

COMMODITY PRICES AND INFLATION RISK*

Anthony Garratt[†]
University of Warwick

Ivan Petrella[‡]
University of Warwick

Friday 15th May, 2020

Abstract

This paper investigates the role of commodity price information when evaluating inflation risk. Using a model averaging approach, we provide strong evidence of in-sample and out-of-sample predictive ability from commodity prices and convenience yields to inflation, establishing clear point and density forecast performance gains when incorporating disaggregated commodities price information. The resulting forecast densities are used to calculate the (ex-ante) risk of inflation breaching defined thresholds that broadly characterise periods of high and low inflation. We find that information in commodity prices significantly enhances our ability to pick out tail inflation events and to characterise the level of risks associated with periods of high volatility in commodity prices.

JEL codes: C32; C53; E37.

Keywords: Inflation risk; Convenience yields; Spot commodity prices; Model averaging; Probability events; Balance of risks and inflation uncertainty.

*We thank Ana Galvao and Simon van Norden and the seminar participants at Warwick Business School, the International Symposium on Forecasting, Santander, June, 2016, the Money, Macro and Finance Conference, King's College, London, September 2017 and the IHS (Institute for Advanced Studies) 1st Vienna Workshop on Econometric Forecasting, February 2018 for useful comments and suggestions.

[†]Warwick Business School, University of Warwick. Email: Anthony.Garratt@wbs.ac.uk

[‡]Warwick Business School, University of Warwick and CEPR. Email: Ivan.Petrella@wbs.ac.uk

“Large swings in food or other commodity prices can have important ramifications for central bankers and their ability to maintain a credible commitment to achieve price stability” Charles Plosser, Former President of the Federal Reserve Bank of Philadelphia¹

1 Introduction

Risks or uncertainty surrounding price stability, both deflationary and inflationary, are a major concern to central banks determining monetary policy, businesses deciding how much to invest, households planning savings and expenditure decisions and more generally are the subject of much public debate. Owing to their high volatility, in particular over the last two decades, commodity prices are often perceived as an important, and commonly cited, source of inflation risk. This is particularly true in central bank settings. For instance, in 2008, worried about the upswing in commodity prices and its potential impact on inflation, the Federal Reserve Chairman Ben Bernanke commented:² “[...] *the latest round of increases in energy prices has added to the upside risks to inflation and inflation expectations.*” In practice, how to model a formal link between commodity prices and inflation remains a much debated, important, but unresolved question. This issue is further complicated by the fact that useful predictions from the perspective of a decision maker, which help assess inflation risk, need to take the form of a density or probability forecast.³ An issue underlined in the same speech quoted above by Ben Bernanke: *“Although theoretical analyses often focus on the case in which policymakers care only about expected economic outcomes and not the uncertainty surrounding those outcomes, in practice policymakers are concerned about the risks to their projections as well as the projections themselves.”*

This paper seeks to measure the risks to US inflation and in particular the contribution of movements in commodity prices to inflation risk. We adopt a model averaging approach to construct the required forecast densities and evaluate the associated predicted risk. Specifically, we incorporate information from disaggregated or individual commodity spot prices and convenience yields using a broad set of linear predictive regressions. Whereas changes in commodity spot prices may be reflected in temporary shifts in inflation, movements in the term structure of commodity price futures – captured by the convenience yields (Fama and French, 1988 and Szymanowska et al., 2014) – are forward looking, possibly reflecting expected inflation, and as such are potentially useful to predict inflation at longer horizons. Hence modeling the link between commodity prices and inflation is complex, requiring a distinction to be made between spot price information and forward looking information (Gospodinov and Ng, 2013, Chinn and Coibion, 2013, Gospodinov, 2016), as well as a break down of the information set into disaggregated commodities (Chen et al., 2014). Moreover, the links between commodity prices and inflation can change depending on whether the movements in commodity prices reflect shifts in demand or supply (see e.g. Kilian, 2009), where the latter can have large idiosyncratic components or be heterogeneous for different commodity groups (Delle Chiaie et al., 2018).⁴

Monetary authorities often seek a measure of inflation that abstracts from volatile and short term movements, in particular core measures of inflation strip out of the aggregate headline inflation series the most volatile components. These components are also those most directly related to the fluctuations of commodity prices (i.e. the food and energy components). Yet, information in commodity prices will still be relevant for predicting core inflation. Given commodities are key inputs in the production process of a large fraction of goods, changes in their prices will be reflected in changes to the marginal cost of production that are naturally passed-through to the aggregate price level (the marginal cost channel). Moreover, commodity price fluctuations can reflect more general price pressure forces, such as aggregate demand (as highlighted for instance by

¹Charles Plosser, President of the Federal Reserve Bank of Philadelphia. “Food or Commodity Price Shocks and Inflation: A Central Banker’s Perspective”. “Food and Water — Basic Challenges to International Stability,” 2009 Global Conference.

²Chairman Ben Bernanke. “Outstanding Issues in the Analysis of Inflation.” Federal Reserve Bank of Boston’s 53rd Annual Economic Conference, June 2008.

³Focusing on point forecasts is justified when the underlying decision problems faced by agents and the government are linear in constraints and quadratic in the loss function; the so-called LQ problem. But, for most decision problems, as argued in Granger and Pesaran (2000a,b), reliance on point forecasts will not be sufficient and probability forecasts will be needed.

⁴Hobijn (2008) highlights the extent of the price change in the overall inflation varies according to the commodity price group or source of the price change.

Kilian (2009) in the context of oil prices). In fact, Knotek and Zaman (2017) show that incorporating both oil and food commodity prices contain relevant information for nowcasting both headline and core inflation.

Using an approximate Bayesian model averaging approach, we find strong evidence of in-sample inflation predictability from information in commodity prices. This result is particularly strong for predictions at medium horizons (6 and 12 months ahead). The in-sample predictability is reflected in substantial out-of-sample gains in point and density forecasts, for both headline and core inflation. Taking the UCSV model popularized by Stock and Watson (2007) as a benchmark,⁵ we provide evidence of large and significant gains in forecasts accuracy, both in point and density forecasts. Moreover, our model produces point forecast that are comparable with those from the Survey of Professional forecasters at a six month forecast horizon, and improve upon them for 1 year ahead forecasts, in particular for core inflation. The forecast densities, across a wide range of model combinations are well calibrated, both in absolute terms and relative to the UCSV benchmark. We document strong evidence of good tail calibration of the densities based on commodity price information and this translates into significant improvements relative to the UCSV model when assessing downward and upward inflation risk.

Our emphasis on probabilistic inflation forecasting motivates an explicit treatment of model uncertainty, both in-sample and out-of-sample, which is required to produce robust density forecasts.⁶ We consider a range of model specifications which differ in their treatment of low frequency moments of inflation, in the spirit of Faust and Wright (2013) and in terms of conditioning variables.⁷ In particular, we look at the importance of the information contained in disaggregated commodity prices by including among the pool of possible regressors either the individual commodity prices or principal components extracted from the entire set of commodity prices (see, e.g., Gospodinov and Ng, 2013). In addition, we document that for many of the commodities standard raw measures of the convenience yields can be severely distorted by the seasonality in futures prices. To address this issues we construct proxies of the convenience yields using information from the entire futures curve (see, e.g. Hevia et al., 2018). The forecast densities from the set of (likely mis-specified) models are combined using the Linear Opinion Pool and the Logarithmic Opinion Pool (see e.g. Jore, Mitchell and Vahey, 2010, Garratt, Mitchell and Vahey, 2014a,b) and different weighting schemes to calculate a single aggregate forecast density. Our modeling choice allows for heterogeneity in the speed and strength of pass-through from different commodities, as well as potentially capturing time variation and instability in their predictive ability. Different commodities may help predict inflation in given periods but not others, and their importance for predicting inflation may depend on other factors becoming more or less relevant at particular points in time. The model average approach allows for the possibility of individual or idiosyncratic effects, not captured by a factor approximation, to be accounted for. In particular, we document that substantial forecast gains will be missed using a limited number of aggregate commodity factors.

We examine inflation risk using the preferred aggregate forecast densities, focusing on inflation tail events, defined as breaches of upper and lower thresholds of interest. When the occurrence of the event is all that matters, irrespective of the size of the deviation from the threshold, inflation risk can be measured as the probability of the event, and therefore we evaluate using the Brier score (Brier 1950) and its decomposition. The resolution component of Brier score decomposition allows us to look more closely into the ability of our density forecast to ‘time’ the events (Murphy, 1973, and Galbraith and van Norden, 2012). We document large gains in terms of Brier scores from our model averaging approach; where the gains are largely attributable to improved and significant resolution of the forecast, i.e. the ability to predict ‘unlikely’ inflation events.

We then provide a detailed analysis, using Kilian and Manganelli’s (2007, 2008) balance of risk measure, assessing inflation risk using a more general loss specification which allows the size of the deviation from the inflation threshold to play a role. Our model produces ex-ante measures of the balance-of-risk, and the associated deflation and excess inflation risks, which are well aligned with the actual inflation realization. In particular, periods of ex-ante deflation (excess inflation) risk anticipate periods where inflation is low (high). Relatedly,

⁵Whereas Stock and Watson (2007) highlights benefit of using an UCSV model for producing accurate point forecasts in inflation, Groen et al. (2013) and Clark and Doh (2014) highlight that stochastic volatility helps with density forecasts.

⁶The usefulness of Bayesian Model Averaging (BMA) for forecasting inflation has been highlighted by Wright (2008) in the context of point forecasting and Rossi and Sekhposyan (2015) for density forecasting.

⁷Lagged dependent variables, changes in spot prices, convenience yields and other exogenous variables

using the information contained in commodity prices minimizes the misperception in risk, i.e. it produces low ex-ante deflation (excess inflation) risk ahead of periods of high (low) inflation. In order to formally assess the potential gains of our model with respect to the UCSV benchmark in gauging inflation risk we propose score functions for the balance of risks and its deflation risk and expected inflation risk components. As is the case for the Kuiper’s Score statistic, our score functions reward ex-ante measures of risks which are aligned with the realization of inflation and penalize ”false signals” (i.e. predictions of low/high inflation risk which turn out to be associated with high/low inflation realization). We find that a decision maker equipped with the model average forecast would be more aware of the risk of high or low inflation when those events materialize compared with the UCSV benchmark. In fact, the baseline model combination produces less volatile and more precise assessment of risk relative to the UCSV benchmark.

Last, we compute an overall measure of inflation uncertainty (unrelated to any specific event) following Rossi and Sekhposyan (2015). Inflation uncertainty is typically overestimated by the UCSV benchmark model, as it fails to reflect the uncertainty associated to inflation outcomes in periods of volatile commodity prices. Overall, we find that information in commodity price significantly enhances our ability to pick out tail inflation events and to characterize the level of risks associated with periods of large swings in commodity prices.

The remainder of this paper is organized as follows. In Section 2 we outline our empirical framework and motivation for our approach, describe the construction of our convenience yield data and define the model set considered in the in and out-of-sample applications. Section 3 reports the in-sample and Section 4 the out-of-sample results. Section 5 reports the calculation and evaluation of inflation risk using probability events, balance of risk measures and an overall inflation uncertainty index. Section 6 concludes.

2 The Empirical Framework

Information in spot commodity prices (see, e.g., Chen et al., 2014) and in the term structure of commodity futures prices, as captured by the convenience yields (Gospodinov and Ng, 2013, GN, hereafter), can potentially be useful when trying to predict inflation. Therefore, in our application, both in and out of sample, we follow GN and Stock and Watson (2007) and adopt the following linear predictive regression framework:

$$\Delta^h p_{t+h} = c_j + \sum_{i=0}^p \phi_{j,i} \Delta p_{t-i} + \sum_{i=0}^{p_q} \varphi_{j,i} q_{j,t-i} + \sum_{i=0}^{p_{cy}} \rho_{j,i} cy_{j,t-i} + \sum_{i=0}^{p_x} \beta_{j,i} \mathbf{x}_{t-i} + e_t \quad (1)$$

where the dependent variable is annualized average inflation over the next h periods, i.e. $\Delta^h p_{t+h} = (1200/h)[\ln(P_{t+h}) - \ln(P_t)]$. The term q_j defines the annual growth rate in the j th real spot commodity price,⁸ cy_j denotes the associated convenience yield, and \mathbf{x} is a vector of non-commodity based conditioning variables. The subscript j denotes the individual commodity or information set considered, where in the application we consider a specification for each of the $j = 1, 2, \dots, n_c$ commodities, where $n_c = 23$, plus a specification where common factors from the panel of all commodities are included as predictors.⁹ The model is estimated using monthly data, for the period 1991m8-2016m3, determined by the availability of the futures contracts. We focus on results using personal consumption expenditure (PCE) to measure inflation, as this is the most closely watched by the Federal Reserve when assessing their monetary policy stance. Both the all items and all items excluding food and energy indices’s are used, defining ”headline” and ”core” inflation respectively. In Appendix E we report a full set of results using consumer price (CPI) inflation, all of which are all qualitatively similar to the PCE results.

The next three sub-sections describe the construction of the commodity information data set, present a preliminary principal component analysis of the data and, last, discuss in greater detail the model set choice.

⁸Note that whether we use real or nominal commodity prices makes little difference to the results as real commodity price movements are dominated by the movements in the nominal prices.

⁹The 23 commodity groups are defined as: Crude Oil, Gas Oil, Heating Oil, Unleaded Gas, Natural Gas, Cocoa, Orange Juice, Coffee, Sugar, Soybean Oil, Corn, Oats, Soybean, Soybean Meal, Wheat, Cotton, Copper, Lumber, Gold, Silver, Feeder Cattle, Live Cattle and Lean Hogs.

2.1 Construction of the convenience yields data

The convenience yields summarizes the information conveyed by the shape of the commodities futures curve. For instance, periods when the convenience yield is positive imply that the discounted futures price lies below the current spot price. Therefore, as highlighted by GN, to the extent that high commodity prices are associated with rising prices, a positive convenience yield is a indicator of high inflation. Following Fama and French (1988) the convenience yield, for commodity j , can be constructed from the slope of the futures curve net of the interest rate:

$$cy_{jt,n} = (1 + i_{t,n}) - (F_{jt,n}/S_{jt}) \quad (2)$$

(where S_{jt} corresponds to the spot price and $F_{jt,n}$ to the futures contract price with maturity to expiration of n months). In practice, when constructing these proxies one needs to deal with two important features of commodity markets: (a) the uneven availability of futures contracts,¹⁰ and (b) the well documented and often pronounced seasonality present in most futures prices (see, e.g., Hevia et al., 2018).

In order to deal with the uneven availability of futures contracts, GN focus on the short end of the futures curve and construct a measure of the convenience yields using the two available contracts with the closest expiration. However, using contracts referring to prices in different months of the year induces pronounced seasonal variation in the convenience yields for many of the commodities under consideration. Inevitably the noise induced by the seasonality in convenience yields substantially weakens their usefulness as predictors of inflationary pressures.¹¹ To address both of the issues highlighted above, we construct convenience yields from interpolated futures curves. Specifically, following Hevia et al. (2018), we model the futures curve imposing Nelson and Siegel (1987)'s parametric restrictions to the loadings and explicitly accounting for stochastic seasonality in the futures prices. Specifically,

$$\log(F_{jt,n}) = \beta_{0,jt} + \frac{n}{12}\beta_{1,jt} + \frac{1 - e^{-\lambda_j n/12}}{\lambda_j}\beta_{2,jt} + \cos\left(\frac{2\pi(t+n)}{12}\right)\gamma_{1,jt} + \sin\left(\frac{2\pi(t+n)}{12}\right)\gamma_{2,jt} + \varepsilon_{jt,n} \quad (3)$$

so that $\beta_{0,jt}$ corresponds to the implied (de-seasonalised) spot price derived from the entire futures curve. The dynamics of the log basis (i.e. the spread between any two futures prices) is captured by the two factors, $\beta_{1,jt}$ and $\beta_{2,jt}$, corresponding respectively to the level and slope of the cost-of-carry (see also Trolle and Schwartz, 2009). We allow the seasonality of the data to evolve stochastically overtime, and this variation is captured by the evolution in $\gamma_{1,jt}$ and $\gamma_{2,jt}$.¹² The common factors are modeled as independent autoregressive processes, therefore the model can be cast in state space and estimated using standard methods (see, e.g., Harvey, 1989). This model is used to interpolate across the missing maturities and retrieve the de-seasonalised futures curve from which we can construct convenience yields as in equation (2). Specifically, we build proxies of short and medium run convenience yield using futures prices 1 and 12 months from the current month, \hat{cy}_{jt}^{1m} and \hat{cy}_{jt}^{12m} respectively. Appendix A provides additional details on the model and its performance in fitting the futures curve for different commodities.

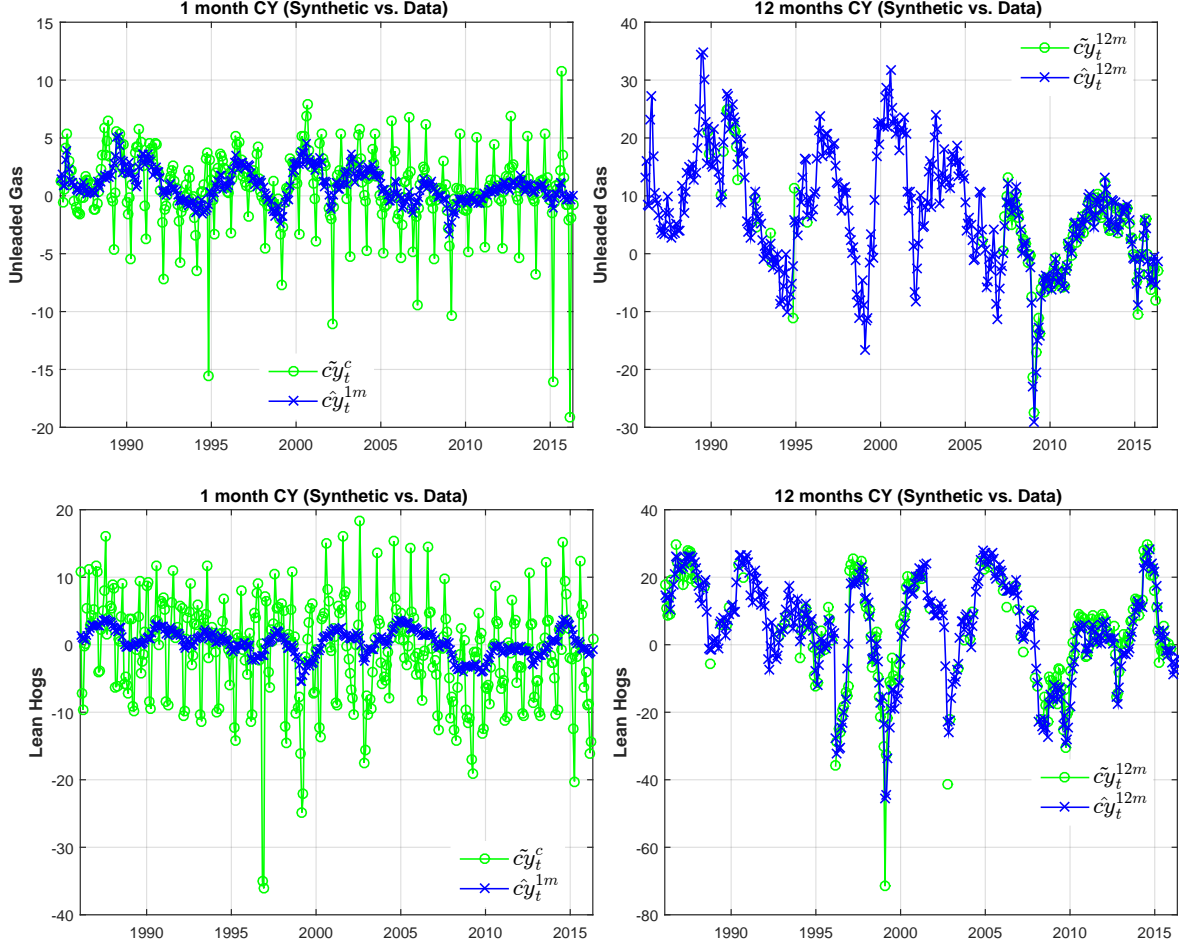
Figure 1 provides us with a visual assessment of the benefits associated with using synthetic futures prices to build cy . As an illustration we focus on two commodities, Unleaded Gas and Lean Hogs. At the short end of the futures curve, the seasonality is quite pronounced. As a consequence, we observe large differences between the proxies constructed from the raw data and those from the synthetic futures, with the dynamic of the former entirely dominated by the seasonality. In contrast, annual convenience yields are not affected by the seasonality as they compare the futures with same expiration month. In this case the model is able to match

¹⁰In fact, futures contracts with expiration in every month of the year are mainly only available for energy commodities and/or only in the late part of the sample under analysis.

¹¹Forecasts based on such proxies would display a clear seasonal behavior in the predicted variable so as to result in a poor forecast for (seasonally adjusted) inflation rates.

¹²The model exploits the fact that the underlying seasonality in the data is common for futures at different maturities. The specific form of the seasonal component was suggested by Hannan, Terrell and Tuckwell (1970). To achieve parsimony, we only allow for a seasonal cycle in the fundamental frequency, i.e. implicitly emphasizing the annual pattern in the seasonality. Moreover, for crude oil and industrial and metal commodities (Cotton, Copper, Lumber, Gold, Silver) for which prices are less evidently seasonal we estimate the same model without the seasonal component.

Figure 1: CONVENIENCE YIELDS



Note: 1 month and 12 month convenience yields for a selection of commodities. Convenience yields are calculated directly from existing futures price contracts (green, -o-), using eq. (2) on the two contracts closest to expiration ($\tilde{c}y_t^c$), and from synthetic futures prices (blue, -x-) with 1 and 12 month maturity ($\hat{c}y_t^{1m}$ and $\hat{c}y_t^{12m}$, respectively).

quite closely the behavior of the annual convenience yields constructed from the raw futures contracts when the data required to construct this proxy are available. Overall, Figure 1 illustrates that our method to construct the convenience yields does not mask the signal in the raw data when there are no issues of unbalancedness and can deal effectively with the seasonality in the underlying futures prices.

2.2 The information in commodity factors

When taking the model to the data, we need to choose how to incorporate the information arising from the large number of disaggregated commodities. A common approach in the literature when faced with many potential predictor variables is to summarize their predictive content using principal components (Stock and Watson, 2002). GN follow this strategy in analyzing the relationship between commodity prices and inflation. Extracting the common factors across different commodities, has the benefit of potentially mitigating the impact of the noise that arises from the construction of the convenience yield proxy. However, this comes at the cost of discarding, ex-ante, information which is idiosyncratic to a single or a subgroup of commodities and can potentially be relevant for predicting future inflation. This will, most likely, be an important limitation for our specific application. Whilst the common factors are generally associated with common shifts in global demand for commodities (Juvenal et al., 2015, and Alquist et al., 2019), supply side factors may well be more idiosyncratic in the sense that they reflect conditions specific to single or groups of commodities. For example, movements in food prices relating to weather conditions or oil price movements relating to increased supply of

oil with the growth in North American shale oil production. In this case it would be appropriate to analyze the disaggregated commodity prices as they potentially incorporate heterogeneous information on inflationary pressure.

In order to understand the extent to which the majority of the information contained in the convenience yields and spot commodity prices can be accounted for by aggregate factors we perform a principal component analysis (PCA) of the data. For the short term the 1-month and the 12-month convenience yield just 37% and 39% of the variation in the data, respectively, is explained by the first two factors.¹³ Increasing the number of factors to three or four, still leaves around two fifths of the overall variation in the data unexplained. A similar story also holds for factors approximating movements in the change in real spot commodity prices (and is in line with the results reported by Delle Chiaie et al., 2018). Two factors explain 48% of the variation and four factors 66.6%.

Overall, the inference we take from the PCA of the data is that commodity spot prices and convenience yields have a weak factor structure and as such using only aggregate factors can potentially miss important information for predicting inflation. For this reason, in our model combination application, we estimate Equation (1) for each of the commodities separately (both their spot price and convenience yield), allowing individual commodities to affect h -step head inflation. In addition we also estimate a model specification which uses common factors of the commodity predictors.

2.3 Model Space for Disaggregated Commodity Prices

We consider a large number of different model specifications based on equation (1). Specifically, different specifications are defined according to which of the two measures of the convenience yield (1 month and 12 months) is used, the maximum number of lag lengths considered and the chosen combination of the conditioning variables. The non-commodity based conditioning variables in \mathbf{x} are: the log of the effective US dollar exchange rate, the level of short term interest rates and the unemployment gap.¹⁴ We use these three variables in five different combinations; each variable separately, all together and no conditioning variables (i.e. \mathbf{x} as an empty set). Finally, we estimate the set of specifications which arise from the variants described above but where we use an inflation-gap dependent variable which takes account of slow variations in trend inflation, following Faust and Wright (2013). We specify the prediction model in terms of an 'inflation gap'.¹⁵ The use of the inflation-gap specification doubles the number of models.

Allowing for the number of commodities, commodity factors, convenience yield measures, lag lengths, non-commodity conditioning variables and the inflation gap dependent variable specifications, the total number of models we consider is $N = 7200$.¹⁶

The ex-post aggregation approach explores a subset of all possible linear regression models available from the given set of potential predictor variables. In this sense it bears some similarities with the complete subset regression based approach of Elliot, Gargano and Timmermann (2013). However, instead of keeping the number of predictors used fixed and randomly trying alternative permutations of the prediction regressors, we define subset of models for each commodity one at a time or all commodities together (using common factors). This drastically reduces the number of subset regressions (containing some but not all commodity types) and allows us instead to consider more explicitly model specification uncertainty.

¹³A detailed analysis is presented in Appendix A. In there we also show that the first PC can be thought as a 'level' factor since the loadings on all the commodities are positive, whereas the second PC can be thought as a 'slope' factor combining with opposite sign the information of the CY in the live meat commodities and the soybean product commodities.

¹⁴The detrending and definition of the unemployment gap use a one-sided Hodrick-Prescott filter (see Stock and Watson (1999)). Using alternative detrending methods, such as Hamilton (2018), does not affect our results.

¹⁵Specifically, the inflation gap is defined as the h -step ahead inflation minus an exponentially smoothed inflation trend: $\tilde{\pi}_t = \kappa \tilde{\pi}_{t-1} + (1 - \kappa) \pi_t$.

¹⁶For each of the inflation and inflation-gap dependent variable specifications, we have the commodities and the factor representation, hence $2 \times (n_c + 1)$ models, where $n_c = 23$. In each case we have two measures of convenience yields, which use the commodity price term structure to represent information on expectations at different horizons. If we set maximum lag order for the dependent variable to $p = 4$, the maximum lag order on the convenience yield, change in the spot commodity price and conditioning variables such that $p_{cy} = p_q = p_x = 2$, and have five combinations on the choice of conditioning variables, then the total number of models is $N = 2 \times (n_c + 1) \times 2 \times (p + 1) \times (p_{cy} + 1) \times 5 = 7200$.

3 Is the information in commodity prices useful for predicting inflation?

Given the difficulty of forecasting inflation, a necessary condition for the success of any out-of-sample forecasting exercise is establishing a degree of in-sample predictability. In this section, we examine whether there is any explanatory power for inflation from commodity spot prices and convenience yields over and above the autoregressive components and other relevant non-commodity conditioning variables.

To take account of the model uncertainty associated with the range of specifications in equation (1), we adopt approximate Bayesian model averaging methods, as described in Garratt *et al.* (2009). Our aim is to compute the probability of hypothesis that commodity spot prices and convenience yields have predictive content for inflation. We treat these probabilities, denoted by $\Pr(H_0|Data)$, as unknown features of interest common across all models:¹⁷

$$\Pr(H_0|Data) = \sum_{k=1}^N \Pr(H_0|Data, M_k) \Pr(M_k|Data). \quad (4)$$

The probabilities can be computed by using approximate Bayesian methods. Thus, overall inference about the hypothesis of interest involves taking a weighted average of the probabilities across all N competing models, M_1, \dots, M_N , with weights given by the posterior model probabilities, which using Bayes' rule can be calculated from the marginal likelihood and a prior. In the results reported in this section we follow Garratt *et al.* (2009) and approximate the posterior model probability using the BIC. Appendix B describes these calculations in detail.¹⁸ In order to take into account of possible structural breaks in the sample under consideration, we re-estimate the model recursively using a short window of 100 observations and compute in-sample probabilities for each of the estimation windows.

Figure 2 summarizes the in-sample evidence regarding whether commodity price information, in the form of spot and the forward looking convenience yield, *have predictive power*, for headline and core inflation, at horizons $h = 1, 6$ and 12 months respectively. The results in Figure 2 provide strong evidence of there being predictive power from convenience yields and changes in real spot commodity prices to both headline and core inflation, over and above its lags and other conditioning variables. The probabilities are almost always one for the two longer prediction horizons, $h = 6$ and $h = 12$. Whereas, at one-month forecast horizon, we find weaker evidence of predictability, with the exception of sporadic episodes in the sample. For instance, during the 2010-12 period we observe higher probabilities of predictability for both headline and core inflation.¹⁹

The results reported in this section underline the importance of including the information from commodity price fluctuations for appropriately assessing inflation risk.

3.1 On the importance of information from disaggregated commodities: in-sample evidence

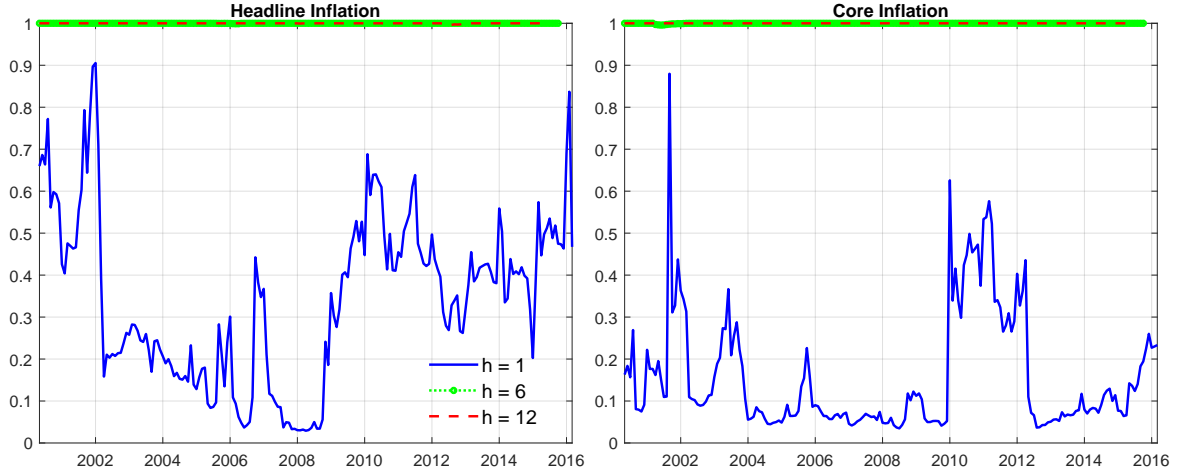
Posterior model probabilities computed for each model enable us to assess the role played by disaggregated commodity prices as compared to the information captured by the aggregate factors. Therefore, in Figure 3, we plot the posterior model probabilities of the sub-set of models which use factor only information. If all models were to carry the same weight one would expect to see the factor models take $1/24$ of the overall weights, at all points in time. Similarly, if the factor models were able to dominate the combinations of individual commodities then we would see a posterior model probability of one. Except in specific episodes, for example 2004-2007, for

¹⁷To be precise for each specific horizon h , we compute $\Pr(H_0|Data) = \Pr(\rho_1 = \dots = \rho_{p_{cy}} = \varphi_1 = \dots = \varphi_{p_q} = 0|Data)$.

¹⁸We stress that, although we adopt a Bayesian approach, it is an approximate one which uses data-based quantities which are familiar to the frequentist econometrician. Moreover, the use of posterior model probabilities (i.e. $p(M_i|Data)$) to select models or do Bayesian model averaging holds true regardless of whether the models are nested or not. Similarly, with our recursive forecasting exercise, we are simply deriving the predictive density at each point in time and then calculating various functions of the resulting densities.

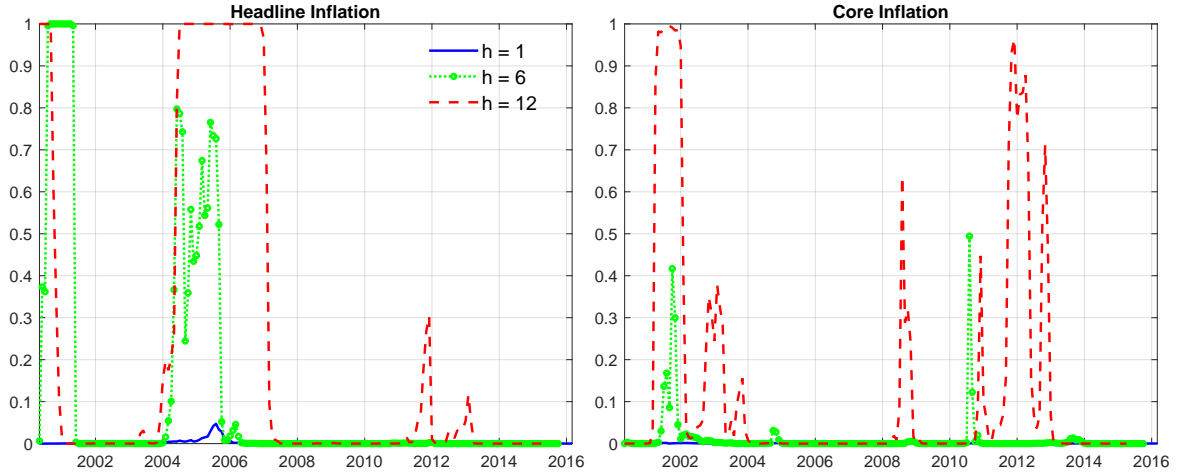
¹⁹It is worth mentioning that these results do not necessarily reflect the predictability from oil prices. In fact, Figure F.7 highlights the changing importance of different commodity groups for inflation predictability over time.

Figure 2: PROBABILITY OF COMMODITY PRICES PREDICTING INFLATION



Notes: The figures reports $1 - \Pr(H_0|Data)$ for different predictive horizons ($h = 1, 6, 12$), calculated following eq. 4. This can be broadly interpreted as the probability that commodity prices (either spot prices or convenience yields) incorporate information useful for the task of predicting inflation.

Figure 3: POSTERIOR PROBABILITY OF MODELS WITH FACTORS



Note: Posterior model probability associated with specifications of model (1) using factors of the commodity information (as opposed to the individual commodity variables). Different model specification are weighted as detailed in (4).

headline inflation at $h = 12$ or more recently in 2010-2014 for core inflation, the posterior model probability of factor based models as opposed to models using disaggregated commodity price information is small. Therefore individual commodity prices contain information that helps to predict inflation beyond the impact of aggregate factors.²⁰

4 Forecasting Inflation Using Commodity Price Information

In this section we assess the ability of our model averaging approach to produce point and density forecasts out-of-sample. Adopting the same model set detailed in Section 2, here we undertake a recursive h -step ahead

²⁰This result is consistent with the weak factor structure of commodity spot prices and convenience yields documented in Section 2.2.

point and density forecasting exercise using a rolling estimation window at each recursion of size 100.²¹ Using a rolling window for model fitting provides a pragmatic approach to handling structural change and to capture shifts in volatility (see, for example, Pesaran and Timmermann, 2004, and Clark, 2011). This involves adopting the direct method of forecasting, where we estimate a different set of models for each forecast horizon h and condition on information known at time t . Assuming natural conjugate priors, it is straightforward to produce forecast densities for inflation using Student- t distributed analytical solutions (see, e.g., Zellner, 1971 and Koop, 2006). Appendix C provides further technical details about how the forecasting is conducted.

Common usage in the macro economic literature (see, among others, Geweke and Amisano, 2011, Kascha and Ravazzolo, 2011, Ranjan and Gneiting, 2010), and ease of calculation motivate us to opt to use the Linear Opinion Pool (LOP) and the Logarithmic Opinion Pool (LogOP) to construct the predictive density, $\Pr(\Delta^h p_{t+h} | Data_t)$. In particular, each opinion pool uses five types of weights: equal weights (EQ), log score (LS) based weights, continuous rank probability score (CRPS) weights, Mean Squared Error (MSE) weights and Bayesian Information Criterion (BIC) weights, as in the in-sample analysis. The MSE weights are proportional to the inverse of the individual models MSE, and as such uses only point forecast information, whereas LS and CRPS weights take into account the ability of individual model to provide good density forecast.²² Moreover, we allow the weights to the different model specifications to vary over time. In particular, we use weights which reflect the (exponentially discounted) past performance of each of the models. Details of the opinion pools and combination weights are provided in Appendix D.

4.1 Point and Density Forecasts

Table 1 reports a range of h -step ahead point and density forecast diagnostics for headline and core inflation respectively.²³ In each case, results are reported for forecast horizons $h = 1, 6$ and 12 . In the first row, for each horizon and each inflation measure, we report the *absolute* values of the various forecast evaluation metrics for the UCSV benchmark model, in the remaining rows, for the model combinations, we report *relative* forecast metrics.

In the first column of Table 1, for headline and core inflation, we report the average values of the root mean squared forecast errors (RMSE) ratios, where a value of less than one favours the combination model over the UCSV benchmark. We find strong evidence that point forecasts are improved by using model combination, for both headline and core inflation. For all forecast horizons we observe RMSE ratios relative to the benchmark UCSV model that are less than one. The gains are large and are often significant in particular at the longer horizons.²⁴

To assess inflation risk requires that we go beyond the point forecasts and place the emphasis on probabilistic density forecasting. In column two Table 1, for headline and core inflation, we report the difference in the average log score of the model combination relative to the UCSV benchmark, and in column three in each case the ratio of the model combinations average CRPS to the average CRPS of the UCSV benchmark. Positive values for

²¹We investigated the effects of using window lengths of 60 and 80 periods, and found the quantitative and qualitative nature of our findings to be robust. An alternative approach would be to estimate our set of models, using a range of different window lengths, and then combine across the model set and window length, as described, for example, in Pesaran and Timmermann (2007). In addition we experimented with an expanding window size, and again found the quantitative and qualitative nature of our findings to be robust, with very similar RMSE/CRPS ratios and log score differences.

²²The intuitive appeal of the logarithmic scoring rule stems from the high score assigned to a density forecast with high probability at the realized value (see e.g. Amisano and Giacomini, 2007). We also include the CRPS which is less sensitive to extreme events (Gneiting and Raftery, 2007, and Panagiotelis and Smith, 2008) and, as such, might be particularly relevant when evaluating the one month ahead inflation forecast.

²³The out-of-sample evaluation period starts in 2000:M7 for $h = 1$, in 2001:M10 for $h = 6$ and 2003:M4 for $h = 12$. All horizons go through to 2016:M3.

²⁴Making exact comparisons of our gains from using commodity price information with the earlier literature requires caution, mainly because of sample differences and the use of different benchmarks. Nevertheless, it is worth noting that the reported gains in our point forecasts are substantially larger than those reported in previous literature. For instance, Chen, Turnovsky and Zivot (2014), who use a different model specification and quarterly data, report gains that are substantially lower than ours. Similarly our forecasting gains are substantially larger than the ones reported by GN, where our increased forecast gains can in part be attributed to the use of commodity specific information as opposed to factors, as we show in Section 4.3. Knotek and Zaman (2017), on the other hand, show that nowcasting models incorporating information on gasoline and food commodity prices, produce larger improvements in the forecast accuracy for US headline and core inflation relative to standard benchmarks, than those reported in Table 1 for $h = 1$.

Table 1: POINT AND DENSITY FORECASTS

	Headline Inflation					Core Inflation				
	h=1					h=1				
	RMSE	Log Score	CRPS	RS Test	K Test	RMSE	Log Score	CRPS	RS Test	K Test
UCSV	2.646	-3.431	1.418	1.032	0.071	1.466	-3.474	0.674	0.790	0.303
LOP Combinations										
Equal weights	0.915	1.120	0.915	1.534	0.000	0.860	1.826	1.006	2.116	0.000
Log score weights	0.901	1.137	0.902	1.561	0.000	0.852	1.859	0.985	2.043	0.000
CRPS weights	0.907	1.131	0.906	1.561	0.000	0.853	1.850	0.989	2.098	0.000
MSE weights	0.898	1.137	0.898	1.647	0.000	0.843	1.841	0.987	2.171	0.000
BIC weights	0.877	1.127	0.866	1.105	0.011	0.882	1.890	0.947	1.149	0.052
LOG Combinations										
Equal weights	0.919	1.081	0.907	1.520	0.000	0.861	1.807	0.950	1.759	0.000
Log score weights	0.904	1.117	0.893	1.534	0.000	0.851	1.893	0.935	1.721	0.000
CRPS weights	0.911	1.094	0.897	1.534	0.000	0.854	1.828	0.937	1.735	0.000
MSE weights	0.902	1.102	0.889	1.571	0.000	0.846	1.822	0.932	1.842	0.000
BIC weights	0.876	1.105	0.877	1.064	0.013	0.882	1.785	0.899	1.086	0.180
h=6										
	RMSE	Log Score	CRPS	RS Test	K Test	RMSE	Log Score	CRPS	RS Test	K Test
UCSV	2.118	-3.417	1.061	0.776	0.866	0.638	-3.216	0.361	0.670	0.647
LOP Combinations										
Equal weights	0.807	1.550	0.848	1.296	0.606	0.882	2.166	1.107	1.880	0.000
Log score weights	0.713	1.622	0.775	1.606	0.270	0.823	2.308	1.003	1.893	0.000
CRPS weights	0.772	1.602	0.808	1.422	0.472	0.847	2.284	1.012	1.679	0.000
MSE weights	0.749	1.611	0.788	1.504	0.323	0.802	2.251	1.036	1.975	0.000
BIC weights	1.204	0.255	1.281	1.983	0.188	1.138	1.786	1.126	1.647	0.250
LOG Combinations										
Equal weights	0.850	0.917	0.863	0.905	0.867	0.917	2.296	0.903	0.616	0.994
Log score weights	0.716	1.699	0.731	0.910	0.784	0.846	2.483	0.811	0.700	0.990
CRPS weights	0.822	1.063	0.822	0.873	0.807	0.887	2.368	0.861	0.634	0.998
MSE weights	0.805	1.088	0.803	1.070	0.677	0.865	2.397	0.841	0.576	0.993
BIC weights	1.202	-1.037	0.803	2.072	0.076	1.137	1.480	0.859	1.841	0.122
h=12										
	RMSE	Log Score	CRPS	RS Test	K Test	RMSE	Log Score	CRPS	RS Test	K Test
UCSV	1.448	-2.887	0.805	1.490	0.179	0.486	-2.307	0.298	1.114	0.799
LOP Combinations										
Equal weights	0.812	1.233	0.854	1.443	0.057	0.890	1.432	1.199	2.061	0.000
Log score weights	0.683	1.363	0.767	1.552	0.000	0.783	1.639	1.061	2.013	0.000
CRPS weights	0.748	1.341	0.788	1.331	0.000	0.827	1.629	1.030	1.911	0.000
MSE weights	0.698	1.351	0.763	1.598	0.004	0.719	1.609	1.069	2.360	0.000
BIC weights	1.136	-1.119	1.303	3.437	0.002	1.336	0.498	1.276	2.412	0.691
LOG Combinations										
Equal weights	0.894	0.709	0.901	1.698	0.235	0.965	1.477	0.887	1.434	0.988
Log score weights	0.689	1.544	0.689	1.103	0.780	0.814	1.860	0.736	1.323	0.960
CRPS weights	0.850	0.995	0.836	1.508	0.438	0.916	1.596	0.827	1.485	0.981
MSE weights	0.821	1.036	0.801	1.089	0.860	0.873	1.758	0.779	1.193	0.994
BIC weights	1.134	-2.047	0.823	3.512	0.000	1.311	-0.198	0.806	2.627	0.591

Note: (i) RMSE - the ratio of the root mean square error (RMSE) of the various model combinations to that of the UCSV model (ii) Log Score - difference in average log score relative to UCSV model over the evaluation period (iii) CRPS - ratio of average continuous rank probability score to UCSV benchmark over the evaluation period. Bold characters denote that the difference with respect to the benchmark model is significant at 10% critical level (using Giacomini and White (2004)'s test for the RMSE and the Kullback-Leibler information criterion (KLIC)-based test, which utilizes the expected difference in the log scores and CRPS of candidate densities; see Bao et al. (2007), Mitchell and Hall (2005) and Amisano and Giacomini (2007)). The last columns corresponding to 'RS Test' and 'K Test' denote the test statistics of Rossi and Sekhposyan (2019) (in bold we report rejection at 10 % level) and p-value of Knappel's (2015) test on the calibration of the density respectively. The entries highlighted in grey correspond to best performing model for each metric/horizon considered.

the difference in average log scores and CRPS ratios of less than one favour the combination over the UCSV benchmark. The forecast density performance of our commodity based model combinations largely mirror those of the point forecasts; we observe a strong and improving performance relative to the UCSV benchmark as the forecast horizon increases. In particular, at forecast horizons $h = 6$ and $h = 12$, for LogOP combinations, we observe large and significant positive average log score differences and CRPS ratio significantly lower than one.²⁵ The results are more mixed at forecast horizon $h = 1$. Column four reports the Rossi and Sekhposyan (2019) test statistic, and column five the p-value of the test in Knoppel (2010), both of which test the optimal calibration properties of the probability integral transforms (PITs) for the densities forecast (see Diebold et al., 1998). Both suggest poor calibration of the forecast densities at forecast horizon $h = 1$ but good calibration for a majority of forecast densities at horizons $h = 6$ and $h = 12$, in particular for LogOP combinations.

Overall, the results in Table 1 suggest that the combination model which produces the best calibrated densities, associated large gains in forecasting performance over the benchmark and more generally over the alternative combinations in particular for $h = 6$ and $h = 12$, is the LogOP log score weighted combination.²⁶ As such in what follows we focus our analysis based on this model, which we refer to as the *baseline* specification.

4.2 A Comparison with the Survey of Professional Forecasters (SPF)

When evaluating forecasts of US inflation a natural comparison is with the set of professional forecasts collated in the Survey of Professional Forecasters (SPF) conducted by the Philadelphia Federal Reserve. Faust and Wright (2013) discuss in detail a number of methods for forecasting inflation and among their conclusions is that judgmental forecasts, which include private sector surveys, are remarkably hard to beat and as such this comparison represents a tough benchmark.

The SPF report quarterly forecasts referring to annualized quarter on quarter inflation. Therefore in order to produce comparable forecasts we re-estimate our models as specified in equation (1), using data at a *quarterly* frequency. The SPF participants receive the survey at the beginning of the second month in any quarter, therefore in order to align the information set of the model to that used by the SPF participants we include information on the conditioning variables corresponding to the last month of the previous quarter.²⁷

Table 2: COMPARISON WITH SURVEY OF PROFESSIONAL FORECASTERS: RMSE RATIOS

	Headline Inflation		Core Inflation	
	$h = 2Q$	$h = 4Q$	$h = 2Q$	$h = 4Q$
SPF	2.080	2.075	0.677	0.677
Baseline	1.004	0.967	0.896	0.954

Note: The first row reports the RMSE values of the SPF forecasts, and the second row the RMSE ratios of the LogOP log score weighted combination RMSE relative to the SPF.

Since our forecasts are “fixed horizon” as opposed to “fixed event”, our comparison is restricted to point forecasts as the SPF asks “fixed horizon” forecast based questions with respect to point forecasts only.²⁸ Table 2 reports the RMSE ratios of the baseline model relative to the SPF. These are for the quarterly horizons where $h = 2Q$ and $4Q$, which (approximately) correspond to our $h = 6$ and $h = 12$ monthly horizons. As the SPF PCE headline and core inflation data are available from 2007q1 this comparison does not cover the same

²⁵The exception to this pattern is for some of the BIC weighted combinations.

²⁶Whereas for point forecasts the best performing opinion pool is linear with the preferred weights typically being based on MSE. Note however, that the LogOP log score weighted combination are often quite close to the best performing point forecast combination, in particular for headline inflation.

²⁷Note that first month of the quarter information for commodity prices and convenience yields, as well as the exchange rate and the interest rate, are potentially available at the time the SPF implements the survey. Including this more up-to-date information does not qualitatively change the results presented in this section.

²⁸The probabilistic questions they ask when constructing histograms, are for headline inflation only and are with respect to current and next year’s annual inflation for each quarterly survey in the year, and as such are “fixed event” based forecasts. Hence using these would involve converting from fixed event to fixed horizon densities, a complication we opt to avoid in this instance. Ganics, Rossi, and Sekhposyan (2020) suggest methods to address this issue.

sample as our monthly forecasts.²⁹ Ratios of less than one favour our baseline model combination. We observe, for headline inflation, approximate equivalence between our quarterly model version and the SPF, whilst for core inflation we observe more consistent gains of approximately 10% and 5% at horizons $h = 2Q$ and $4Q$ respectively. Hence for point forecast performance, arguably a strength of the SPF survey, our approach is at least equivalent.³⁰

4.3 On the importance of information from disaggregated commodities: out-of-sample evidence

In this section we evaluate how models including only factors of the commodity price data as predictors perform against the overall model combination, i.e. how important is the idiosyncratic information contained in the disaggregated commodity prices for forecasting inflation. To answer this question, in Table 3, we compare the RMSE ratios, log score differences and CRPS ratios of a factor only commodity benchmark combination, which therefore does not use individual commodities price information, with our baseline commodity combined model, which includes individual commodity spot prices and convenience yields.³¹

Table 3: COMPARISON WITH FACTORS ONLY COMBINATION

	Headline Inflation			Core Inflation		
	RMSE Ratio	Diff. Log Score	Ratio CRPS	RMSE Ratio	Diff. Log Score	Ratio CRPS
$h = 1$	1.007	-0.013	0.999	0.983	0.052	0.980
$h = 6$	0.764	0.310	0.765	0.867	0.128	0.867
$h = 12$	0.774	0.260	0.785	0.866	0.112	0.880

Note: Relative performance of the LogOP log score weight combinations, with idiosyncratic information from the individual commodities versus those with factors or principal components. Values of the RMSE and CRPS ratios less than 1 and positive values in the difference between log scores suggest that the idiosyncratic information from the individual commodities improve the out of sample performance of the models. Bold numbers report RMSE and CRPS ratios significantly less than 1 or log score differentials significantly different from zero.

The results suggest that using the factor only based model results in a worsening out-of-sample forecast performance at horizons $h = 6$ and $h = 12$, but with approximate equal performance at forecast horizon $h = 1$. We observe RMSE ratios of less than one in five out of six cases, with an average value of 0.859. The log score differences are, with the exception of headline inflation at $h = 1$, always positive, and are generally significant. Similarly, at forecast horizons $h = 6$ and $h = 12$, the CRPS ratios are less than one, with an average gain of 37%, all of which are significant. An explanation of this result may well relate to the ability of disaggregate commodities to vary, idiosyncratically, that allows for more predictive power relative to using factors which conflate these individual movements. The factor loadings, as observed in Section 2, are concentrated on agriculture, food items and also energy. The results are also consistent with the in-sample evidence, lending further support in favour of incorporating the information using disaggregated commodities, as opposed to a focus on the common dynamics across the different commodity market as captured by aggregate factors.

5 Assessing Inflation Risk

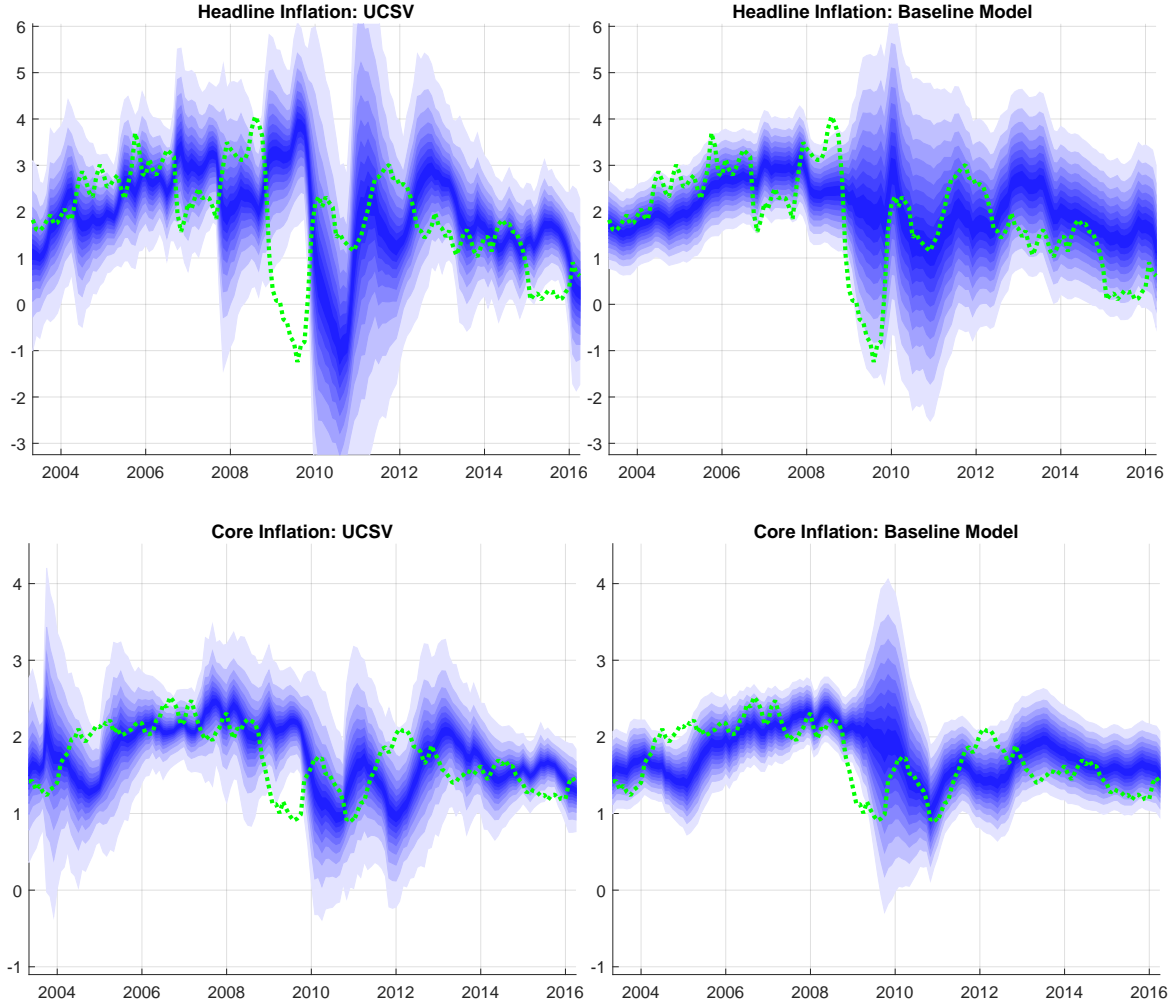
Having established that the model combinations including information from commodity prices produce well calibrated density forecasts, in particular for $h = 6$ and $h = 12$, in this section we use the combined inflation forecast densities to assess inflation risk and to analyse specific economic events of interest, events that may well lie in the tails of the distribution. Before proceeding, it is of interest to visualise how the forecast densities for

²⁹CPI headline inflation point forecasts are available from 1981q3, but the core CPI measure along with PCE headline and core were available from 2007q1 onwards and hence we opt for the common period.

³⁰Note also that in Appendix E, Table F.1, we report the same comparison with respect to all the combined model. We observe in some cases, in particular the LOP MSE weighted model, even stronger performance relative to the SPF.

³¹In both cases we use the LogOP with log score weights combination. Specifically, for the factor only model we consider a combination of possible model specifications considering model uncertainty in terms of conditioning variables and lags specification.

Figure 4: INFLATION DENSITY FORECASTS: $h = 12$



Note: The green line plots observed inflation and the blue shaded area are 5% to 95% intervals of the forecast density, each with a 5% width.

the model combination evolve overtime and how they compare to the densities produced by the UCSV model. In Figure 4 we plot the inflation density forecasts at horizon $h = 12$, in the form of a ‘fan chart’ (i.e. plotting the 5% to 95%, confidence intervals, with a 5% width).³²

The plots reveal a number of observations which, as we will see, are relevant to our results reported in the later part of this section. First, the baseline models density is more ‘condensed’ in comparison to the UCSV density up until the 2008 crisis period. After this point uncertainty or the variance of the distribution increases significantly. Therefore, although we do not explicitly model time variation in the volatility of the residuals, the resulting densities display a shift in volatility. This is partly due to the use of a ‘short’ rolling window in the estimation of the model, but most importantly to the fact that the commodity prices and convenience yields, i.e. the predictors in our setting, are subject to large variations around the last recession.³³ Second, the change overtime in the baseline densities variance is quite smooth, in particular when compared to movements in the UCSV density. In contrast, the variance of the UCSV density is generally smaller, and is subject to

³²Appendix E reports the same density plot for 12 month ahead CPI inflation, where the qualitative points highlighted here are also true, as is the case when examining forecast horizon $h = 6$.

³³Most commodity prices display substantial falls from mid 2008 to mid 2009, and then bounce back in late 2010 and early 2011. For example, between July 2008 and July 2009, the all commodity price index fell by 43.9% (Energy and Food fell by 52% and 22.3% respectively), but thereafter increased such that by December 2010 it was 41% higher than the July 2009 value (Energy and Food rose by 39.4% and 29.0% respectively). By April 2011 the all commodities price index was some 70.2% higher than its July 2009 value (Energy and Food had risen by 77.6% and 40.6% respectively). We also observe substantial shifts in the futures curves in most commodity markets, from being steeply upward sloped (i.e. contango) ahead of the crisis to being in deep backwardation in the midst of the recession and periods thereafter. With substantial differences in the timing for different commodities.

Table 4: EVALUATION AND CALIBRATION OF SECTIONS OF THE DENSITIES

		RS Test: Density Calibration					
		Headline Inflation			Core Inflation		
		Left	Center	Right	Left	Center	Right
h = 6	UCSV	0.765	0.615	0.773	0.863	0.524	0.703
	Baseline	0.763	0.910	0.777	0.383	0.505	0.700
h = 12	UCSV	1.209	1.116	1.046	0.976	0.587	1.163
	Baseline	1.103	0.671	1.046	0.970	0.468	1.323
		Ratio threshold weighted CRPS					
		Headline Inflation			Core Inflation		
		Left	Center	Right	Left	Center	Right
h = 6	UCSV	0.253	0.289	0.369	0.080	0.106	0.115
	Baseline	0.711	0.780	0.686	0.794	0.835	0.783
h = 12	UCSV	0.189	0.217	0.281	0.072	0.090	0.095
	Baseline	0.624	0.795	0.585	0.661	0.773	0.671

Note: The upper panel reports the density calibration test statistics of Rossi and Sekhposyan (2019). The lower panel reports the threshold weighted (tw) CRPS test statistic, as in Gneiting and Ranjan (2011). The first row refers to the *absolute* values for the UCSV benchmark, whereas the second row reports the ratios of baseline LogOP log score weighted to the UCSV benchmark model. In both panels, bold entries are significant at the 10% level.

abrupt changes overtime. Such abrupt movements, coupled with the fact we observe periods where inflation lies considerably outside the range of the UCSV density, for example the 2008 crisis period, suggests that the UCSV density may not do well in terms of identifying the timing of events and in assessing the risk. Interestingly, while the increased dispersion in the density forecast is largely concentrated in the recession period (i.e. 2008-2010) for core inflation, the larger dispersion extends to the end of the sample for headline inflation. Comparing the baseline model combination with the UCSV, it is clear that the former is quicker to pick up the shifts in volatility, in particular for core inflation. Overall, the baseline model combination, in part reflecting the high variability in the commodity markets, shows sharp movements in the crisis, and adapts well to the inflation environment in the decade post recession, characterized by sustained inflation and higher uncertainty, especially for headline inflation. In contrast, the UCSV density increases its variance or uncertainty by less, but shows large and more abrupt shifts in the mean.

In spite of sharp differences between the resulting forecasts, both densities remain well calibrated. Evidence for this is in the upper part of Table 4, which reports the calibration test of Rossi and Sekhposyan (2019), as applied to different parts of the density forecast. They highlight that both densities are well calibrated, at the centre of the distribution as well as in the two ‘tail’ regions. However, threshold-weighted CRPS statistics (see Gneiting and Ranjan, 2011), which evaluate density forecasts emphasising the tail regions of interest, reported in the lower part of Table 4, enable us to discriminate between the baseline model combination and benchmark. According to this metric the baseline model combination produces more precise density forecasts, with threshold weighted CRPS ratios all less than one, where the gains range from around 20% to 40%, and all are significant.

5.1 Probability Events and the Brier Score

Thus far we have used statistically based criteria to evaluate our forecast densities. However, forecasts are arguably best judged in terms of how well they perform the task for which they were produced, or alternatively in terms of the decision makers loss function. This may take the form of an expected loss with respect to a particular event, possibly defined by a policy framework, or more generally calculating a measure of uncertainty or risk at particular points in time. In our case, we examine events of interest which correspond to large deviations from the central bank inflation target. Measuring the risk of an event, as outlined in Kilian and Manganello (2007, 2008), has two basic requirements. The first is that the measure of risk is related to the

probability distribution of the underlying random variable. The second, that any measure of risk must be linked to the preferences of the economic agent.

We begin our analysis by considering a limiting but interesting case where we set parameters in the loss functions (defined in section 5.2) so that the measures of 'risk' reduce to the probabilities of exceeding the target at either end of the specified range. The benefit of looking at this simple case is that event probabilities can be easily evaluated and decomposed using well understood metrics (see Gneiting, 2008, and Galbraith and van Norden, 2012). Policymakers need to account for forecast uncertainty when making decisions and therefore the probability that specific inflation events, considered to be of policy relevance, are of interest. Moreover, whilst good calibration or reliability of a density forecast is clearly desirable, other features of the density such as resolution which measures the ability to distinguish between relatively high-probability and low-probability events, i.e. the ability of correctly 'timing' the occurrence of the events, is highly desirable from a practical perspective. In this section we evaluate the performance of the forecast probabilities, of a number inflation events, calculated using our baseline forecast densities. The evaluation uses the well known Brier score (Brier, 1950), which measures whether the forecast probability of the event matches when the event occurred or did not occur.

To help interpret the Brier score statistics, we calculate the well-known decomposition of the Brier score into uncertainty, reliability (calibration) and resolution (see Murphy, 1973). Uncertainty measures the observed frequency of the event in question over the evaluation period, and is the same for all specifications, being a function only of the observations and not the forecasts. As described in Galbraith and van Norden (2012), reliability measures the difference between the actual forecast probabilities of an event and the observed frequency of the event.³⁴ In contrast, resolution measures the ability to distinguish between relatively high-probability and low-probability events. If resolution is high, then the conditional expectation of the outcomes differs substantially from its unconditional mean and the probability of an event occurring is high when the event eventually materializes. Low Brier scores are preferred and as Resolution enters negatively into the decomposition, high resolution is preferred as it lowers the Brier score.³⁵

We examine three probability events for headline (core) inflation: less than 1% (1.3%), between 1% and 3% (1.3% and 2.3%) and greater than 3% (2.3%), at annualized rates.³⁶ The Brier scores and the components of its decomposition are reported in Table 5 for our UCSV benchmark and baseline combined model. As uncertainty is the same in all cases, following Galbraith and van Norden (2012), we scale the Brier score, reliability and resolution measures by the uncertainty measure. Lower Brier scores and reliability and higher resolution (as it enters the decomposition with a negative sign) are preferred.

The Brier score for the baseline commodity model average approach is lower than that for the UCSV benchmark, in all cases. In particular we observe large Brier score differences for headline and core inflation, for both $h = 6$ and $h = 12$ forecast horizons, in the upper and lower tail events. Average percentage differences of the baseline Brier scores are some 40-50% lower than those of the UCSV benchmark. The Brier score decomposition provides some insight as to the determinants of the score gains of the probability event forecasts. We observe that the Brier score gains of the baseline over the UCSV benchmark arises from both preferred reliability or calibration but importantly from higher and mostly significant resolution performance. While we observe poor reliability and reject the null of good calibration for both models, the baseline model combination

³⁴For example, suppose 20% is the probability of a specific inflation event (say inflation less say 2%), and the model (system) predicts that inflation falls below this same threshold for 20% of the observations, then we have a perfect match with the data, and reliability (calibration) is perfect, with this term being zero.

³⁵The choice of the number of bins is relevant when constructing the reliability (*Rel*) and resolution (*Res*) terms of the Brier Score decomposition. Uncertainty (*Unc*) is a function of the data only. Therefore, given $BS = Unc + Rel - Res$, we (i) calculate the *BS* for a given set of (zero/one) data outcomes and probabilities, for a given event, which does not require the use of bins, then (ii) nominate the number of bins to construct *Rel* and *Res*, and then calculate the *BS* from the decomposition. The preferred number of bins was then selected, according to which gives an accurate approximation to *BS* calculated in (i). Experimenting with the choice of number of bins suggested an accurate measure (to within three decimal places) when choosing 20 bins. Accuracy ceased to improve beyond 20 bins, and therefore the results are robust to this choice.

³⁶As core inflation fluctuates over a narrower range in the data, we consider less 'extreme' thresholds for the events. For core inflation we also center the event at 1.8%, which is roughly the average of the series over the sample considered. For the core measure of inflation, at forecast horizon $h = 12$, we note the small number of occurrences of the event of interest, and as such caution should be applied when interpreting the results.

Table 5: SCALED BRIER SCORES FOR INFLATION EVENTS

		Headline Inflation			Core Inflation		
		Event: $\pi \leq 1\%$			Event: $\pi \leq 1.3\%$		
		BS	Rel.	Res.	BS	Rel.	Res.
h = 6	UCSV	1.312	0.391	0.079	1.372	0.541	0.169
	Baseline	0.892	0.110	0.212	1.076	0.189	0.112
h = 12	UCSV	1.352	0.451	0.099	1.359	0.449	0.091
	Baseline	0.824	0.157	0.332	0.910	0.190	0.279
		Event: $1\% \leq \pi \leq 3\%$			Event: $1.3\% \leq \pi \leq 2.3\%$		
		BS	Rel.	Res.	BS	Rel.	Res.
h = 6	UCSV	1.222	0.328	0.105	1.167	0.264	0.095
	Baseline	1.085	0.155	0.097	1.082	0.121	0.039
h = 12	UCSV	1.327	0.466	0.140	1.592	0.695	0.102
	Baseline	1.132	0.328	0.195	1.018	0.122	0.103
		Event: $\pi \geq 3\%$			Event: $\pi \geq .23\%$		
		BS	Rel.	Res.	BS	Rel.	Res.
h = 6	UCSV	1.133	0.302	0.168	1.160	0.276	0.117
	Baseline	1.001	0.152	0.150	0.928	0.087	0.167
h = 12	UCSV	1.406	0.680	0.273	1.707	0.800	0.094
	Baseline	1.089	0.221	0.132	1.248	0.517	0.234

Note: The test for correct reliability/calibration and for resolution follows that outlined in Galbraith and van Norden (2012). To test calibration, we estimate the model $x_i = a + b\hat{p}_i + c\hat{p}_i^2 + e_i$ where x_i is a zero/one indicator that the event happened and \hat{p}_i the estimated probability of the event. We jointly test $H_0 : a = 0, b = 1, c = 0$ with χ_3^2 distributed Wald statistic, using robust (Newey-West standard errors). The test of zero resolution is of $H_0 : b = 0$ and $c = 0$ with χ_2^2 Wald statistic. Bold highlighted entries reject the null hypothesis at the 10% level of significance.

displays a statistically different from zero resolution in nine of the twelve baseline cases reported. Hence the commodity average model substantially dominates in terms of resolution potentially providing densities with the ability to identify specific events with a high degree of confidence.

5.2 The Balance of Risks

In this section we move away from the idea that characterising risk can be stated in terms of the probability of exceeding nominated thresholds. The extent to which a threshold is exceeded matters and risk is thought of as a combination of the probability of the event and the severity of the events outcome. In particular we follow Kilian and Manganeli (2007) and adopt the following loss function:

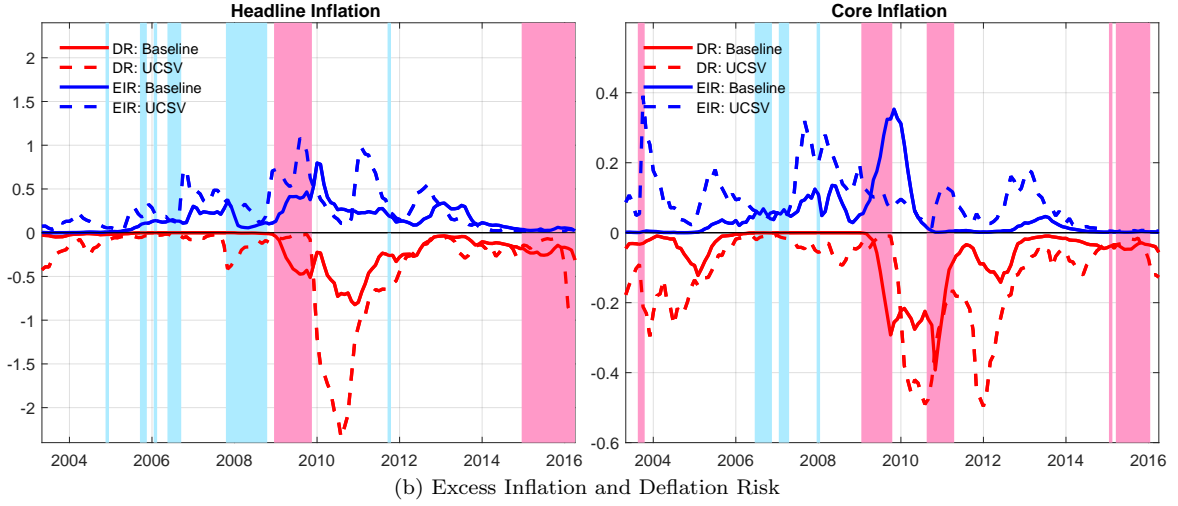
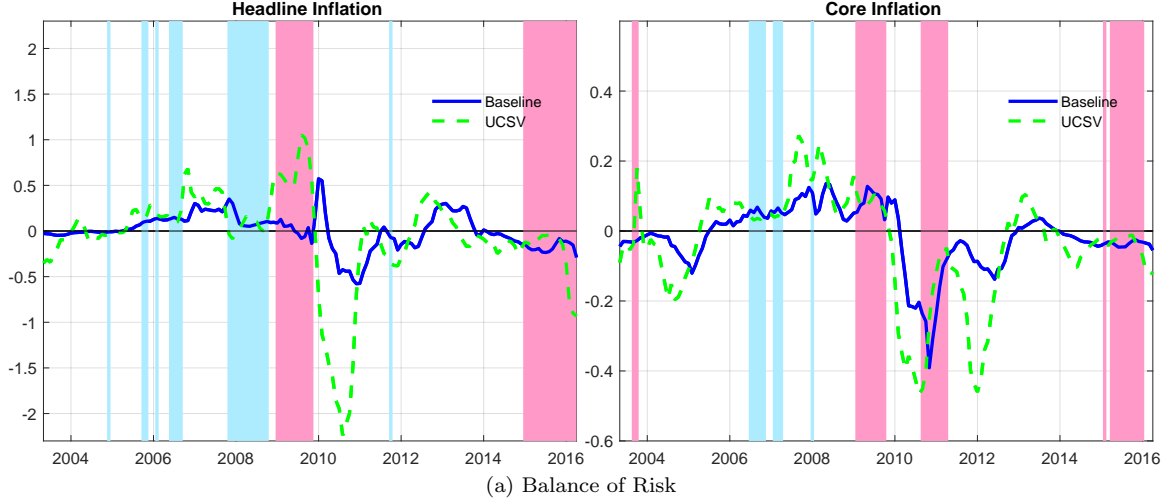
$$l(\pi) = \begin{cases} \frac{1}{2}(\underline{\pi} - \bar{\pi})^\alpha & \text{if } \pi < \underline{\pi}, \\ 0 & \text{if } \underline{\pi} < \pi \leq \bar{\pi} \\ \frac{1}{2}(\bar{\pi} - \pi)^\beta & \text{if } \pi > \bar{\pi} \end{cases} \quad (5)$$

where π denotes the inflation rate over the horizon of interest, the parameters $\alpha \geq 0$ and $\beta \geq 0$ measures the degree of risk aversion of the economic agent. Given the loss function (5) deflation risk (DR) and excessive inflation risk (EIR) are defined as:

$$\begin{aligned} DR_\alpha(\pi) &\equiv - \int_{-\infty}^{\underline{\pi}} (\underline{\pi} - \pi)^\alpha dF(\pi), & \alpha > 0 \\ EIR_\beta(\pi) &\equiv \int_{\bar{\pi}}^{\infty} (\pi - \bar{\pi})^\beta dF(\pi), & \beta > 0. \end{aligned}$$

The Balance of Risks (BR) measures the trade-off between the risk of high and low inflation and is effectively a weighted average of the risks associated with the two events, with the weights depending on the preference

Figure 5: INFLATION RISK: $h = 12$



Note: The upper panel reports Kilian and Manganelli (2007)'s measure of the balance of risk for 12 months ahead inflation forecast. Using the quadratic loss function which gives equal weight to deviations above and below predefined thresholds the balance of risk is equal to the sum of the expected excess inflation and expected deflation: $BR_{1,1} = DR_1 + EIR_1$, where $DR_1 = E(\pi - \pi | \pi < \pi) \times \Pr(\pi < \pi)$ and $EIR_1 = E(\pi - \pi | \pi > \pi) \times \Pr(\pi > \pi)$. For headline (core) inflation $\pi = 3\%$ (2.3%) and $\pi = 1\%$ (1.3%). Blue (pink) shades corresponds to periods where inflation turned out to be above (below) the stated threshold. The lower panel reports deflation risk (DR) in red and excess inflation risk (EIR) in blue.

parameters of the economics agent:

$$BR_{\alpha-1, \beta-1}(\pi) \equiv \frac{\alpha}{2} DR_{\alpha-1}(\pi) + \frac{\beta}{2} EIR_{\beta-1}(\pi). \quad (6)$$

For instance, adopting a symmetric quadratic loss function (i.e., with $\alpha = \beta = 2$), deflation risk is defined as expected deflation, which is the product of the expected shortfall of inflation given that inflation falls below the lower threshold and the probability of falling below the lower threshold ($DR_1 = E(\pi - \pi | \pi < \pi) \times \Pr(\pi < \pi)$). Similarly, excessive inflation risk is defined as the expected excessive inflation, i.e. the product of the excess in inflation given that inflation is above the upper threshold and the probability of breaching the upper threshold ($EIR_1 = E(\pi - \pi | \pi > \pi) \times \Pr(\pi > \pi)$). The BR is equal to the sum of the two risks.³⁷ Operating in an uncertain environment, a central bank whose preferences are defined by (5) should optimally set the policy rate so that the marginal increase in the loss associated with above and below target inflation are equalized, i.e. the BR is zero.

³⁷The BR is not an indicator of overall risk (which is measured by expected loss), but is an indicator of the optimality of the distribution of risks. If the BR is negative, then it is tilted towards deflation; if positive, the risks are towards excessive inflation. See Kilian and Manganelli, 2007, for additional details.

The upper panel of Figure 5 plots the BR for headline and core inflation at forecast horizon $h = 12$, calculated using the baseline model combination and the UCSV benchmark model using a symmetric quadratic loss function. The lower panel plots the corresponding (linear) deflationary or downside risks and the excessive inflationary or upside risk (DR_1 and EIR_1). The light blue shaded areas are the periods in the evaluation sample where headline (core) inflation exceeded 3% (2.3%), and the pink shaded area are where headline (core) inflation was less than 1% (1.3%).

The figures highlight that in the first five and last five years of the sample, the BR for the two models are in broad agreement. During these periods, the BR for UCSV is, overall, more volatile containing positive (negative) 'spikes' typically lagging periods where inflation was above (below) the selected thresholds. The signals from the two models are quite different during the last recessions and the subsequent recovery (i.e. in the 2007-2011 sample). Specifically, for headline inflation the UCSV model wrongly predicts a positive BR (i.e. excess inflation risk larger than the deflation risk) in the middle of the recession, when in fact headline inflation reached its lowest values over the sample. The baseline model combination, in contrast, correctly predicts a BR leaning towards the positive side at the end of 2007, reflecting the high commodity prices at the time of the forecast. The subsequent period of low inflation is instead characterized by overall high uncertainty (with the dispersion of the density forecast increasing quite dramatically, as documented in Figure 5) and a roughly equal BR. The commodity price baseline combination also displays a spike in 2010 which coincides with the recovery in inflation after the recession (although in this case inflation did not breach the upper threshold, despite getting close to it). A similar picture emerges for core inflation. The UCSV model generally tends to be more volatile and to lag as oppose to predict periods of high and low inflation. The model combination on the other hand provides ex-ante measures of the BR that are in line with the actual realizations, with the only exception being the low inflation outcome during the recession where the model predicts a positive BR for inflation, but core inflation turned out to be particularly weak.

The lower panel of Figure 5 decomposes the BR into the excess inflation and deflationary risks. The UCSV model displays substantial volatility and is generally characterized by more erratic behavior, predicting high levels of deflation (excess inflation) risk in periods where inflation turned out to be high (low). On the other hand, the baseline model combination shows a higher ability in timing periods of high and low inflation, also the baseline model tends to be less prone to 'false signals'. For instance DR is effectively zero, both for headline and core inflation, between 2005-2008 in correspondence to the spiraling commodity prices, so that the BR was essentially just a reflection of excess inflation risk. Similarly, the model attributes zero EIR in the post recession period, which has been characterized by a prolonged period of weak inflation.

5.2.1 Assessing the difference in ex-ante risk

In this section we seek to quantify and assess the differences between the BR for our baseline model and the UCSV benchmarks for a wider range of risk profiles and forecast horizons. If (policy) decisions are driven by the (ex-ante) risk assessment, we should want to give a high score to the risk which correctly 'times' the realization of the event. Therefore, we would want ex-ante expected downside risk to be high when realized inflation is below the lower threshold, and low when above the upper threshold.³⁸ To operationalize this idea, define an indicator function $\mathbb{I}(x)$ which takes the value 1 when condition x is satisfied and zero otherwise. A score function for the deflation risk can then be defined as:

$$-DR_{t+h|t}(\alpha, \underline{\pi}) [\mathbb{I}(\pi_{t+h} < \underline{\pi}) - \kappa \mathbb{I}(\pi_{t+h} \geq \underline{\pi})] \quad (7)$$

and $\kappa \geq 0$. Note, as DR is by construction negative we multiply by minus one so that we maximise the value of the score. For $\kappa = 0$, the score rule (7) assigns high scores to models that produce an ex-ante high deflation

³⁸Expecting the assessed risk to be in line with the data realization is consistent, for instance, with the standard evaluation of probability forecasts using the Brier Score. In that case, for a model to have a high score, it should predict a high probability (ideally 1) when the event occurs, and low a probability (ideally 0) when the event does not occur. This is equivalent to saying that one expects to have high inflation/deflation risks (in the Brier Score case fully summarized with the probability) coincide with high/low inflation episodes.

risk when the ‘deflation’ event materializes (i.e. when $\pi_{t+h} < \underline{\pi}$). A key feature of the scoring rule, when $\kappa > 0$, is that the occurrence of ‘false alarms’ (i.e. high ex-ante deflation risk when the event does not materialize) are penalized. Similarly, a score rule for excessive inflation risk can be defined as:

$$EIR_{t+h|t}(\beta, \bar{\pi}) [\mathbb{I}(\pi_{t+h} > \bar{\pi}) - \kappa \mathbb{I}(\pi_{t+h} \leq \bar{\pi})] \quad (8)$$

For the BR, we need to take into account three possible outcomes; inflation within the bounds, above the upper bound and below the lower bound. We therefore define a scoring rule as:

$$BR_{t+h|t}(\alpha, \beta, \bar{\pi}) [\mathbb{I}(\pi_{t+h} > \bar{\pi}) - \mathbb{I}(\pi_{t+h} < \underline{\pi})] - \kappa |BR_{t+h|t}|(\alpha, \beta, \bar{\pi}) \mathbb{I}(\underline{\pi} \leq \pi_{t+h} \leq \bar{\pi}) \quad (9)$$

In this case, with $\kappa = 0$, models that produce a positive (and high) BR in periods of above target inflation and negative (and low) balance of risk in periods of below target inflation are preferred. With $\kappa > 0$, ‘false alarms’ are penalized, where the last term assigns a negative score to model that produce values of the BR different than zero any time inflation is within the target bands.

These score statistics are similar in spirit to the well known Kuiper score (KS). The KS statistic is defined as the difference between the “hit” rate (e.g. the proportion of times you correctly predict inflation above a threshold) and the “false alarm” (e.g. the proportion of times you incorrectly predict inflation below the same threshold). Therefore the chosen score function collapse to the Kuiper score (a) when we set $\kappa = 1$ (i.e. when a wrong prediction is weighted equally to a good prediction) and (b) when the risks are translated into indicators of possible events (i.e. where, for example, periods where DR is say 1 or 10 are treated as equivalent, namely as periods of low expected inflation). In our evaluation we consider different values for the penalty $\kappa = 1$, but most importantly we “weight the hit” by the perceived risk. This choice reflects the underlying assumption that the decisions (e.g. policy actions) triggered by the risk assessment are ‘proportional’ to the perceived risk. Therefore big misses are associated with larger (policy) mistakes.

Table 6: EVALUATION OF EX-ANTE INFLATION RISK

	Headline Inflation				Core Inflation			
	$h = 6$		$h = 12$		$h = 6$		$h = 12$	
	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 0.5$	$\kappa = 1$
Neutral								
$BR_{1,1}$	0.144	0.247	0.116	0.184	0.033	0.063	0.029	0.055
DR_1	0.115	0.231	0.104	0.195	0.028	0.059	0.035	0.070
EIR_1	0.001	0.035	0.036	0.084	0.015	0.037	0.023	0.047
Inflation Adverse								
$BR_{1,2}$	0.392	0.714	0.304	0.576	0.045	0.090	0.076	0.139
EIR_2	0.129	0.409	0.177	0.416	0.024	0.061	0.060	0.126
Deflation Adverse								
$BR_{2,1}$	1.361	2.351	0.719	1.306	0.073	0.146	0.070	0.157
DR_2	0.736	1.584	0.469	0.926	0.046	0.105	0.065	0.146

Note: The Table reports the difference between the score statistics, as defined by equations (7), (8) and (9), of the baseline and UCSV benchmark models. A positive number indicates the baseline has a higher score (is preferred) and those in bold are significantly greater than zero at the 10% level of significance. For the neutral case we adopt $\alpha = \beta = 2$, in the inflation adverse case we set $\alpha = 2$, $\beta = 3$ and for the deflation adverse case $\alpha = 3$, $\beta = 2$.

Table 6 reports the difference between the score statistics, as defined by equations (7), (8) and (9), of the baseline and UCSV benchmark models. For headline and core inflation, at forecast horizons $h = 6$ and $h = 12$, we analyse ‘expected’ (i.e. forecast/ex-ante) risk for three loss functions agents or decision makers may have (i) inflation neutral ($\alpha = \beta = 2$), in the sense of equally weighting loss associated with breaches of upper and lower inflation thresholds (ii) inflation adverse ($\alpha = 2$ and $\beta = 3$), where we give higher weight to losses incurred when we exceed the upper inflation threshold and (iii) deflation aversion ($\alpha = 3$ and $\beta = 2$), where we give higher weight to a breach of the lower inflation threshold. Positive differences favour the baseline model.

In all cases we observe positive score differentials between the baseline models combinations with commodity

prices and the UCSV benchmark. Moreover the score differentials are significantly different from zero in a majority of cases. This evidence is particularly strong at forecast horizon $h = 12$, for both headline and core inflation. Therefore, a decision maker equipped with the model average forecast would be aware of the risk of high or low inflation when those events materialize. Relying on the UCSV benchmark, in contrast, would project a misleading picture of the underlying inflation risk. The better performance of the baseline model also reflects its ability of avoiding “false signals” (as highlighted in Figure 5 for the neutral case). As we increase the cost of false signals, by increasing κ from 0.5 to 1.0, we observe an approximate doubling of the score gains.

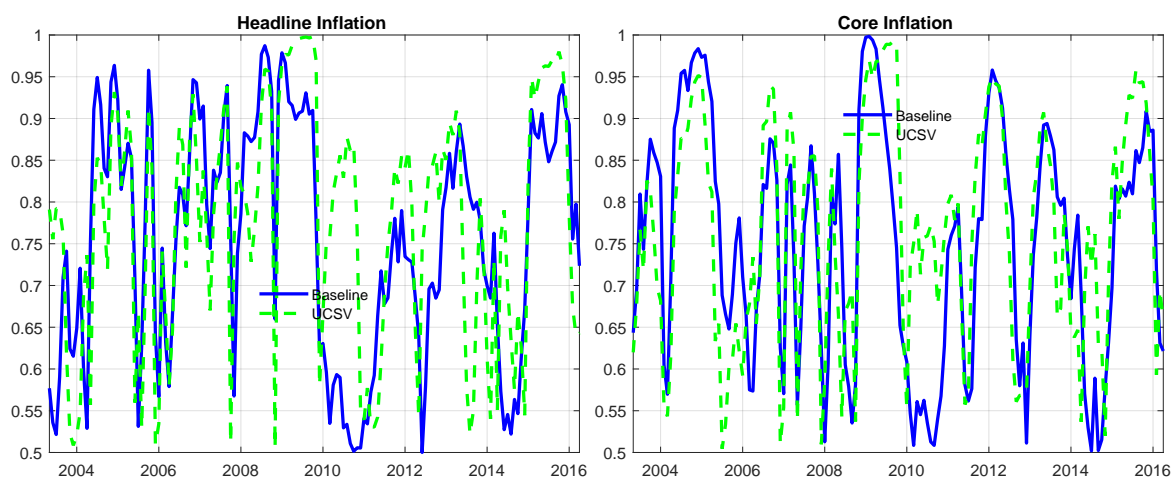
5.3 Measuring Inflation Uncertainty

In the last part of our analysis we compute a measure of inflation uncertainty. Following Rossi and Sekhposyan (2015), we use the information contained in the whole of our predictive densities, as opposed to the BR measure which analyzes the risks associated with breaching specified inflation thresholds or targets. Specifically, we measure uncertainty by comparing the realized forecast errors with the conditional distribution of forecast errors.³⁹ If the realized forecast error lies in the tail of the distribution, we conclude that the realization was difficult to predict, and therefore the environment with respect to that variable is very uncertain. Figure 6 plots (normalized) inflation uncertainty for a 1 year forecast horizon for our baseline average model and the UCSV benchmark model. The normalized uncertainty measure takes a value between 0.5 and 1: values close to 1 are associated with high levels of uncertainty, whereas values close to 0.5 suggest that uncertainty is minimal (i.e. the observed outcome is at the center of the predicted density and therefore did not constitute a surprise).

In Table 1 we have documented that accounting for the information in commodity prices allows us to produce better calibrated density forecasts. Figure 6 highlights that the densities from the UCSV benchmark model, as with the BR, suggests a highly volatile and therefore potentially misleading picture for the underlying uncertainty. Using a UCSV benchmark model density will, on average, produce higher uncertainty relative to that suggested from using the baseline commodity model, for some sub-periods.

For example, during the mid 2008 period, where commodity price inflation was high but then fell very sharply (the IMF Crude oil price index, measuring the average of Brent, WTI and RAC oil prices, was high at 247 in 2008m6 but then fell to 77.7 in 2008m12) we observe high levels of uncertainty for both the baseline

Figure 6: INFLATION UNCERTAINTY: $h = 12$



Note: The figure reports Rossi and Sekhposyan (2015)’s measure of uncertainty calculated for the 12 months ahead inflation forecast. This corresponds to the value of the CDF of inflation forecast evaluated at the value of the (ex-post) realized inflation and it is normalized to be between 0.5 and 1. In all panels, the blue line corresponds to the baseline model average, whereas the green broken line corresponds to the UCSV benchmark.

³⁹The reported uncertainty index is based on cumulative density of the forecast errors evaluated at the actual realized forecast error. In their specific application Rossi and Sekhposyan (2015) compare the forecast error to the unconditional distribution. In this section we also follow their suggestion of normalizing the uncertainty indicator and in doing so we do not distinguish upside from downside uncertainty.

for headline and core inflation, and the UCSV benchmark model. But for headline inflation, and to a lesser extent core inflation, we see a continued high level of uncertainty from the UCSV model which is not reflected in the baseline uncertainty measure corresponding to the pick up in inflation in 2010, which was expected given the quick rebound in commodity prices after the latest recession. Therefore not including commodity price information leads, in this case, to a substantial overestimation of uncertainty. Decomposing this into its upside and downside components highlights that the higher overall uncertainty reflects over prediction of the upside uncertainty, whereas by including commodity price information, the model combination predicts a strong rebound in inflation and almost no upside uncertainty.⁴⁰

6 Conclusion

This paper uses combined or averaged inflation forecast densities, incorporating information from commodity prices, to calculate and assess inflation risk. Using a large set of simple (and possibly mis-specified) linear regressions we provide evidence of a high probability of in-sample predictability from commodity prices to inflation. Using the same set of models, we undertake an out-of-sample analysis where we consider point and density forecast combinations which we find to have strong predictive performance. Brier score statistics and their decompositions highlight the ability of the model to predict with high probability the timing of episodes of high and low inflation as documented by a statistically significant resolution for the aggregate densities. In particular for $h = 6$ and $h = 12$ forecast horizons, there is strong evidence to suggest that the models including commodity price information have statistically significant and quantitatively larger ability to discriminate low probability events when compared to the UCSV benchmark. We extend the Balance of risk analysis of Kilian and Manganelli (2007, 2008) by developing scores measures for the balance of risk, deflation risk and excessive inflation risk. The scores suggest commodity prices have useful informational content when assessing the risk of either deflation and inflation, as they produce a far less volatile signal of the inflation risk when compared to the UCSV model. Measures of overall inflation uncertainty, which are separate from any defined inflation event, suggest that an UCSV benchmark would over estimate the degree of uncertainty which is systematically larger for the UCSV in particular in periods of large swings in commodity prices.

Overall we find substantial gains from using information from commodity prices when assessing probability event and inflation risk. In particular, aside of the gains in point forecast, which have previously been documented in the literature, we find large gains in probabilistic forecast. Therefore, movements in commodity prices provide important information for assessing inflation dynamics and risk, in particular for predicting shifts of inflation outside the conventional bands around the central bank target.

⁴⁰See Figure F.8 in Appendix F.

References

- [1] Alquist, R., S. Bhattarai and O. Coibion (2019), “Commodity-Price Comovement and Global Economic Activity,” *Journal of Monetary Economics*, forthcoming.
- [2] Amisano, G. and R. Giacomini (2007), “Comparing Density Forecasts via Likelihood Ratio Tests”, *Journal of Business and Economic Statistics*, 25, 177-190.
- [3] Bao, Y., T-H. Lee and B. Saltoglu (2007), “Comparing Density Forecast Models”, *Journal of Forecasting*, 26, 203-225.
- [4] Brier, G.W. (1950), “Verification of Forecasts Expressed in Terms of Probability”, *Monthly Weather Review*, 78, 1-3.
- [5] Clark, T.E. and T. Doh (2014), “Evaluating Alternative Models of Trend Inflation”, *International Journal of Forecasting*, 30, 426-449.
- [6] ChenY., S. Turnovsky and E. Zivot (2014), “Forecasting Inflation Using Commodity Price Aggregates”, *Journal of Econometrics*, 183, 117-134.
- [7] Chinn,M.D., and O. Coibion (2013), “The Predictive Content of Commodity Futures”, *Journal of Futures Markets*, 34, 607-636.
- [8] Clark Iain J. (2014) “Commodity Option Pricing: A Practitioner’s Guide.” John Wiley & Sons Ltd.
- [9] Clark, T.E. (2011), “Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility”, *Journal of Business and Economic Statistics*, 29, 327-341.
- [10] Delle Chiaie, S., L. Ferrara and D. Giannone (2018), “Common Factors of Commodity Prices,” *CEPR Discussion Paper 12767*.
- [11] Diebold, F.X., T.A. Gunther and A.S. Tay (1998), “Evaluating Density Forecasts; with applications to financial risk management”, *International Economic Review*, 39, 863-83.
- [12] Elliot, G., A. Gargano and A. Timmermann (2013), “Complete Subset Regressions”, *Journal of Econometrics*, 3177, 357-373.
- [13] Fama, E. F. & French, K. R (1988), “Business Cycles and the Behavior of Metals Prices,” *Journal of Finance*, 43(5), 1075-93.
- [14] Faust, J. and J.H. Wright (2013), “Forecasting Inflation” G. Elliot, C.W.J. Granger and A. Timmermann, eds., *Handbook of Economic Forecasting Vol 2A*, North-Holland, 3-56.
- [15] Fernandez, C., Ley, E. and Steel, M. (2001). “Benchmark priors for Bayesian model averaging”, *Journal of Econometrics*, 100, 381-427.
- [16] Galbraith, J.W, and S. van Norden (2012), “Assessing GDP and Inflation Probability Forecasts Derived from the Bank of England Fan Charts”, *Journal of Royal Statistical Society, Series A*, 175, 713-727.
- [17] Ganics, G., Rossi, B. and T. Sekhposyan (2020), “From Fixed-Event to Fixed-Horizon Density Forecasts: Obtaining Measures of Multi-Horizon Uncertainty From Survey Density Forecasts ”, *Barcelona GSE Working Paper: 1142*
- [18] Garratt, A., Mitchell, J. and S.P. Vahey (2014a), “Measuring Output Gap Nowcast Uncertainty ”, *International Journal of Forecasting*, 30, 2, 268-279.
- [19] Garratt, A., Mitchell, J. and S.P. Vahey (2014b), “Probabilistic Forecasting for Inflation Warnings from the Federal Reserve”, *WBS Working Paper No. 7*.

- [20] Garratt, A., Koop, G., Mise, E. and S. Vahey (2009), “Real-Time Prediction with Monetary Aggregates in the Presence of Model Uncertainty”, *Journal of Business Economics and Statistics*, 27, 480-491.
- [21] Genest, C. and J. Zidek (1986), “Combining Probability Distributions: a critique and an annotated bibliography”, *Statistical Science*, 1, 114-135.
- [22] Giacomini, R. and H. White (2006), “Tests of Conditional Predictive Ability”, *Econometrica*, 74, 1545-1578.
- [23] Gneiting, T. (2008) “Probabilistic Forecasting”, *Journal of the Royal Statistical Society, Ser A*, 171, 319-321.
- [24] Gneiting, T. and A. Raftery (2007) “Strictly Proper Scoring Prediction and Estimation”, *Journal of the American Statistical Association*, 102, 359-378.
- [25] Gneiting, T. and R. Ranjan (2011) “Comparing Density Forecasts Using Threshold and Quantile Weighted Scoring Rules”, *Journal of Business and Economics Statistics*, 29, 411-422.
- [26] Gospodinov, N. and S. Ng (2013), “Commodity Prices, Convenience Yields, and Inflation”, *Review of Economics and Statistics*, 95, 206-219.
- [27] Gospodinov, N. (2016). “The role of commodity prices in forecasting U.S. core inflation”, *FRB Atlanta Working Paper 2016-5*, Federal Reserve Bank of Atlanta.
- [28] Granger, C.W.J. and M.H. Pesaran (2000a). “Decision Theoretic Approach to Forecast Evaluation”, in W.S. Chan, W.K. Li and H. Tong (eds.), *Statistics and Finance: An Interface*, 261-278. Imperial College Press : London
- [29] Granger, C.W.J. and M.H. Pesaran (2000b). “Economic and Statistical Measures of Forecast Accuracy”, *Journal of Forecasting*, 19, 537-560.
- [30] Groen, J.J.J, Paap, R. and F. Ravazzolo (2013). “Real-time Inflation Forecasting in a Changing World”, *Journal of Business and Economic Statistics*, 31, 29-44.
- [31] Geweke, J. and G. Amisano (2011), “Optimal Prediction Pools”, *Journal of Econometrics*, 164, 130-141.
- [32] Hall, S.G. and J. Mitchell (2007), “Density Forecast Combination”, *International Journal of Forecasting*, 23, 1-13.
- [33] Hamilton, J. D. (2018), “Why You Should Never Use the Hodrick-Prescott Filter”, *Review of Economics and Statistics*, 100, 831-843.
- [34] Hannan, E J, Terrell, R D and Tuckwell, N E (1970). “The Seasonal Adjustment of Economic Time Series”, *International Economic Review*, vol. 11(1), pages 24-52, February.
- [35] Harvey, A. C. (1989). “Forecasting, Structural Time Series Models and the Kalman Filter,” *Cambridge University Press*.
- [36] Hevia, C., Petrella, I. and Sola, M. (2018). “Risk Premia and Seasonality in Commodity Futures”, *Journal of Applied Econometrics*, 33, 853-873.
- [37] Hobijn, B. (2008), “Commodity Price Movements and PCE Inflation”, *Current Issues in Economics and Finance*, Federal Reserve Bank of New York, 14, Number 8, 1-7.
- [38] 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.

- [39] Juvenal L. and I. Petrella (2015). "Speculation in the Oil Market", *Journal of Applied Econometrics*, 30, 621-649.
- [40] Kascha, C. and F. Ravazzolo (2011), "Combining Inflation Density Forecasts", *Journal of Forecasting*, 29, 231-250.
- [41] Karstanje, D., van der Wel, M. and D. van Dijk (2015), "Common Factors in Commodity Futures Curves", available at <https://personal.eur.nl/vanderwel/workingpapers.html>.
- [42] Kilian, L. and S. Manganelli (2007), "Quantifying the Risk of Deflation", *Journal of Money Credit and Banking*, 39, 561-590.
- [43] Kilian, L. and S. Manganelli (2008), "The Central Banker as a Risk Manager: Estimating the Federal Reserve's Preferences Under Greenspan", *Journal of Money Credit and Banking*, 40, 1103-1129.
- [44] Kilian, L. (2009), "Not all Oil Price Shocks are alike: Disentangling demand and supply shocks in the crude oil market", *American Economic Review*, 99, 1053-1069.
- [45] Koop, G. (2003). *Bayesian Econometrics*, John Wiley and Sons, Chichester.
- [46] Knuppel, M. (2015), "Evaluating the Calibration of Multi-Step Ahead Density Forecasts Using Raw Moments", *Journal of Business Economics and Statistics*, 33, 270-281.
- [47] Knotek, E.S. and S.Zaman (2017), "U.S. Headline and Core Inflation", *Journal of Money Credit and Banking*, 49,2 931-968.
- [48] 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.
- [49] Murphy, A.H. (1973), "A New Vector Partition of the Probability Score", *Journal of Applied Meteorology*, 12, 595-600.
- [50] Nelson, C. R. and Siegel, A. F. (1987). "Parsimonious Modeling of Yield Curves", *The Journal of Business*, 60, 473-489.
- [51] Panagiotelis, A. and M. Smith (2008), "Bayesian Density Forecasting of Intraday Electricity Prices Using Multivariate Skew t-distributions", *International Journal of Forecasting*, 24, 710-727.
- [52] Pesaran, M. H. & Timmermann, A. (2004), "How costly is it to ignore breaks when forecasting the direction of a time series?," *International Journal of Forecasting*, vol. 20(3), pages 411-425.
- [53] Pesaran, M. H. & Timmermann, A. (2007), "Selection of Estimation Window in the Presence of Breaks", *Journal of Econometrics*, 137, 411-425.
- [54] Pindyck, R (1993), "The Present Value Model of Rational Commodity Pricing", *Economic Journal*, 103, 511-530.
- [55] Proietti, T. (2000). "Comparing seasonal components for structural time series models", *International Journal of Forecasting*, 16, 247-260.
- [56] Ranjan, R. and T. Gneiting (2010), "Combining Probability Forecasts", *Journal of the Royal Statistical Society Series B*, 72, 71-91.
- [57] Rossi, B. and T. Sekhposyan (2015), "Macroeconomic Uncertainty Indices Based on Nowcast and Forecast Error Distributions", *American Economic Review: Papers and Proceedings*, 105, 650-655.

- [58] Rossi, B. and T. Sekhposyan (2019), “Alternative Tests for Correct Specification of Conditional Predictive Densities”, *Journal of Econometrics*, 208, 638-657.
- [59] Schwarz, G. (1978). “Estimating the dimension of a model”, *Annals of Statistics*, 6, pp. 461-464.
- [60] Szymanowska, M., F. Roon, T. Nijman & Rob Goorbergh, (2014), “An Anatomy of Commodity Futures Risk ”, *Journal of Finance*, 69, 453-482.
- [61] Stock, J.H. and M.W. Watson (1999). “Forecasting inflation”, *Journal of Monetary Economics*, 44, 293-335.
- [62] Stock, J.H. and M.W. Watson (2002). “Macroeconomic Forecasting Using Diffusion Indexes”, *Journal of Business and Economic Statistics*, 20, 147-162.
- [63] Stock, J.H. and M.W. Watson (2007) “Why Has U.S. Inflation Become Harder to Forecast?”, *Journal of Money, Credit and Banking*, 39, 3-33.
- [64] Trolle, A. B. & Schwartz, E. S. (2009). “Unspanned Stochastic Volatility and the Pricing of Commodity Derivatives”, *Review of Financial Studies*, 22, 4423-4461.
- [65] Wright, J. H. (2008). “Bayesian Model Averaging and exchange rate forecasts”, *Journal of Econometrics*, 146, 329-341.
- [66] Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley and Sons.

Commodity Prices and Inflation Risk Supplementary Material

Anthony Garratt

Warwick Business School, University of Warwick

Email: Anthony.Garratt@wbs.ac.uk

Ivan Petrella

Warwick Business School, University of Warwick

Email: ivan.petrella@wbs.ac.uk

A Constructing Convenience Yields

As part of our analysis we are concerned with the appropriate empirical measure of a convenience yield and then subsequent use of this measure when predicting and forecasting inflation. In this appendix we describe the construction of short-end (1-month) and annual (12-month) convenience yield measures, which address the issues of the uneven availability of commodity futures prices within the calendar year and the strong seasonality present in most commodity prices. The aim is to extract commodity specific convenience yields which are not affected by spurious seasonal dynamics and therefore enable the best prediction and forecasting performance. To do so we use the full set of term structure information available in the futures curve of each commodity.

Define the basis as $F_{jt,n} - S_{jt}$, where S_{jt} and $F_{jt,n}$ denote the spot and futures price of commodity j for delivery at time $t + n$ respectively and let $i_{t,n}$ be the nominal interest earned between periods t and $t + n$. The return to purchasing a commodity at time t and selling at time $t + n$ has two components (i) the opportunity cost of the forgone interest when having to borrow in order to buy a commodity but less (ii) a convenience yield gained as a result of physically holding the asset or inventory, (less storage cost). Convenience yields arise because commodity inventories held can have a productive value e.g. wheat acting as an input to producing another commodity (flour for example) or from holding inventories to meet unexpected demand as in the Pindyck (1993) model of rational commodity pricing, where convenience yields play the role of dividends that anticipate future changes in spot commodity prices. From our perspective the importance of including information from the convenience yields as a predictor of inflation arises from the fact that those variables contain forward looking information on the developments in the commodity markets.

Fama and French (1988) highlight the link between spot prices, futures and convenience yields:

$$F_{jt,n} - S_{jt} = S_{jt}i_{t,n} - CY_{jt,n} \quad (\text{A.1})$$

where $CY_{jt,n}$ denotes the convenience yield associated with the holding of a futures contract expiring in n months. Therefore we can retrieve $cy_{jt,n} = CY_{jt,n}/S_{jt}$ as:

$$cy_{jt,n} = (1 + i_{t,n}) - (F_{jt,n}/S_{jt}). \quad (\text{A.2})$$

The convenience yield defined in equation (A.2) can potentially be calculated at different maturities. GN focus on the short end of the futures curve and, for each commodity j , construct a proxy of the convenience yield as

$$\tilde{cy}_{jt}^c = (1 + i_t)^{\frac{\tau_{2,t} - \tau_{1,t}}{12}} - (F_{jt,\tau_{2,t}}/F_{jt,\tau_{1,t}}) \quad (\text{A.3})$$

where i_t is the annualized three month US Treasury bill, $F_{jt,\tau_{1,t}}$ is the price of the futures contract closest to expiration (which is meant in this case to proxy for the spot price) and $F_{jt,\tau_{2,t}}$ is the second available futures price. However, there are two drawbacks of constructing a proxy for the convenience yield following equation (A.3).

First, the uneven availability of futures prices within the calendar year forces the investigator to infer the slope of the futures curve from the two closest contracts, which are not always of equal distance overtime or to the current month of the year where the convenience yield is calculated. Note that generally only for energy commodities futures contracts are issued for each month of the year, therefore $\tau_{1,t} = 1$ and $\tau_{2,t} = 2 \forall t$. Whereas, for most other commodities (food, grains, industrials, metals and livestock and metals) there are, generally, only a limited number of contracts issued for specific months of the year (see e.g. Clark, 2014).¹ Second, and arguably the main issue when considering the appropriate data to be working with for prediction and forecasting, is how to deal with the well documented and often pronounced seasonality present in most of the futures prices (see e.g. Karstanje et al., 2015, and Hevia et al., 2018). In this case, the dynamics of the ratio of the prices of two adjacent contracts (and therefore the dynamics of \tilde{cy}_{jt}^c) can

¹For example, there are only 5 contracts issued per year for Corn futures (with expiration in the months of March, May, July, September and December). This means that the second term in (A.3) for November refers to the ratio of the price between the December and March contracts. Whereas in January and February, one would be looking at the ratio between the price of the March and May contracts. Effectively, in these instances, equation (A.3) assumes that the futures curve is flat across adjacent maturities.

be easily dominated by the seasonal pattern in the futures prices, and as such can generate spurious seasonal dynamic in the proxy for the convenience yields. Forecasts based on such proxies would display a clear seasonal behavior in the predicted variable so as to result in a poor forecast for (seasonally adjusted) inflation rates.

One resolution to the problem of seasonality, is to define an annual convenience yield using futures prices contracts with the same expiration maturity corresponding to the same calendar month of the year, i.e.

$$\tilde{c}y_{jt}^{12m} = (1 + i_t) - (F_{jt,12+\tau_{1,t}}/F_{jt,\tau_{1,t}}) \quad (\text{A.4})$$

therefore accounting for the seasonality in futures prices (if one assumes that this changes little over time). However, we are unable to construct continuous time series of the annual convenience yields following (A.4), as for a range of commodities in our sample, the number of traded contracts does not always span to the next year. This is certainly the case for most commodities in the 90s, but it is still an issue with most commodities, for example in the livestock group, even in the later part of the sample.² To deal with these issues and to create an appropriate convenience yield data set, we instead propose to model the dynamics of the entire futures commodity curve.

For each commodity j , the cost-of-carry relation can be used to define the per-period log or percentage basis (see e.g. Szymanowska et al., 2014). We model the cost of carry curve for a generic commodity, as a function of two common factors, plus a (stochastic) seasonal component. Adding more than a single factor allows us to distinguish between the slope convenience yield curve at the short and long end of the maturity spectrum. We follow Trolle and Schwartz (2009) and parameterise the forward cost of carry curve similar to the Nelson and Siegel (1987) specification, and model the seasonal variation in futures prices as outlined in Hevia et al. (2018). This model is then used to retrieve the de-seasonalised futures curve, $\hat{F}_{jt,1}^{SA}$, from which we can construct convenience yields as in equation (A.2), at the 1-month short end and 12-month annual maturities of the futures curve, $\hat{c}y_{jt}^{1m}$ and $\hat{c}y_{jt}^{12m}$ respectively. The next sections provides details on the data and on the model used for the futures curve.

A.1 Commodity Data

The commodities futures are constructed from the underlying futures contracts published from the relevant exchanges, which are freely accessible from the web provider Quandl.³ We consider 23 commodities among the energy, food, grains, industrials, metals and livestock commodity classes. The sample period for futures prices varies across commodity class, but to aid the approximation of the time series properties of the factors, we use the longest sample of data available for that specific commodity, where the earliest starts in December 1983 and where all go through to March 2016. We include contracts with maturities up to 18 months, although available maturities have varied over time.⁴ We drop from our sample the contracts closest to expiration and label a 1 month futures contract those that expire in the month after the next month, and likewise for the longer maturities. We impose this convention for two reasons. First, delivery for contracts that are about to expire can be made as early as six days after the last trading day. Second, futures contracts become very illiquid a couple of weeks before expiration. Passive traders usually roll forward contracts at the beginning of their expiration month. We have also investigated whether the fit of the model is robust to the exclusion of futures contracts (in particular those of long maturities) that are affected by liquidity issues. We obtain very similar estimates for the factors if we were to exclude (i) contracts with a monthly return that equals zero and (ii) contract with measured volume below a small threshold.

²The number of traded contracts increases over the sample, in particular during the second half of the 2000's an increase in the interest in commodity options was associated with larger numbers of commodity futures issued at longer maturities. The final column in Table A.1 gives the total number of available annual contracts that would allow the construction of (A.4). Often these are low numbers as a proportion of the total sample size.

³<https://www.quandl.com/collections/futures>.

⁴For instance, for heating oil in the early part of our sample, contracts were available with maturities up to 12 months. New contracts appeared in 1991 with maturities up to 18 months. For the agricultural commodities there are only contracts with deliveries for some of the months of the year. In particular, there are 5 traded months per year for corn and wheat and 7 traded months per year for soybean.

A.2 Model Description

For each commodity j , the cost-of-carry relation can be used to define the per-period log or percentage basis as $F_{jt,\tau} = S_{jt} \exp[y_{jt,\tau}]$. Following Hevia et al. (2018) we model the basic curve imposing Nelson and Siegel (1987)'s parametric restrictions to the loadings (see also Trolle and Schwartz, 2009) and explicitly account for stochastic seasonality in the futures prices. Specifically, denoting with $f_{jt,\tau}$ the log of futures price expiring in τ months, we model the futures curve to be a function of two factors (level and slope), plus a seasonal component:

$$f_{jt,\tau} = \beta_{0,jt} + \frac{\tau}{12}\beta_{1,jt} + \frac{1 - e^{-\lambda_j\tau/12}}{\lambda_j}\beta_{2,jt} + \cos\left(\frac{2\pi(t+\tau)}{12}\right)\gamma_{1,jt} + \sin\left(\frac{2\pi(t+\tau)}{12}\right)\gamma_{2,jt} + \varepsilon_{jt,\tau} \quad (\text{A.5})$$

so that $\beta_{0,jt}$ corresponds to the implied (de-seasonalized) spot price derived from the entire futures curve. The dynamics of the log basis is captured by the two factors, $\beta_{1,jt}$ and $\beta_{2,jt}$, corresponding respectively to the level and slope of the cost-of-carry. We allow the seasonality of the data to evolve stochastically overtime, and this variation is captured by the evolution in $\gamma_{1,jt}$ and $\gamma_{2,jt}$.⁵ Specifically we assume that the seasonal pattern of variation in the data is common across maturities. $\varepsilon_{jt,\tau}$ is a measurement error, which we assume to be *iid*. The joint dynamic of the factors follow a restricted VAR(1). The dynamic of the factors is characterized by the VAR(1):

$$\begin{bmatrix} \beta_{j,t} \\ \gamma_{j,t} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_j \\ 0 \end{bmatrix} + \begin{bmatrix} \Phi_j & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \beta_{j,t-1} \\ \gamma_{j,t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{j\beta,t} \\ \mathbf{u}_{j\gamma,t} \end{bmatrix} \quad (\text{A.6})$$

where $E(\mathbf{u}_{j\beta,t}, \mathbf{u}_{j\gamma,t}) = 0$ and $\mathbf{u}_{j\beta,t} \sim N(0, \Omega_{j,\beta})$, $\mathbf{u}_{j\gamma,t} \sim N(0, \Omega_{j,\gamma})$. We further simplify the model and assume that both Φ_j , $\Omega_{j,\beta}$ and $\Omega_{j,\gamma}$ are diagonal matrices. In the implementation we restrict $\Omega_{j,\gamma}$ so to depend only on a scalar factor, specifically $\Omega_{j,\gamma} = \sigma_{j,\gamma}^2 I$ (see e.g. Harvey, 1989). The model is estimated using the Kalman Filter which deals with the unbalanced nature of the data (see Harvey, 1989). The fit of the model (A.5)-(A.6) can be gauged using the average R^2 for each of the maturity included in the model. We note that the model first very well all maturities for all of the commodities. For instance if looking at the average R^2 for the contracts with 1 to 3 months maturities and at the average fit for 1 to 12 months maturity futures, we find that the lowest value of the average R^2 is 0.97, across all 23 commodities, thereby highlighting that the model provides a very good fit of the entire futures curve, for all commodities. The implication is that the proxies constructed from the seasonally adjusted futures prices (i.e. eq. A.7-A.8) will be subject to limited approximation error.

With this model one can easily retrieve a synthetic or seasonally adjusted futures price at each maturity, specifically

$$\widehat{S}_{jt}^{SA} = \exp\left(\widehat{\beta}_{0,jt|t}\right) \quad (\text{A.7})$$

$$\widehat{F}_{jt,\tau}^{SA} = \widehat{S}_{jt}^{SA} \exp\left[\frac{\tau}{12}\widehat{\beta}_{1,jt|t} + \frac{1 - e^{-\lambda_j\tau/12}}{\lambda_j}\widehat{\beta}_{2,jt|t}\right] \quad (\text{A.8})$$

where $\widehat{\beta}_{i,jt|t}$ ($i = 0, 1, 2$) are retrieved using the Kalman filter.⁶ We then use the seasonally adjusted spot and futures prices (A.7)-(A.8) to calculate convenience yields as in equation (A.2), at the 1-month short end and

⁵The specific form of the seasonal component was suggested by Hannan, Terrell and Tuckwell (1970) as an alternative to standard dummy variable methods of modeling seasonality (see Proietti, 2000, for a link between the two representations). The loading matrix associated with the seasonal component is time varying but predetermined. To achieve parsimony, we only allow for a seasonal cycle in the fundamental frequency, i.e. implicitly emphasizing the annual pattern in the seasonality. Moreover, for crude oil and industrial and metal commodities (Cotton, Copper, Lumber, Gold, Silver) for which prices are less evidently seasonal we estimate the same model without the seasonal component.

⁶We prefer using the real-time (i.e. one-sided) factor estimates from the Kalman filter, as opposed to the factors obtained with the Kalman smoother, as they are only a function of current and past information (hence the timing $t|t$). Note the difference between the real-time filter and smoother factors are however very small in the context of our application.

12-month annual maturities of the futures curve:

$$\hat{cy}_{jt}^{1m} = (1 + i_t)^{\frac{1}{12}} - (\hat{F}_{jt,1}^{SA}/\hat{S}_{jt}^{SA}) \quad (\text{A.9})$$

$$\hat{cy}_{jt}^{12m} = (1 + i_t) - (\hat{F}_{jt,12}^{SA}/\hat{S}_{jt}^{SA}) \quad (\text{A.10})$$

Allowing for a two factor representation of the cost-of-carry implies that we are not imposing perfect collinearity between the log basis calculated at different maturities. In fact differences in the short and annual convenience yield are a function of the relative magnitude of the level and slope factors of the cost-of-carry curve and the commodity specific decaying weight, λ_j , which is estimated from the data.

As further means of assessing the seasonally adjusted convenience yields, \hat{cy}_{jt}^{1m} and \hat{cy}_{jt}^{12m} , relative to the closest measures used by GN, \tilde{cy}_{jt}^c , and the annual solution to seasonality if all futures prices were available, \tilde{cy}_{jt}^{12m} , Table A.1 reports some correlations between these various measures. If one focuses on the commodities which do not display seasonal variations, for instance crude oil and copper. The first column reports the correlation between cy proxies constructed looking at the short end of the futures curve and the 12 months cy . The correlation between the two is strong, but is far from being perfect, which suggests that the short end of the futures curve and the long end might contain different information so that it is potentially useful to consider the two separately in the task of predicting inflation. The second and third column report the correlation between the convenience yields constructed from the raw data and the one constructed using synthetic data (appropriately de-seasonalized). For the 12 months cy this is very high for all commodities. This is because this measure, comparing contracts with same month of expiration, is not affected by the seasonality in the underlying prices. It is worth noting though that the unavailability of annual contracts is such that a (continuous) time series for the entire period under investigation cannot be constructed for most commodities. Looking at the short end of the curve, we note that the correlation between the (seasonally adjusted) cy proxy constructed from synthetic data and raw data is strong for the nonseasonal commodities, and falls drastically for the seasonal commodities. In fact, for the latter the dynamics of the cy is dominated by seasonal fluctuations, as illustrated for instance in Figure 1.

Table A.1: CONVENIENCE YIELDS

	Corr($\tilde{cy}_{jt}^c, \tilde{cy}_{jt}^{12m}$)	Corr($\hat{cy}_{jt}^{1m}, \tilde{cy}_{jt}^c$)	Corr($\hat{cy}_{jt}^{12m}, \tilde{cy}_{jt}^{12m}$)	# Annual
Crude Oil	0.915	0.866	0.991	335
Gas Oil	0.850	0.866	0.998	158
Heating Oil	0.652	0.707	0.998	307
Unleaded Gas	0.476	0.478	0.988	128
Natural Gas	0.529	0.501	0.987	290
Cocoa	0.895	0.921	0.986	383
Orange Juice	0.902	0.952	0.995	347
Coffee	0.939	0.952	0.993	388
Sugar	0.784	0.841	0.990	369
Soybean Oil	0.802	0.918	0.989	313
Corn	0.684	0.752	0.985	384
Oats	0.813	0.854	0.989	152
Soybean	0.586	0.693	0.977	339
Soybean Meal	0.604	0.779	0.987	264
Wheat	0.727	0.773	0.991	270
Cotton	0.694	0.473	0.949	383
Copper	0.791	0.875	0.995	369
Lumber	0.723	0.911	0.959	82
Gold	0.905	0.808	0.993	311
Silver	0.791	0.728	0.991	319
Feeder Cattle	–	0.739	–	0
Live Cattle	0.602	0.577	0.961	117
Lean Hogs	0.442	0.461	0.957	245

Note: \tilde{cy}_{jt}^c corresponds to the convenience yields calculated from the two closest available futures (as outlined in eq. A.3), \tilde{cy}_{jt}^{12m} corresponds to the convenience yields calculated from futures contracts referring to the same month of the year (i.e. with 12 months distance, as outlined in eq. A.4). \hat{cy}_{jt}^{1m} and \hat{cy}_{jt}^{12m} correspond to the 1m and 12m convenience yields calculated from the entire futures curve. The last column denoted with ‘# Annual’ reports the total number of available annual contracts that allow the construction of \tilde{cy}_{jt}^{12m} .

A.3 Principal component analysis of commodity prices and convenience yields

In this section we provide additional evidence on the factor structure in the underlying commodity data. Extracting the common factors across different commodities, has the benefit of potentially mitigating the impact of the noise that arises from the construction of the convenience yield proxy. However, this comes at the cost of discarding, ex-ante, information which is idiosyncratic to a single or a subgroup of commodities and can potentially be relevant for predicting future inflation.

Table A.2: PRINCIPAL COMPONENT ANALYSIS OF CONVENIENCE YIELDS

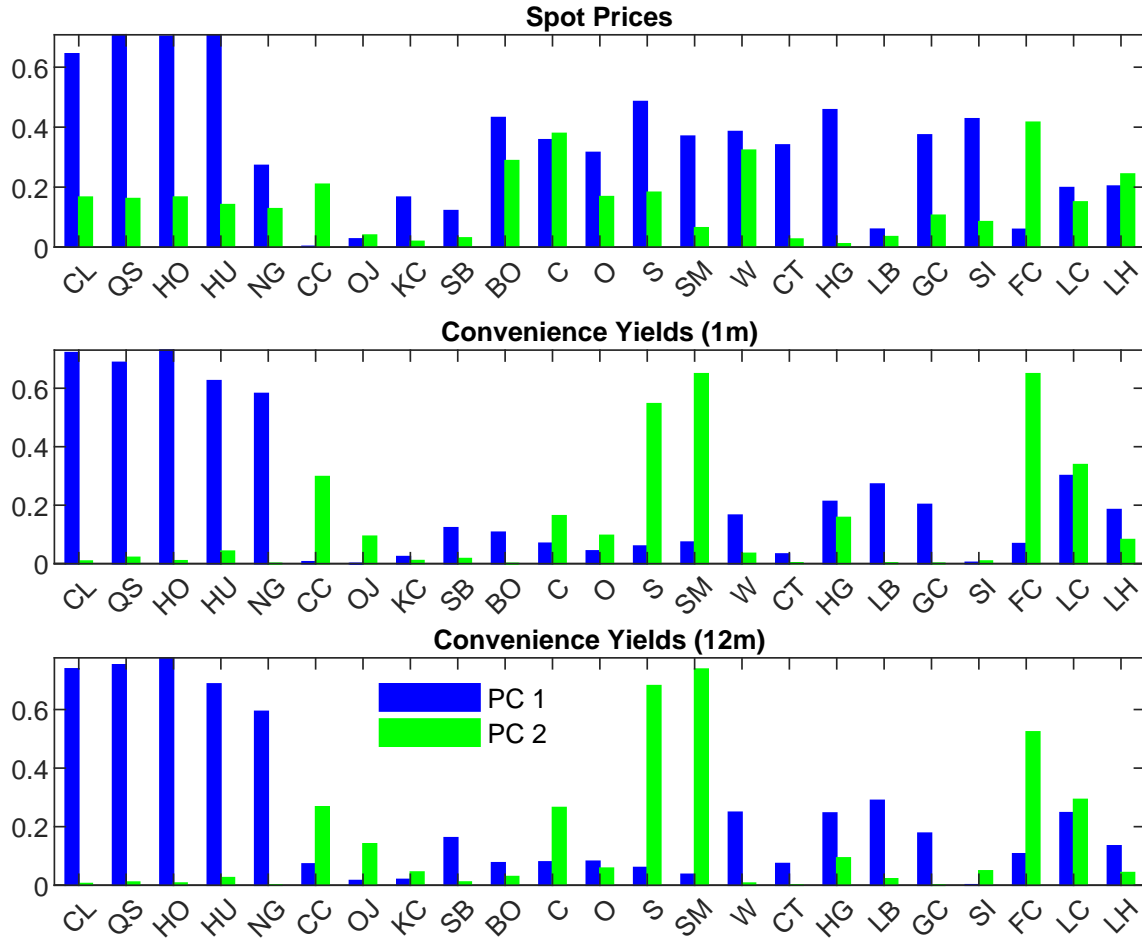
All Commodities	Cumulative R^2		
	q_{jt}	$\hat{c}y_{jt}^{1m}$	$\hat{c}y_{jt}^{12m}$
PC 1	0.333	0.231	0.246
PC 2	0.484	0.371	0.389
PC 3	0.591	0.476	0.496
PC 4	0.666	0.566	0.590
Energy			
PC 1	0.865	0.795	0.837
PC 2	0.979	0.907	0.941
Food			
PC 1	0.394	0.324	0.345
PC 2	0.641	0.619	0.613
Grain			
PCA 1	0.699	0.396	0.444
PC 2	0.830	0.611	0.652
Livestock			
PC 1	0.687	0.598	0.551
PC 2	0.917	0.884	0.869
Metals and Industrials			
PC 1	0.516	0.328	0.326
PC 2	0.716	0.578	0.575

Note: R^2 of estimated principal components for the annual growth in the real spot prices (q_{jt}), 1 month and 12 months convenience yields ($\hat{c}y_{jt}^{1m}$ and $\hat{c}y_{jt}^{12m}$, respectively), estimated for all commodities and for each commodity groups. The commodity groups are defined as: *Energy*: Crude Oil, Gas Oil, Heating Oil, Unleaded Gas, Natural Gas, *Food*: Cocoa, Orange Juice, Coffee, Sugar *Grains*: Soybean Oil, Corn, Oats, Soybean, Soybean Meal, Wheat, *Industrials*: Cotton, Copper, Lumber. *Precious Metals*: Gold, Silver, *Livestock and meats*: Feeder Cattle, Live Cattle, Lean Hogs. The metals and industrial group merges together two groups which otherwise would be too small. Sample period: 1991m8 - 2016m3.

In order to understand the extent to which the majority of the information contained in the convenience yields and spot commodity prices can be accounted for by aggregate factors we perform a principal component analysis (PCA) of the data (see Table A.2). Delle Chiaie et al. (2018) has highlighted that spot commodity prices display a weak factor structure, with a large fraction of the commodity price dynamic being specific to different commodity groups and only part of the variation in the data common accross different commodity groups. The first column of Table A.2 reiterates these observations. Two factors explain 48% and four factors 66.6% of the variation in the data. Whereas the commonality in the data is much stronger within commodity groups. A similar picture emerges when we investigate the dynamic of the convenience yields with even more marked results suggesting a weak facture structure in the data. Specifically, for the short term 1-month convenience yield, $\hat{c}y_{jt}^{1m}$, and the 12-month convenience yield, $\hat{c}y_{jt}^{12m}$, just 37% and 39% of the variation in the data, respectively, is explained by the first two factors. Increasing the number of factors to three or four, still leaves around two fifths of the overall variation in the data unexplained.

Figure A.1 reports the R^2 associated to the first two principal components of the data for each commodity in our dataset. It highlights a very similar structure in the factors' structure for both spot prices and convenience yields. The first principal component loads heavily on the energy commodities, explaining up to almost 70% in the variability of these data (both for spot prices and convenience yields) and to a lesser extent on the metals (both copper and precious metals). For spot prices the first principal component also loads on agricultural

Figure A.1: EXPLANATORY POWER OF THE FIRST TWO PRINCIPAL COMPONENTS

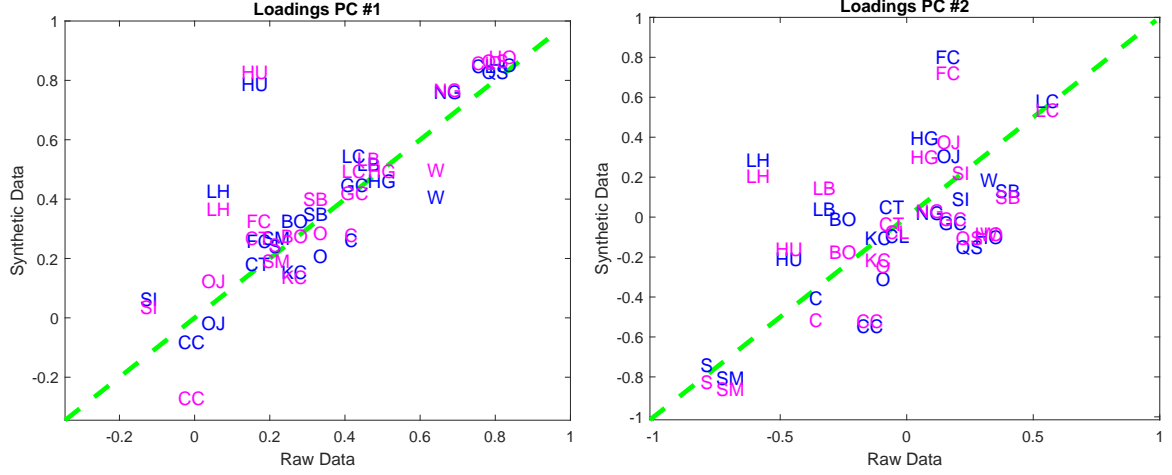


Note: R^2 of estimated principal components for the annual growth in the real spot prices (upper panel), 1 month and 12 months convenience yields (middle and lower panel, respectively). The commodity are ordered in groups and are defined as: *Energy*: Crude Oil (CL), Gas Oil (QS), Heating Oil (HO), Unleaded Gas (HU), Natural Gas (NG), *Food*: Cocoa (CC), Orange Juice (OJ), Coffee (KC), Sugar (SB) *Grains*: Soybean Oil (BO), Corn (C), Oats (O), Soybean (S), Soybean Meal (SM), Wheat (W), *Industrials*: Cotton (CT), Copper (HG), Lumber (LB). *Precious Metals*: Gold (GC), Silver (SI), *Livestock and meats*: Feeder Cattle (FC), Live Cattle (LC), Lean Hogs (LH). Sample period: 1991m8 - 2016m3.

commodities. The second principal components is instead related to the variation in live meat prices and soybean products. This picture is overall similar to the one painted by GN, using their alternative proxy of the convenience yields. In fact figure A.2 plots the loadings for each commodity associated with the two principal components extracted from the three convenience yield proxies. In most cases the loadings are of similar magnitude regardless of the specific proxy of the convenience yield, with the largest deviation corresponding to the most seasonal commodities (such as Lean Hogs, LH, and Feeder Cattle, FC) for which using a convenience yield contaminated by the seasonality shows up into a smaller loading on the second principal component. Interestingly the plot also highlights that whether the first PC can be thought as a 'level' factor since all loadings are positive, the second PC is instead a 'slope' factor, combining with opposite sign the information of the CY for the live meat and the soybean product commodities.

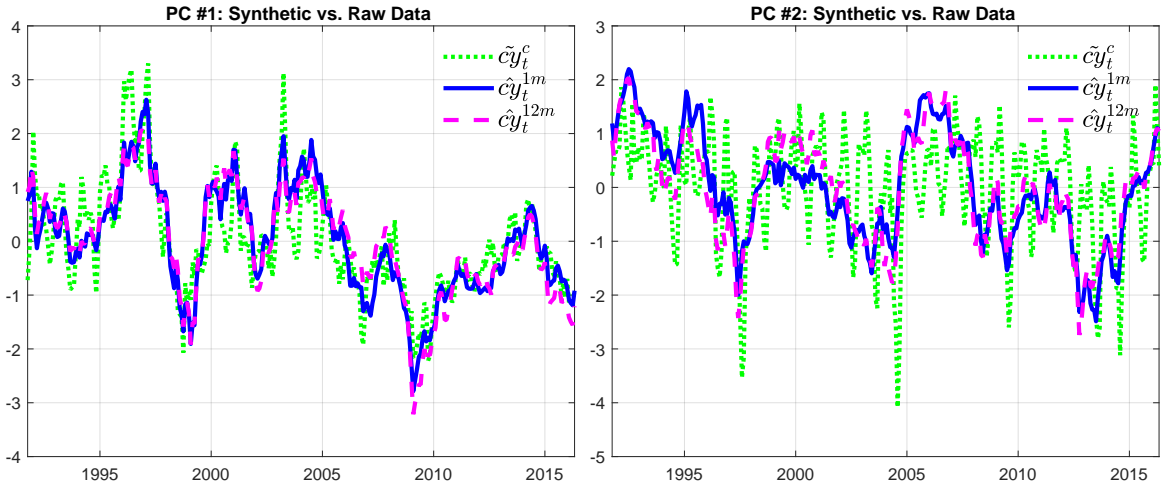
A potential advantage of using common factors of the convenience yields proxies is that, in so far as the seasonal variation in the data is not common across many series so that potentially they could cancel each other out, they should be less affected by the seasonal variation in the underlying data. To investigate this conjecture, in Figure A.3, we plot the first two principal components extracted from different proxies of the convenience yields. We observe that the common factors of the convenience yield proxies constructed from the two closest available contracts, in fact display substantial seasonal variation. This is most obviously the case for second principal component (in the right hand figure), which loads mainly on agricultural and livestock commodities,

Figure A.2: COMPARISON OF LOADINGS ASSOCIATED TO THE PCs



Note: Comparison of the loadings associated to the first two principal component extracted from different convenience yields proxies. On the y-axis we measure the loadings on the synthetic data 1 and 12 month maturities (denoted with the magenta and green colour, respectively) whereas on the x-axis we report the loadings associated with the proxy of the CY calculated from the two closest contracts.

Figure A.3: COMPARISON OF PC



Note: Comparison of the principal component extracted from different convenience yields proxies. Convenience yields are calculated directly from existing futures price contracts on the two contracts closest to expiration ($\tilde{c}y_t^c$), and from synthetic futures prices (blue, -x-) with 1 and 12 month maturity ($\hat{c}y_t^{1m}$ and $\hat{c}y_t^{12m}$, respectively).

which display pronounced seasonal variation, and to a lesser extent for the first principal component (in the left hand figure) which instead loads mainly on energy commodities. In contrast, the principal components extracted from the synthetic 1 and 12 months convenience yields do not display any obvious seasonal dynamics.

Overall, the inference we take from the PCA of the data is that commodity spot prices and convenience yields have a weak factor structure and as such using only aggregate factors can potentially miss important information for predicting inflation.

B Details on the Computation of the Probability of Predictive Content

Using Bayes rule, the posterior model probability can be written as:

$$\Pr(M_k|Data) \propto \Pr(Data|M_k) \Pr(M_k), \quad (B.1)$$

where $\Pr(Data|M_k)$ is referred to as the marginal likelihood and $\Pr(M_k)$ the prior weight attached to this model—the prior model probability. Both of these quantities require prior information. Given the controversy attached to prior elicitation, $\Pr(M_k)$ is often simply set to the non-informative choice where, *a priori*, each model receives equal weight. We will adopt this choice in our empirical work. Similarly, the Bayesian literature has proposed many benchmark or reference prior approximations to $\Pr(Data|M_k)$ which do not require the researcher to subjectively elicit a prior (see, e.g., Fernandez, Ley and Steel, 2001). Here we use the Schwarz or Bayesian Information Criterion (BIC). Formally, Schwarz (1978) presents an asymptotic approximation to the marginal likelihood of the form:

$$\ln \Pr(Data|M_k) \approx l - \frac{K \ln(T)}{2}. \quad (B.2)$$

where l denotes the log of the likelihood function evaluated at the maximum likelihood estimate (MLE), K denotes the number of parameters in the model and T is the sample size. The previous equation is proportional to the BIC commonly used for model selection. Hence, it selects the same model as BIC. The exponential of the previous equation provides weights proportional to the posterior model probabilities used in BMA i.e.

$$\Pr(M_k|Data) = w_j = \frac{\exp(BIC_{kU})}{\sum_{k=1}^N \exp(BIC_{kU})} \quad (B.3)$$

This means that we do not have to elicit an informative prior and is familiar to non-Bayesians. It yields results which are closely related to those obtained using many of the benchmark priors used by Bayesians (see Fernandez, Ley and Steel, 2001). With regards to $\Pr(H_0|Data, M_k)$, we avoid the use of subjective prior information and use the standard non-informative prior. Thus, the posterior is proportional to the likelihood function and MLEs are used as point estimates.

The next step is to calculate the “feature of interest” in every model. Using the same type of logic relating to BICs described above (that is, BICs can be used to create approximations to Bayesian posterior model probabilities), we calculate the BICs for M_k (the unrestricted model) and the restricted model (that is, the model with either of the set of restrictions described in the probability statements of interest) call these BIC_U and BIC_R , respectively. Some basic manipulations of the results noted in above says that, as an example, for the first probability of interest:

$$\Pr(Restriction|Data, M_k, h) = \frac{\exp(BIC_R)}{\exp(BIC_R) + \exp(BIC_U)}. \quad (B.4)$$

i.e. the “probability that information in commodity prices and convenience yields has no predictive content for inflation” for one model, M_k and horizon h .

C Predictive Densities for the Linear Projections

Let the linear projections be written as:

$$Y = XB + U \quad (C.1)$$

where Y is a $T \times 1$ matrix of observations on inflation and X is an appropriately defined matrix of lags of the dependent variable, convenience yields, the change in the spot price of the relevant commodity price and the conditioning variables (it could also include a constant and deterministic terms etc.), B are the coefficients and U is the error matrix, characterized by a variance term.

Based on these T observations, Zellner (1971, pages 233-236) derives the predictive distribution (using a common noninformative prior) for out-of-sample observations, W generated according to the same model:

$$W = ZB + V, \quad (C.2)$$

where B is the same B as in (C.1), V has the same distribution as U , etc. (see Zellner, 1971, chapter 8 for details of definitions, etc.). Crucially, Z is assumed to be known. In this setup, the predictive distribution is multivariate Student-t (see page 235 of Zellner, 1971). Analytical results for predictive means, variances and probabilities such as $\Pr(\Delta p_{t+h} < a\%)$ can be directly obtained using the properties of the multivariate Student-t distribution. For other predictive features of interest, predictive simulation, involving simulating from this multivariate Student-t can be done in a straightforward manner.

The previous material assumed Z is known. How can we handle Z in our case? In the case of one period ahead prediction, $h = 1$, then Z is known. That is, in (C.1), if Y contains information available at time t , then X will contain information dated $t - 1$ or earlier. Hence, in (C.2) if W is a $t + 1$ quantity to be forecast, then Z will contain information dated t or earlier. But what about the case of h period ahead prediction where $h > 1$? Then Z is not known. But, following common practice, we can simply adopt the direct method of forecasting and estimate a different regression for each h . For $h = 1$ work with a standard linear regression as described above. For $h > 1$, still work with a linear regression as in (C.1), except let Y contain information at time t , but let X only contain information through period $t - h$ (i.e. let X contain lags of explanatory variables lagged at least h periods). All these predictive densities will be multivariate Student-t and, hence, their properties can be evaluated (either analytically or simply by simulating from the multivariate Student-t predictive density)

The preceding describes how we derive h step ahead predictive densities. Note that this is an approximate Bayesian strategy and, thus, the resulting predictive densities will not fully reflect parameter uncertainty. We justify this approximate approach through a need to keep the computational burden manageable given the very large number of models we are dealing with, noting that we have $N = 7200$ models for each horizon and recursion. Thus, it is important to make modeling choices which yield analytical posterior and predictive results. If we had to use posterior simulation, the computational burden would have been overwhelming.

D Details on Forecast Density Combination

We combine the out-of-sample density forecasts from the N linear regression models using the Linear Opinion Pool (LOP) and the Logarithmic Opinion Pool (LogOP). Given $i = 1, \dots, N$ linear regression specifications in equation (1), the LOP aggregate inflation density is given by:

$$p(\pi_\tau) = \sum_{i=1}^N w_{i,\tau} g(\pi_\tau | I_{i,\tau}), \quad \tau = \underline{\tau}, \dots, \bar{\tau}, \quad (D.1)$$

where $g(\pi_\tau | I_{i,\tau})$ is the one step ahead inflation forecast density from model i , conditional on the information set $I_{i,\tau}$. The non-negative weights, $w_{i,\tau}$, in this finite mixture sum to unity, where the weights change with each recursion in the evaluation period $\tau = \underline{\tau}, \dots, \bar{\tau}$.⁷

The ensemble density for inflation, defined by using the LogOP, can be expressed in its geometric form as:

$$p^{LogOP}(\pi_\tau) = \frac{\prod_{i=1}^N g(\pi_\tau | I_{i,\tau})^{w_{i,\tau}}}{\int \prod_{i=1}^N g(\pi_\tau | I_{i,\tau})^{w_{i,\tau}} d\pi_\tau}, \quad \tau = \underline{\tau}, \dots, \bar{\tau}, \quad (D.2)$$

where the denominator is a constant that ensures that the ensemble density is a proper density. The LogOP is linear in its logarithmic form, where a feature of the logarithmic combination is that it preserves the t -distribution form of the combination given t component densities.

To construct average or aggregate inflation density forecasts using the LOP and LogOP we require weights, $w_{i,\tau}$. In this application we use five different types of weight. First, we adopt equal weights (EQ), where $w_{i,\tau}^{EQ} = 1/N$. Second, we compute log score (LS) based weights, which use the log score, which reflects the fit of the individual linear regression

⁷The restriction that each weight is positive could be relaxed; see Genest and Zidek (1986).

forecast densities for inflation. Following Amisano and Giacomini (2007), Hall and Mitchell (2007) and Jore, Mitchell and Vahey (2010) we use the logarithmic score to measure density fit for each model through the evaluation period. The intuitive appeal of the logarithmic scoring rule stems from the high score assigned to a density forecast with high probability at the realized value. The logarithmic score of the i^{th} density forecast, $\ln g(\pi'_\tau | I_{i,\tau})$, is the logarithm of the probability density function $g(\cdot | I_{i,\tau})$, evaluated at the realized inflation outturn, π'_τ . Specifically, the recursive log score (LS) weights for the one step ahead densities take the form:

$$w_{i,\tau}^{LS} = \frac{\exp[\phi_{i,\tau}^{LS}]}{\sum_{i=1}^N \exp[\phi_{i,\tau}^{LS}]}, \quad \tau = \underline{\tau}, \dots, \bar{\tau} \quad (\text{D.3})$$

where $\phi_{i,\tau}^{LS} = \omega \phi_{i,\tau-1}^{LS} + (1 - \omega) \times \ln g(\pi'_\tau | I_{i,\tau})$, $\phi_{i,0}^{LS} = 0$ and we set $\omega = 0.9$, which attaches a smaller weight to the most recent logarithmic score. In the empirical exercise we adopt a training period of size κ or a minimum number of periods deemed to be sufficient for the log score accumulation to contain sufficient information and as such when we evaluate we use weights where $\tau - \kappa$ to $\tau - 1$ comprises the rolling window. It is important to note that the weight on the various specifications varies through time. Hence, the aggregate potentially exhibits greater flexibility than any single linear regression specification (in which the individual model parameters are recursively updated).

The third set of weights considered are Continuous Rank Probability Score (CRPS) weights (Gneiting and Raftery (2007)), defined as:

$$w_{i,\tau}^{CRPS} = \frac{[\phi_{i,\tau}^{CRPS}]^{-1}}{\sum_{i=1}^N [\phi_{i,\tau}^{CRPS}]^{-1}}, \quad \tau = \underline{\tau}, \dots, \bar{\tau} \quad (\text{D.4})$$

where $\phi_{i,\tau}^{CRPS} = \omega \phi_{i,\tau-1}^{CRPS} + (1 - \omega) \times CRPS_{i,\tau}$, $\phi_{i,0}^{CRPS} = 0$ and if we assume the predictive distribution is Gaussian with mean μ and variance σ^2 , the $CRPS_{i,\tau}$ can be defined as:

$$CRPS_{i,\tau} = \sigma \left[\frac{1}{\sqrt{pi}} - 2\phi\left(\frac{\pi'_\tau - \mu}{\sigma}\right) - \frac{\pi'_\tau - \mu}{\sigma} \left(2\Phi\left(\frac{\pi'_\tau - \mu}{\sigma}\right) - 1 \right) \right]$$

Among others Gneiting and Raftery (2007) and Panagiotelis and Smith (2008) have argued that the CRPS, which rewards predictive densities with high probability near and at the out turn, provides a robust metric of density forecasts. The logarithmic score involves as harsh penalty for low probability events and thus is sensitive to extreme events, whereas the CRPS is less sensitive to extreme events.

The fourth set of weights used are proportional to the inverse of the individual models Mean Square Error (MSE), and as such uses only point forecast information (where the point forecast is taken as the mean value of the density), ignore the correlation across forecast errors and are given by:

$$w_{i,\tau}^{MSE} = \frac{[\phi_{i,\tau}^{MSE}]^{-1}}{\sum_{i=1}^N [\phi_{i,\tau}^{MSE}]^{-1}}, \quad \tau = \underline{\tau}, \dots, \bar{\tau}$$

where $\phi_{i,\tau}^{MSE} = \omega \phi_{i,\tau-1}^{MSE} + (1 - \omega) \times MSE_{i,\tau}$, $\phi_{i,0}^{MSE} = 0$.

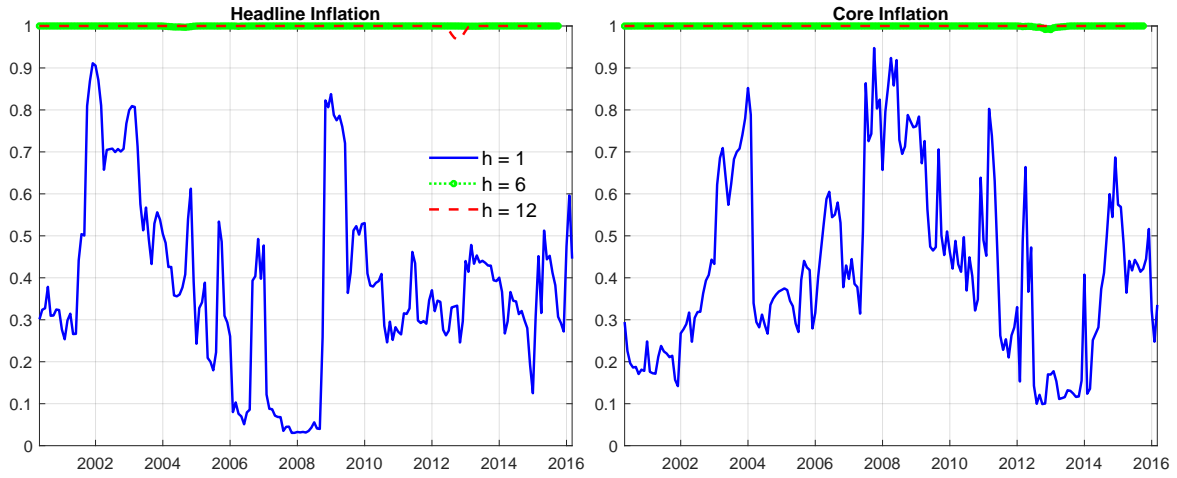
Finally, the fifth set of weights are the Bayesian Information criterion (BIC), comparable to those use in-sample, which take the form:

$$w_{i,\tau}^{BIC} = \frac{\exp[\phi_{i,\tau}^{BIC}]}{\sum_{i=1}^N \exp[\phi_{i,\tau}^{BIC}]}, \quad \tau = \underline{\tau}, \dots, \bar{\tau} \quad (\text{D.5})$$

where $\phi_{i,\tau}^{BIC} = \omega \phi_{i,\tau-1}^{BIC} + (1 - \omega) \times \ln g(\pi'_\tau | I_{i,\tau})$, $\phi_{i,0}^{BIC} = 0$ and we set $\omega = 0.9$, which attaches a smaller weight to the most recent BIC.

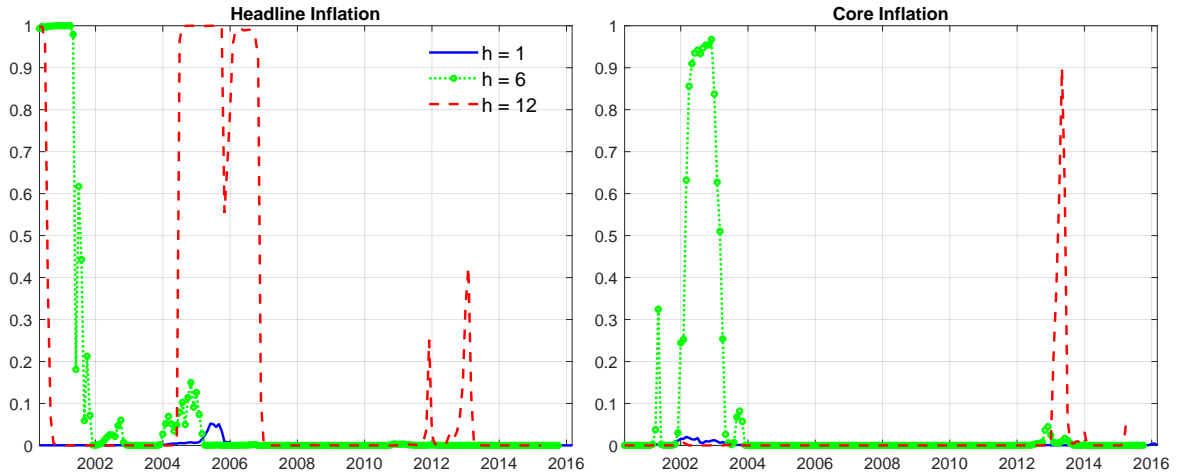
E Assessing the model on CPI inflation

Figure E.1: PROBABILITY OF COMMODITY PRICES PREDICTING (CPI) INFLATION



Notes: The figures reports $1 - \Pr(H_0|Data)$ for different predictive horizons ($h = 1, 6, 12$), calculated following eq. 4. This can be broadly interpreted as the probability that commodity prices (either spot prices or convenience yields) incorporate information useful for the task of predicting inflation.

Figure E.2: POSTERIOR PROBABILITY OF MODELS WITH FACTORS (CPI INFLATION)



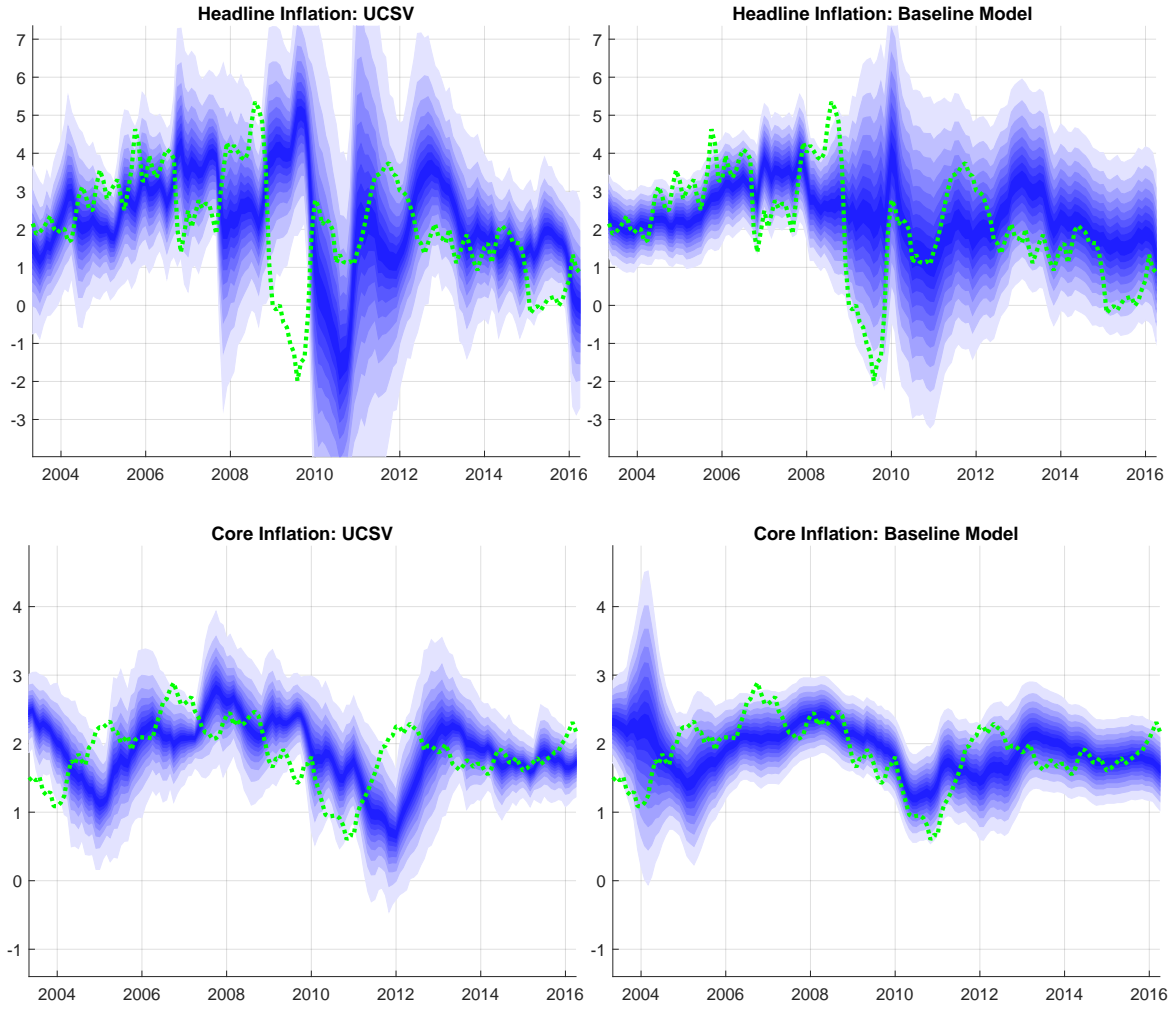
Note: Posterior model probability associated with specifications of model (1) using factors of the commodity information (as opposed to the individual commodity variables). Different model specification are weighted as detailed in (4).

Table E.1: POINT AND DENSITY FORECASTS (CPI INFLATION)

	Headline (CPI) Inflation					Core (CPI) Inflation				
	h=1					h=1				
	RMSE	Log Score	CRPS	RS Test	K Test	RMSE	Log Score	CRPS	RS Test	K Test
UCSV	3.863	-3.407	2.032	1.193	0.003	0.994	-3.497	0.552	0.651	0.799
LOP Combinations										
Equal weights	0.880	0.766	0.887	1.568	0.000	0.921	1.981	1.072	1.904	0.000
Log score weights	0.866	0.784	0.875	1.595	0.000	0.914	2.041	1.026	1.676	0.000
CRPS weights	0.872	0.777	0.878	1.595	0.000	0.913	2.025	1.038	1.731	0.000
MSE weights	0.864	0.785	0.870	1.609	0.000	0.898	1.989	1.059	1.941	0.000
BIC weights	0.869	0.777	0.857	1.117	0.013	0.947	2.141	0.957	0.658	0.948
LOG Combinations										
Equal weights	0.884	0.724	0.882	1.509	0.000	0.926	2.101	0.955	1.471	0.000
Log score weights	0.869	0.760	0.868	1.568	0.000	0.918	2.141	0.935	1.354	0.000
CRPS weights	0.877	0.738	0.873	1.540	0.000	0.917	2.132	0.937	1.443	0.000
MSE weights	0.868	0.747	0.865	1.568	0.000	0.906	2.109	0.941	1.531	0.000
BIC weights	0.869	0.752	0.857	1.042	0.011	0.947	2.143	0.910	0.663	0.986
h=6										
	RMSE	Log Score	CRPS	RS Test	K Test	RMSE	Log Score	CRPS	RS Test	K Test
UCSV	2.897	-3.474	1.453	0.860	0.476	0.571	-3.278	0.324	0.894	0.513
LOP Combinations										
Equal weights	0.793	1.286	0.828	1.286	0.753	0.956	2.136	1.339	2.111	0.000
Log score weights	0.703	1.357	0.761	1.534	0.389	0.908	2.300	1.211	2.160	0.000
CRPS weights	0.761	1.333	0.793	1.402	0.614	0.925	2.268	1.197	1.787	0.000
MSE weights	0.739	1.337	0.775	1.418	0.432	0.873	2.174	1.303	2.200	0.000
BIC weights	1.130	0.132	1.210	1.944	0.068	1.120	1.723	1.224	2.336	0.063
LOG Combinations										
Equal weights	0.836	0.684	0.862	1.002	0.834	1.002	2.361	1.008	0.754	0.913
Log score weights	0.710	1.428	0.732	0.815	0.901	0.915	2.542	0.900	0.730	0.998
CRPS weights	0.809	0.846	0.823	0.887	0.846	0.963	2.410	0.962	0.911	0.933
MSE weights	0.792	0.875	0.803	0.847	0.822	0.935	2.408	0.943	0.713	0.902
BIC weights	1.130	-0.987	0.809	2.210	0.018	1.122	1.538	0.975	2.478	0.026
h=12										
	RMSE	Log Score	CRPS	RS Test	K Test	RMSE	Log Score	CRPS	RS Test	K Test
UCSV	1.952	-3.007	1.068	1.616	0.082	0.605	-2.901	0.350	1.237	0.779
LOP Combinations										
Equal weights	0.774	1.095	0.821	1.338	0.138	0.825	1.801	1.182	2.300	0.000
Log score weights	0.661	1.230	0.740	1.511	0.000	0.759	2.045	1.083	2.226	0.000
CRPS weights	0.717	1.198	0.760	1.283	0.006	0.774	2.003	1.039	1.796	0.000
MSE weights	0.673	1.200	0.742	1.512	0.019	0.667	1.899	1.141	2.286	0.000
BIC weights	0.968	-0.811	1.193	3.538	0.028	0.955	0.761	1.023	2.626	0.237
LOG Combinations										
Equal weights	0.852	0.661	0.879	1.580	0.516	0.903	1.860	0.915	1.220	0.730
Log score weights	0.675	1.397	0.683	1.158	0.672	0.739	2.385	0.691	0.812	0.681
CRPS weights	0.811	0.912	0.819	1.367	0.759	0.820	2.037	0.815	1.268	0.855
MSE weights	0.782	0.945	0.785	0.879	0.970	0.762	2.078	0.765	0.794	0.928
BIC weights	0.961	-1.586	0.812	3.663	0.008	0.956	0.425	0.802	2.690	0.169

Note: (i) RMSE - the ratio of the root mean square error (RMSE) of the various model combinations to that of the UCSV) model (ii) Log Score - difference in average log score relative to UCSV model over the evaluation period (iii) CRPS - ratio of average continuous rank probability score to UCSV benchmark over the evaluation period. Bold characters denote that the difference with respect to the benchmark model is significant at 10% critical level (using Giacomini and White (2004)'s test for the RMSE and the Kullback-Leibler information criterion (KLIC)-based test, which utilizes the expected difference in the log scores and CRPS of candidate densities; see Bao et al. (2007), Mitchell and Hall (2005) and Amisano and Giacomini (2007)). The last columns corresponding to 'RS Test' and 'K Test' denote the test statistics and p-value of Rossi and Sekhposyan (2019) and Knoppel(2015)'s tests on the calibration of the density. The entries highlighted in grey correspond to best performing model for each metric/horizon considered.

Figure E.3: INFLATION DENSITY FORECASTS: $h = 12$



Note: The green line plots observed inflation and the blue shaded area are 5 to 95% intervals of the forecast density, each with a 5 percent width.

Table E.2: COMPARISON TO SURVEY OF PROFESSIONAL FORECASTERS (CPI INFLATION)

	Headline (CPI) Inflation		Core (CPI) Inflation	
	$h = 2Q$	$h = 4Q$	$h = 2Q$	$h = 4Q$
SPF	2.453	2.481	0.595	0.597
Baseline	1.055	1.028	0.861	0.902

Note: The table reports the RMSE ratios of the LogOP log score weighted combination, relative to the SPF.

Table E.3: CALIBRATION AND EVALUATION OF SECTIONS OF DENSITY (CPI INFLATION)

		Quantile Scores					
		Headline (CPI) Inflation			Core (CPI) Inflation		
		Left	Center	Right	Left	Center	Right
h = 6	UCSV	0.351	0.408	0.517	0.070	0.097	0.103
	Baseline	0.721	0.787	0.682	0.894	0.922	0.880
h = 12	UCSV	0.262	0.313	0.394	0.073	0.105	0.099
	Baseline	0.638	0.814	0.577	0.609	0.614	0.655

		Density Calibration					
		Headline (CPI) Inflation			Core (CPI) Inflation		
		Left	Center	Right	Left	Center	Right
h = 6	UCSV	0.787	0.707	0.987	1.219	0.721	1.257
	Baseline	0.458	0.815	0.484	0.374	0.503	0.730
h = 12	UCSV	1.213	0.863	1.324	1.497	1.146	1.353
	Baseline	1.091	0.416	1.158	0.437	0.472	0.812

Note: The upper panel reports the density calibration test statistics of Rossi and Sekhposyan (2019). The lower panel reports the threshold weighted (tw) CRPS test statistic, as in Gneiting and Ranjan (2011). The first row refers to the *absolute* values for the UCSV benchmark, whereas the second row reports the ratios of baseline LogOP log score weighted to the UCSV benchmark model. In both panels, bold entries are significant at the 10% level.

Table E.4: SCALED BRIER SCORES FOR INFLATION EVENTS (CPI INFLATION)

		Headline (CPI) Inflation				Core (CPI) Inflation			
		Event: $\pi \leq 1\%$				Event: $\pi \leq 1.5\%$			
		BS	Unc.	Rel.	Res.	BS	Unc.	Rel.	Res.
h = 6	UCSV	1.447	1.000	0.549	0.102	1.135	1.000	0.379	0.244
	Baseline	0.957	1.000	0.082	0.125	1.064	1.000	0.225	0.161
h = 12	UCSV	1.365	1.000	0.452	0.087	1.420	1.000	0.566	0.146
	Baseline	0.803	1.000	0.176	0.373	0.884	1.000	0.248	0.364
		Event: $1\% \leq \pi \leq 3\%$				Event: $1.5\% \leq \pi \leq 2.5\%$			
		BS	Unc.	Rel.	Res.	BS	Unc.	Rel.	Res.
h = 6	UCSV	1.205	1.000	0.325	0.120	1.445	1.000	0.596	0.151
	Baseline	1.017	1.000	0.125	0.109	1.171	1.000	0.238	0.067
h = 12	UCSV	1.281	1.000	0.414	0.134	1.632	1.000	0.760	0.128
	Baseline	1.157	1.000	0.348	0.191	0.980	1.000	0.263	0.283
		Event: $\pi \geq 3\%$				Event: $\pi \geq 2.5\%$			
		BS	Unc.	Rel.	Res.	BS	Unc.	Rel.	Res.
h = 6	UCSV	1.057	1.000	0.254	0.197	1.377	1.000	0.511	0.134
	Baseline	0.943	1.000	0.121	0.178	1.083	1.000	0.191	0.108
h = 12	UCSV	1.321	1.000	0.561	0.240	1.839	1.000	1.018	0.179
	Baseline	1.065	1.000	0.229	0.164	1.271	1.000	0.576	0.306

Note: The test for correct reliability/calibration and for resolution follows that outlined in Galbraith and van Norden (2012). To test calibration, we estimate the model $x_i = a + b\hat{p}_i + c\hat{p}_i^2 + e_i$ where x_i is a zero/one indicator that the event happened and \hat{p}_i the estimated probability of the event. We jointly test $H_0 : a = 0, b = 1, c = 0$ with χ^2_3 distributed Wald statistic, using robust (Newey-West standard errors). The test of zero resolution is of $H_0 : b = 0$ and $c = 0$ with χ^2_2 Wald statistic. The entries highlighted in grey correspond to best performing model for each metric/horizon considered.

Table E.5: EVALUATION OF EX-ANTE INFLATION RISK

	Headline (CPI) Inflation				Core (CPI) Inflation			
	h = 6		h = 12		h = 6		h = 12	
	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 0.5$	$\kappa = 1$
Neutral								
$BR_{1,1}$	0.261	0.367	0.199	0.286	-0.010	0.005	0.034	0.068
DR_1	0.185	0.356	0.148	0.276	-0.016	-0.010	0.036	0.067
EIR_1	-0.045	-0.008	0.033	0.103	0.011	0.031	0.017	0.036
Inflation Adverse								
$BR_{1,2}$	0.505	0.874	0.421	0.818	-0.003	0.026	0.042	0.093
EIR_2	-0.077	0.420	0.138	0.562	0.014	0.042	0.031	0.065
Deflation Adverse								
$BR_{2,1}$	3.137	4.370	1.403	2.247	-0.040	-0.019	0.059	0.130
DR_2	1.348	2.900	0.800	1.576	-0.028	-0.011	0.050	0.106

Note: The Table reports the difference between the score statistics, as defined by equations (7), (8) and (9), of the baseline and UCSV benchmark models. A positive number indicates the baseline has a higher score (is preferred) and those in bold are significantly greater than zero at the 5% level of significance. For the neutral case we adopt $\alpha = \beta = 2$ and $a = 0.5$, in the inflation adverse case we set $\alpha = 2$, $\beta = 3$ and $a = 0.5$ and for the deflation adverse case $\alpha = 3$, $\beta = 2$ and $a = 0.5$.

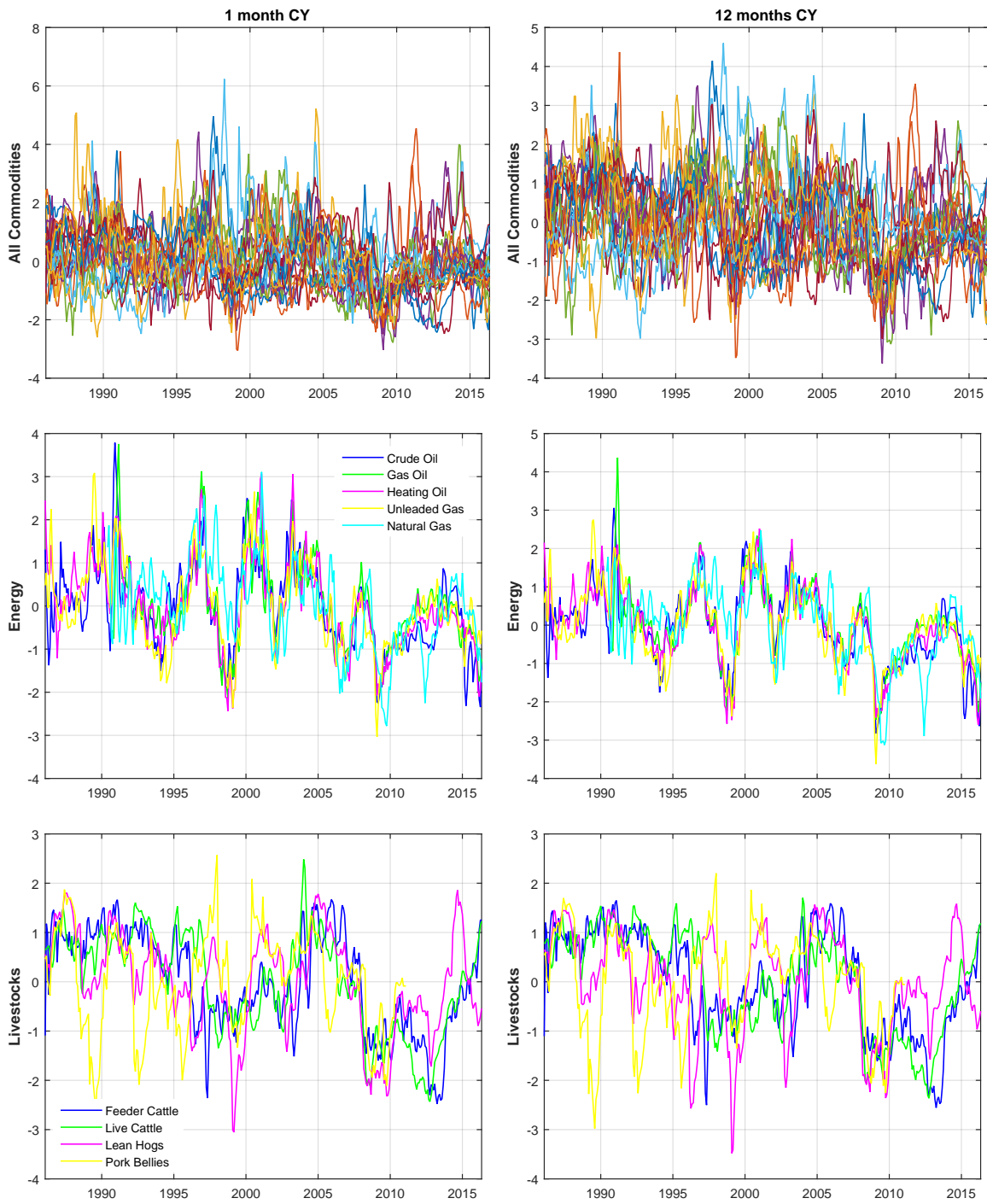
F Additional Tables and Figures

Table F.1: COMPARISON TO SURVEY OF PROFESSIONAL FORECASTERS

	Headline Inflation		Core Inflation	
	$h = 2Q$	$h = 4Q$	$h = 2Q$	$h = 4Q$
SPF	2.080	2.075	0.677	0.677
LOP Combinations				
Equal weights	1.061	1.031	0.903	0.943
Log score weights	0.995	0.954	0.874	0.905
CRPS weights	1.029	0.986	0.879	0.912
MSE weights	1.004	0.944	0.836	0.853
BIC weights	1.349	1.477	1.058	1.336
LOG Combinations				
Equal weights	1.073	1.055	0.937	1.011
Log score weights	1.004	0.967	0.896	0.954
CRPS weights	1.045	1.016	0.909	0.973
MSE weights	1.024	0.980	0.879	0.934
BIC weights	1.024	0.980	0.879	0.934

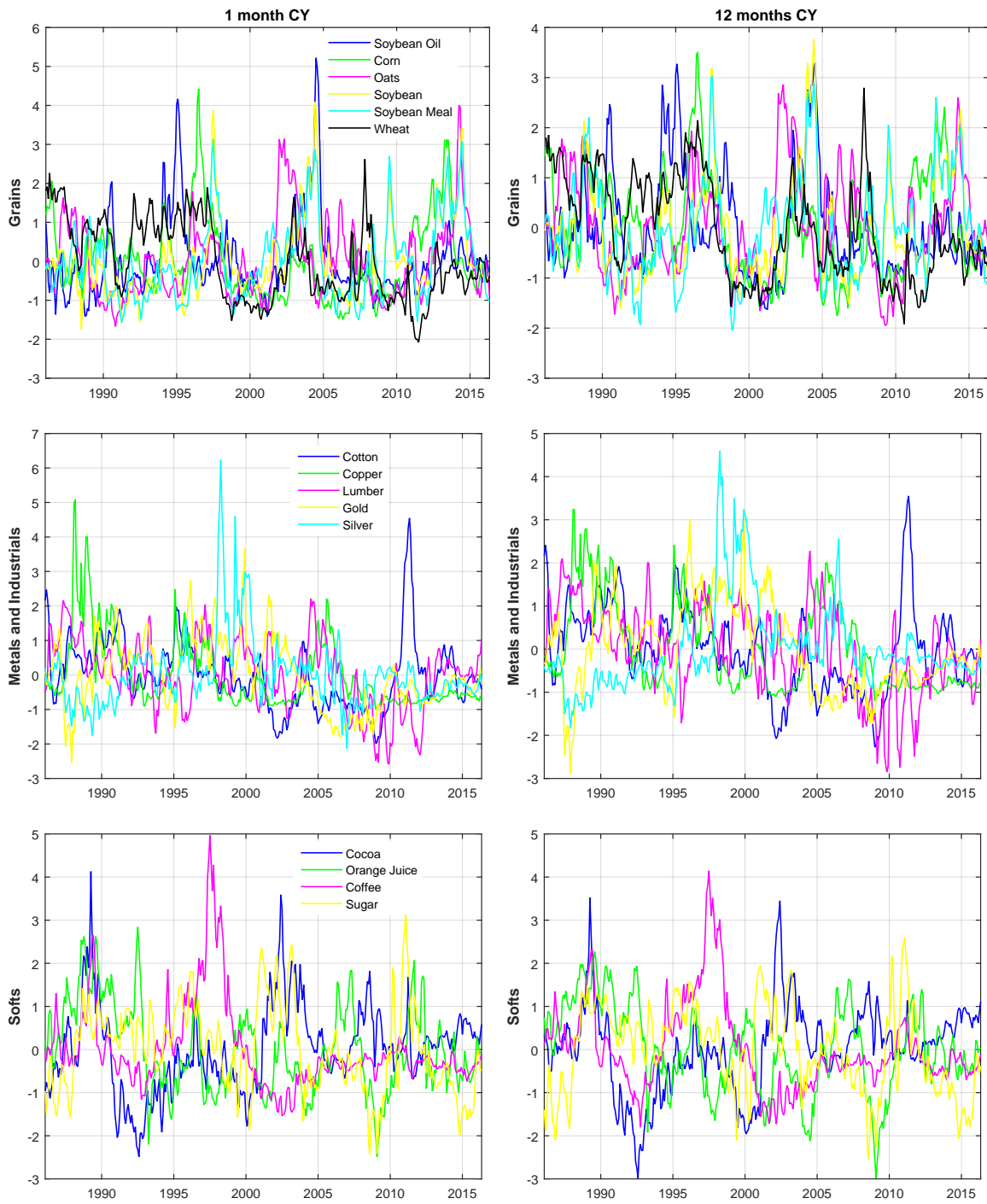
Note: The first row reports the *absolute* RMSE values of the SPF forecasts, and the second row the RMSE ratios of the LogOP log score weighted combination RMSE relative to the SPF.

Figure F.1: CONVENIENCE YIELDS BY COMMODITY GROUP



Note: Convenience yields (normalized) for each commodity group. The 1 month and 12 months convenience yields are calculated from synthetic futures prices as outlined in Appendix A.

Figure F.2: CONVENIENCE YIELDS BY COMMODITY GROUP



Note: Convenience yields (normalized) for each commodity group. The 1 month and 12 months convenience yields are calculated from synthetic futures prices as outlined in Appendix A.

Figure F.3: POSTERIOR PROBABILITY OF MODELS WITH 1M (vs 12M) CONVENIENCE YIELD

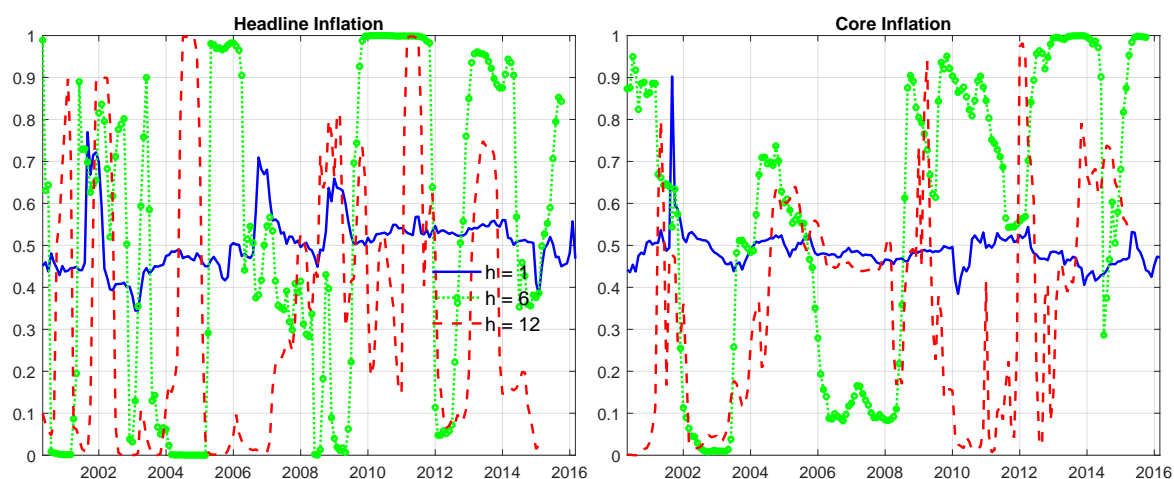


Figure F.4: POSTERIOR PROBABILITY OF AR MODELS (vs GAP)

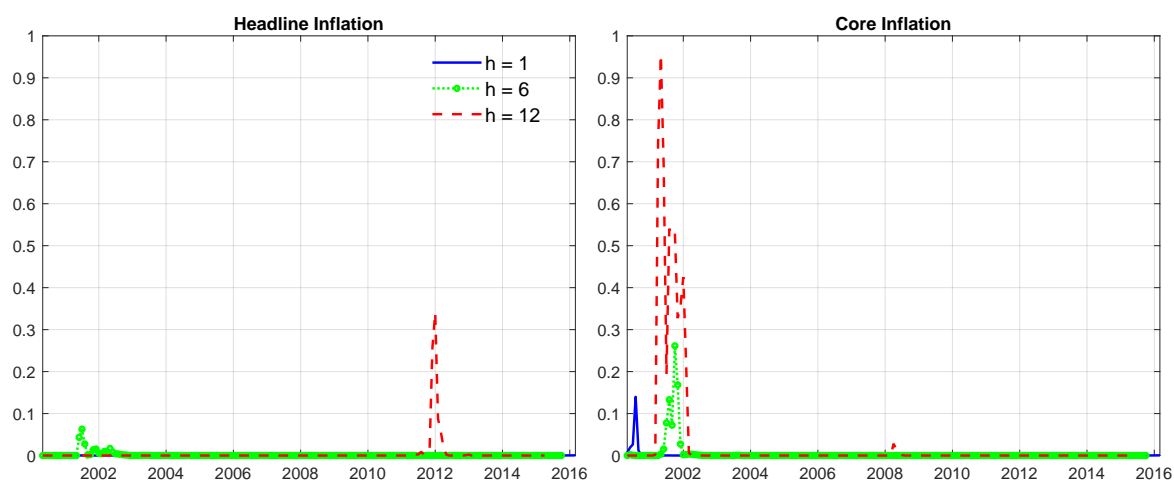


Figure F.5: PROBABILITY THAT COMMODITY PRICES PREDICT INFLATION: AR MODELS

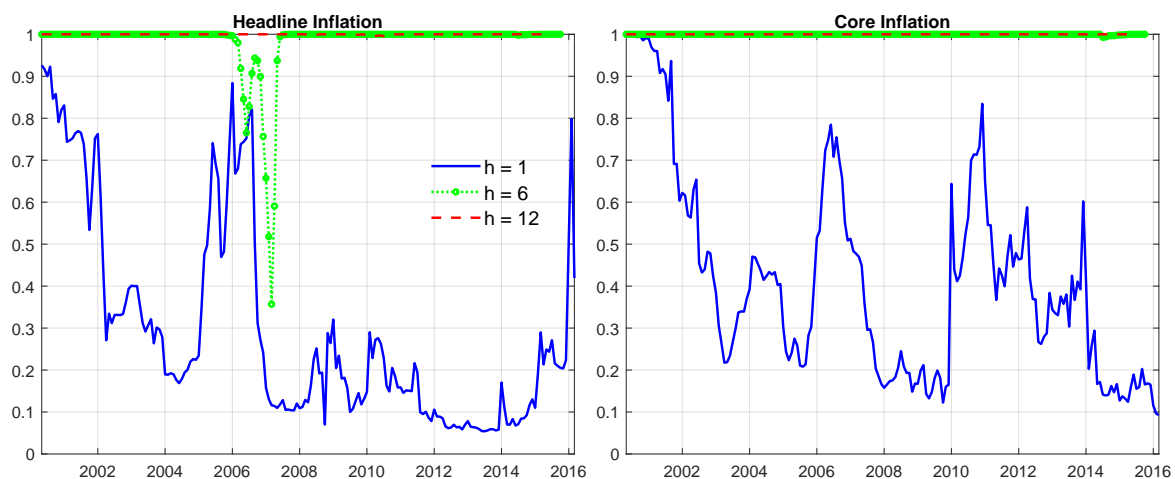


Figure F.6: PROBABILITY THAT COMMODITY PRICES PREDICT INFLATION: GAP MODELS

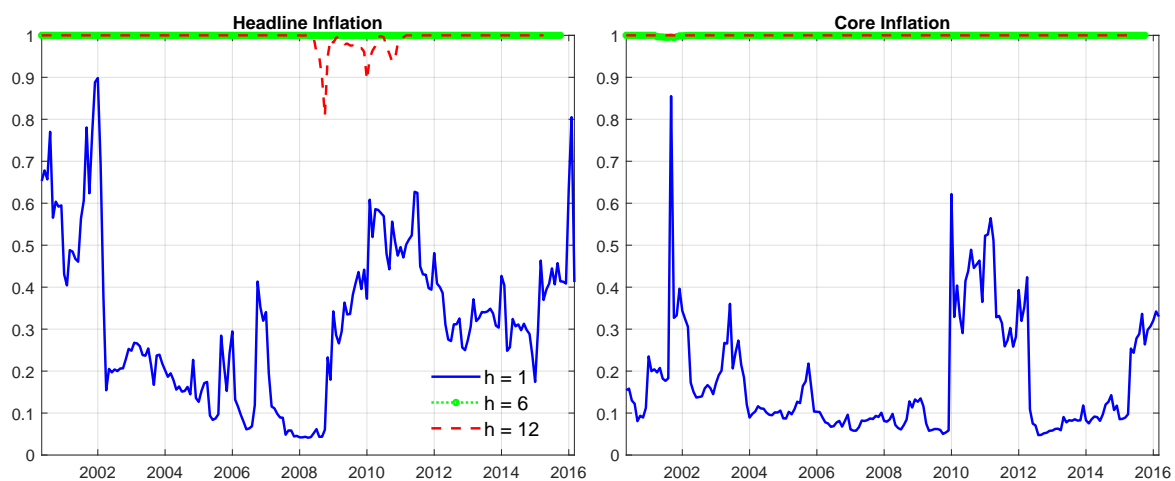


Figure F.7: MODEL WEIGHTS

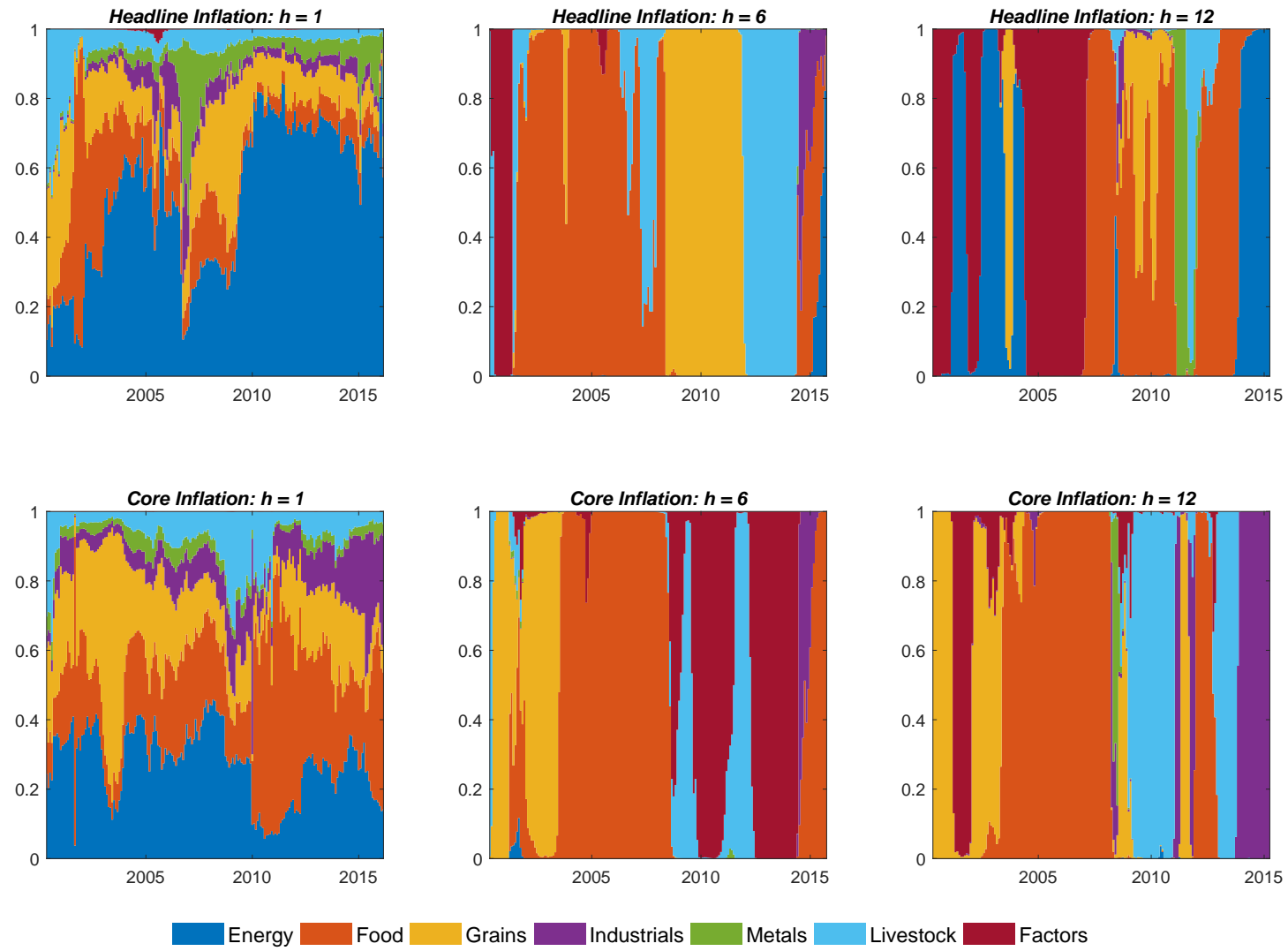
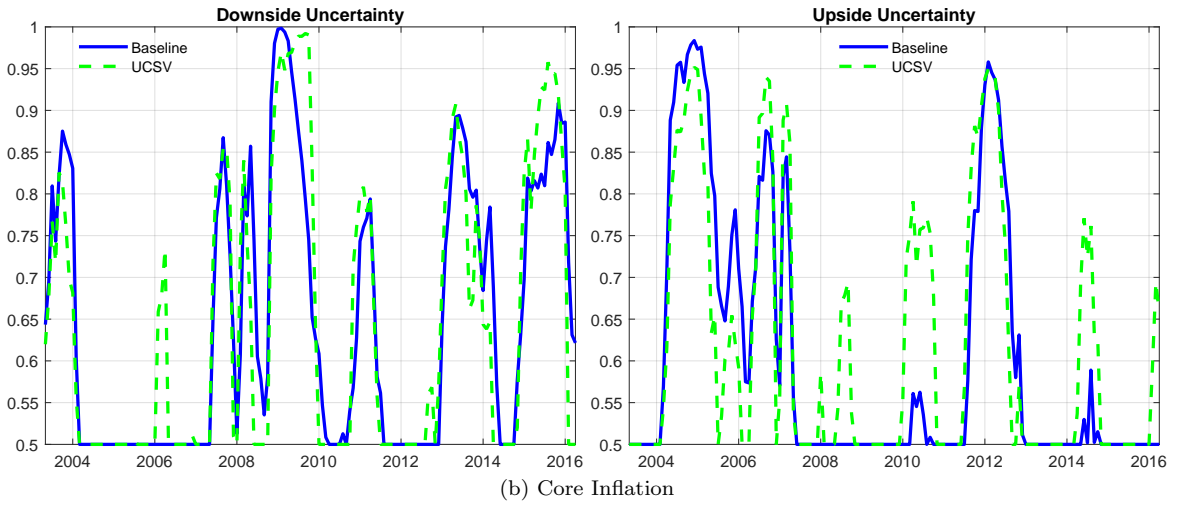
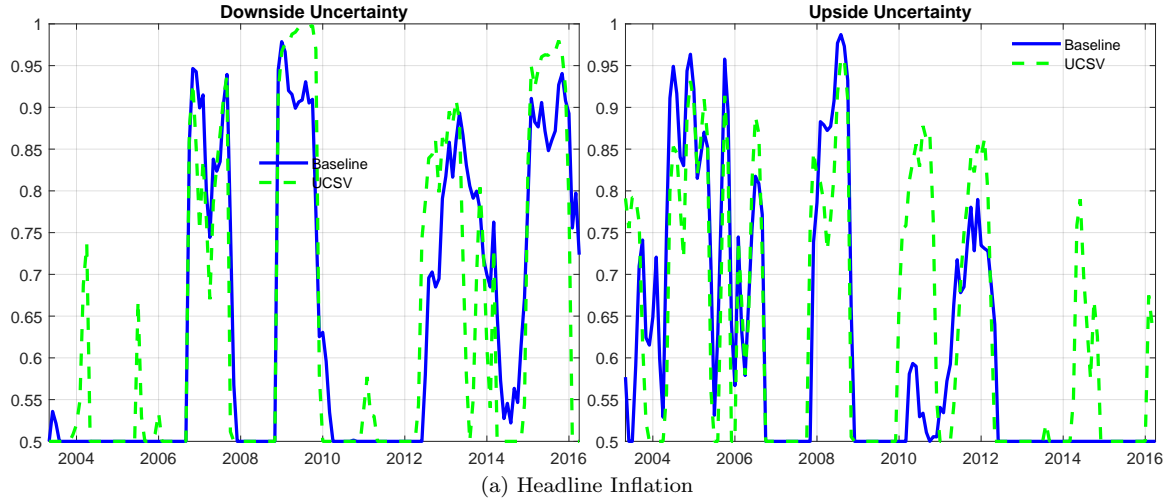


Figure F.8: UPSIDE/DOWNSIDE UNCERTAINTY: $h = 12$



Note: The Figure reports the upside and downside uncertainty decomposition for headline inflation (panel (a)) and core inflation (panel (b)). Those are calculated following Rossi and Sekhposyan (2015). In all panels, the blue line corresponds to the baseline model overage, whereas the green broken line corresponds to the autoregressive models.