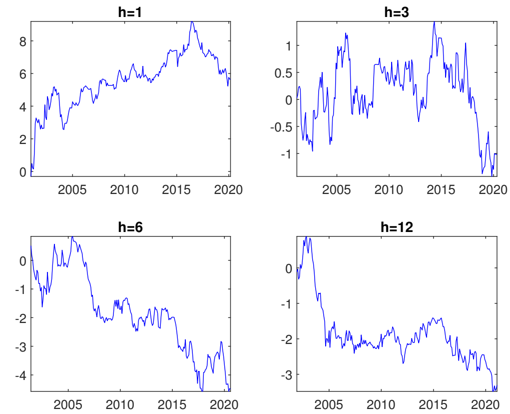
(c) Non-Fuel



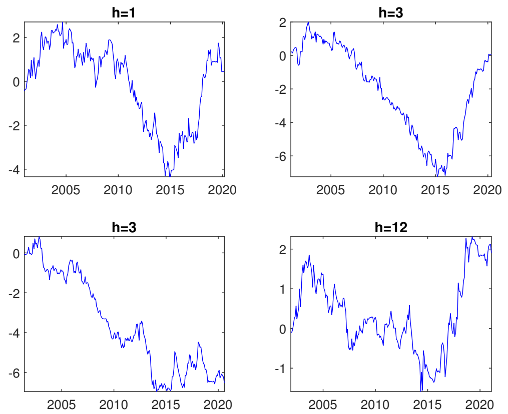
(d) Agriculture & Fertilisers



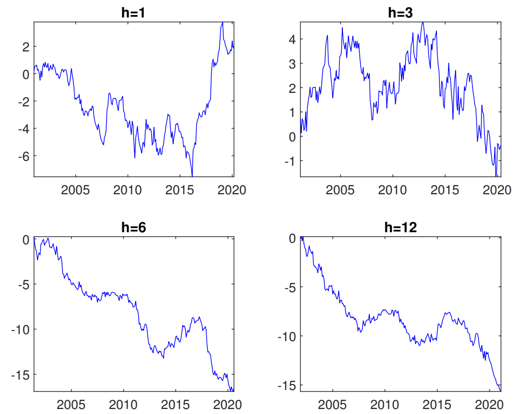
(e) Metals



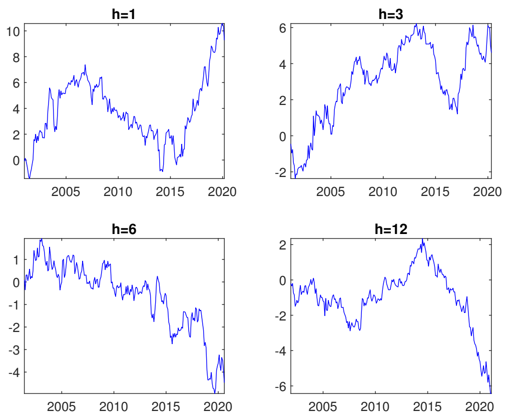
(f) Oil



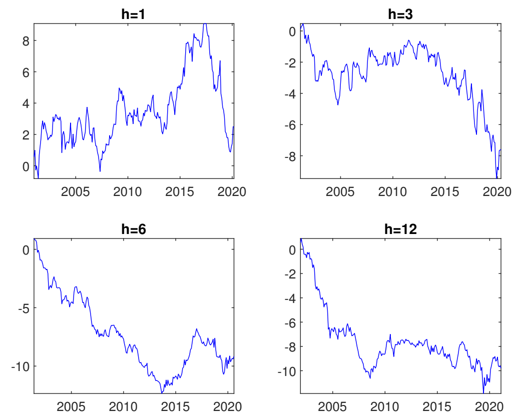
(g) Coal, Gas & Propane



(h) Food

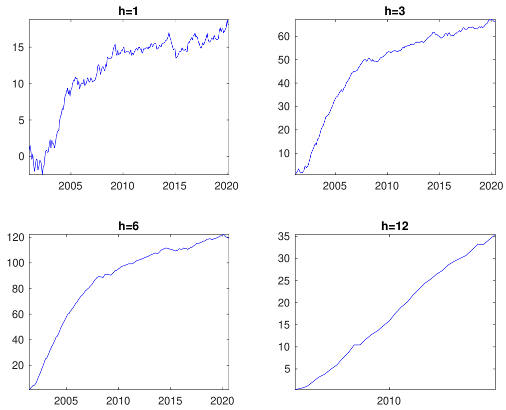


(i) Beverages



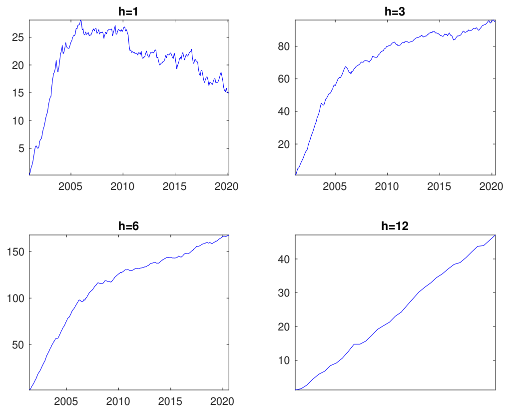
(j) Agriculture Raw Materials

Figure 8: Cumulative differences in log-predictive likelihood of the model allowing for both time-varying slope and volatility (tvp) relative to the model that only allows for time-varying slope. Increases in the statistic denote dates in which tvp outperforms the alternative.

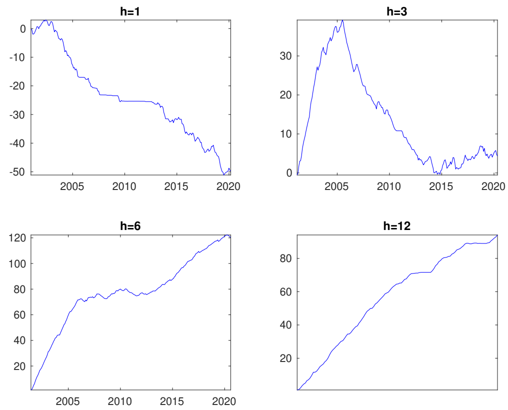


(a) All Commodity

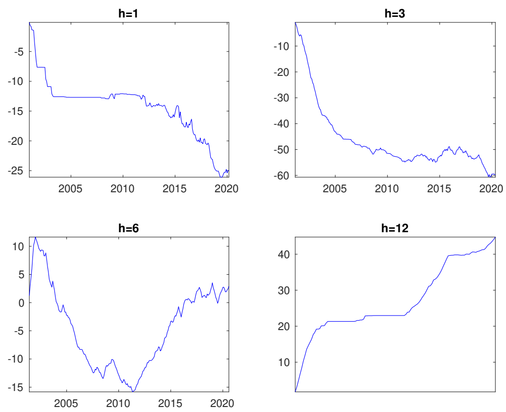
(b) Fuel



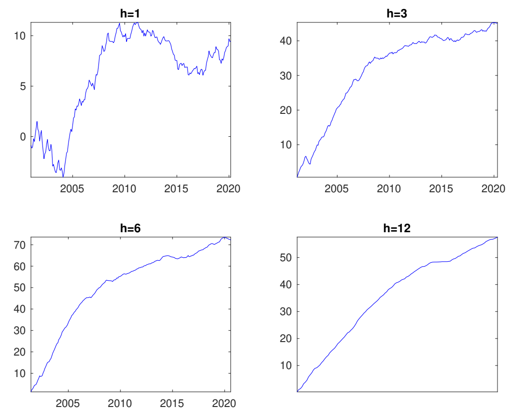
(c) Non-Fuel



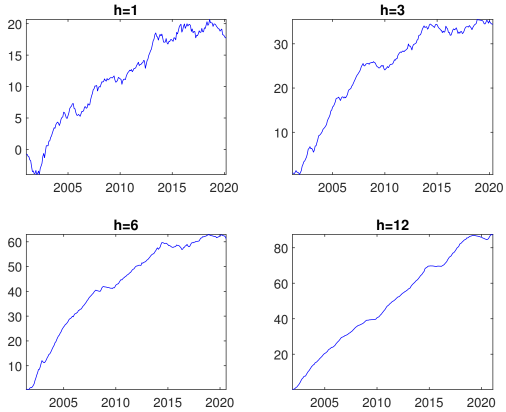
(d) Agriculture & Fertilisers



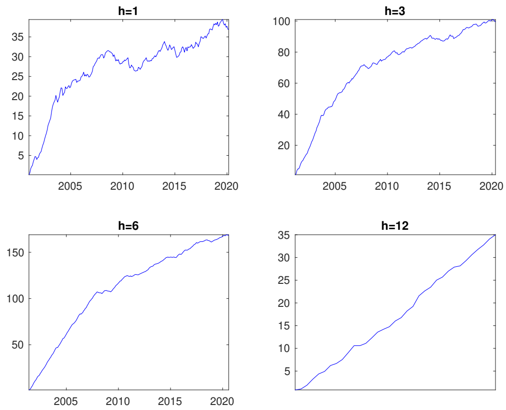
(e) Metals



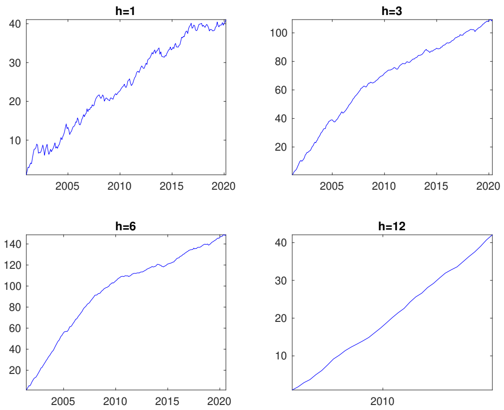
(f) Oil



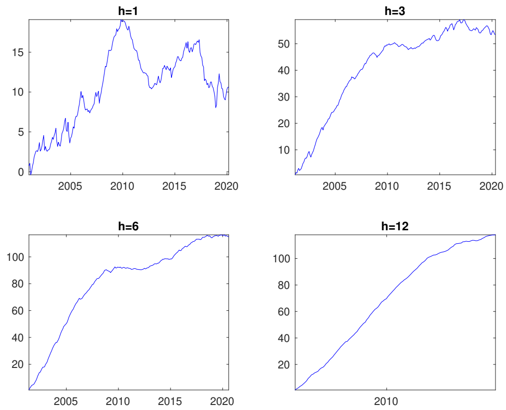
(g) Coal, Gas, & Propane



(h) Food



(i) Beverages



(j) Agriculture Raw Materials

In conclusion, this exercise shows that incorporating time-varying volatility is the key ingredient to obtain a more accurate calibration of density forecasts. Our findings are broadly in line with the macroeconomic literature that, similarly, emphasizes the importance of time-evolving second-order moments for forecast improvements.[24]

6.3 Economic evaluation

So far, we have described various (purely) statistical criteria that can be used to evaluate the competing forecast models. However, for practical purposes related to portfolio allocation, an evaluation based on economic criteria might also be interesting. An economic evaluation of the competing models is obtained following Marquering and Verbeek (2004), Campbell and Thompson (2008), and Welch and Goyal (2008). In particular, we calculate realized utility gains for a mean-variance investor. We first compute the average utility for a mean-variance investor with relative risk aversion parameter γ who allocates her portfolio on a monthly basis between commodities and risk-free bills using forecasts of commodity returns based on the competing forecasting specifications under investigation. For the forecast of the variance of commodity returns, and similar to Campbell and Thompson (2008), we assume that the investor estimates the variance using a rolling window of monthly returns. We compute the average utility for a mean-variance investor who forecasts commodity returns using an individual forecasting specification as specified in Table 3. At the end of period t the investor will decide to allocate the following share of her portfolio in period $t + 1$:

$$w_{j,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{j,t+1|t}}{\hat{\sigma}^2_{t+1|t}}\right) \tag{9}$$

and realizes an average utility level of

$$\hat{\nu}_j = \hat{\mu}_j - \left(\frac{1}{2}\right)\gamma\hat{s}^2_j \tag{10}$$

where $\hat{\sigma}^2$ is the rolling-window estimate of the variance of commodity returns, and $\hat{\mu}_j$ and $\hat{s}^2_j$ are the sample mean and variance, respectively, over the out-of-sample period for the return on the portfolio formed using forecasts based on an individual forecasting method indexed by j.

[24]Among other contributions, Clark (2011) illustrates how introducing stochastic volatility to BVAR models improves real-time accuracy of U.S. macroeconomic density forecasts, Jore et al. (2010) show that allowing for discrete breaks in variances improved U.S. density forecasts made in the Great Moderation period, while Marcellino et al. (2016) also document that introducing stochastic volatiilty to a mixed-frequency factor model contributes to higher accuracy of euro-area GDP forecasts.

The utility gain (or certainty equivalent return) can be interpreted as the portfolio management fee that an investor would be willing to pay to have access to the additional information available in a forecast method j relative to the information coming from the benchmark model alone. We express the utility gain in average annualized percentage return. [25] The constant dynamic factor model (*dfm*) is considered to be the benchmark model and, as such, positive values indicate that the alternative competing specifications perform better. We report results for $\gamma = 3$; the results are qualitatively similar for $\gamma = 6$. In the context of mean-variance analysis, another commonly used measure of economic value is the Sharpe ratio. This measure, which is defined as the ratio between the mean return of a portfolio and its standard deviation, summarizes the mean–variance trade-off of a given investment strategy. Despite evidence (e.g. Han (2006)) that Sharpe ratios tend to overestimate the conditional risk an investor faces at each point in time, thereby underestimating the performance of dynamic trading strategies, we report Sharpe ratios as a complement to the reported utility differences. Following Della Corte et al. (2009) and Lo (2002), we adjust for serial correlation in the monthly returns by multiplying the monthly Sharpe ratios by the adjustment factor

$$\frac{12}{\sqrt{12 + 2\sum_{k=1}^{11}(12-k)\rho_k}} \tag{11}$$

where ρ_k is the autocorrelation coefficient of returns at lag k.

Table 7 presents results for the economic significance of the competing forecasting models. In particular, we report utility gains $\bar{U}$ and Sharpe ratios for four main forecasting strategies (*dfm*, *dfmBlocks*, *tvp*, *tvpBlocks*). We use the baseline constant-coefficient *dfm* as the benchmark model. Three main results emerge. *First*, introducing a time-varying or/and a hierarchical factor structure to the baseline (time-invariant) dynamic factor model enhances its economic value. For all ten commodity indices, one of three alternative competing factor specifications delivers higher Sharpe ratios and positive utility. *Second*, focusing on the relevance of capturing

[25]The impact of transaction costs is an essential consideration in assessing the profitability of trading strategies. Depending on the cost of trading and the extent of weights fluctuation, different trading strategies could be more costly to implement. A realistic evaluation of the profitability of competing strategies should, therefore, take into account the effect of transaction costs. In general, making an accurate determination of the size of transaction costs is challenging because it involves three (main) factors: (i) whether the investor is an individual or institutional, (ii) the precise value of the transaction, and (iii) whether the broker belongs to a brokerage firm or whether she engages in direct internet trading. We account for the effect of transaction costs by computing the relevant performance measures for the investor's realized returns net of transaction costs. A wide range of estimates has been used in empirical studies, ranging from transaction costs between 0.1% and 2%. We set the proportional transaction costs to 1%, representing a high-cost regime.

time-evolving dynamics, strategies based on time varying parameter factor models yield (on average) higher Sharpe ratios than the benchmark constant parameter model (*dfm*). In particular, this strong performance is mostly driven by the time-varying hierarchical model (*tvpBlocks*) which improves upon the benchmark for all ten commodity indices. With the exception of the *Beverages* index, utility gains are mostly in agreement with the documented performance based on Sharpe ratios. *Lastly*, accurately capturing commodity heterogeneity appears to be particularly important for economic gains. Precisely, our economic evaluation criteria suggest that, for 9 out of 10 commodity indices, the model that delivers the highest economic gains is either the constant-coefficient or the time-varying block-structure specification. As previously, the only exception is the *Beverages* index for which a standard time-varying factor model yields the highest economic value. Nonetheless, it becomes easily apparent that time-variation helps to 'reveal' the economic value of hierarchical structure models.

Table 7: For each strategy in the row header, we report the Sharpe ratio S_R and utility gains $\bar{U}$ (monthly utility changes are annualized) of an investor using any of the three predictive models (*dfmBlocks*, *tvp*, *tvpBlocks*) relative to an investor using the benchmark constant-coefficient factor model (*dfm*). Underlined values denote the best performing strategy for each statistic.

	All	Fuel	NonFuel	AgrFer	Metals	Oil	CGP	Food	Beve	AgrRaw
dfm										
S_R	0.06	0.04	0.12	0.13	0.09	0.04	0.01	0.08	0.05	0.10
$\bar{U}$	-	-	-	-	-	-	-	-	-	-
dfmBlocks										
S_R	0.09	<u>0.09</u>	0.10	0.14	0.08	<u>0.10</u>	0.02	0.07	0.03	<u>0.15</u>
$\bar{U}$	0.09	<u>0.31</u>	-0.03	0.01	-0.03	<u>0.33</u>	0.11	-0.05	-0.08	<u>0.05</u>
tvp										
S_R	0.10	0.06	0.11	0.15	0.05	0.07	0.03	0.14	<u>0.07</u>	0.10
$\bar{U}$	0.08	0.12	0.00	0.03	-0.13	0.05	0.19	0.07	<u>0.03</u>	-0.01
tvpBlocks										
S_R	<u>0.14</u>	0.08	<u>0.17</u>	<u>0.22</u>	<u>0.11</u>	0.07	<u>0.04</u>	<u>0.17</u>	0.06	0.14
$\bar{U}$	<u>0.15</u>	0.15	<u>0.04</u>	<u>0.09</u>	<u>0.07</u>	0.13	<u>0.28</u>	<u>0.13</u>	-0.01	0.01

7 Conclusions

Recent research has shown that commodity prices exhibit substantial co-movement, which can be captured by few "common" factors. These factors are broadly related to global *demand for commodity* shocks, which are pervasive across *all* commodity prices, and idiosyncratic (commodity-specific) supply shocks. A separate literature has stressed how the composition of underlying structural shocks that drive commodity prices has changed over time, with the importance of global demand shocks becoming higher since 2000 due to strong growth in emerging economies, potentially resulting in unstable unconditional correlations across commodity prices. These two findings suggest that (i) forecast accuracy for the price of a *given commodity* could benefit from the information contained in *other commodity* prices and that (ii) dealing with potential structural breaks could also improve forecast accuracy.

In this paper we investigate the merits of constructing forecasts for key commodity prices and indices from models that make use of a large information set and that can deal with structural breaks. Among others, we consider large TVP dynamic factor models and TVP *hierarchical* dynamic factor models that impose the presence of specific blocks on the factor model structure of commodity prices, as well as their constant-coefficient counterparts. Given that standard estimation methods for small-dimensional models fail in a data-rich environment, we adopt estimation techniques that allow us to tackle the issue of dimensionality.

Overall, we find that the out-of-sample predictability of commodity prices varies substantially across economic states and different commodities. In terms of point forecast accuracy, competing models exhibit similar performance, with significant gains being documented throughout different periods. Focusing on the potential significance of adopting a specific block structure, we document how this feature can, in some cases, improve density forecast accuracy at short-term horizons. In contrast, we find that modeling parameter time variation exhibits greater predictive content, with gains being consistent across commodities and forecast horizons. In particular, we illustrate how appropriately modeling time-varying volatility drives the bulk of density forecast improvements coming from time-varying specifications. Lastly, trading strategies based on the various forecast models show that controlling for the high degree of commonalities leads to higher Sharpe ratios, and to higher values for investors.

A Appendix

Our Appendix contains a description of the 10 commodity indices and 68 individual commodity prices, and explores additional density forecast evaluation exercises for the specifications presented in the main section.

A.1 Data

Table A1: Commodity Price Indices

Description (Mnemonic)	Weight
All Commodity (PALLFNF)	100.0
Fuel (PNRG)	59.1
Non-Fuel (PNFUEL)	40.9
Agriculture and Fertilizers (N/A)	36.4
Metals (PMETA)	22.7
Oil (POILAPSP)	28.6
Coal, Natural Gas and Propane (N/A)	12.3
Food (PFOOD)	27.8
Beverages (PBEVE)	2.3
Agricultural Raw Materials (PRAWM)	4.3

Table A2: Individual Commodity Prices

Data Type	Commodity	Commodity.Description
USD	PALUM	Aluminum, 99.5% minimum purity, LME spot price, CIF UK ports, US$ per metric ton
USD	PBANSOP	Bananas, Central American and Ecuador, FOB U.S. Ports, US$ per metric ton
USD	PBARL	Barley, Canadian no.1 Western Barley, spot price, US$ per metric ton
USD	PBEEF	Beef, Australian and New Zealand 85% lean fores, CIF U.S. import price, US cents per pound
USD	PCOALAU	Coal, Australian thermal coal, 12,000- btu/pound, less than 1% sulfur, 14% ash, FOB Newcastle/Port Kembla, US$ per metric ton
USD	PCOALSA	Coal South African export price US$ per metric ton
USD	PCOCO	Cocoa beans, International Cocoa Organization cash price, CIF US and European ports, US$ per metric ton
USD	PCOFFOTM	Coffee, Other Mild Arabicas, International Coffee Organization New York cash price, ex-dock New York, US cents per pound
USD	PCOFFROB	Coffee, Robusta, International Coffee Organization New York cash price, ex-dock New York, US cents per pound
USD	PROIL	Rapeseed oil, crude, fob Rotterdam, US$ per metric ton
USD	PCOPP	Copper, grade A cathode, LME spot price, CIF European ports, US$ per metric ton
USD	PCOTTIND	Cotton, Cotton Outlook 'A Index', Middling 1-3/32 inch staple, CIF Liverpool, US cents per pound
USD	PFSHMEAL	Fishmeal, Peru Fish meal/pellets 65% protein, CIF, US$ per metric ton
USD	PGNUTS	Groundnuts (peanuts), 40/50 (40 to 50 count per ounce), cif Argentina, US$ per metric ton
USD	PHIDE	Hides, Heavy native steers, over 53 pounds, wholesale dealer's price, US, Chicago, fob Shipping Point, US cents per pound
USD	PIORECR	China import Iron Ore Fines 62% FE spot (CFR Tianjin port), US dollars per metric ton
USD	PLAMB	Lamb, frozen carcass Smithfield London, US cents per pound
USD	PLEAD	Lead, 99.97% pure, LME spot price, CIF European Ports, US$ per metric ton
USD	PLOGORE	Soft Logs, Average Export price from the U.S. for Douglas Fir, US$ per cubic meter
USD	PLOGSK	Hard Logs, Best quality Malaysian meranti, import price Japan, US$ per cubic meter
USD	PMAIZMT	Maize (corn), U.S. No.2 Yellow, FOB Gulf of Mexico, U.S. price, US$ per metric ton
USD	PNGASEU	Natural Gas, Netherlands TTF Natural Gas Forward Day Ahead, US$ per Million Metric British Thermal Unit
USD	PNGASJP	Natural Gas, Indonesian Liquefied Natural Gas in Japan, US$ per Million Metric British Thermal Unit
USD	PNGASUS	Natural Gas, Natural Gas spot price at the Henry Hub terminal in Louisiana, US$ per Million Metric British Thermal Unit
USD	PNICK	Nickel, melting grade, LME spot price, CIF European ports, US$ per metric ton
		Crude Oil (petroleum), simple

Data Type	Commodity	Commodity.Description
USD	POLVOIL	Olive Oil, extra virgin less than 1% free fatty acid, ex-tanker price U.K., US$ per metric ton
USD	PORANG	Oranges, miscellaneous oranges CIF French import price, US$ per metric ton
USD	PPOIL	Palm oil, Malaysia Palm Oil Futures (first contract forward) 4-5 percent FFA, US$ per metric ton
USD	PPORK	Swine (pork), 51-52% lean Hogs, U.S. price, US cents per pound.
USD	PPOULT	Poultry (chicken), Whole bird spot price, Ready-to-cook, whole, iced, Georgia docks, US cents per pound
USD	PRICENPQ	Rice, 5 percent broken milled white rice, Thailand nominal price quote, US$ per metric ton
USD	PRUBB	Rubber, Singapore Commodity Exchange, No. 3 Rubber Smoked Sheets, 1st contract, US cents per pound
USD	PSALM	Fish (salmon), Farm Bred Norwegian Salmon, export price, US$ per kilogram
USD	PSAWMAL	Hard Sawnwood, Dark Red Meranti, select and better quality, C&F U.K port, US$ per cubic meter
USD	PSAWORE	Soft Sawnwood, average export price of Douglas Fir, U.S. Price, US$ per cubic meter
USD	PSHRI	Shrimp, No.1 shell-on headless, 26-30 count per pound, Mexican origin, New York port, US$ per kilogram
USD	PSMEA	Soybean Meal, Chicago Soybean Meal Futures (first contract forward) Minimum 48 percent protein, US$ per metric ton
USD	PSOIL	Soybean Oil, Chicago Soybean Oil Futures (first contract forward) exchange approved grades, US$ per metric ton
USD	PSOYB	Soybeans, U.S. soybeans, Chicago Soybean futures contract (first contract forward) No. 2 yellow and par, US$ per metric ton
USD	PSUGAISA	Sugar, Free Market, Coffee Sugar and Cocoa Exchange (CSCE) contract no.11 nearest future position, US cents per pound
USD	PSUGAUSA	Sugar, U.S. import price, contract no.14 nearest futures position, US cents per pound (Footnote: No. 14 revised to No. 16)
USD	PSUNO	Sunflower oil, Sunflower Oil, US export price from Gulf of Mexico, US$ per metric ton
USD	PTEA	Tea, Mombasa, Kenya, Auction Price, US cents per kilogram, From July 1998,Kenya auctions, Best Pekoe Fannings. Prior, London auctions, c.i.f. U.K. warehouses
USD	PTIN	Tin, standard grade, LME spot price, US$ per metric ton
USD	PURAN	Uranium, NUEXCO, Restricted Price, Nuexco exchange spot, US$ per pound
USD	PWHEAMT	Wheat, No.1 Hard Red Winter, ordinary protein, Kansas City, US$ per metric ton
USD	POATS	Oats, Generic 1st 'O' Future, USD/bushel
USD	PSORG	Sorghum, U.S., Number 2 yellow, fob Gulf of Mexico USD cents per pound
USD	PWOOLC	Wool, coarse, 23 micron, Australian Wool Exchange spot quote, US cents per kilogram
USD	PWOOLF	Wool, fine, 19 micron, Australian Wool Exchange spot quote, US cents per kilogram
USD	PZINC	Zinc, high grade 98% pure, US$ per metric ton

Data Type	Commodity	Commodity.Description
USD	PLMMODY	Molybdenum, 57 to 63% purity contained in roasted molybdenum concentrate, LME spot price, USD/ton
USD	PCOBA	Cobalt, U.S. cathodes, spot
USD	PGOLD	Gold, Fixing Committee of the London Bullion Market Association, London 3 PM fixed price, USD/troy ounce
USD	PSILVER	Silver, London Bullion Market Association, USD/troy ounce
USD	PPALLA	Palladium, LME spot price, USD/troy ounce
USD	PPLAT	Platinum, LME spot price, USD/troy ounce
USD	PPROPANE	North American Spot LPG, Propane Price/Mont Belvieu LST
USD	PUREA	US Gulf NOLA, Urea Granular Spot Price, USD/ST
USD	PPOTASH	Potassium Chloride, Standard Grade: FOB Vancouver Spot Price, USD/metric tonne
USD	PDAP	Diammonium phosphate, US Gulf NOLA DAP Epxort Spot Price per MT, USD/metric tonne
USD	PTOMATO	Monthly average consumer prices in metropolitan France, Tomatoes (1 Kg), EUR
USD	PMILK	USDA Class 3 Milk Spot Price, USD/cwt
USD	PCHANA	MCX India, Chana Spot INR/100 Kgs
USD	PAPPLE	Monthly average consumer prices in metropolitan France, Apples (1 Kg), EUR

A.2 Dynamic factor models

We retain the factor model representation given in equations 1 and 2. For the idiosyncratic components ϵ_{it}, we impose the following decomposition

$$\epsilon_{it} = \sum_{j=1}^{K} l_{ij} g_{jt} + \nu_{it} \tag{A.1}$$

$$l_{ij} = \begin{cases} \neq 0, & \text{if } i \in j \\ 0, & \text{otherwise} \end{cases} \tag{A.2}$$

where g_{jt} is the vector of block factors, l_{ij} the block factor loadings and ν_{it} the purely idiosyncratic component. Equation A.2 describes how the block structure is imposed: whenever the commodity i does not belong to the block j, the associated factor loadings l_{ij} are set equal to zero. The block diagonal matrix is of the form:

$$\begin{bmatrix} l_{11} & 0 & \dots & 0 \\ 0 & l_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & l_{KK} \end{bmatrix}$$

where K denotes the different blocks specified in Table 1.

Moreover, the block factors g_{jt} and the purely idiosyncratic component ν_{it} are assumed to follow independent autoregressive processes

$$g_{jt} = \phi_j g_{jt-1} + w_{jt} \quad w_{jt} \sim \text{i.i.d}N(0, 1) \tag{A.3}$$

$$\nu_{it} = \rho_i \nu_{it-1} + e_{it} \quad e_{it} \sim \text{i.i.d}N(0, \sigma_i^2) \tag{A.4}$$

A.3 Estimation of time-varying specifications

Let $x_{it} = (x_{1t}, ..., x_{nt})'$ be an $n-$ dimensional vector of variables that follows a dynamic factor model of the form:

$$x_{it} = \lambda_{it} f_t + \epsilon_{it}, \tag{A.5}$$

$$f_t = B_t f_{t-1} + \eta_t, \tag{A.6}$$

where f_t is the $k \times 1$ vector of factors, λ_{it} is the $n \times k$ factor loadings, B_t is a $k \times k$ matrix of $VAR(1)$ coefficients and ϵ_{it} and η_t are disturbance terms. It is further assumed that $\epsilon_t \sim N(0, V_t)$ and $\eta_t \sim N(0, Q_t)$ where V_t and Q_t are the $n \times n$ and $k \times k$ diagonal covariance matrices respectively. Note that the ϵ_{it} are uncorrelated with both f_t and η_t at all leads and lags. In order to complete the description of the TVP-DFM model we need to define how the time-varying parameters evolve. We allow λ_t and β_t to evolve as driftless random walks:

$$\lambda_t = \lambda_{t-1} + u_t \quad u_t \sim N(0, R_t), \tag{A.7}$$

$$\beta_t = \beta_{t-1} + v_t \quad v_t \sim N(0, W_t). \tag{A.8}$$

The model has a standard state space representation where equations A.5 are the measurement equations and A.6 to A.8 are the state equations. The state vector f_t, λ_t, β_t are estimated via the Kalman smoother, provided that an estimate of the covariances, V_t, Q_t, R_t, W_t is available. We assume that errors across blocks of equations are uncorrelated, i.e. that u_t and v_t are *i.i.d.* errors, uncorrelated with each other as well as with ϵ_t and η_t at all leads and lags.[26] The model covariances are estimated using a standard forgetting factor algorithm. First, R_t and W_t evolve as follows:

$$R_t = \left(\frac{1-\theta_R}{\theta_R}\right) P^{\lambda}_{t-1/t-1},$$
$$W_t = \left(\frac{1-\theta_W}{\theta_W}\right) P^{\beta}_{t-1/t-1},$$

where $P^{\lambda}_{t-1/t-1}$ and $P^{\beta}_{t-1/t-1}$ are the estimated covariance matrices of the unobserved state vectors λ_t and β_t in the model. The smoothing parameters θ_R and θ_W are set at 0.96. The matrices V_t and Q_t are estimated by suitably discounting past squared one step ahead prediction

[26] See, for instance, Cooley (1971); Koop and Korobilis (2012)

errors:

$$\widehat{V}_t = \kappa_v \widehat{V}_{t-1} + (1 - \kappa_v)\epsilon_t \epsilon_t' \tag{A.9}$$
$$\widehat{Q}_t = \kappa_Q \widehat{Q}_{t-1} + (1 - \kappa_Q)\eta_t \eta_t'$$

where ϵ_t is the vector that collects the measurement errors in equation A.5 and κ_v and κ_Q are also set at 0.96.[27]

[27] This framework is flexible enough to enable the estimation of the degree of evolution of the model's parameters from the data, therefore allowing for modeling different degrees of time variation and, when necessary, no variation at all. Results based on such an approach do not improve upon our baseline time-varying specifications. Moreover, we have estimated specifications with different pre-specified values than the ones reported above. None of these additional specifications enhance predictability and, therefore, results are not reported.

A.4 Density scores evaluation criteria

Suppose the forecasting exercise aims at comparing two density forecasts of models $i = 1, 2$

$$\log(p_{i,h,t}(y_{t+h}|F_{i,t-1})), \quad i = 1, 2 \tag{A.10}$$

where $p_{i,h,t}(*)$ denotes the predictive likelihood of model i at horizon h, y is the vector of target variable(s), $F_{i,t-1}$ is the information set of model i available at time t. The KLIC differential between them is the expected difference in their log-predictive likelihood. In particular, we will focus on the cumulative differences between the two likelihoods

$$S_{j,h} = \sum_{t=1}^{T-h} [\log(p_{1,h,t}(y_{t+h}|F_{1,t-1})) - \log(p_{2,h,t}(y_{t+h}|F_{2,t-1}))] \tag{A.11}$$

When comparing two different predictive densities, the average difference between their logarithms is inherently related to their KLIC distance. Among alternative models, choosing the one with the highest log-predictive likelihood is equivalent to selecting the model with the minimal KLIC distance.[28] To assess whether any detected differences in the log-likelihoods of the models under investigation are statistically significant, we employ the equal predictive ability test proposed by Amisano and Giacomini (2007).

A.5 Additional density forecast evaluations

A.5.1 Additional log-score differentials

Although the focus of our work is identifying the benefits of accounting for time-variation and commodity-specific co-movement for the purpose of forecasting, Figure A.1 presents a density forecasting comparison between the model that combines both of these features (*tvpBlocks*) and the random walk benchmark. Overall, the strictly positive and increasing line provides strong evidence in favour of the predictive power of the time-varying hierarchical specification.

In the main section presenting our main forecasting results we have already established that introducing time-variation to dynamic factor models is particularly beneficial to density forecasting accuracy. As a robustness check, here we illustrate whether this result holds for hierarchical factor models as well. Figure A.2 presents an illustration of the cumulative log-score differential between the time-varying hierarchical model (*tvpBlocks*) and its constant-coefficient counterpart (*dfmBlocks*). An increasing and positive-values line signals a better performance of

[28]See the discussion in Hall and Mitchell (2007) and Geweke and Amisano (2010).

the *tvpBlocks* specification. Looking at the overall picture, it becomes apparent that time-varying dynamics help revealing the predictive content of hierarchical factor structures. At the shortest forecast horizon of one-month ahead, *Non-fuel* commodity indices such as *Food*, *Beverages* and *Agricultural Raw Materials* benefit the most. Gains continue to be strong for longer horizons, albeit of smaller magnitude. At the longest horizon of one-year ahead the largest density forecast improvements are documented for the *Agricultural Raw Materials* index, while the index that seems to benefit the least is *Oil*. In summary, findings presented in both the main section and appendix emphasize the important role of appropriately capturing structural instabilities in delivering accurate density forecasts.

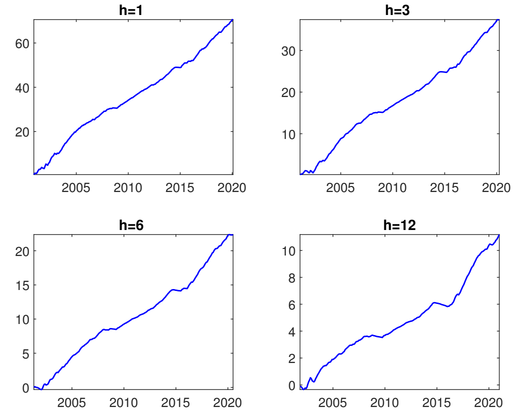
Table A3: Equal predictive ability test.

	All	Fuel	NF	AF	Metals	Oil	CGP	Food	Beve	AR
tvpBlocks										
$h = 1$	-10.45*	-6.86*	-11.38*	-8.24*	-5.58*	-7.83*	-7.81*	-10.28*	-12.50*	-11.16*
$h = 3$	-9.63*	-5.29*	-10.20*	-7.47*	-6.91*	-5.03*	-7.48*	-8.37*	-9.89*	-7.80*
$h = 6$	-9.88*	-5.83*	-6.34*	-8.80*	-4.61*	-6.05*	-4.97*	-8.36*	-13.54*	-8.65*
$h = 12$	-9.87*	-5.58*	-5.09*	-7.00*	-4.38*	-4.09*	-5.06*	-4.54*	-7.22*	-9.18*

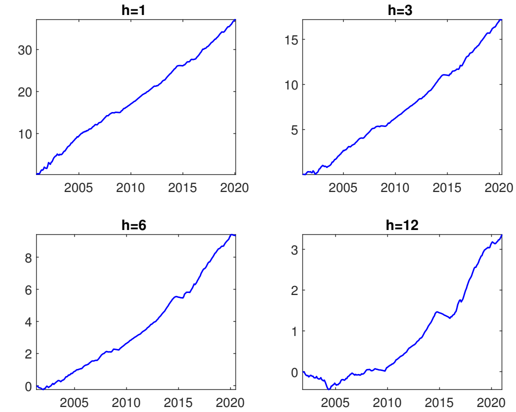
[1] The table reports the t-statistics for the null hypothesis that the model under investigation has the same predictive ability as the dynamic factor model with constant coefficients. Positive values with one asterisk (*) denote higher log-predictive likelihood at 5% significance level.

[2] NF correspond to the Non-Fuel index, AF to the Acriculture & Fertilizers, CGP to the Coal, Natural Gas & Propane, and AR to the Agriculture Raw Materials.

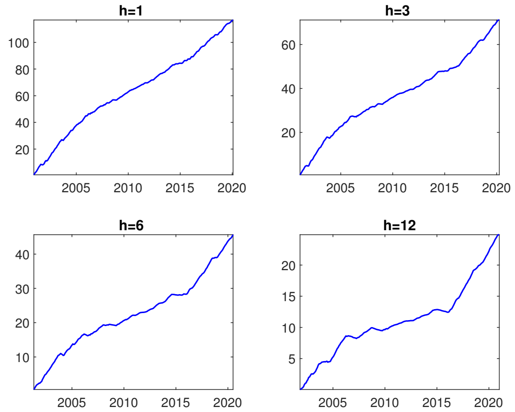
Figure A.1: Cumulative differences in log-predictive likelihood of the time-varying hierarchical dynamic factor model (tvpBlocks) relative to the random walk benchmark. Increases in the statistic denote dates in which tvpBlocks outperforms the alternative.



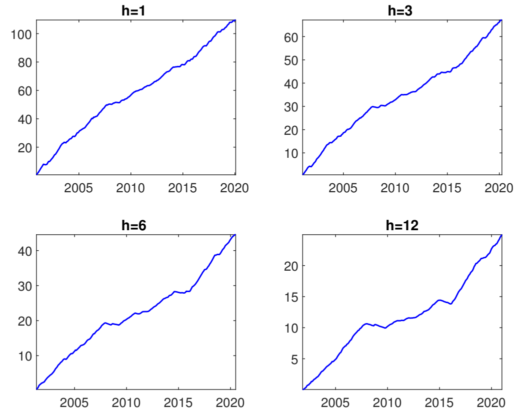
(a) All Commodity



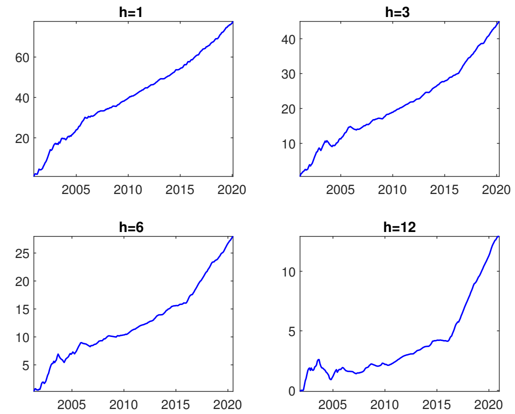
(b) Fuel



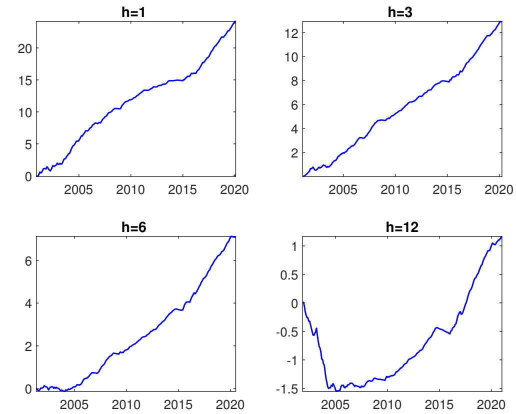
(c) Non-Fuel



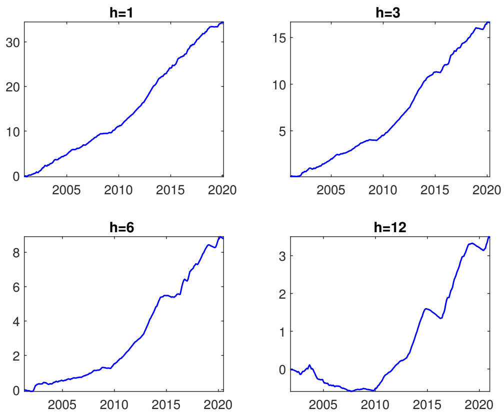
(d) Agriculture & Fertilizers



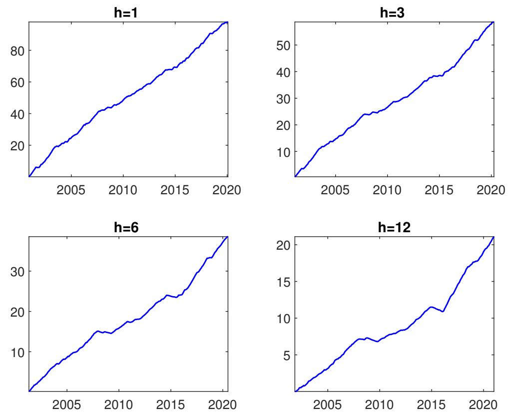
(e) Metals



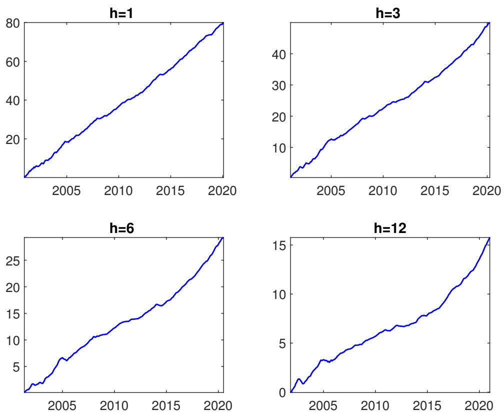
(f) Oil



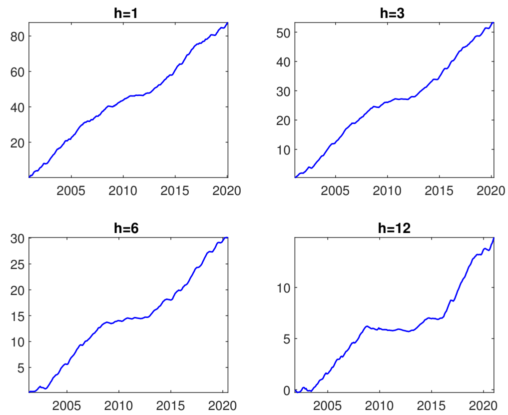
(g) Coal, Natural Gas & Propane



(h) Food

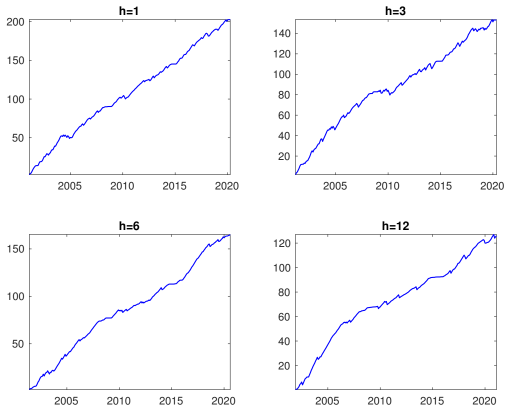


(i) Beverages

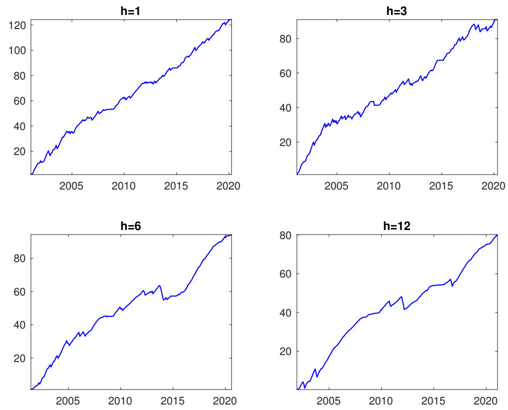


(j) Agriculture Raw Materials

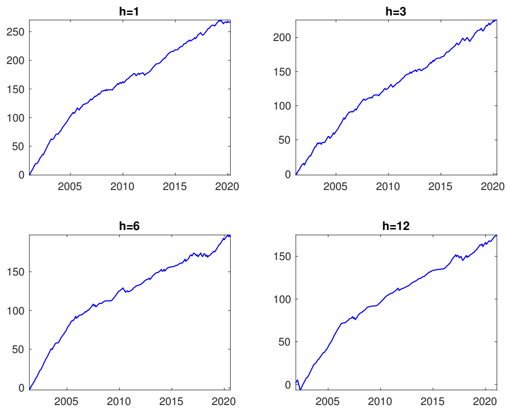
Figure A.2: Cumulative differences in log-predictive likelihood of the time-varying hierarchical dynamic factor model (tvpBlocks) relative to its constant-coefficient counterpart (dfmBlocks). Increases in the statistic denote dates in which tvpBlocks outperforms the alternative.



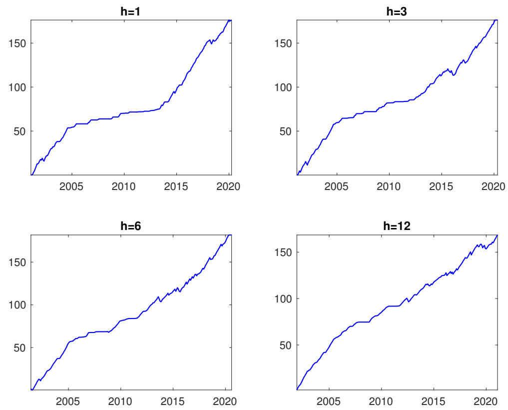
(a) All Commodity



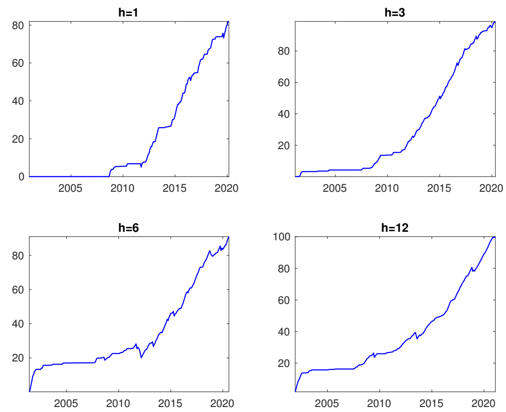
(b) Fuel



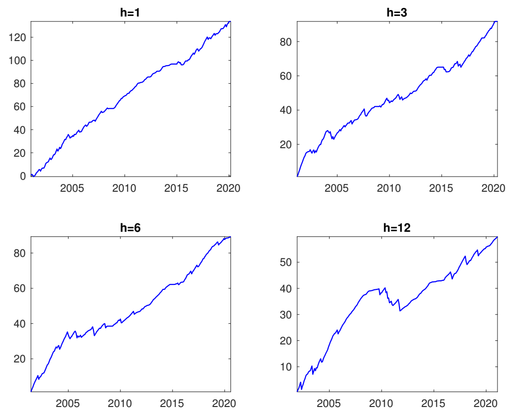
(c) Non-Fuel



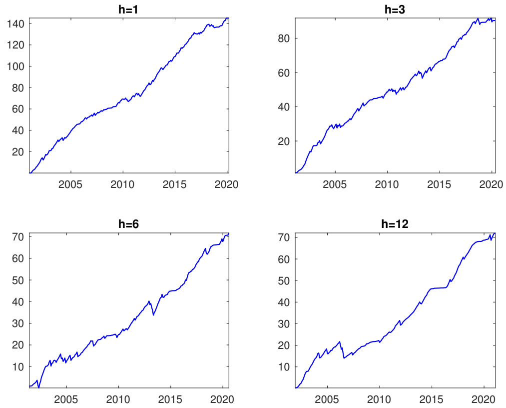
(d) Agriculture & Fertilizers



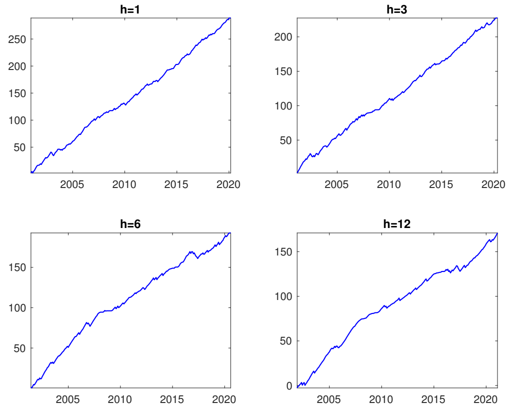
(e) Metals



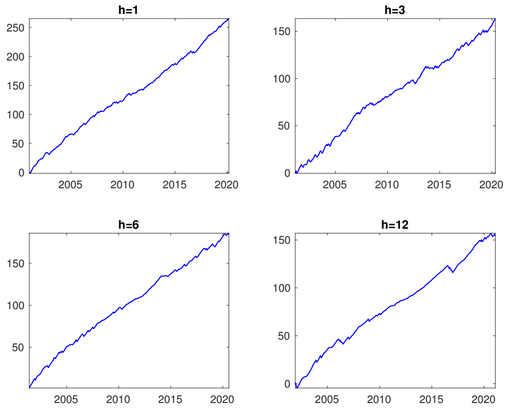
(f) Oil



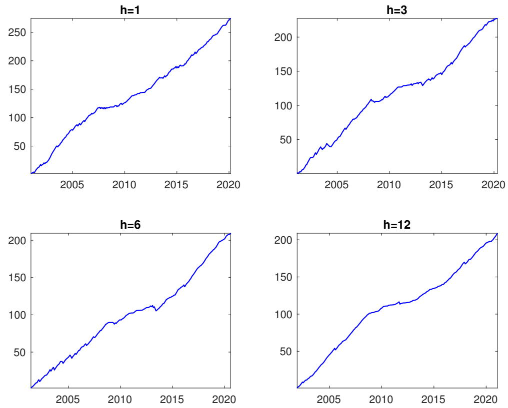
(g) Coal, Natural Gas & Propane



(h) Food



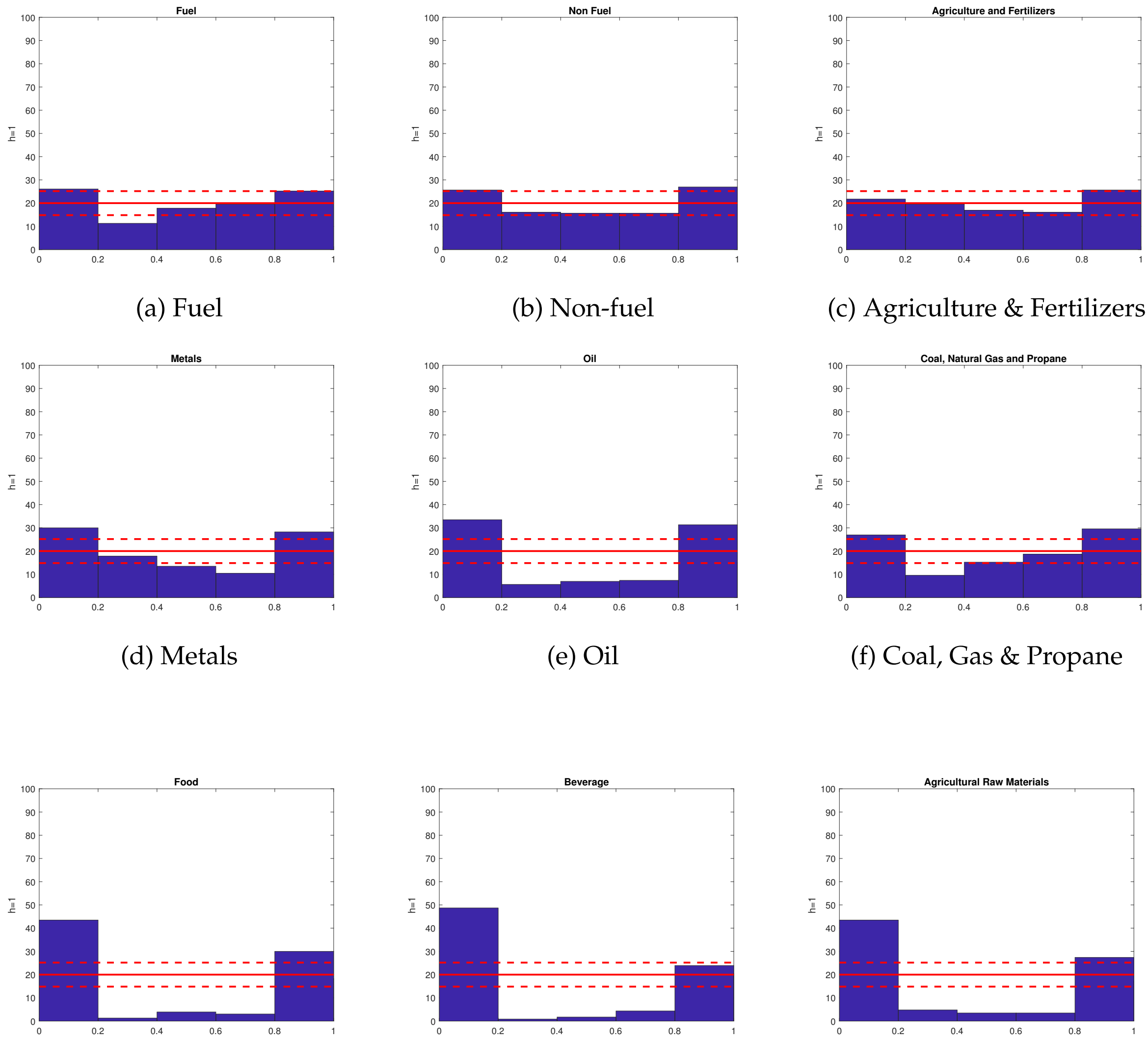
(i) Beverages



(j) Agriculture Raw Materials

A.5.2 Additional PITs ($h = 1$)

Figure A.3: Probability density functions of the PITs for the *dfm* model for one-month ahead forecasts of the nine remaining commodity indices. The dashed lines are 95% confidence intervals, constructed using a normal approximation to a binomial distribution, as per Diebold et al. (1998)

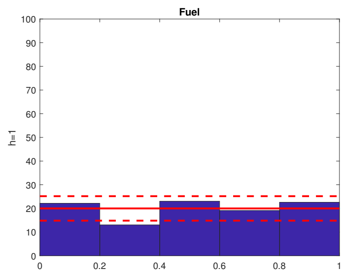


(a) Fuel (b) Non-fuel (c) Agriculture & Fertilizers

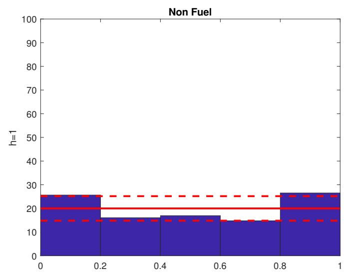
(d) Metals (e) Oil (f) Coal, Gas & Propane

(g) Food (h) Beverages (i) Agricultural Raw Materials

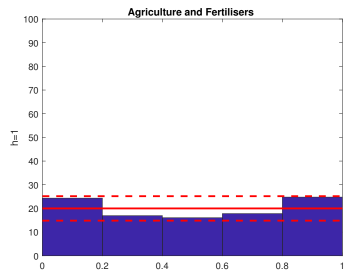
Figure A.4: Probability density functions of the PITs for the *dfmBlocks* model for one-month ahead forecasts of the nine remaining commodity indices. The dashed lines are 95% confidence intervals, constructed using a normal approximation to a binomial distribution, as per Diebold et al. (1998)



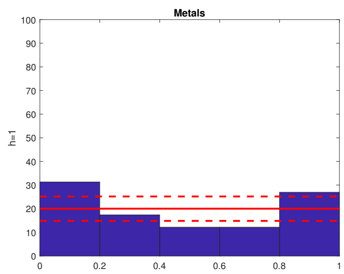
(a) Fuel



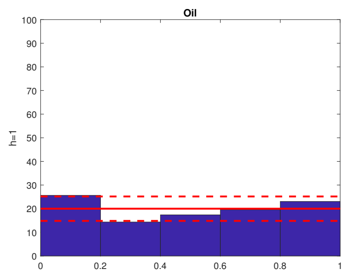
(b) Non-fuel



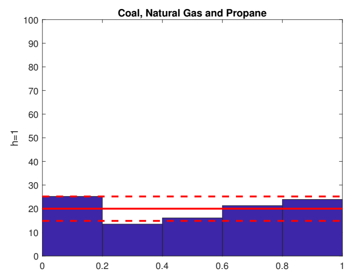
(c) Agriculture & Fertilizers



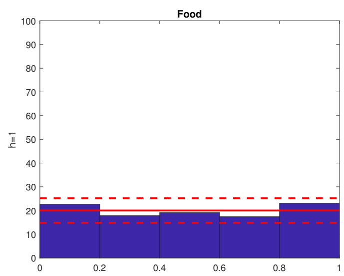
(d) Metals



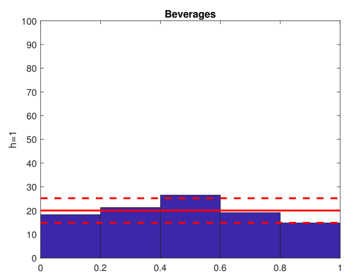
(e) Oil



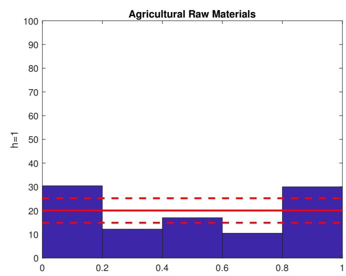
(f) Coal, Gas & Propane



(g) Food

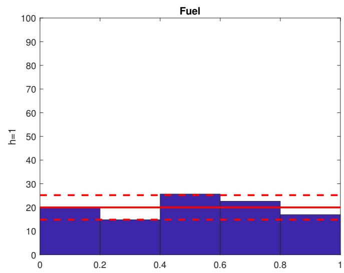


(h) Beverages

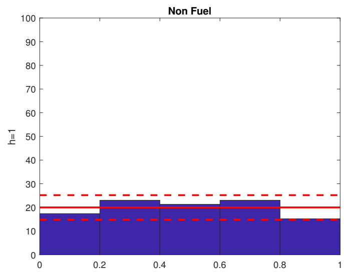


(i) Agricultural Raw Materials

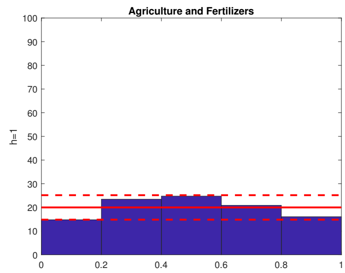
Figure A.5: Probability density functions of the PITs for the *tvp* model for one-month ahead forecasts of the nine remaining commodity indices. The dashed lines are 95% confidence intervals, constructed using a normal approximation to a binomial distribution, as per Diebold et al. (1998)



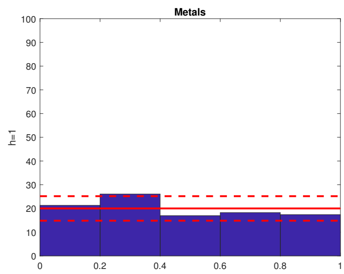
(a) Fuel



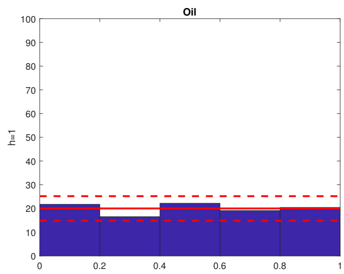
(b) Non-fuel



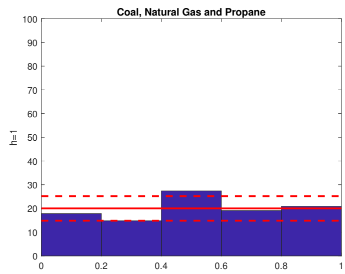
(c) Agriculture & Fertilizers



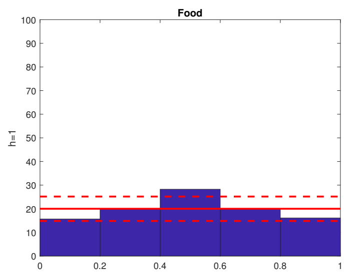
(d) Metals



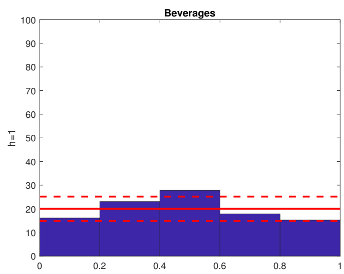
(e) Oil



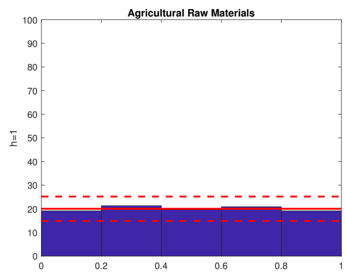
(f) Coal, Gas & Propane



(g) Food

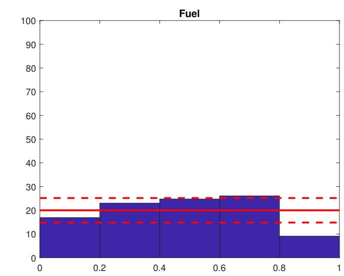


(h) Beverages

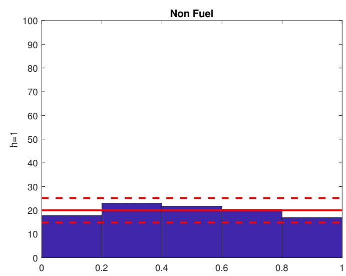


(i) Agricultural Raw Materials

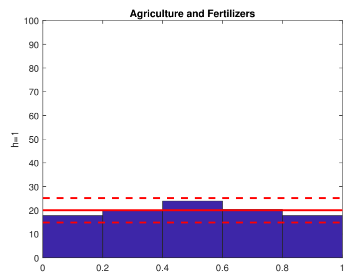
Figure A.6: Probability density functions of the PITs for the *tvpBlocks* model for one-month ahead forecasts of the nine remaining commodity indices. The dashed lines are 95% confidence intervals, constructed using a normal approximation to a binomial distribution, as per Diebold et al. (1998)



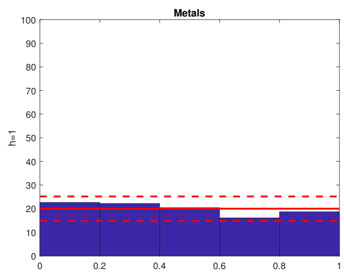
(a) Fuel



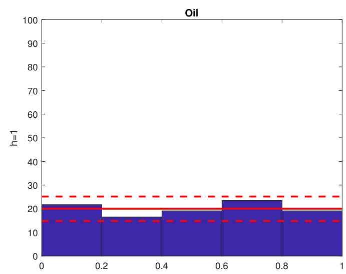
(b) Non-fuel



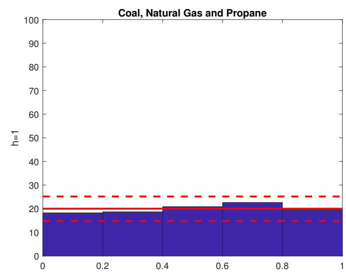
(c) Agriculture & Fertilizers



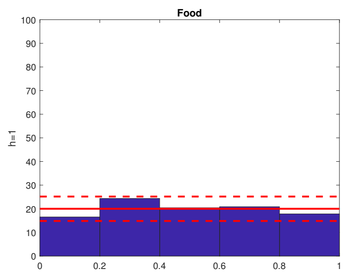
(d) Metals



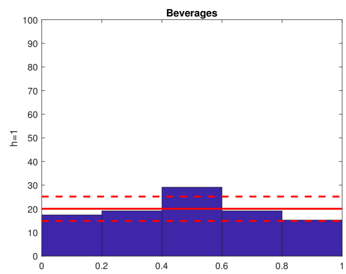
(e) Oil



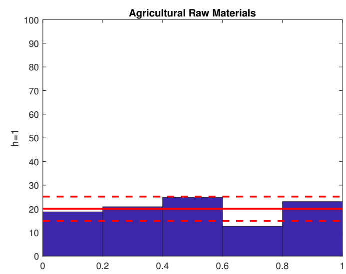
(f) Coal, Gas & Propane



(g) Food



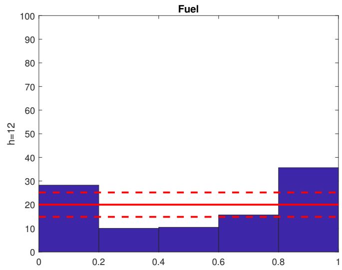
(h) Beverages



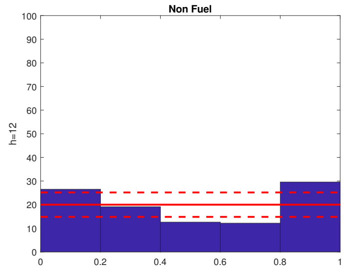
(i) Agricultural Raw Materials

A.5.3 Additional PITs ($h = 12$)

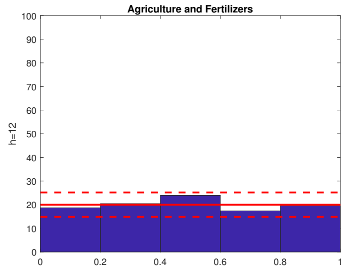
Figure A.7: Probability density functions of the PITs for the *dfm* model for one-year ahead forecasts of the nine remaining commodity indices. The dashed lines are 95% confidence intervals, constructed using a normal approximation to a binomial distribution, as per Diebold et al. (1998)



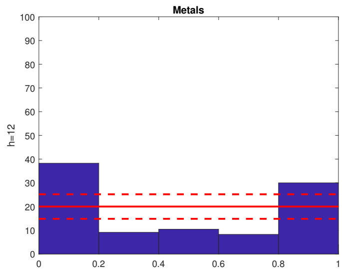
(a) Fuel



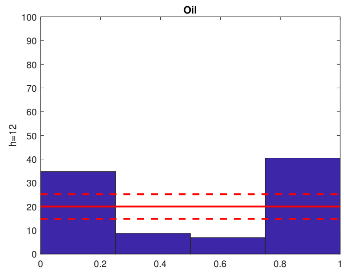
(b) Non-fuel



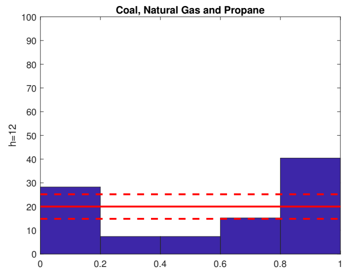
(c) Agriculture & Fertilizers



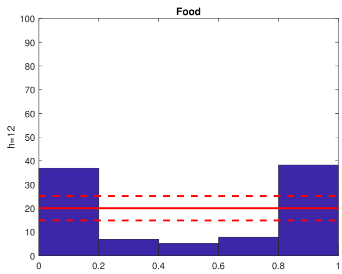
(d) Metals



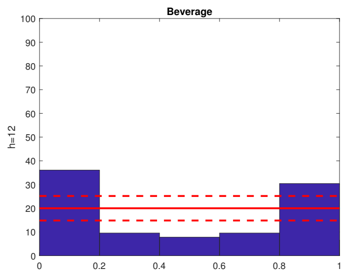
(e) Oil



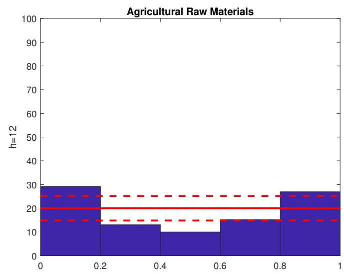
(f) Coal, Gas & Propane



(g) Food

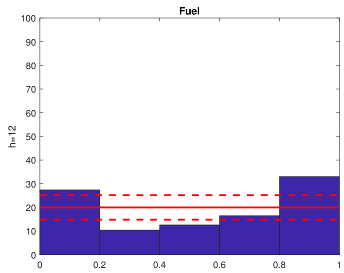


(h) Beverages

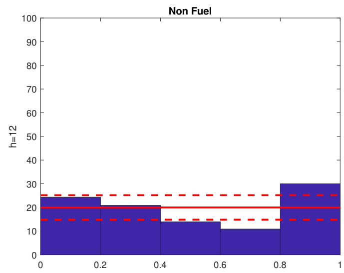


(i) Agricultural Raw Materials

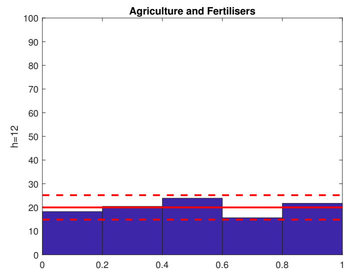
Figure A.8: Probability density functions of the PITs for the *dfmBlocks* model for one-year ahead forecasts of the nine remaining commodity indices. The dashed lines are 95% confidence intervals, constructed using a normal approximation to a binomial distribution, as per Diebold et al. (1998).



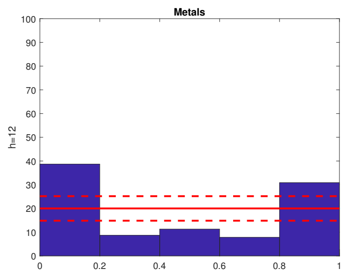
(a) Fuel



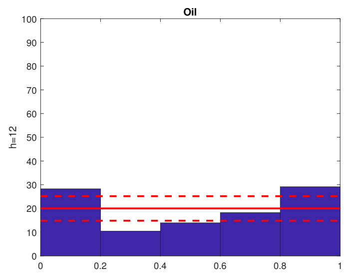
(b) Non-fuel



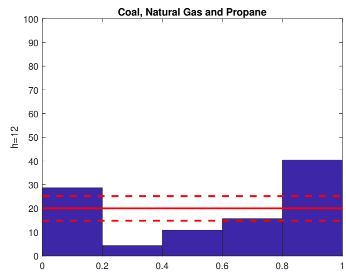
(c) Agriculture & Fertilizers



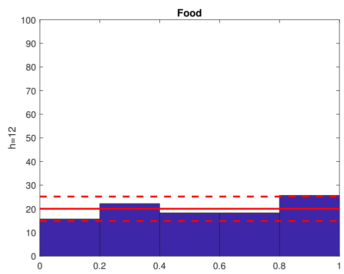
(d) Metals



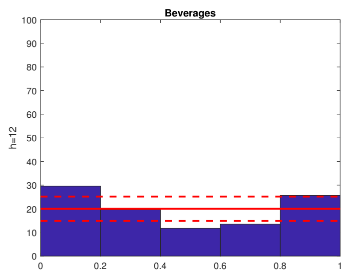
(e) Oil



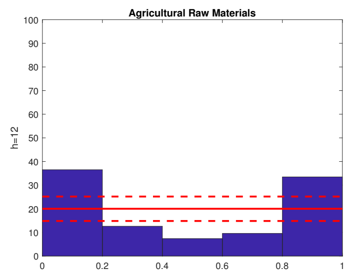
(f) Coal, Gas & Propane



(g) Food

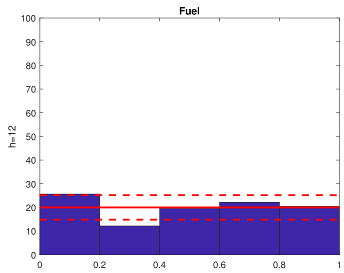


(h) Beverages

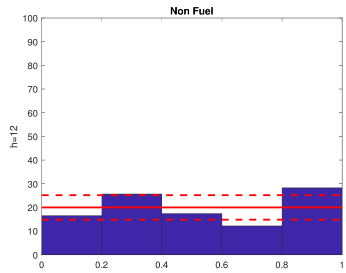


(i) Agricultural Raw Materials

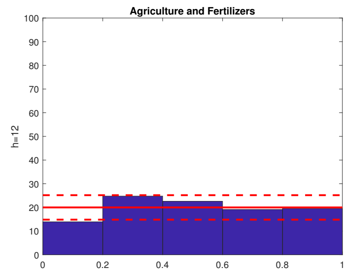
Figure A.9: Probability density functions of the PITs for the *tvp* model for one-year ahead forecasts of the nine remaining commodity indices. The dashed lines are 95% confidence intervals, constructed using a normal approximation to a binomial distribution, as per Diebold et al. (1998)



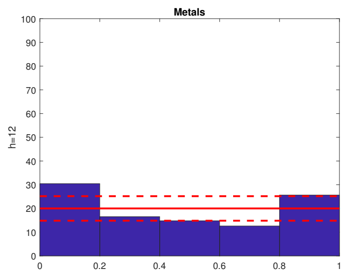
(a) Fuel



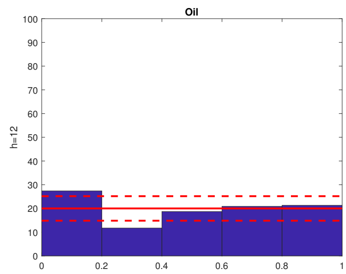
(b) Non-fuel



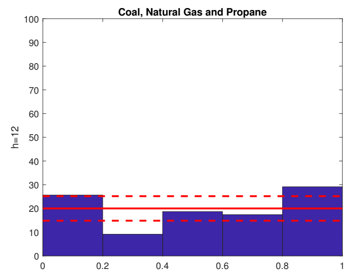
(c) Agriculture & Fertilizers



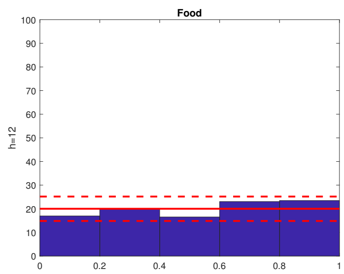
(d) Metals



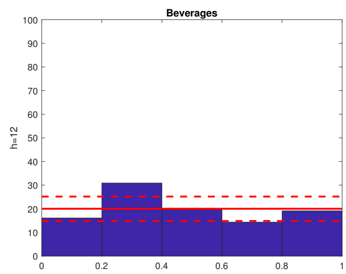
(e) Oil



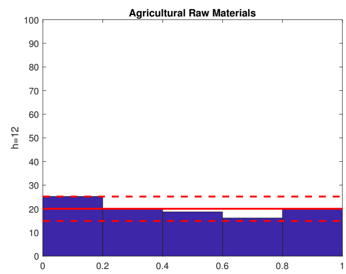
(f) Coal, Gas & Propane



(g) Food

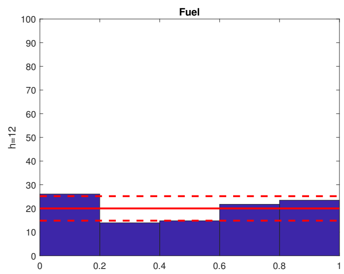


(h) Beverages

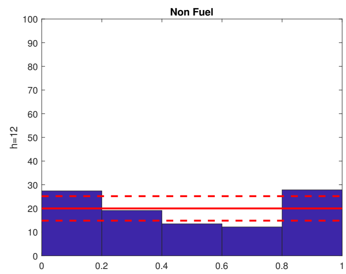


(i) Agricultural Raw Materials

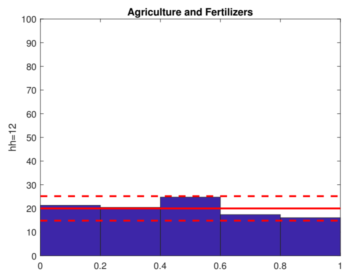
Figure A.10: Probability density functions of the PITs for the *tvpBlocks* model for one-year ahead forecasts of the nine remaining commodity indices. The dashed lines are 95% confidence intervals, constructed using a normal approximation to a binomial distribution, as per Diebold et al. (1998)



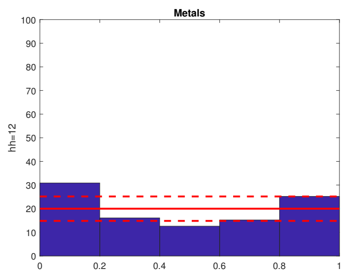
(a) Fuel



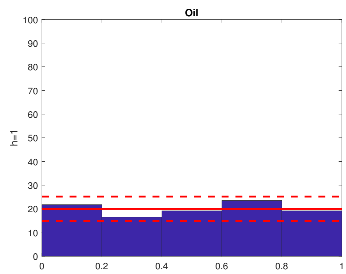
(b) Non-fuel



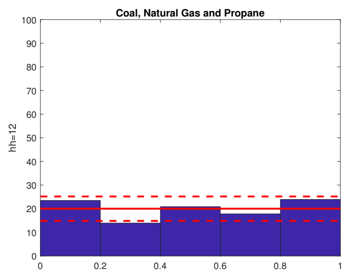
(c) Agriculture & Fertilizers



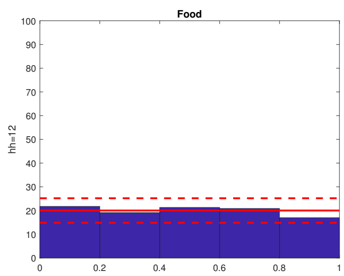
(d) Metals



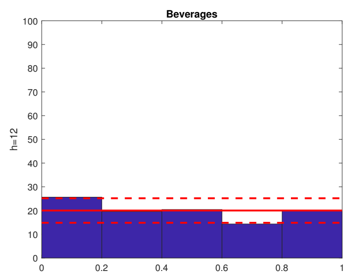
(e) Oil



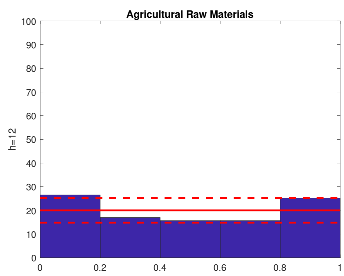
(f) Coal, Gas & Propane



(g) Food



(h) Beverages



(i) Agricultural Raw Materials

References

Abbate, A. and M. Marcellino (2014): "Modelling and forecasting exchange rates with time-varying parameter models," *Journal of Economic Literature*.

Alquist, R., S. Bhattarai, and O. Coibion (2014): "Commodity-price comovement and global economic activity," Tech. rep., National Bureau of Economic Research.

——— (2020): "Commodity-price comovement and global economic activity," *Journal of Monetary Economics*, 112, 41–56.

Amisano, G. and R. Giacomini (2007): "Comparing density forecasts via weighted likelihood ratio tests," *Journal of Business & Economic Statistics*, 25, 177–190.

Bai, J. and S. Ng (2006): "Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions," *Econometrica*, 74, 1133–1150.

Baumeister, C., D. Korobilis, and T. K. Lee (2020): "Energy markets and global economic conditions," *The Review of Economics and Statistics*, 1–45.

Berkowitz, J. (2001): "Testing density forecasts, with applications to risk management," *Journal of Business & Economic Statistics*, 19, 465–474.

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

Blanchard, O. J. and J. Gali (2007): "The Macroeconomic Effects of Oil Shocks: Why are the 2000s so different from the 1970s?" .

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.

Byrne, J. P., B. M. Ibrahim, and R. Sakemoto (2019): "Carry trades and commodity risk factors," *Journal of International Money and Finance*, 96, 121–129.

Campbell, J. Y. and S. B. Thompson (2008): "Predicting excess stock returns out of sample: Can anything beat the historical average?" *The Review of Financial Studies*, 21, 1509–1531.

Casoli, C. and R. J. Lucchetti (2022): "Permanent-Transitory decomposition of cointegrated time series via Dynamic Factor Models, with an application to commodity prices," *The Econometrics Journal*.

Chen, S.-L., J. D. Jackson, H. Kim, and P. Resiandini (2014): "What drives commodity prices?" *American Journal of Agricultural Economics*, 96, 1455–1468.

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

Clark, T. E. (2011): "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility," *Journal of Business & Economic Statistics*, 29, 327–341.

Cogley, T. and T. J. Sargent (2005): "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic dynamics*, 8, 262–302.

Cooley, T. F. (1971): "Estimation in the Presence of Sequential Parameter Variation, unpublished Ph. D," Ph.D. thesis, dissertation, University of Pennsylvania.

Corradi, V. and N. R. Swanson (2006): "Predictive density evaluation," *Handbook of economic forecasting*, 1, 197–284.

Del Negro, M. and C. Otrok (2008): "Dynamic factor models with time-varying parameters: measuring changes in international business cycles," *FRB of New York Staff Report*.

Della Corte, P., L. Sarno, and I. Tsiakas (2009): "An economic evaluation of empirical exchange rate models," *The review of financial studies*, 22, 3491–3530.

Delle Chiaie, S., L. Ferrara, and D. Giannone (2022): "Common factors of commodity prices," *Journal of Applied Econometrics*, 37, 461–476.

Denton, F. T. (1985): "Data mining as an industry," *The Review of Economics and Statistics*, 124–127.

Diebold, F. X., T. A. Gunther, and A. Tay (1997): "Evaluating density forecasts," .

Diebold, F. X., J. Hahn, and A. S. Tay (1999): "Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns on foreign exchange," *Review of Economics and Statistics*, 81, 661–673.

Diebold, F. X. and R. S. Mariano (1995): "Com paring predictive accu racy," *Journal of Business and Economic Statistics*, 13, 253–263.

Doz, C., D. Giannone, and L. Reichlin (2011): "A two-step estimator for large approximate dynamic factor models based on Kalman filtering," *Journal of Econometrics*, 164, 188–205.

D'Agostino, A. and D. Giannone (2012): "Comparing alternative predictors based on large-panel factor models," *Oxford bulletin of economics and statistics*, 74, 306–326.

Gargano, A. and A. Timmermann (2014): "Forecasting commodity price indexes using macroeconomic and financial predictors," *International Journal of Forecasting*, 30, 825–843.

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

Gonçalves, S. and B. Perron (2014): "Bootstrapping factor-augmented regression models," *Journal of Econometrics*, 182, 156–173.

Gonçalves, S., B. Perron, and A. Djogbenou (2017): "Bootstrap prediction intervals for factor models," *Journal of Business & Economic Statistics*, 35, 53–69.

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

Han, Y. (2006): "Asset allocation with a high dimensional latent factor stochastic volatility model," *The Review of Financial Studies*, 19, 237–271.

Harvey, D., S. Leybourne, and P. Newbold (1997): "Testing the equality of prediction mean squared errors," *International Journal of forecasting*, 13, 281–291.

Hong, H. and M. Yogo (2012): "What does futures market interest tell us about the macroeconomy and asset prices?" *Journal of Financial Economics*, 105, 473–490.

Inoue, A. and L. Kilian (2005): "In-sample or out-of-sample tests of predictability: Which one should we use?" *Econometric Reviews*, 23, 371–402.

Jore, A. S., J. Mitchell, and S. P. Vahey (2010): "Combining forecast densities from VARs with uncertain instabilities," *Journal of Applied Econometrics*, 25, 621–634.

Kapetanios, G., M. Marcellino, and F. Venditti (2019): "Large time-varying parameter VARs: A nonparametric approach," *Journal of Applied Econometrics*, 34, 1027–1049.

Kilian, L. (2009): "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *American Economic Review*, 99, 1053–69.

Knüppel, M. (2015): "Evaluating the calibration of multi-step-ahead density forecasts using raw moments," *Journal of Business & Economic Statistics*, 33, 270–281.

Koop, G. and D. Korobilis (2010): *Bayesian multivariate time series methods for empirical macroeconomics*, Now Publishers Inc.

——— (2012): "Forecasting inflation using dynamic model averaging," *International Economic Review*, 53, 867–886.

——— (2013): "Large time-varying parameter VARs," *Journal of Econometrics*, 177, 185–198.

Krugman, P. (2008): "Running out of planet to exploit," *New York Times*, 21, 17–18.

Lo, A. W. (2002): "The statistics of Sharpe ratios," *Financial analysts journal*, 58, 36–52.

Marcellino, M., M. Porqueddu, and F. Venditti (2016): "Short-term GDP forecasting with a mixed-frequency dynamic factor model with stochastic volatility," *Journal of Business & Economic Statistics*, 34, 118–127.

Marquering, W. and M. Verbeek (2004): "The economic value of predicting stock index returns and volatility," *Journal of Financial and Quantitative Analysis*, 39, 407–429.

Mitchell, J. and S. G. Hall (2005): "Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR 'fan'charts of inflation," *Oxford bulletin of economics and statistics*, 67, 995–1033.

Moench, E., S. Ng, and S. Potter (2013): "Dynamic hierarchical factor models," *Review of Economics and Statistics*, 95, 1811–1817.

Mumtaz, H. and P. Surico (2012): "Evolving international inflation dynamics: world and country-specific factors," *Journal of the European Economic Association*, 10, 716–734.

Peersman, G., S. K. Rüth, and W. Van der Veken (2021): "The interplay between oil and food commodity prices: Has it changed over time?" *Journal of International Economics*, 133.

Pesenti, P. A. and J. J. Groen (2011): *1. Commodity Prices, Commodity Currencies, and Global Economic Developments*, University of Chicago Press.

Pindyck, R. S. and J. J. Rotemberg (1990): "Do stock prices move together too much?" *NBER Working Paper*.

Poncela, P., E. Senra, and L. P. Sierra (2014): "Common dynamics of nonenergy commodity prices and their relation to uncertainty," *Applied Economics*, 46, 3724–3735.

Primiceri, G. E. (2005): "Time varying structural vector autoregressions and monetary policy," *The Review of Economic Studies*, 72, 821–852.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010): "Out-of-sample equity premium prediction: Combination forecasts and links to the real economy," *The Review of Financial Studies*, 23, 821–862.

Riggi, M. and F. Venditti (2015): "The time varying effect of oil price shocks on euro-area exports," *Journal of Economic Dynamics and Control*, 59, 75–94.

Vansteenkiste, I. (2009): "How important are common factors in driving non-fuel commodity prices? A dynamic factor analysis," .

Welch, I. and A. Goyal (2008): "A comprehensive look at the empirical performance of equity premium prediction," *The Review of Financial Studies*, 21, 1455–1508.

West, K. D. (1996): "Asymptotic inference about predictive ability," *Econometrica: Journal of the Econometric Society*, 1067–1084.

Yin, L. and L. Han (2015): "Co-movements in commodity prices: Global, sectoral and commodity-specific factors," *Economics Letters*, 126, 96–100.

Acknowledgements

We want to thank for their comments an anonymous referee, Andrea De Polis, participants at the Computational and Financial Econometrics 2019 conference (UK), the International Association for Applied Econometrics Annual Conference 2019 (Cyprus), the Symposium of the Society for Nonlinear Dynamics and Econometrics 2020 (Virtual) and the Vienna Workshop on Economic Forecasting 2020 (Virtual).

Anastasia Allayioti
European Central Bank, Frankfurt am Main, Germany; email: Anastasia.Allayioti@ecb.europa.eu

Fabrizio Venditti
Bank of Italy, Rome, Italy; email: fabrizio.venditti@bancaditalia.it

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

PDF ISBN 978-92-899-6381-7 ISSN 1725-2806 doi:10.2866/810813 QB-AR-24-018-EN-N