# STRATHCLYDE

## DISCUSSION PAPERS IN ECONOMICS

# REAL-TIME DENSITY NOWCASTS OF US INFLATION: A MODEL-COMBINATION APPROACH

## BY

## EDWARD S KNOTEK AND SAEED ZAMAN

## NO 20-15

## DEPARTMENT OF ECONOMICS
## UNIVERSITY OF STRATHCLYDE, GLASGOW

# Real-Time Density Nowcasts of U.S. Inflation: A Model-Combination Approach[*]

Edward S. Knotek II[1]

Saeed Zaman[2]

Original version: May 29, 2019
This version: October 20, 2020

## Abstract

We develop a flexible modeling framework to produce density nowcasts for U.S. inflation at a trading-day frequency. Our framework: (1) combines individual density nowcasts from three classes of parsimonious mixed-frequency models; (2) adopts a novel flexible treatment in the use of the aggregation function; and (3) permits dynamic model averaging via the use of weights that are updated based on learning from past performance. Together these features provide density nowcasts that can accommodate non-Gaussian properties. We document the competitive properties of the nowcasts generated from our framework using high-frequency real-time data over the period 2000-2015.

Keywords: mixed-frequency models, inflation, density nowcasts, density combinations

JEL classifications: C15, C53, E3, E37

## 1. Introduction

Inflation developments are of interest to policymakers, forecasters, financial market participants, and the general public. This interest includes not only the point forecast but also the range of potential inflation outcomes and their probability of occurring—i.e., the density forecast. Building on the literature that finds that the accuracy of multistep point forecasts can be improved by conditioning on high-quality point nowcasts, Krüger, Clark, and Ravazzolo (2017) and Tallman and Zaman (2020) document that conditioning quarterly macroeconomic models with both nowcast means and nowcast densities leads to improvements in the accuracy of multistep point and density forecasts, especially for inflation.[3] Realizing these gains in practice requires relatively accurate nowcast means and nowcast densities for inflation. Previous research by Modugno (2013), Monteforte and Moretti (2013), Breitung and Roling (2015), Knotek and Zaman (2017), and Clement (2017) has developed mixed-frequency approaches to nowcast U.S. inflation, with an exclusive focus on point nowcast accuracy. In this paper, we develop a flexible framework that uses model-combination strategies with three classes of mixed-frequency models to generate highly accurate point and density nowcasts for U.S. inflation.

The past two decades have seen considerable growth in the density forecasting and density nowcasting literature.[4] Density forecasts help to illuminate the balance of risks around point forecasts (e.g., Rossi, 2014; Mazzi, Mitchell, and Montana, 2014) and strengthen the "credibility" of forecasts (Wright, 2019). However, recent work on density nowcasting has focused on real GDP growth and other indicators of real economic activity. The paper that comes closest to the notion of inflation density nowcasting is Garratt, Mitchell, and Vahey (2014), but there the nowcasts are one-step-ahead forecasts from models estimated with quarterly data; intra-quarterly daily, weekly, and monthly data are not used, and hence the density nowcast estimates are largely unchanged during the quarter.

We contribute to the density nowcasting literature by proposing a flexible framework to produce both point and density nowcast estimates for U.S. headline and core inflation measures:

---

[3] E.g., Faust and Wright (2013) and Knotek and Zaman (2019) find that conditioning quarterly macro models with more accurate jumping-off points from external nowcasts improves multistep forecast accuracy.

[4] See Tay and Wallis (2000) for a survey on density forecasting, including its application in macroeconomics and finance, and Aastveit et al. (2018) for a more recent survey on density forecasting and density combinations.

CPI inflation, core CPI inflation, PCE inflation, and core PCE inflation. Our flexible framework is based on model combinations across three classes of mixed-frequency models that previous research has shown produce high-quality point nowcasts for inflation; for one of the model classes, we develop a procedure to generate the density nowcasts. By using mixed-frequency models, we can produce inflation density nowcasts at a trading-day frequency that take advantage of high-frequency data and update as information accumulates over the course of a month or a quarter.

We use model combinations because the characterization of uncertainty from a single (possibly misspecified) model could be too restrictive. In addition, combining density estimates across a range of models provides a flexible density that can potentially accommodate non-Gaussian features such as skewness and kurtosis, which may more closely approximate the true density.[5] Inspired by previous research on density forecast combinations (e.g., Bache et al., 2011; Aastveit et al., 2014; Garratt, Mitchell, and Vahey, 2014), we combine the density nowcasts in a two-stage procedure.[6] In stage 1, density nowcasts coming from different model specifications within each of the three model classes are combined. In stage 2, we combine across the three stage 1 combinations to form a "grand" combination.

Combining densities requires a functional form for the aggregation and weights to apply to the different densities. Previous research has used either the linear opinion pool or the logarithmic opinion pool as the functional form for aggregation, with some researchers using both methods and presenting results for the approach that is more accurate over some evaluation period. Instead of enforcing a particular functional form for aggregation at the outset, we devise and implement a novel flexible aggregation strategy that lets the data dynamically determine which of the two functional forms it prefers. A potential advantage of this flexibility is that it allows for the possibility of switching between the two functional forms, at different points within a month or a quarter based on the nowcast origin or at different points in time during the

---

[5] Alternatively, one could flexibly characterize uncertainty using a single mixed-frequency model featuring stochastic volatility and estimate the model with Bayesian methods, as in Carriero, Clark, and Marcellino (2015) for real GDP growth or Koop, McIntyre, and Mitchell (2020) for U.K. regional indicators.
[6] Our general strategy for combining densities from many models is similar in spirit to the approach in Aastveit, et al. (2014) for density nowcasting GDP, but with many differences in implementation. Chernis and Sekkel (2018) employ a similar two-stage procedure to produce point nowcast combinations for real economic indicators.

sample as more observations become available. This novel flexible aggregation strategy can be used broadly in multistep forecasting applications when combining point or density forecasts.

The literature has considered a variety of weighting schemes to use when combining density forecasts, in part because no single scheme has been shown to work "best" under all circumstances. The bulk of the density combination literature has considered a limited number of weighting schemes—although Krüger (2015) and Ganics (2017) are exceptions—with the scheme based on recursive updating of past performance using the log score metric being the most popular. In contrast, we consider a relatively large number of weighting schemes, ranging from equal weights to schemes based on past predictive performance to "optimal" schemes that optimize some loss function over a historical sample. In most cases, these weights dynamically update over time to learn from past performance.

Using high-frequency real-time data over the evaluation period 2000-2015, we conduct a comprehensive set of out-of-sample inflation density nowcasting exercises to assess our flexible framework using a variety of inflation measures, inflation rates, weighting strategies, and mixed-frequency model classes. This examination reveals that combining individual densities generally helps improve density nowcast accuracy, and as information accumulates over the course of a month or a quarter, the accuracy of the combined density nowcasts and the associated point nowcasts steadily improves. We also document evidence of dynamic model switching, which highlights the importance of combining estimates from a range of models to circumvent the instability issues from a single model. But it matters how the densities are combined: not all combination methods improve accuracy compared with the best-performing individual densities. The grand combinations based on our flexible aggregation strategy and the log score weighting scheme, which relies on past predictive performance, or the "optimal" weighting scheme of Conflitti, De Mol, and Giannone (2015) are among the best performing in terms of relative accuracy for headline inflation and are well calibrated. The Ganics (2017) weighting scheme, which optimizes the calibration fit, produces the best calibrated densities for headline inflation, but its relative accuracy is inferior to the log score or Conflitti, De Mol, and Giannone (2015) weighting schemes. In the case of core inflation, all weighting schemes generate comparable accuracy of point and density nowcasts. Overall, the accuracy of the implied point nowcasts

4

from the grand combination matches the accuracy of the best performing mixed-frequency model of Knotek and Zaman (2017) and is more accurate in the case of core PCE inflation.

Our empirical results indicate evidence of both time-varying skewness and kurtosis in the predictive densities of inflation measures, suggesting that asymmetries and fat tails are an empirical feature of inflation data. Combination methods that produce density estimates derived from a richer set of models tend to display a higher degree of skewness and kurtosis in the predictive densities. We also find evidence of time-varying variances (i.e., uncertainty in the nowcast estimates) that echo the broader patterns reported elsewhere in the inflation uncertainty literature using stochastic volatility models (e.g., Carriero, Clark, and Marcellino, 2019; Knotek, Zaman, and Clark, 2015).

Finally, we conduct a horse race with the Survey of Professional Forecasters (SPF) for point and density nowcast accuracy. Our grand combination's density nowcasts provide superior point and density nowcasts for CPI inflation and PCE inflation. For core CPI inflation and core PCE inflation, our grand combination's nowcasting performance is competitive with the SPF. The ability of our proposed framework to generate highly accurate point and density nowcasts of inflation is a useful outcome for practitioners.

The paper proceeds as follows. Section 2 describes the mixed-frequency models. Section 3 discusses the combination methods to combine individual densities. Section 4 describes the real-time data. Section 5 discusses the nowcast evaluation strategy. Section 6 presents empirical results. Section 7 compares the accuracy of the combined density nowcasts to SPF. Section 8 concludes.

## 2. Mixed-Frequency Models

Building on the literature that has shown that relatively parsimonious approaches dominate more sophisticated approaches for nowcasting and near-term forecasting of inflation (e.g., Koop and Korobilis, 2012; Knotek and Zaman, 2017), we consider three classes of mixed-frequency models that relate aggregate inflation to its components and to a limited number of

other indicators.[7] We briefly discuss the models and procedures for constructing their density nowcasts here; the supplementary appendix provides further details.

## 2.1 Deterministic Model Switching (DMS)

Knotek and Zaman (2017) construct a mixed-frequency model for inflation point nowcasting that relies on a small number of variables and combines univariate and multivariate regressions. The model uses disaggregate and aggregate variables to construct nowcasts for the aggregate, but the disaggregate information is used only if it is available and deemed useful.[8] The latter aspect gives rise to time-varying coefficients that change in a deterministic fashion based on the available information set, which we label as deterministic model switching (DMS).

Monthly inflation rates $\pi_t$ are modeled via a general representation:

$$A_{s(t)} \, \boldsymbol{Z_t} = B_{s(t)} + C_{s(t)} \boldsymbol{X_t} + \sum_{j=1}^{J} D_{j,s(t)} \boldsymbol{Z_{t-j}} + \varepsilon_{s(t)}. \tag{1}$$

The coefficient matrices $A$, $B$, $C$, and $D_j$ can potentially vary over time with the information set $s(t)$. For headline inflation, $\boldsymbol{Z}_t$ is a vector of aggregates—CPI inflation, $\pi_t^{\text{CPI}}$, and PCE inflation, $\pi_t^{\text{PCE}}$—and $\boldsymbol{X}_t$ is a vector of disaggregate components comprising gasoline inflation ($\pi_t^{\text{Gasoline}}$), food inflation ($\pi_t^{\text{Food}}$), and core inflation rates ($\pi_t^{\text{Core CPI}}$ and $\pi_t^{\text{Core PCE}}$). To illustrate the type of deterministic model switching that occurs, for example, the nowcast for $\pi_t^{\text{PCE}}$ in month $t$ is a function of actual $\pi_t^{\text{CPI}}$ in month $t$ via $A_{s(t)}$, if $\pi_t^{\text{CPI}}$ is available from statistical agencies; if not, the nowcast for $\pi_t^{\text{PCE}}$ in month $t$ is a function of disaggregates' nowcasts included in $\boldsymbol{X}_t$ via $C_{s(t)}$. If the vector of disaggregates is incomplete for month $t$, then $C_{s(t)}=0$. High-frequency data on gasoline and oil prices are used to construct a gasoline inflation nowcast via an auxiliary model, while food inflation nowcasts are derived using a univariate AR specification. Core inflation rates can also be modeled using equation (1), where no disaggregates are used and nowcasts are formed either via univariate AR models or bridge regressions if $\pi_t^{\text{Core CPI}}$ is available for month $t$

---

[7] These model classes have also been applied in other contexts, such as nowcasting GDP growth. We restrict attention to models that are well understood by econometricians and economic forecasters, but one could extend the model pool to include a larger and broader set of model classes, including machine learning models.
[8] Hendry and Hubrich (2011) and Ravazzolo and Vahey (2014) focus on multistep inflation forecasting using disaggregates.

while $\pi_t^{\text{Core PCE}}$ is not. This mixed-frequency model switches between univariate and multivariate regressions depending on the available information within a month or a quarter.

We innovate on the DMS approach for constructing inflation point nowcasts in two ways. First and foremost, we devise and implement parametric block wild bootstrap algorithms to produce density nowcasts for the DMS framework. When the DMS selects the multivariate regression model that uses disaggregates, the density nowcasts are constructed by (1) constructing density estimates for each of the three disaggregates (core inflation, food inflation, and gasoline inflation); and (2) combining the density estimates using the weights in $C_{s(t)}$ to construct the density nowcast for aggregate inflation, similar to the combination approaches in Ravazzolo and Vahey (2014) and Tallman and Zaman (2017) but in this case with an application to nowcasting. While Knotek and Zaman (2017) estimate the model using short rolling windows, which leads to very flexible parameters and incorporates changing volatility in a parsimonious way, we consider density combinations that allow for a variety of rolling or expanding estimation windows, as discussed below.

## 2.2 Mixed Data Sampling (MIDAS)

Monteforte and Moretti (2013) generate inflation point nowcasts using a MIDAS model with leads, which is a reduced-form regression relating a low-frequency variable to high-frequency variable(s).[9] In our application, oil and gasoline prices act as the high-frequency leads, while monthly inflation is at a lower frequency. To prevent parameter proliferation, MIDAS works with distributed lag polynomial operators that reduce the estimation to a smaller number of parameters. The model is estimated using nonlinear least squares.

The MIDAS model with leads for inflation at time $t+h$, $\pi_{t+h}$, takes the form

$$\pi_{t+h} = \alpha_{(h)} + \sum_{J=0}^{P(M)-1} \chi_{j+1,(h)} \, \pi_{t-j} + \sum_{J=0}^{P(M)-1} \gamma_{j+1,(h)} \, Z_{t-j} +$$

$$\beta_h \sum_{J=0}^{P(HF)-1} \omega_{P(HF)-j} \left( \theta_{(h)}^{HF} \right) X_{P(HF)-j,t+1}^{HF} + e_{t+h} \qquad (2)$$

---

[9] Ghysels, Santa-Clara, and Valkanov (2005, 2006) popularized these models; Clements and Galvão (2008) is an influential paper on the application of MIDAS to macroeconomic forecasting. These models are increasingly used for nowcasting macroeconomic indicators across the globe (e.g., see Allan et al., 2014, for the economy of Scotland).

where *Z* includes other monthly variables; *P(M)* is the number of lags of the monthly regressors (we use 1); and *P(HF)* is the number of high-frequency observations, $X^{HF}_{1,t+1},....., X^{HF}_{P(HF),t+1}$ in month *t*+1 (i.e., the target nowcast month). The coefficients are independently estimated for each forecast horizon (*h*).[10] The assumption $\sum_{J=0}^{P(HF)-1} \omega_{P(HF)-j}\left(\theta^{HF}_{(h)}\right) = 1$ helps identify $\beta_h$.

We extend Monteforte and Moretti's (2013) work on inflation point nowcasts by generating inflation density nowcasts from MIDAS models. These density nowcasts are constructed by drawing errors from a normal distribution with a standard deviation coming from past residuals after rescaling to correct the variance. Aastveit, Foroni, and Ravazzolo (2017) show that this approach is slightly inferior to the block wild bootstrap in their application when nowcasting GDP, but it provides a substantial computational advantage in our exercises.

## 2.3 Dynamic Factor Model (DFM)

Building on Giannone, Reichlin, and Small (2008), mixed-frequency dynamic factor models (DFMs) are widely used for nowcasting. Modugno (2013) uses a DFM to generate inflation point nowcasts from a data set comprising monthly, weekly, and daily data by extracting a common factor at a daily frequency via the estimation method of Bańbura and Modugno (2014). The DFM takes the form

$$y_t = Cf_t + \varepsilon_t \, , \, \varepsilon_t \sim N(0,\Sigma) \tag{3}$$

where $y_t$ is a vector of observations, *C* is a matrix of loadings, $\varepsilon_t$ is a vector of idiosyncratic components, and $f_t$ is a vector of unobserved common components that follows

$$B f_t = A(L)f_{t-1} + \mu_t, \mu_t \sim N(0,Q) \tag{4}$$

where *B* and *A(L)* are coefficient matrices that capture factor dynamics. The estimated latent daily factor(s) aggregate to weekly and monthly factors, which are used to construct nowcast estimates for monthly variables, including inflation.

We extend the previous work using DFMs for inflation point nowcasts in order to generate density nowcasts for U.S. inflation. Our density nowcasts are constructed using a standard parametric bootstrapping procedure for factor models, similar to Aastveit et al. (2014). This bootstrapping procedure accounts for factor, parameter, and shock uncertainty.

---

[10] In our exercises, *h* ranges from 1 to 2 for nowcasting monthly inflation and from 1 to 4 for nowcasting quarterly inflation.

## 2.4 Mixed-Frequency Model Space

Each of the three mixed-frequency model classes requires assumptions about certain elements, such as the size of the rolling windows in the DMS model, the polynomial specification in the MIDAS model, or the number of lags of the factors in the DFM. To account for this uncertainty within each model class, we consider many different specifications, listed in Table 1. We employ 132 model variants distributed unequally across our three model classes, with 108 in the DMS class, 12 in the MIDAS class, and 12 in the DFM class.[11] We combine density nowcasts within each model class in "stage 1" combinations, and we then combine the stage 1 combinations into a "grand" combination in stage 2.

## 3. Combination Methods

Simple combinations of point forecasts have a long history (see Bates and Granger, 1969) and often perform well, even when compared with weighting schemes that minimize some loss function (e.g., Clark and McCracken, 2010). In contrast, the density forecast combination literature has generally documented performance gains from optimal weighting schemes and schemes based on past predictive performance over the use of equal weights.[12] Combining candidate density estimates requires both a functional form to use in combining the densities and a mechanism for deriving the weights to place on each density. We present a novel functional form to aggregate our candidate density nowcasts and consider a variety of weighting schemes in this context. To be clear about our approach, in stage 1 we use a particular functional form and weighting scheme to combine the individual densities within a model class, for each of the three model classes; and then in stage 2, we use the same functional form and weighting scheme to combine the densities from the three model classes into the grand combination.

---

[11] In preliminary results, we considered 36 specifications for the MIDAS model class (12 each for Beta, BetaNN, and Almon polynomials). This combination performed similarly to the combination in the paper, but the latter approach greatly reduced computing time. For the DFM class we initially used combinations with one or two factors, giving us 24 specifications. This combination slightly underperformed the combination reported in the paper using only one factor, consistent with DFM inflation point nowcasting accuracy findings in Modugno (2013).

[12] For some examples, see Hall and Mitchell (2007), Jore, Mitchell, and Vahey (2010), Bache et al. (2011), Bjornland et al. (2011), and Aastveit et al. (2014). Kascha and Ravazzolo (2010) and Garratt, Mitchell, and Vahey (2011) provide counterexamples.

### 3.1 Functional Forms for Aggregation

We consider three functional forms or aggregation methods: the linear opinion pool, the logarithmic opinion pool, and a novel flexible method that combines the previous two functional forms in a data-dependent way.

The linear opinion pool is the weighted linear combination of individual component densities and is widely used (e.g., Bache et al., 2011; Aastveit et al., 2014; Mazzi, Mitchell, and Montana, 2014; Ravazzolo and Vahey, 2014). If there are $M$ models, then

$$p_{\tau,t,h}^{LIN}(y_t) = \sum_{i=1}^{M} w_{\tau,t,i,h} f_{\tau,t,i,h}(y_t \mid I_{\tau,t,i}) \tag{5}$$

where $p_{\tau,t,h}^{LIN}(y_t)$ is the combined linear pool predictive density for variable $y$ at the point in time $\tau$ within month $t$ for forecast horizon $h$.[13] The density forecast from model $i$, $f_{\tau,t,i,h}(y_t \mid I_{\tau,t,i})$, is conditional on information set $I_{\tau,t,i}$ which can differ across models. The potentially time-varying, nonnegative weights $w_{\tau,t,i,h}$ are recursively updated at each forecast origin based on some criteria that we discuss below and sum to 1.

This linear form of aggregation implies that if all the individual component densities are distributed normally with different mean and variance, then the combined density will be nonnormal (or a mixture-normal). The larger the number of individual densities, the more flexible the resulting density obtained from the combination. The nonnormal characteristic for the combined density via the linear pool is desirable if the unknown true density is nonnormal. An advantage of a mixture-normal distribution is that it permits skewness and kurtosis.

The logarithmic opinion pool is the geometric weighted average of the individual component densities:

$$p_{\tau,t,h}^{LOG}(y_t) = \frac{\prod_{i=1}^{M} f_{\tau,t,i,h}(y_t \mid I_{\tau,t,i})^{w_{\tau,t,i,h}}}{\int \prod_{i=1}^{M} f_{\tau,t,i,h}(y_t \mid I_{\tau,t,i})^{w_{\tau,t,i,h}} dy_t} \tag{6}$$

where $p_{\tau,t,h}^{LOG}(y_t)$ is the combined log pool predictive density for variable $y$ at the point in time $\tau$ within month $t$ for forecast horizon $h$. (To economize on notation, we omit $h$ subscripts going forward.) Importantly, the combination based on this geometric functional form assigns a zero

---

[13] As is common in nowcasting applications, we produce and evaluate nowcasts at different points in time within a given month; we use the notation $\tau$ to capture the former and $t$ to capture the latter.

probability to a region if any single individual density assigns a zero probability to that region. This may be an undesirable feature because a single incorrectly specified density can significantly influence the specification of the combined density (see Bjornland et al., 2011).[14]

In this paper, we propose a novel flexible aggregation method that is in the spirit of "letting the data speak" about which of the two functional forms—the linear pool or the logarithmic pool—is preferred.[15] Specifically, instead of taking a stand on a particular functional form at the outset, we allow flexibility in letting the data determine which of the two functional forms is preferred at every point in time. A potential advantage of this flexibility is that it allows for the possibility of dynamically switching between the two functional forms, at different points in time in the sample ($t$ in our notation above) and at different points in time within a month or a quarter ($\tau$ in our notation above) to take advantage of the differing information sets and the high-frequency data flow for nowcasting applications. For example, at the beginning of a month, when uncertainty around the point nowcast for that month would be expected to be higher, since much of the underlying source information is not yet available, the data may prefer the linear opinion pool, while later in the month the data may prefer the log opinion pool. This flexible aggregation method could be applied in more general multistep forecasting applications where the functional form is allowed to vary based on the forecasting horizon.

We implement this flexible aggregation method by determining, for each point in time $\tau$ within each target month to be nowcasted $t$, which of the two functional forms has historically produced more accurate densities. We denote the start and end of the sample by $T_0$ and $T$, respectively, and we let $D_\tau$ denote the normal data release lags (in number of months) at the point in time $\tau$, which captures the delay in calculating the historical density accuracies.[16] We initialize the flexible functional form by using the linear pool, i.e., if $t \leq T_0 + D_\tau - 1$,

$$functional\ form_{\tau,t} = Linear\ Pool \tag{7}$$

Subsequently,

---

[14] Figure A1 and Figure A2 in the appendix visually contrast the properties of the linear and log opinion pools.
[15] Knüppel and Krüger (2019) and Garratt, Henckel, and Vahey (2019) propose approaches to modify the linear opinion pool but do not consider the hybrid approach that we pursue.
[16] In our empirical application, the delay parameter $D$ is dependent on $\tau$, the position within the month when the nowcast is being made. Because we use the third monthly inflation release as the "true" inflation reading for a particular month, the delay $D=4$ early in the month when nowcasting monthly inflation, because the most recent "true" inflation reading comes from four months prior to the current month being nowcasted. The delay $D=3$ late in the month, however, as the statistical agencies have released prior months' inflation numbers by that point.

$$functional\ form_{\tau,t} = \begin{cases} Linear\ Pool & if\ LIN_{\tau,t} \geq LOG_{\tau,t} \\ Logarithm\ Pool & if\ LIN_{\tau,t} < LOG_{\tau,t} \end{cases} \tag{8}$$

for $t = T_0 + D_\tau, ..., T$, with

$$LIN_{\tau,t} = \frac{\sum_{T_0}^{t-D_\tau} \log\left( p_{\tau,t}^{LIN}(y_t = y_t^o) \right)}{t - D_\tau - T_0 + 1} \tag{9}$$

$$LOG_{\tau,t} = \frac{\sum_{T_0}^{t-D_\tau} \log\left( p_{\tau,t}^{LOG}(y_t = y_t^o) \right)}{t - D_\tau - T_0 + 1} \tag{10}$$

where the values for $p^{LIN}(\cdot)$ and $p^{LOG}(\cdot)$ are computed from equation (5) and equation (6), respectively, and $y_t^o$ is the observed value in month $t$, which is independent of the point in time when the nowcast was made ($\tau$) and the nowcast or forecast horizon ($h$).

## 3.2 Weighting Schemes

We consider five weighting schemes, ranging from equal weights to schemes based on average past predictive performance to schemes based on optimization of a specific loss function.

1. Equal weights. Simply, each density $i$ gets a weight

$$w_{\tau,t,i} = 1 / M \tag{11}$$

in the construction of the combination density nowcast. Aastveit et al. (2018) characterize such a combination as a "restrictive finite mixture."

2. Log-score weights. The logarithmic score (log-score) is the logarithm of the density forecast evaluated at the observation and is widely used to assess the accuracy of density forecasts. Accordingly, it makes sense to derive weights based on the past nowcast performance of the candidate densities using the log-score metric, with more accurate candidate densities receiving larger weights.[17] The weights $w_{\tau,t,i}$ are computed by averaging the past predictive

---

[17] This approach amounts to "learning from past mistakes" and is widely used in the density combination literature due to its simplicity (e.g., Gerard and Nimark, 2008; Jore, Mitchell, and Vahey, 2010; Kascha and Ravazzolo, 2010; Bjornland et al., 2011; Garratt et al., 2011; Aastveit et al., 2014; Beckmann et al., 2020). While we use the entire expanding history, we also explored the strategy of computing the weights over rolling 12-month periods, to "learn from recent mistakes" in computing the average score. The density nowcasting accuracy results were similar. We thank Gary Koop for suggesting this exercise.

performance using an expanding window for nowcast evaluation. We initialize the weights by setting them to equal weights (equation 11) if $t \le T_0 + D_\tau - 1$. For $t = T_0 + D_\tau, \ldots, T$:

$$w_{\tau,t,i} = \frac{\exp\left[\sum_{T_0}^{t-D_\tau} \log f_{\tau,t,i}(y_t = y_t^o)\right]}{\sum_{j=1}^{M} \exp\left[\sum_{T_0}^{t-D_\tau} \log f_{\tau,t,j}(y_t = y_t^o)\right]} \tag{12}$$

3. Continuous ranked probability score (CRPS). The CRPS is a widely used alternative metric for assessing density forecasts that is also based on past predictive performance. It is popular because it is more robust to outliers compared with the log-score metric and it rewards densities that have probability mass closer to the actual observation. The CRPS score for density nowcast $i$ at the point in time $\tau$ within month $t$, $CRPS_{\tau,t,i}$, is the squared difference between the CDF of the density forecast and the CDF of the actual realizations:

$$CRPS_{\tau,t,i} = \int_{-\infty}^{+\infty} (F_{\tau,t,i}(y_t) - 1\{y_t \ge y_t^o\})^2 \, dy \tag{13}$$

where $F_{\tau,t,i}(y_t)$ is the CDF of density nowcast $i$, $f_{\tau,t,i}(y_t)$, and $1\{y_t \ge y_t^o\}$ is the indicator function equal to 1 if $y_t \ge y_t^o$ and 0 otherwise.[18] The smaller the difference between the two cumulative distributions, the more accurate is the density nowcast, and the more weight that density nowcast $i$ should receive, such that weights depend on the inverse CRPS. We initialize the weights by setting them to equal weights $w_{\tau,t,i} = 1/M$ if $t \le T_0 + D_\tau - 1$. For $t = T_0 + D_\tau, \ldots, T$, the weights are computed based on an expanding window of historical predictive performance:[19]

$$w_{\tau,t,i} = \frac{\left[\sum_{T_0}^{t-D_\tau} CRPS_{\tau,t,i}^{-1}\right]}{\sum_{j=1}^{M}\left[\sum_{T_0}^{t-D_\tau} CRPS_{\tau,t,j}^{-1}\right]} \tag{14}$$

4. Conflitti, De Mol, and Giannone (2015) iterative algorithm. Following Hall and Mitchell (2007), the optimal vector of weights $W_{\tau,t}^* = (w_{\tau,t,1}^*, \ldots, w_{\tau,t,M}^*)$ minimizes the Kullback-Leibler information criterion (KLIC) divergence, $KLIC_{\tau,t} = (1/t)\sum_{s=1}^{t}[\log g_s(y_s = y_s^o)$ $- \log f_{\tau,s}(y_s = y_s^o, W_{\tau,s})]$, where $f_{\tau,t}(\cdot, W_{\tau,t})$ is the combined density nowcast across the $M$

---

[18] The CDF of the observation is a Heaviside step function, which takes a value of 0 for all values of the density that are less than the actual realization and 1 for all values of the density that are greater than or equal to the realization.
[19] We also explored the strategy of computing weights using a 12-month rolling window. The results for density nowcasting accuracy were similar.

individual densities at time $t$ which is a function of weights $W_{\tau,t}$ and $g_t(\cdot)$ is the true but unknown density corresponding to the actual realizations $y_t^o$. The solution to the optimization problem, subject to constraints $w_{\tau,t,i} \geq 0 \ \forall \ i$ and $\sum_{i=1}^{M} w_{\tau,t,i} = 1$ in each period, is

$$W_{\tau,t}^{*} = \arg\max_{W_{\tau,s}} \frac{1}{t - T_0 + 1} \sum_{s=T_0}^{t} [\log f_{\tau,s}(y_s = y_s^o, W_{\tau,s})]. \tag{15}$$

Hall and Mitchell (2007) and Amisano and Geweke (2011) note that this optimization can be solved using numerical search algorithms.[20] Conflitti, De Mol, and Giannone (2015) propose an iterative solution to the above optimization problem that is computationally feasible for combining density estimates for large $M$. Specifically, they break the objective function (equation 15) into a set of auxiliary functions that can be easily maximized in an iterative fashion to solve for the optimal weights each period subject to the constraints. The maximization of auxiliary functions is convenient because it is simply a sum of $M$ terms, and when each term is a function of a single weight, there are a total of $M$ weights. The algorithm is initialized with equal weights, $w_{\tau,t,i}^{(0)} = (1/M)$. If $t \leq T_0 + D_\tau - 1$, then $w_{\tau,t,i}^{*} = w_{\tau,t,i}^{(0)}$. For $t = T_0 + D_\tau, \ldots, T$, then

$$w_{\tau,t,i}^{(k+1)} = w_{\tau,t,i}^{(k)} \frac{1}{t - D_\tau - T_0 + 1} \sum_{s=T_0}^{t-D_\tau} \frac{f_{\tau,s,i}(y_s = y_s^o)}{\sum_{j=1}^{M} w_{\tau,s,j}^{(k)} f_{\tau,s,j}(y_s = y_s^o)} \tag{16}$$

If $(w_{\tau,t,i}^{(k+1)} - w_{\tau,t,i}^{(k)}) \leq \varepsilon$, then $w_{\tau,t,i}^{*} = w_{\tau,t,i}^{(k+1)}$. The constraints on the weights are satisfied at every iteration so long as the weights are initially equal. We denote this combination as CMG.

5. Ganics (2017) optimization based on calibration fit. Ganics (2017) proposes an approach to derive optimal weights based on the calibration fit of the model using the probability integral transform (PIT). If the preference is for well-calibrated densities irrespective of the user's loss function, then intuitively it makes sense to devise a weighting strategy that directly accounts for the calibration fit of each individual (model) density forecast. Ganics (2017) illustrates via Monte Carlo applications and an empirical application that a combination approach relying on the calibration fit to derive the weights not only yields well-calibrated densities but also leads to superior density forecasts in terms of log-score. The paper examines the efficacy of

---

[20] Pauwels and Vasnev (2016) perform an in-depth analysis of this optimization procedure and document a number of practical recommendations.

three popular metrics to evaluate the calibration fit—the Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling (AD) statistics—and finds that the three metrics perform comparably with a slight advantage for the scheme based on the AD statistic in terms of improved calibration of the combination. Accordingly, we consider the weighting strategy that optimizes the calibration fit based on the AD metric.

The Ganics optimization-based weights are calculated as follows.[21] For a combined density nowcast $f_{\tau,t}(y_t, W_{\tau,t})$ that is a function of a vector of weights $W_{\tau,t}$, compute the spread between the CDF of the perfect uniform distribution and the empirical CDF of the PIT $z_{\tau,t}$ corresponding to the combined density nowcast, $\Omega_{\tau,t}(r, W_{\tau,t}) = 1[z_{\tau,t} \leq r] - r$, where $r \in [0,1]$ denotes the quantile of the combined density nowcast. Next, compute an average of the spread over the forecast evaluation sample,

$$\Upsilon_{\tau,t}(r, W_{\tau,t}) = \frac{1}{t - T_0} \sum_{s=T_0}^{t-1} \Omega_{\tau,s}(r, W_{\tau,s}) \tag{17}$$

The AD statistic, which is used as the objective function for the combined density, is defined as

$$AD_{\tau,t}(W_{\tau,t}) = \int_\rho \frac{\Upsilon_{\tau,t}(r, W_{\tau,t})}{r(1-r)} dr \tag{18}$$

for $\rho \in [0,1]$. Since a lower value for the AD statistic is preferred to a higher value, we again initialize the Ganics weights as equal weights $w_{\tau,t,i}^* = 1/M$ if $t \leq T_0 + D_\tau - 1$; otherwise, for $t = T_0 + D_\tau, \ldots, T$, the Ganics weights solve the minimization problem

$$W_{\tau,t}^* = \arg\min_{W_{\tau,t}} AD_{\tau,t}(W_{\tau,t}) \tag{19}$$

## 4. Real-Time Data

Under the assumption that density nowcasts are most informative when they are made in real time, our analysis relies on the real-time data that would have been available to forecasters in the past. Our mixed-frequency model-combination framework can generate nowcasts on a daily basis. To keep the results manageable, we assess nowcasting performance for each

---

[21] We are grateful to Greg Ganics for sharing computer code.

month's inflation reading at six representative dates. Table 2 provides details about the information flow corresponding to these representative dates.[22]

To facilitate direct comparisons with the point nowcast accuracy results in Knotek and Zaman (2017), we use the same data set and data transformations. The monthly inflation rate is defined as $\pi_t = 100(P_t / P_{t-1} - 1)$, where $P_t$ is the price index in month $t$. The 12-month trailing (year-over-year) inflation rate is $\pi_{t,t-12} = 100(P_t / P_{t-12} - 1)$. Quarterly annualized inflation rates are $\pi_T^Q = 100[(P_T^Q / P_{T-1}^Q)^4 - 1]$, with $P_T^Q = (1/3)(P_{T,t=1} + P_{T,t=2} + P_{T,t=3})$ the price index for quarter $T$ and $P_{T,t=k}$ is the price level in the k-th month of quarter $T$. The mixed-frequency models forecast monthly inflation rates, which are used to back out the corresponding price indices to calculate the 12-month trailing inflation rate and quarterly annualized inflation rate.

We nowcast U.S. headline and core inflation rates in both the consumer price index (CPI) and the personal consumption expenditures price index (PCE), which are monthly series. All three model classes use higher-frequency data on gasoline and oil prices. The DMS and DFM models also use data on monthly food and gasoline inflation. The real-time vintages for the monthly PCE price index and core PCE price index begin in June 2000 and come from the Federal Reserve Bank of Saint Louis' Archival Federal Reserve Economic Data (ALFRED). The real-time vintages for CPI inflation, core CPI inflation, and food CPI inflation going back to November 1996 are also from ALFRED. The real-time data for gasoline CPI inflation beginning in January 1999 come from Haver Analytics and ALFRED. Weekly retail gasoline prices (for all grades) going back to the start of 1993 are obtained from the Energy Information Administration (EIA).[23] Daily Brent crude oil spot prices going back to 1987 are obtained from the Financial Times via Haver Analytics. The seasonally adjusted CPI gasoline series is used to compute seasonal factors that are then applied to retail gasoline prices to adjust them for seasonality.

Additional data are included in the estimation of the DFM. The additional weekly data include the prices of diesel fuel, regular-grade retail gasoline, mid-grade retail gasoline, and premium-grade retail gasoline (from the EIA). The additional daily variables include the foodstuffs price index from the Commodity Research Bureau (CRB), the grains price index from

---

[22] We also produce results for quarterly inflation rates, shown in the appendix. Table A1 shows the seven representative dates that we use for the quarterly nowcasting exercise and the available information.
[23] The EIA publishes weekly gasoline prices every Monday.

Standard & Poor's (S&P), the fats and oils price index from CRB, the raw sugar price from the International Sugar Organization, the raw industrials price index from CRB, the agricultural commodities price index from S&P, the textiles and fibers price index from CRB, the industrial metals price index from S&P, steel scrap prices from the Foundation for International Business & Economic Research, the 10-year Treasury note constant maturity yield and the 3-month Treasury bill rate from the Federal Reserve Board, the S&P 500 stock price index as reported in the *Wall Street Journal*, and the nominal trade-weighted exchange value of the dollar against major currencies from the Federal Reserve Board. These data are all downloaded from Haver Analytics.

To facilitate a horse race between nowcasts produced from our modeling framework and those reported in the Survey of Professional Forecasters (SPF), we also download the historical SPF nowcast estimates from the Federal Reserve Bank of Philadelphia's real-time database.

Following Tulip (2009) and Knotek and Zaman (2017), we treat the third inflation reading for each month or quarter as our measure of "truth." The third estimate has the advantage that it incorporates more complete source data than earlier estimates, but it usually abstracts from methodological revisions, which would have been difficult to predict in real time.

## 5. Nowcast Evaluation

We use a range of metrics to evaluate our inflation density nowcasts by examining both the absolute accuracy and the relative accuracy of the density nowcasts along with the accuracy of the implied point nowcasts. Absolute accuracy tests for the calibration fit of the density estimate. Density forecasts are considered well calibrated or correctly specified when the density forecasts match the distribution of the observations over a large sample. The preference is for densities that are well calibrated. We assess the calibration properties of the density nowcasts using probability integral transforms (PITs), originally proposed by Diebold, Gunther, and Tay (1998), and interval forecasts (i.e., 70% prediction intervals).

The PIT $z_{\tau,t}$ is the CDF of the predictive density nowcast $p_{\tau,t}(y_t)$ from the point in time $\tau$ when the nowcast is made within month $t$ evaluated at the actual data realization $y_t^o$:

$$z_{\tau,t} = \int_{-\infty}^{y_t^o} p_{\tau,t}(u)\, du, t = T_0, \dots, T \tag{23}$$

Intuitively, the PIT indicates in which region (i.e., percentile) of the density nowcast the actual realizations fall (see Gerard and Nimark, 2008). A realization that falls at the middle of the density nowcast would be assigned a PIT value of 0.5, while a realization at the $10^{th}$ percentile would be assigned a value of 0.1. Density nowcasts that are well calibrated have PITs that are uniformly distributed across observations $t$; in a large sample, the actual realizations would be expected to span the entire region of the density nowcast with a probability matching the probability implied by the density nowcast. Therefore, a visual assessment for calibration can be performed by plotting PITs in the form of a histogram along with the uniform U(0,1) distribution. Correctly specified density estimates resemble rectangles (i.e., flat histograms), while severe departures from uniformity suggest calibration failure. In addition to this visual assessment, we conduct a battery of formal statistical assessments using Pearson's Chi-squared test (of uniformity and independence), the Berkowitz (2001) test (of normality of the inverse normal of the PITs), the Kolmogorov-Smirnov test (of uniformity), the Anderson-Darling test (of uniformity that specifically puts more weight on deviations between the empirical CDF of the PITs and the CDF of the uniform distribution in the tails), and the Knüppel (2015) test.[24] The Knüppel test allows for simultaneous testing for both uniformity and independence of the PITs.

Interval forecasts are another popular metric to gauge the calibration of the density forecasts (e.g., Clark, 2011; Carriero, Clark, and Marcellino, 2015; Tallman and Zaman, 2020). Accordingly, we compute the empirical 70% prediction intervals (i.e., the coverage rates) of the density nowcasts, which are defined as the difference between the $85^{th}$ and $15^{th}$ percentiles of the density nowcasts. We compare the empirical 70% coverage rates with a nominal value of 70% to assess the extent to which the density nowcast estimates are correctly calibrated.

The PITs from two or more competing density nowcasts can all be uniformly distributed and hence appear correctly specified, but we would not be able to distinguish whether one density is of higher quality (i.e., more accurate) than the others based on PITs alone. Relative accuracy involves comparing competing density estimates for their quality based on numerical scores. The density forecast that assigns a higher density (i.e., a higher probability mass) at the actual realization gets a higher numerical score and is considered more accurate. Conditional on

---

[24] See Hall and Mitchell (2007) and Rossi and Sekhposyan (2014) for details about the exact implementation of these tests. In implementing the calibration metrics, we benefitted from Matlab code made available on the websites of Barbara Rossi, Tatevik Sekhposyan, and Malte Knüppel.

obtaining well-calibrated density nowcasts, the density nowcast with the highest numerical score is preferred. Scoring metrics such as log score and CRPS are two widely used scoring rules that allow for the ranking of rival density nowcasts. As discussed earlier, the log score is the logarithm of the probability density function (corresponding to the density nowcast) evaluated at the actual realization. The higher the log score, the more accurate the density nowcast. The CRPS is the difference between the predicted and the realized cumulative distributions. Smaller CRPS values imply more accurate density nowcasts.

We also examine point nowcasts based on the mean of the (combined) density nowcasts. We assess point nowcast accuracy via the standard metric of root mean squared error (RMSE):

$$RMSE_{\tau,t} = \sqrt{\frac{\Sigma_{t=T_0}^{T}(y_t^o - E(p_{\tau,t}(y_t)))^2}{T-T_0+1}} \tag{24}$$

where $E\big(p_{\tau,t}(y_t)\big)$ refers to the mean of the density nowcast, $p_{\tau,t}(y_t)$, and $T - T_0 + 1$ is the size of the forecast evaluation sample.

## 6. Empirical Results Using Real-Time Data

We perform a comprehensive investigation of nowcast combination methods with multiple mixed-frequency inflation nowcasting approaches to produce density and point nowcasts of U.S. inflation. Overall, we examine three combination methods; five weighting schemes; four inflation measures (headline and core inflation, CPI and PCE inflation); three inflation rates (12-month trailing inflation, month-over-month inflation, and quarterly inflation); and multiple intraperiod points at which we generate our nowcasts (six distinct points for each monthly inflation reading, and seven distinct points for each quarterly inflation reading). In short, we end up with a massive number of results.

To keep the discussion of the results and the length of the paper manageable, we focus on a subset of results. Specifically, our discussion focuses entirely on results corresponding to combinations derived from our novel flexible aggregation strategy.[25] The main text also considers inflation nowcast accuracy for 12-month trailing inflation rates.[26] We first briefly

---

[25] The results based on this flexible strategy are equivalent to the ones based on the linear opinion pool, because in our empirical exercises the linear opinion pool always performs better than the log opinion pool and hence the flexible strategy always selects the linear opinion pool over the log opinion pool as the aggregation function.
[26] See the appendix for results for month-over-month inflation rates and quarterly annualized inflation rates.

examine the accuracy of the individual density nowcasts from one model in each of the three mixed-frequency model classes and then the accuracy of the stage 1 combinations. We then compare results for grand combinations across all five weighting schemes, followed by a comparison of the accuracy of the grand combination constructed using the log-score weighting scheme with its three component densities (DMS, MIDAS, and DFM). Finally, we examine the time-varying properties of the grand combination.

### 6.1 Density Nowcasts from Mixed-Frequency Model Classes

We first establish that density nowcasts obtained from a single specification within each of the three model classes are incorrectly calibrated with relatively few exceptions. The exact specifications we consider for the single DMS model, the single MIDAS model, and the single DFM model follow the baseline models in each class in Knotek and Zaman (2017).

Figure 1 plots the density nowcast PITs from these three mixed-frequency models for the four inflation measures, based on nowcasts made at two distinct points in time: using the available data through the end of the month preceding the target nowcast month (case 1), and using the available data as of day 22 of the target nowcast month (case 4). We find these points to be broadly representative of our results without showing every case. In general, the arrival of additional data during the target month very marginally improves the calibration of the density nowcasts, as is evident by comparing the proximity to uniformity in the top and bottom panels.

Each of these three individual mixed-frequency models fails one or more of the necessary statistical tests of calibration fit for at least one of the inflation measures of interest. For example, the DFM generates well-calibrated densities for CPI inflation (in cases 1 and 4) and PCE inflation (in case 4) but fails to do so for both core CPI and core PCE, as evidenced by notable departures from uniformity. The DMS model has difficulties producing well-calibrated densities overall, although as more information accumulates for the target month, the calibration of the CPI density nowcasts produced from the DMS model improves significantly (e.g., case 4). The calibration of the MIDAS model's CPI density nowcasts also improves considerably with the arrival of additional information; by contrast, the core PCE inflation density nowcasts are fairly well calibrated both early and late in the targeted month.

The requirement of correctly specified density nowcasts is a necessary condition. So with these individual specifications generally failing this important requirement, we omit a discussion of relative accuracy performance. In summary, the density nowcasts produced from a single model specification have difficulties capturing the true degree of uncertainty around the point nowcasts.

We next consider the stage 1 combinations that produce density nowcasts within each class of mixed-frequency models. Figure 2 plots the PITs from the combined density nowcasts from each model class. The plots indicate some improvements in the calibration fit of the combinations compared with the respective individual specifications, especially for the DFM model class. However, in some instances there is evidence of a slight deterioration in the calibration fit. For example, the calibration fit of the combined density nowcasts from the MIDAS class for core PCE inflation in case 4 worsens compared with the single specification as shown in Figure 1. This latter result highlights an important aspect of density combination, which is that if the candidate densities are all individually well calibrated, then their combination via the linear opinion pool may suffer calibration failure due to an increased variance (i.e., the combination overestimates uncertainty). Fortunately, in our application, the density nowcasts for core PCE inflation from the 12 different MIDAS specifications are similar. Therefore, the variance of the resulting combination does not increase enough to cause issues with the calibration fit; i.e., the disagreement component, disagreement about the mean, is small.

In addition to some improvements in calibration fit, the relative accuracy of the stage 1 combinations in terms of log scores is at least as accurate as, and often more accurate than, the respective individual specifications (see Figure A3 in the appendix). Taken together, these results suggest that there are gains from combining models within a model class, but there remain deficiencies in the calibration fit of these within-model-class combinations.

## 6.2  Comparison across Grand Combinations

We next explore whether the calibration fit of the density nowcasts can be improved further by combining the stage 1 combination density nowcasts from each of the three mixed-

frequency model classes into a stage 2 "grand" combination.[27]  Figure 3 plots the PITs across various grand combinations based on the five weighting schemes listed above.  Immediately, the figure demonstrates that there are gains from combining densities across mixed-frequency model classes as evidenced by PITs that are now closer to the uniform distribution.  More precisely, combining densities via our two-stage procedure fixes the defective stage 1 combination density nowcasts.  A close inspection of the PIT histograms across all four inflation measures, and in both cases 1 and 4, reveals that all five grand combination density estimates are better calibrated compared with the density estimates from either the individual specifications or those formed by combining specifications only within a model class.

Table 3 reports the formal statistical assessment of the calibration fit.  Among the five combinations we consider, we find that the combination based on the weighting scheme proposed by Ganics generates the best-calibrated densities, with the smallest number of rejections of the null hypothesis of correct calibration.[28]  The density estimates based on the log score and CMG weighting schemes are the next best combinations, with somewhat higher numbers of rejections of the null.  The ability of the Ganics weighting scheme to produce the best-calibrated densities makes intuitive sense, as his approach is based on direct optimization of the calibration fit of the candidate densities.

Beyond calibration fit, we are also interested in relative accuracy.  Figure 4 assesses relative accuracy via the log score (panel a) and CRPS (panel b) metrics.  In general, the relative accuracy of the grand combinations improves as additional information arrives over the course of the target nowcast month, as is evident by steadily increasing average log scores and declining CRPS values.  The figure shows that the grand combination based on the Ganics weighting scheme generates inferior density nowcasts for CPI inflation and PCE inflation compared with both the log score and CMG weighting schemes.  In the case of core CPI inflation, all weighting scheme combinations perform comparably in terms of relative accuracies.  For core PCE inflation, the log score and Ganics weighting schemes are a touch worse than the other schemes when using log score as the relative accuracy metric, but the log score weights are just as

---

[27] As a reminder, the same weighting scheme is used to form both the stage 1 and stage 2 combinations.  Figure A4 in the appendix plots the nowcasts coming from the grand combination using real-time data at two representative dates (case 1 and case 4) for each month along with the actual outcomes.

[28] We omit the results for the equal weights scheme and the CRPS weighting scheme to economize on space in the table; in general, these weighting schemes were inferior to those shown, which is visible in Figure 3.

accurate as other weighting schemes when using CRPS to judge relative accuracy. Given that the density nowcast estimates based on log score and CMG weighting schemes generate superior relative density nowcast accuracy and satisfactory calibration fit across inflation measures, we generally favor these schemes over the other weighting schemes.

The patterns observed for the relative accuracy scores of the density nowcasts echo the point nowcast accuracy results reported in Knotek and Zaman (2017). Knotek and Zaman (2017) document that a single version of the DMS model was substantially more accurate then competing MIDAS or DFM models in nowcasting CPI inflation and PCE inflation, while all three models were competitive in nowcasting core CPI inflation and core PCE inflation. In terms of the relative accuracy of the density nowcasts, combination schemes such as log score that put more weight on the DMS model class generate more accurate density nowcasts for CPI inflation and PCE inflation, even though the calibration fit of the density nowcasts from the DMS model class is inferior to both the DFM and MIDAS model classes for these inflation measures. However, the mean of the density nowcasts—i.e., the point nowcasts—from the DMS model class is substantially more accurate than the means coming from either the DFM or the MIDAS model classes. When evaluated using the log score metric, which is considered a broader measure of density accuracy (e.g., Clark, 2011), the DMS model class's more accurate density nowcast mean more than offsets its slightly poorer calibration fit, resulting in a higher log score.

In contrast, the Ganics weighting scheme, which focuses on calibration fit, assigns large weights to the DFM and MIDAS model classes for nowcasting CPI inflation and PCE inflation (see Figure A5 in the appendix for an example), because these two model classes produce better-calibrated densities than the DMS model class. The equal weights scheme and the CRPS-based weighting scheme also tend to put more weight on the DFM and MIDAS model classes. Because these two model classes produce substantially inferior point nowcasts compared with the DMS model class, the additional weight assigned to them results in lower log scores and higher CRPS values.

In addition to absolute and relative accuracy in a density sense, Figure 5 plots the RMSE of the implied point nowcasts corresponding to the grand combinations. The results for point nowcast accuracy echo the results for density nowcast accuracy shown in Figure 4. As information accumulates over the course of the month and we move from case 1 to case 6, the

accuracy of the point nowcasts steadily improves and the RMSEs steadily decline. In the case of CPI inflation and PCE inflation, the combinations using log score weights produce lower RMSEs than the other weighting schemes. For core CPI inflation and core PCE inflation, the differences in RMSEs across weighting schemes are very small and all combinations generally perform comparably.

## 6.3 Comparing the Grand Combination with Its Underlying Component Densities

We noted above that the grand combination helps improve the calibration of the density nowcasts, which is an important objective. We also compare the accuracy of the grand combination to its three component densities—DMS, MIDAS, and DFM—when both the stage 1 and stage 2 combinations are made using the log score weighting scheme to assess the extent to which relative accuracy and point accuracy gains are coming from combining densities across mixed-frequency model classes.

Figure 6 plots the density nowcast relative accuracy comparison using the log score and CRPS metrics. These figures indicate that the grand combination's nowcasts are generally among the best performing, especially when assessed using the CRPS metric. However, when evaluated using the log score metric, there are some instances for CPI inflation and PCE inflation where the DMS combination is significantly more accurate than the grand combination. Indeed, the DMS combination alone is often quite competitive with the grand combination. Only in the case of core PCE inflation with the log score metric does the MIDAS combination outperform the DMS combination.

Figure 7 compares the point nowcast accuracy of the grand combination with the DMS combination and the single DMS specification from Knotek and Zaman (2017).[29] The point nowcast accuracy of the grand combination is similar to the single DMS specification for CPI inflation, PCE inflation, and core CPI inflation. In the case of core PCE inflation, the grand combination is more accurate than the single DMS specification, with notably lower RMSEs for cases 1 through 4; for cases 3 and 4, the accuracy gains are statistically significant at the 10%

---

[29] We omit the MIDAS and DFM combinations because the DMS combination is substantially more accurate for CPI inflation and PCE inflation, and all combinations are competitive for core PCE inflation and core CPI inflation. The grand combination and DMS combination use the log score weighting scheme.

significance level.[30]  We view these results as desirable because it satisfies an important objective of our model-combination approach, which is that we want the point nowcasting accuracy of the combined density to be at least as good as the single DMS specification from Knotek and Zaman (2017).

Given that the point accuracy of the DMS combination is similar to the accuracy of the grand combination for all four inflation measures, one could focus only on combinations from the DMS model class.  However, we view combining density nowcasts across the three mixed-frequency model classes as more desirable for at least two reasons.  First, given the preference for well-calibrated densities, the grand combination comes closer to being correctly specified than the DMS combinations.  Second, while the DMS model class performed well over our evaluation sample, there is no guarantee that it will continue to do so going forward.  The grand combination based on a richer set of models provides better insurance in the face of future uncertainties about model specifications and will be more robust to structural instabilities.[31]

## 6.4 Time-Varying Properties of the Grand Combination: Weights, Uncertainty, Skewness, and Kurtosis

The blend of adaptive, possibly time-varying weights and the use of a flexible aggregation strategy could yield time-varying estimates of the variance (uncertainty), skewness (asymmetry), and kurtosis in the inflation density nowcasts.  Figure 8 plots the properties of the density nowcasts for the four inflation measures for cases 1 and 4 during the target nowcast month: the evolution of the weights applied to the stage 1 combinations in making the stage 2 grand combination (top row); the evolution of the uncertainty (i.e., the volatility) around the point nowcasts, measured as the width of the 70% prediction intervals of the density nowcasts (second row); estimates of skewness (third row); and estimates of kurtosis (fourth row).[32]

We note the following items from the figure.  First, in the case of CPI inflation and PCE inflation, the DMS model class quickly dominates the DFM and MIDAS model classes,

---

[30] Statistical significance is based on the Diebold-Mariano and West test (with the truncation lag parameter h-1 for the HAC variance estimator), and two-sided standard normal critical values.
[31] Figures A6, A7, A8, and A9 in the appendix report similar results for month-over-month inflation and quarterly annualized inflation rates.
[32] Skewness=0 and kurtosis=3 for a variable that is normally distributed, so departures from these values suggest evidence for asymmetric distributions if skewness is different from zero or fat tails if kurtosis is greater than 3.

receiving much or nearly all of the weight for most of our evaluation sample. However, it is worth pointing out that even if a single model class receives a nearly 100% weight in the stage 2 combination, that model class is nevertheless a stage 1 combination of many individual model specifications. These stage 1 combinations almost always include more than one model specification (see Figures A10, A11, and A12 in the appendix). In the case of core CPI inflation and core PCE inflation, there is considerably more variation regarding which model class receives the most weight, although the MIDAS model class tended to be the best performing late in the sample and so received the most weight.

Second, we highlight the fast model-switching behavior of the weighting scheme based on the log-score. When the differences in the density accuracy among the candidate densities are in the moderate to large range, then the log score metric discriminates among densities rather sharply by heavily penalizing the poor performers. This sharp distinction implies that the density nowcast assessed as the most accurate gets a very high score, which translates into a substantially higher average score and, in turn, significantly higher weight in the combination, possibly resulting in a fast-switching pattern.[33] Recent research generally views this feature of fast model-switching as desirable, and so a framework that allows for it is viewed favorably (see Beckmann et al., 2020).

Third, as information accumulates over the course of the target nowcast month, the precision of the density nowcasts improves, and there is a shift lower in the uncertainty estimates in case 4 compared with case 1.[34]

Fourth, there are visible movements in the uncertainty estimates, more so in the case of headline inflation than core inflation. It is also evident that the profiles of the uncertainty estimates differ across inflation measures and across monthly cases, reflecting different nowcast origins within a month for a given inflation measure. Interestingly, the patterns seen in the case of quarterly inflation provide stronger evidence of shifts in uncertainty over time (see Figure A13 in the appendix). Moreover, the broader movements in the uncertainty estimates implied by our density nowcasts for quarterly inflation are similar to uncertainty estimates reported elsewhere in the literature using stochastic volatility models (e.g., Knotek, Zaman, and Clark, 2015).

---

[33] Aastveit et al. (2014) document a similar characteristic for the log score weighting scheme in their real GDP nowcasting application.

[34] See the narrowing of the prediction intervals in case 4 compared with case 1 in Figure A4 in the appendix.

Fifth, there is evidence of both time-varying skewness and kurtosis in the density nowcasts. There have been extended stretches in which skewness differed from zero and kurtosis differed from three, consistent with departures from Gaussianity, along with occasional spikes in both measures.[35] Overall, the density nowcast estimates generated from our two-stage combination process featuring a flexible aggregation strategy with the log score weighting scheme can adapt in a time-varying manner to accommodate non-Gaussian features such as asymmetry and/or heavy tails and are doing an adequate job of capturing uncertainty around future inflation outcomes.[36]

## 7. Comparison with the Survey of Professional Forecasters

Given the ability of their respondents to use a variety of high-frequency data sources in a flexible fashion, nowcasts coming from surveys of professional forecasters are a difficult benchmark to beat. We test the nowcasting performance of our grand combination against the inflation nowcasts provided by the Survey of Professional Forecasters (SPF), in terms of both point nowcasting performance and density nowcasting performance. In doing so, we match our model's real-time information set to the survey dates from the SPF each quarter.

For point nowcasting, we compare the median SPF response to the mean of the combined density nowcast coming from the grand combination using the log score weighting scheme and the flexible aggregation strategy. For density nowcasting, we compare our grand combination with *estimated* survey density nowcasts formed using a normal distribution, whose mean is set equal to the median SPF point nowcast and whose variance is set to match the variance of the past historical errors of the SPF point nowcasts over a short rolling window.

Estimates of survey density nowcasts based on historical errors have been shown to be a good benchmark, especially for inflation.[37] The Federal Open Market Committee uses historical

---

[35] Figure A14 in the appendix illustrates the stage 2 grand combination as of case 1 for nowcasting the target month of January 2001; the resulting CPI inflation and PCE inflation grand combination densities are noticeably fat tailed and asymmetric compared with the densities for core CPI inflation and core PCE inflation.

[36] Figures A15, A16, and A17 in the appendix plot the weights and higher-order moments when using the CMG, Ganics, and CRPS weighting schemes, respectively.

[37] Krüger, Clark, and Ravazzolo (2017) and Tallman and Zaman (2019) document competitive nowcasting performance, including the good calibration fit of the density nowcasts of inflation constructed through this simple procedure. The procedure's use of a short rolling window in computing the variance of the past historical errors is a simple and convenient way to incorporate the changing variance of the density estimates instead of explicitly

forecast errors to provide an estimate of the uncertainty surrounding the outlook in the Summary of Economic Projections (see Reifschneider and Tulip, 2019). While the SPF does provide some density forecasts by combining individual respondents' density forecasts, we favor the historical errors approach because Clements (2018) shows that the survey projection's second moments are inferior to simple statistical models. In addition, the SPF only reports fixed-event density forecasts for core PCE inflation and core CPI inflation, which limits their comparability to our results to the fourth quarter of each year.

Table 4 reports the results from the out-of-sample nowcasting horse race between our grand combination using real-time data and the SPF for point and density nowcasts. The evaluation period runs from 2000Q4 through 2015Q2 for CPI inflation, and 2007Q1 through 2015Q2 for core CPI inflation, PCE inflation, and core PCE inflation.

The point nowcasts implied by the grand combination are substantially more accurate (i.e., have lower RMSEs) than the SPF for both CPI inflation and PCE inflation, and the gains are statistically significant. For these two inflation measures, the density nowcasts from the grand combination are substantially more accurate than the simple SPF-based benchmark density nowcasts, as indicated by significantly higher log scores.

For core CPI inflation and core PCE inflation, both the point accuracy and the density accuracy of the grand combination are competitive with the SPF. As noted above, there is only limited evidence of skewness and kurtosis in the predictive distributions for core inflation, suggesting that flexible density estimates are not far from the normality assumption embedded in the estimated SPF density. In the case of core inflation, we see our framework's ability to adapt in a dynamic fashion to produce an approximately normal distribution as a testament to the benefits of our flexible approach.

## 8. Conclusion

We develop a flexible framework based on model combinations to produce density nowcasts for U.S. inflation. By combining individual density nowcasts from three classes of parsimonious mixed-frequency models, this framework generates nowcasts at a trading-day

---

modeling stochastic volatility, a point also emphasized by Ganics, Rossi, and Sekhposyan (2018), who find support for this simple benchmark.

frequency and updates as information accumulates over the course of a month or a quarter. We propose a novel flexible aggregation strategy to combine the density nowcasts both within and across model classes. We complement this flexible aggregation strategy with an examination of a variety of dynamic model averaging approaches, where the weights used to combine the nowcasts can be updated based on learning from past performance. An important feature of this proposed framework is its ability to accommodate non-Gaussian and time-varying properties of variance, skewness, and kurtosis in the density nowcast estimates. These dynamic features are essential in improving the accuracy of density nowcasts for headline inflation.

Our flexible framework allows us to incorporate a range of recent density combination methods proposed in the literature in a comprehensive empirical examination. Overall, using high-frequency, real-time data over the period 2000-2015, we show that the grand combination from our approach can generate highly accurate density nowcasts, but the combination method matters for the accuracy of the combined density nowcasts. Density combinations based on our novel flexible aggregation strategy using the log score weighting scheme, which relies on past predictive performance, or the "optimal" weighting scheme proposed by Conflitti, De Mol, and Giannone (2015) are among the best performing in terms of relative accuracy and are well calibrated. The Ganics (2017) weighting scheme produces the best calibrated densities for headline inflation, but its relative accuracy is inferior to the log score and the Conflitti, De Mol, and Giannone (2015) weighting schemes. In the case of core inflation, all combination methods perform comparably.

In a horse race with the Survey of Professional Forecasters, the grand combination's density nowcasts provide superior point and density nowcasts for CPI inflation and PCE inflation. For core CPI inflation and core PCE inflation, our grand combination's nowcasting performance is competitive with the SPF. The ability of our proposed framework to generate highly accurate point and density nowcasts of inflation is a useful outcome for practitioners.

Our empirical findings should serve as a guide to practitioners about the combination methods that may or may not work for nowcasting U.S. inflation. Our study provides further evidence that, when it comes to density combinations, there is no single best procedure; rather, it is important to examine combination methods on a case-by-case basis.

## 9. References

Aastveit, Knut Are, Claudia Foroni, and Francesco Ravazzolo. 2017. "Density Forecasts With Midas Models." *Journal of Applied Econometrics* 32 (4): 783–801. https://doi.org/10.1002/jae.2545.

Aastveit, Knut Are, Karsten R. Gerdrup, Anne Sofie Jore, and Leif Anders Thorsrud. 2014. "Nowcasting GDP in Real Time: A Density Combination Approach." *Journal of Business & Economic Statistics* 32 (1): 48–68. https://doi.org/10.1080/07350015.2013.844155.

Aastveit, Knut Are, James Mitchell, Francesco Ravazzolo, and Herman van Dijk. 2018. "The Evolution of Forecast Density Combinations in Economics." 18-069/III. Tinbergen Institute Discussion Papers. Tinbergen Institute. https://ideas.repec.org/p/tin/wpaper/20180069.html.

Allan, Grant, Gary Koop, Stuart McIntyre, and Paul Smith. 2014. "Nowcasting Scottish GDP Growth." 1411. Working Papers. University of Strathclyde Business School, Department of Economics. https://ideas.repec.org/p/str/wpaper/1411.html.

Amisano, Gianni, and John Geweke. 2011. "Optimal Prediction Pools." *Journal of Econometrics* 164 (1): 130–41. https://doi.org/10.1016/j.jeconom.2011.02.017.

Bache, Ida Wolden, Anne Sofie Jore, James Mitchell, and Shaun P. Vahey. 2011. "Combining VAR and DSGE Forecast Densities." *Journal of Economic Dynamics and Control* 35 (10): 1659–70. https://doi.org/10.1016/j.jedc.2011.04.006.

Bańbura, Marta, and Michele Modugno. 2014. "Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data." *Journal of Applied Econometrics* 29 (1): 133–60. https://doi.org/10.1002/jae.2306.

Bates, J. M., and C. W. J. Granger. 1969. "The Combination of Forecasts." *Operational Research Quarterly* 20 (4): 451–68. https://doi.org/10.2307/3008764.

Beckmann, Joscha, Gary Koop, Dimitris Korobilis, and Rainer Alexander Schüssler. 2020. "Exchange Rate Predictability and Dynamic Bayesian Learning." *Journal of Applied Econometrics* 35 (4): 410–21. https://doi.org/10.1002/jae.2761.

Berkowitz, Jeremy. 2001. "Testing Density Forecasts, With Applications to Risk Management" *Journal of Business and Economic Statistics* 19: 465-474. https://www.jstor.org/stable/1392281.

Bjørnland, Hilde C., Karsten Gerdrup, Anne Sofie Jore, Christie Smith, and Leif Anders Thorsrud. 2011. "Weights and Pools for a Norwegian Density Combination." *The North American Journal of Economics and Finance* 22 (1): 61–76. https://doi.org/10.1016/j.najef.2010.09.001.

Breitung, Jörg, and Christoph Roling. 2015. "Forecasting Inflation Rates Using Daily Data: A Nonparametric MIDAS Approach: Nonparametric MIDAS Forecasting." *Journal of Forecasting* 34 (7): 588–603. https://doi.org/10.1002/for.2361.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino. 2015. "Realtime Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178 (4): 837–62. https://doi.org/10.1111/rssa.12092.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino. 2019. "Large Bayesian Vector Autoregressions with Stochastic Volatility and Non-Conjugate Priors." *Journal of Econometrics*, Big Data in Dynamic Predictive Econometric Modeling, 212 (1): 137–54. https://doi.org/10.1016/j.jeconom.2019.04.024.

Chernis, Tony, and Rodrigo Sekkel. 2018. "Nowcasting Canadian Economic Activity in an Uncertain Environment." 2018–9. Staff Discussion Paper. Bank of Canada. https://www.bankofcanada.ca/2018/08/staff-discussion-paper-2018-9/.

Clark, Todd E. 2011. "Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility." *Journal of Business & Economic Statistics* 29 (3): 327–41. https://doi.org/10.1198/jbes.2010.09248.

Clark, Todd E., and Michael W. McCracken. 2010. "Averaging Forecasts from VARs with Uncertain Instabilities." *Journal of Applied Econometrics* 25 (1): 5–29. https://doi.org/10.1002/jae.1127.

Clements, Michael P. 2018. "Are Macroeconomic Density Forecasts Informative?" *International Journal of Forecasting* 34 (2): 181–98. https://doi.org/10.1016/j.ijforecast.2017.10.004.

Clements, Michael P, and Ana Beatriz Galvão. 2008. "Macroeconomic Forecasting With Mixed-Frequency Data: Forecasting Output Growth in the United States." *Journal of Business & Economic Statistics* 26 (4): 546–54. https://doi.org/10.1198/073500108000000015.

Conflitti, Cristina, Christine De Mol, and Domenico Giannone. 2015. "Optimal Combination of Survey Forecasts." *International Journal of Forecasting* 31 (4): 1096–1103. https://doi.org/10.1016/j.ijforecast.2015.03.009.

Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay. 1998. "Evaluating Density Forecasts with Applications to Financial Risk Management." *International Economic Review* 39 (4): 863–83. https://doi.org/10.2307/2527342.

Faust, Jon, and Jonathan H. Wright. 2013. "Chapter 1 - Forecasting Inflation." In *Handbook of Economic Forecasting*, edited by G. Elliott and A. Timmermann, 2, Part A:2–56. Elsevier. https://doi.org/10.1016/B978-0-444-53683-9.00001-3.

Ganics, Gergely Akos. 2017. "Optimal Density Forecast Combinations." 1751. Working Papers. Banco de España. https://ideas.repec.org/p/bde/wpaper/1751.html.

Ganics, Gergely Akos, Barbara Rossi, and Tatevik Sekhposyan. 2018. "From Fixed-Event to Fixed-Horizon Density Forecasts: Professional Forecasters' View on Multi-Horizon Uncertainty."

Garratt, Anthony, Timo Henckel, and Shaun P. Vahey. 2019. "Empirically-Transformed Linear Opinion Pools." 2019–47. CAMA Working Papers. Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University. https://ideas.repec.org/p/een/camaaa/2019-47.html.

Garratt, Anthony, James Mitchell, and Shaun P. Vahey. 2014. "Measuring Output Gap Nowcast Uncertainty." *International Journal of Forecasting* 30 (2): 268–79. https://doi.org/10.1016/j.ijforecast.2013.07.012.

Garratt, Anthony, James Mitchell, Shaun P. Vahey, and Elizabeth C. Wakerly. 2011. "Real-Time Inflation Forecast Densities from Ensemble Phillips Curves." *The North American Journal of Economics and Finance* 22 (1): 77–87. https://doi.org/10.1016/j.najef.2010.09.003.

Gerard, Hugo, and Kristoffer Nimark. 2008. "Combining Multivariate Density Forecasts Using Predictive Criteria." rdp2008-02. RBA Research Discussion Papers. Reserve Bank of Australia. https://ideas.repec.org/p/rba/rbardp/rdp2008-02.html.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov. 2005. "There Is a Risk-Return Trade-off after All." *Journal of Financial Economics* 76 (3): 509–48. https://doi.org/10.1016/j.jfineco.2004.03.008.

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov. 2006. "Predicting Volatility: Getting the Most out of Return Data Sampled at Different Frequencies." *Journal of Econometrics* 131 (1–2): 59–95. https://doi.org/10.1016/j.jeconom.2005.01.004.

Giannone, Domenico, Lucrezia Reichlin, and David Small. 2008. "Nowcasting: The Real-Time Informational Content of Macroeconomic Data." *Journal of Monetary Economics* 55 (4): 665–76. https://doi.org/10.1016/j.jmoneco.2008.05.010.

Hall, Stephen G., and James Mitchell. 2007. "Combining Density Forecasts." *International Journal of Forecasting* 23 (1): 1–13. https://doi.org/10.1016/j.ijforecast.2006.08.001.

Hendry, David F., and Kirstin Hubrich. 2011. "Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate." *Journal of Business & Economic Statistics* 29 (2): 216–27. https://doi.org/10.1198/jbes.2009.07112.

Jore, Anne Sofie, James Mitchell, and Shaun P. Vahey. 2010. "Combining Forecast Densities from VARs with Uncertain Instabilities." *Journal of Applied Econometrics* 25 (4): 621–34. https://doi.org/10.1002/jae.1162.

Kascha, Christian, and Francesco Ravazzolo. 2010. "Combining Inflation Density Forecasts." *Journal of Forecasting* 29 (1–2): 231–50. https://doi.org/10.1002/for.1147.

Knotek, Edward S., II, and Saeed Zaman. 2017. "Nowcasting U.S. Headline and Core Inflation." *Journal of Money, Credit and Banking* 49 (5): 931–68. https://doi.org/10.1111/jmcb.12401.

Knotek, Edward S., II, and Saeed Zaman. 2019. "Financial Nowcasts and Their Usefulness in Macroeconomic Forecasting." *International Journal of Forecasting* 35 (4): 1708–24. https://doi.org/10.1016/j.ijforecast.2018.10.012.

Knotek, Edward S., II, Saeed Zaman, and Todd E. Clark. 2015. "Measuring Inflation Forecast Uncertainty." *Economic Commentary (Federal Reserve Bank of Cleveland)*, March, 1–6. https://doi.org/10.26509/frbc-ec-201503.

Knüppel, Malte. 2015. "Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments." *Journal of Business & Economic Statistics* 33 (2): 270–81. https://doi.org/10.1080/07350015.2014.948175.

Knüppel, Malte, and Fabian Krüger. 2019. "Forecast Uncertainty, Disagreement, and the Linear Pool." 28/2019. Discussion Papers. Deutsche Bundesbank. https://ideas.repec.org/p/zbw/bubdps/282019.html.

Koop, Gary, and Dimitris Korobilis. 2012. "Forecasting Inflation Using Dynamic Model Averaging." *International Economic Review* 53 (3): 867–86. https://doi.org/10.1111/j.1468-2354.2012.00704.x.

Koop, Gary, Stuart McIntyre, and James Mitchell. 2020. "UK Regional Nowcasting Using a Mixed Frequency Vector Auto-regressive Model with Entropic Tilting." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 183 (1): 91–119. https://doi.org/10.1111/rssa.12491.

Krüger, Fabian. 2014. "Combining Density Forecasts under Various Scoring Rules: An Analysis of UK Inflation."

Krüger, Fabian, Todd E. Clark, and Francesco Ravazzolo. 2017. "Using Entropic Tilting to Combine BVAR Forecasts With External Nowcasts." *Journal of Business & Economic Statistics* 35 (3): 470–85. https://doi.org/10.1080/07350015.2015.1087856.

Marsilli, Clément. 2017. "Nowcasting US Inflation Using a MIDAS Augmented Phillips Curve." *International Journal of Computational Economics and Econometrics* 7 (1/2): 64. https://doi.org/10.1504/IJCEE.2017.080649.

Mazzi, Gian Luigi, James Mitchell, and Gaetana Montana. 2014. "Density Nowcasts and Model Combination: Nowcasting Euro-Area GDP Growth over the 2008-09 Recession." *Oxford Bulletin of Economics and Statistics* 76 (2): 233–56. https://doi.org/10.1111/obes.12015.

Modugno, Michele. 2013. "Now-Casting Inflation Using High Frequency Data." *International Journal of Forecasting* 29 (4): 664–75. https://doi.org/10.1016/j.ijforecast.2012.12.003.

Monteforte, Libero, and Gianluca Moretti. 2013. "Real-Time Forecasts of Inflation: The Role of Financial Variables: Real-Time Forecasts of Inflation." *Journal of Forecasting* 32 (1): 51–61. https://doi.org/10.1002/for.1250.

Pauwels, Laurent L., and Andrey L. Vasnev. 2016. "A Note on the Estimation of Optimal Weights for Density Forecast Combinations." *International Journal of Forecasting* 32 (2): 391–97. https://doi.org/10.1016/j.ijforecast.2015.09.002.

Ravazzolo, Francesco, and Shaun P. Vahey. 2014. "Forecast Densities for Economic Aggregates from Disaggregate Ensembles." *Studies in Nonlinear Dynamics & Econometrics* 18 (4): 367–81. https://doi.org/10.1515/snde-2012-0088.

Reifschneider, David, and Peter Tulip. 2019. "Gauging the Uncertainty of the Economic Outlook Using Historical Forecasting Errors: The Federal Reserve's Approach." *International Journal of Forecasting* 35 (4): 1564–82. https://doi.org/10.1016/j.ijforecast.2018.07.016.

Rossi, Barbara. 2014. "Density Forecasts in Economics, Forecasting and Policymaking." 37. Els Opuscles Del CREI. Centre de Recerca en Economia International (CREI). http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.722.6744.

Rossi, Barbara, and Tatevik Sekhposyan. 2014. "Evaluating Predictive Densities of US Output Growth and Inflation in a Large Macroeconomic Data Set." *International Journal of Forecasting* 30 (3): 662–82. https://doi.org/10.1016/j.ijforecast.2013.03.005.

Tallman, Ellis W., and Saeed Zaman. 2017. "Forecasting Inflation: Phillips Curve Effects on Services Price Measures." *International Journal of Forecasting* 33 (2): 442–57. https://doi.org/10.1016/j.ijforecast.2016.10.004.

Tallman, Ellis W., and Saeed Zaman. 2020. "Combining Survey Long-Run Forecasts and Nowcasts with BVAR Forecasts Using Relative Entropy." *International Journal of Forecasting* 36 (2): 373–98. https://doi.org/10.1016/j.ijforecast.2019.04.024.

Tay, Anthony S., and Kenneth F. Wallis. 2000. "Density Forecasting: A Survey." *Journal of Forecasting* 19 (4): 235–54. https://doi.org/10.1002/1099-131X(200007)19:4<235::AID-FOR772>3.0.CO;2-L.

Tulip, Peter. 2009. "Has the Economy Become More Predictable? Changes in Greenbook Forecast Accuracy." *Journal of Money, Credit and Banking* 41 (6): 1217–31. https://doi.org/10.1111/j.1538-4616.2009.00253.x.

Wright, Jonathan H. 2019. "Some Observations on Forecasting and Policy." *International Journal of Forecasting* 35 (3): 1186–92. https://doi.org/10.1016/j.ijforecast.2019.04.003.

Table 1: Model Space: Mixed-Frequency Model Classes and Specifications

| Model Class | Modeling Options | Number of Models |
|---|---|---|
| DMS | Autoregressive (AR) Lags = [1, 2] | 108 |
| | Estimation window core inflation = [24, 36] months | |
| | Estimation window headline inflation = [24, 36, 84] months | |
| | Estimation window oil-gasoline error correction = [60, 72, 84] months | |
| | Number of years to use for computing seasonal factors = [3, 5, 7] years | |
| MIDAS | Estimation window = [5-year, 7-year, 10-year rolling, expanding] | 12 |
| | Polynomial option = [Beta] | |
| | High-frequency data = [Daily only, weekly only, both daily and weekly] | |
| DFM | Number of factors = 1 | 12 |
| | Number of lags = [1, 2, 3, 4, 5, 6] | |
| | Estimation window = [Expanding, 5-year rolling] | |
| Total | | 132 |

Notes: DMS is deterministic model switching; MIDAS is mixed data sampling; and DFM is dynamic factor model. See the text and the appendix for details.

Table 2: Representative Dates for Monthly Nowcasting Performance

| Case | Date | Information Set (Example: Nowcasting target month is January) | Target Month Horizon |
|---|---|---|---|
| 1 | Last day of the previous month | December 31: Have CPI and PCE through November; high-frequency information through December 31 | CPI: h=2<br>PCE: h=2 |
| 2 | Day 8 of the target month | January 8: Have CPI and PCE through November; high-frequency information through the end of the first week of January, which includes weekly retail gasoline reading for the first week of January | CPI: h=2<br>PCE: h=2 |
| 3 | Day 15 of the target month | January 15: Receive CPI for December and have PCE through November; high-frequency information through end of second week of January, which includes two weekly retail gasoline readings for January | CPI: h=1<br>PCE: h=2 |
| 4 | Day 22 of the target month | January 22: Have CPI for December and PCE through November; high-frequency information through end of third week of January, which includes three weekly retail gasoline readings for January | CPI: h=1<br>PCE: h=2 |
| 5 | Last day of the target month | January 31: Have CPI for December and receive PCE for December; high-frequency information for all of January, which includes all four weekly retail gasoline readings for January | CPI: h=1<br>PCE: h=1 |
| 6 | Day 15 of the following month | February 15: Receive CPI for January and have PCE through December; high-frequency information for all of January | CPI: h=n/a<br>PCE: h=1 |

Table 3: Calibration Diagnostics

| Nowcasting Case | Berk. | Chi-Sq. | AD | KS | KL | 70% Cov. Rate |
|---|---|---|---|---|---|---|
| CPI: Grand Combination Based on Log Score | | | | | | |
| Case 1 | 0.80 | **0.05** | **0.02** | 0.35 | 0.26 | 60.67 |
| Case 2 | 0.23 | 0.98 | 0.42 | 0.91 | 0.74 | 72.47 |
| Case 3 | 0.50 | 0.38 | 0.74 | 0.81 | 0.79 | 69.66 |
| Case 4 | **0.00** | 0.33 | 0.36 | 0.50 | 0.38 | 77.53 |
| Case 5 | **0.01** | **0.05** | 0.44 | 0.18 | 0.09 | 76.97 |
| CPI: Grand Combination Based on CMG | | | | | | |
| Case 1 | 0.14 | 0.92 | 0.63 | 0.95 | 0.99 | 70.22 |
| Case 2 | **0.00** | 0.28 | 0.36 | 0.20 | 0.50 | 74.16 |
| Case 3 | **0.05** | 0.28 | 0.37 | 0.52 | 0.56 | 76.40 |
| Case 4 | **0.01** | 0.14 | 0.37 | 0.17 | 0.37 | 77.53 |
| Case 5 | **0.00** | **0.03** | 0.07 | 0.13 | 0.07 | 80.90 |
| CPI: Grand Combination Based on Ganics | | | | | | |
| Case 1 | 0.10 | 0.55 | 0.37 | 0.76 | 0.22 | 71.35 |
| Case 2 | **0.01** | 0.75 | 0.39 | 0.56 | 0.44 | 71.35 |
| Case 3 | 0.07 | **0.02** | 0.35 | 0.38 | 0.29 | 77.53 |
| Case 4 | 0.11 | 0.56 | 0.36 | 0.48 | 0.43 | 76.40 |
| Case 5 | **0.04** | 0.23 | 0.39 | 0.17 | 0.21 | 75.84 |
| Core CPI: Grand Combination Based on Log Score | | | | | | |
| Case 1 | 0.32 | **0.02** | 0.42 | 0.30 | 0.20 | 65.73 |
| Case 2 | 0.17 | 0.22 | 0.43 | 0.22 | 0.30 | 61.80 |
| Case 3 | 0.52 | 0.41 | 0.36 | 0.46 | 0.43 | 64.61 |
| Case 4 | 0.39 | 0.38 | 0.35 | 0.43 | 0.46 | 65.17 |
| Case 5 | 0.33 | 0.28 | 0.37 | 0.37 | 0.47 | 65.73 |
| Core CPI: Grand Combination Based on CMG | | | | | | |
| Case 1 | **0.02** | 0.70 | 0.42 | 0.24 | 0.53 | 67.42 |
| Case 2 | **0.01** | 0.46 | 0.09 | 0.16 | 0.42 | 67.98 |
| Case 3 | 0.41 | 0.72 | 0.36 | 0.44 | 0.52 | 65.17 |
| Case 4 | 0.38 | 0.43 | 0.35 | 0.60 | 0.53 | 66.29 |
| Case 5 | 0.35 | 0.60 | 0.35 | 0.58 | 0.53 | 66.85 |
| Core CPI: Grand Combination Based on Ganics | | | | | | |
| Case 1 | 0.08 | 0.16 | 0.06 | 0.10 | 0.35 | 65.73 |
| Case 2 | 0.17 | 0.61 | 0.08 | 0.19 | 0.36 | 65.73 |
| Case 3 | 0.44 | **0.05** | 0.36 | 0.26 | 0.47 | 63.48 |
| Case 4 | 0.25 | 0.23 | 0.42 | 0.14 | 0.39 | 63.48 |
| Case 5 | 0.37 | 0.14 | 0.39 | 0.30 | 0.28 | 65.17 |

## Table 3: Calibration Diagnostics (continued)

| Nowcasting Case | Berk. | Chi-Sq. | AD | KS | KL | 70% Cov. Rate |
|---|---|---|---|---|---|---|
| **PCE: Grand Combination Based on Log Score** | | | | | | |
| Case 1 | 0.53 | **0.02** | **0.01** | 0.06 | 0.14 | 59.66 |
| Case 2 | **0.05** | 0.30 | **0.01** | **0.01** | 0.34 | 65.91 |
| Case 3 | 0.20 | 0.46 | 0.37 | 0.24 | 0.37 | 68.75 |
| Case 4 | 0.09 | 0.11 | 0.39 | 0.45 | 0.43 | 72.73 |
| Case 5 | **0.05** | 0.18 | **0.03** | 0.11 | 0.51 | 69.89 |
| Case 6 | 0.49 | 0.19 | **0.01** | 0.14 | 0.13 | 63.07 |
| **PCE: Grand Combination Based on CMG** | | | | | | |
| Case 1 | 0.23 | 0.51 | 0.40 | 0.31 | 0.73 | 67.61 |
| Case 2 | **0.02** | 0.48 | 0.07 | **0.03** | 0.57 | 69.89 |
| Case 3 | 0.07 | **0.01** | 0.36 | 0.26 | 0.37 | 71.02 |
| Case 4 | **0.03** | 0.07 | 0.37 | 0.37 | 0.28 | 71.02 |
| Case 5 | **0.00** | 0.11 | 0.07 | 0.09 | 0.51 | 70.45 |
| Case 6 | **0.04** | 0.82 | 0.39 | 0.27 | 0.72 | 65.34 |
| **PCE: Grand Combination Based on Ganics** | | | | | | |
| Case 1 | 0.39 | 0.56 | 0.45 | 0.54 | 0.94 | 68.18 |
| Case 2 | **0.05** | 0.74 | 0.42 | 0.21 | 0.68 | 69.89 |
| Case 3 | 0.93 | 0.33 | 0.50 | 0.67 | 0.54 | 67.05 |
| Case 4 | 0.91 | 0.76 | 0.57 | 0.47 | 0.50 | 67.05 |
| Case 5 | 0.13 | 0.47 | 0.41 | 0.52 | 0.29 | 73.86 |
| Case 6 | 0.07 | 0.63 | 0.36 | 0.68 | 0.71 | 65.34 |
| **Core PCE: Grand Combination Based on Log Score** | | | | | | |
| Case 1 | 0.26 | **0.01** | 0.36 | 0.28 | 0.44 | 63.07 |
| Case 2 | 0.43 | 0.78 | 0.36 | 0.38 | 0.61 | 64.20 |
| Case 3 | 0.17 | 0.06 | **0.01** | 0.12 | 0.12 | 59.09 |
| Case 4 | 0.06 | **0.01** | **0.00** | 0.25 | 0.16 | 60.80 |
| Case 5 | 0.42 | 0.49 | 0.38 | 0.67 | 0.55 | 66.48 |
| Case 6 | 0.77 | **0.05** | **0.04** | 0.28 | 0.26 | 63.07 |
| **Core PCE: Grand Combination Based on CMG** | | | | | | |
| Case 1 | **0.00** | 0.70 | 0.63 | 0.90 | 0.76 | 67.61 |
| Case 2 | **0.00** | 0.94 | 0.55 | 0.87 | 0.67 | 67.61 |
| Case 3 | **0.04** | 0.40 | 0.67 | 0.90 | 0.89 | 69.32 |
| Case 4 | 0.06 | 0.25 | 0.64 | 0.82 | 0.90 | 68.18 |
| Case 5 | 0.17 | 0.70 | 0.36 | 0.21 | 0.67 | 66.48 |
| Case 6 | 0.21 | 0.94 | 0.40 | 0.83 | 0.75 | 67.05 |
| **Core PCE: Grand Combination Based on Ganics** | | | | | | |
| Case 1 | **0.00** | 0.64 | 0.44 | 0.83 | 0.57 | 67.05 |
| Case 2 | **0.00** | 0.68 | 0.51 | 0.79 | 0.79 | 67.05 |
| Case 3 | 0.52 | 0.70 | 0.35 | 0.71 | 0.63 | 65.34 |
| Case 4 | 0.46 | 0.88 | 0.36 | 0.71 | 0.49 | 64.77 |
| Case 5 | 0.97 | 0.88 | 0.37 | 0.80 | 0.73 | 69.32 |
| Case 6 | 0.54 | 0.51 | 0.36 | 0.83 | 0.70 | 64.20 |

Notes: Entries except for those in the final column are p-values. "Berk" is the Berkowitz test. "Chi-Sq" is the Pearson Chi-Squared test. "AD" is the Anderson-Darling test. "KS" is the Kolmogorov-Smirnov test. "KL" is the Knüppel test. Entries in **bold** indicate rejection of the null hypothesis of correctly calibrated density nowcasts at a 5% significance level. "70% Cov. Rate" shows the 70% empirical coverage rates.

Table 4: Nowcasting Comparison with the Survey of Professional Forecasters

| | CPI | Core CPI | PCE | Core PCE |
|---|---|---|---|---|
| **Point Nowcast Comparison** | | | | |
| Grand combination RMSE | 1.040 | 0.569 | 0.793 | 0.525 |
| SPF (median) RMSE | 1.429 | 0.577 | 1.089 | 0.504 |
| Ratio, avg. SPF MSE/Grand MSE | 1.888 | 1.025 | 1.883 | 0.922 |
| *GW p-values* | **0.010** | 0.881 | **0.002** | 0.617 |
| **Density Nowcast Comparison** | | | | |
| Grand combination log score (Grand LS) | -1.370 | -0.811 | -1.064 | -0.735 |
| SPF log score (SPF LS) | -1.913 | -1.519 | -1.572 | -0.780 |
| Relative, SPF LS – Grand LS | -0.543 | -0.709 | -0.508 | -0.045 |
| DM type test *p-values* | **0.000** | 0.185 | **0.000** | 0.525 |

Notes: The grand combination uses real-time data available through the SPF survey date for each quarter. The SPF density nowcasts are based on historical forecast errors; see the text for details. The CPI exercise uses real-time data from 2000Q4 through 2015Q2. The core CPI, PCE, and core PCE exercises use real-time data from 2007Q1 (the first available SPF estimate) through 2015Q2. The DM type test reports the results of a test for equal predictive accuracy based on testing whether the constant term in the regression of the differences in the log score on the constant is statistically different from zero.

Figure 1: Comparison of PITs across Single Specifications of Mixed-Frequency Model Classes
(a) Case 1
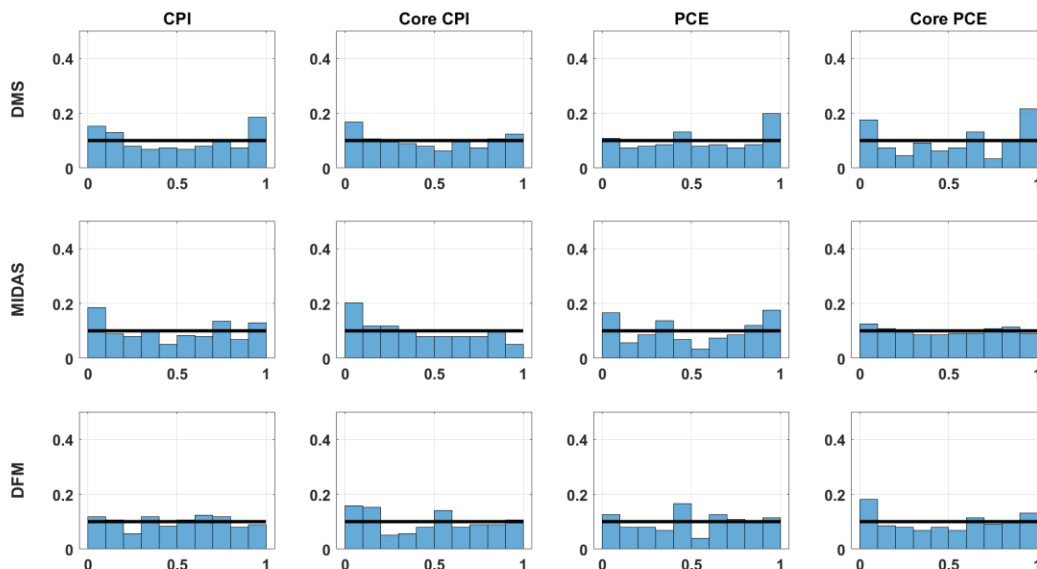


(b) Case 4
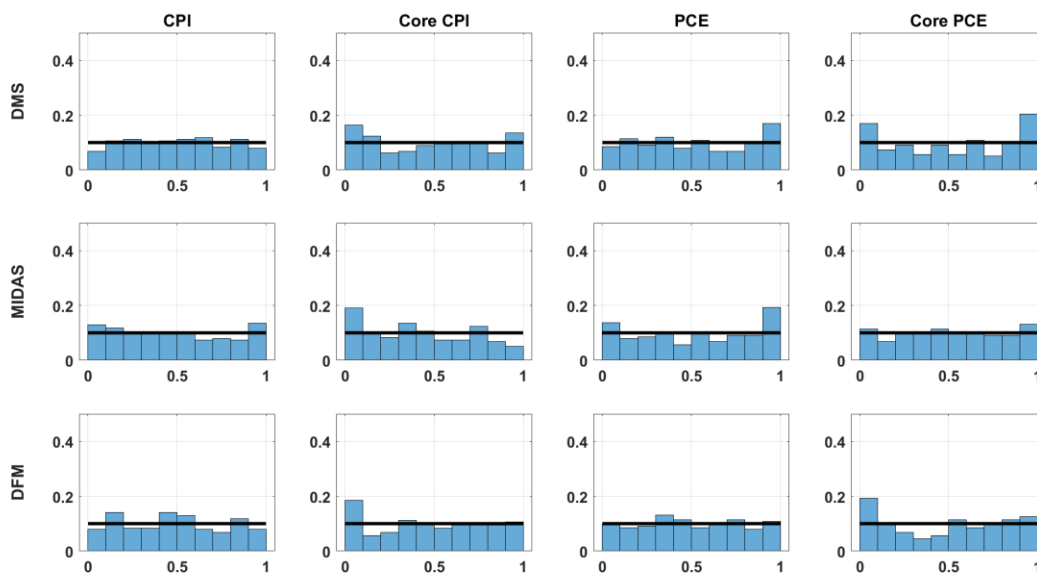


Notes: The figure plots histograms of the empirical distribution of the PITs for single specifications of the DMS, MIDAS, and DFM model classes (blue bars) and the uniform U(0,1) distribution (black lines), generated at either the last day of the month preceding the target nowcast month (case 1) or day 22 of the target nowcast month (case 4). The x-axis shows the decile bins and the y-axis shows the percentage of observations falling within each decile bin. The nowcast evaluation sample spans September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure 2: Comparison of PITs across Stage 1 Combinations within Model Classes
(a) Case 1
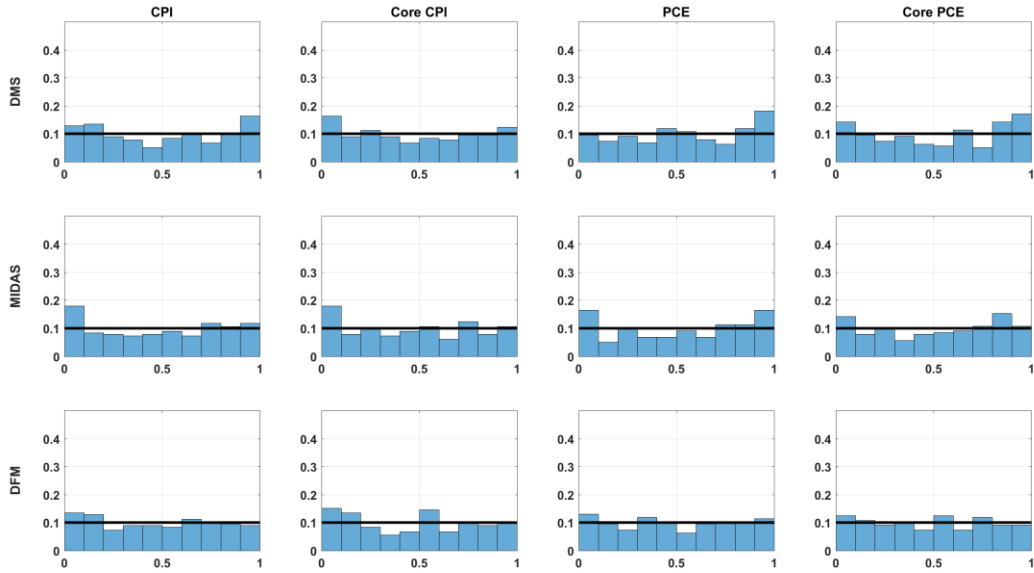


(b) Case 4
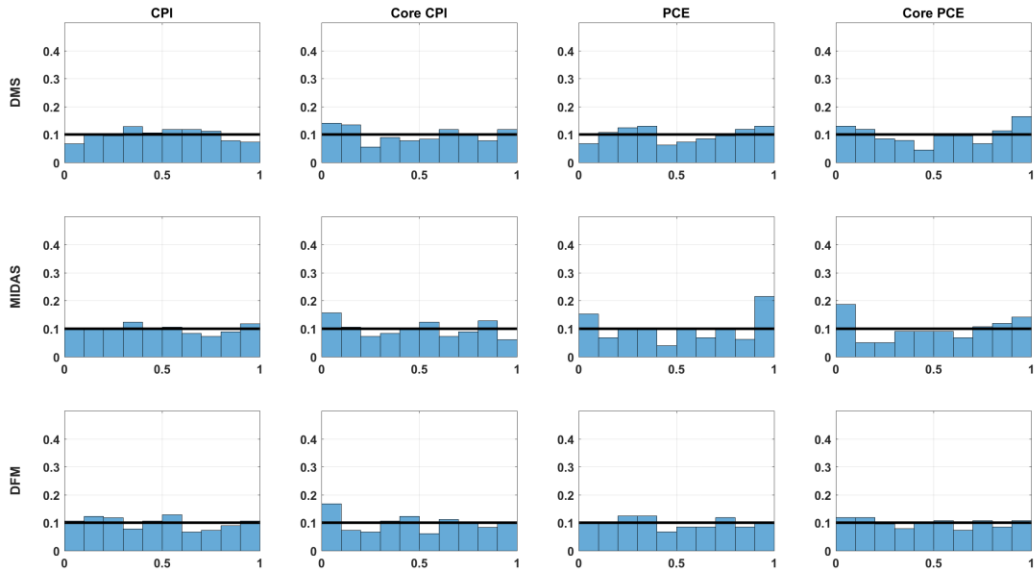


Notes: The figure plots histograms of the empirical distribution of the PITs for stage 1 combinations within the DMS, MIDAS, and DFM model classes (blue bars) and the uniform (0,1) distribution (black lines), generated at either the last day of the month preceding the target nowcast month (case 1) or day 22 of the target nowcast month (case 4). The x-axis shows the decile bins and the y-axis shows the percentage of observations falling within each decile bin. The nowcast evaluation sample spans September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

## Figure 3: Comparison of PITs across Grand Combinations
### (a) Case 1
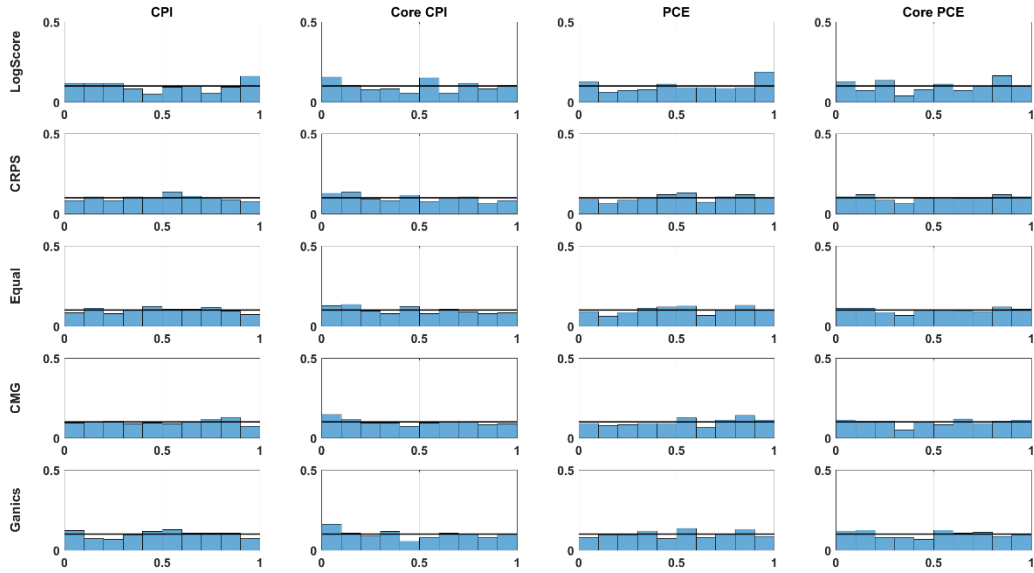


### (b) Case 4
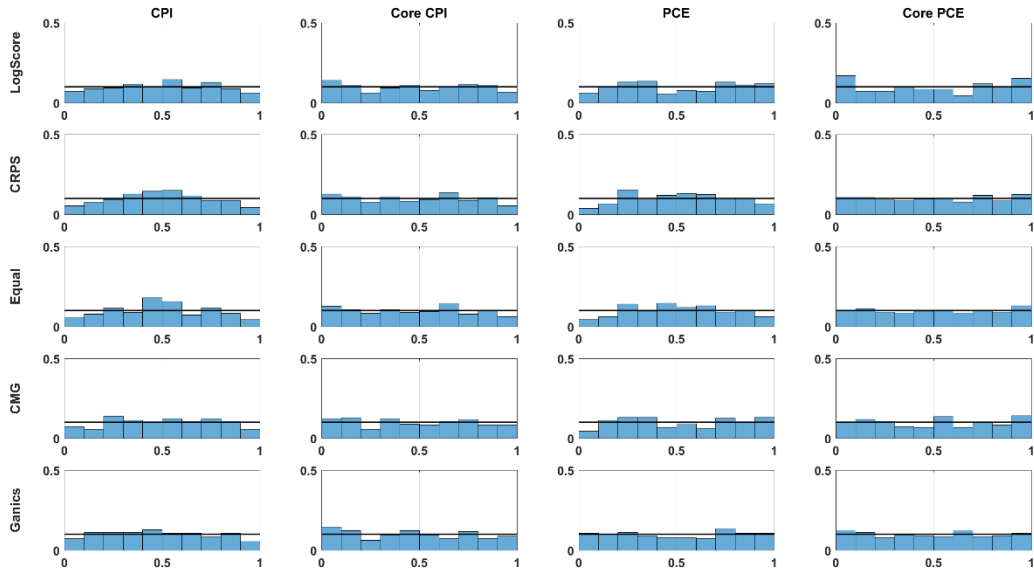


Notes: The figure plots histograms of the empirical distribution of the PITs for stage 2 combinations across the DMS, MIDAS, and DFM model classes (blue bars) and the uniform (0,1) distribution (black lines), generated at either the last day of the month preceding the target nowcast month (case 1) or day 22 of the target nowcast month (case 4). The x-axis shows the decile bins and the y-axis shows the percentage of observations falling within each decile bin. The nowcast evaluation sample spans September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

41

## Figure 4: Density Performance Comparisons across Grand Combinations

### (a) Relative accuracy based on log score
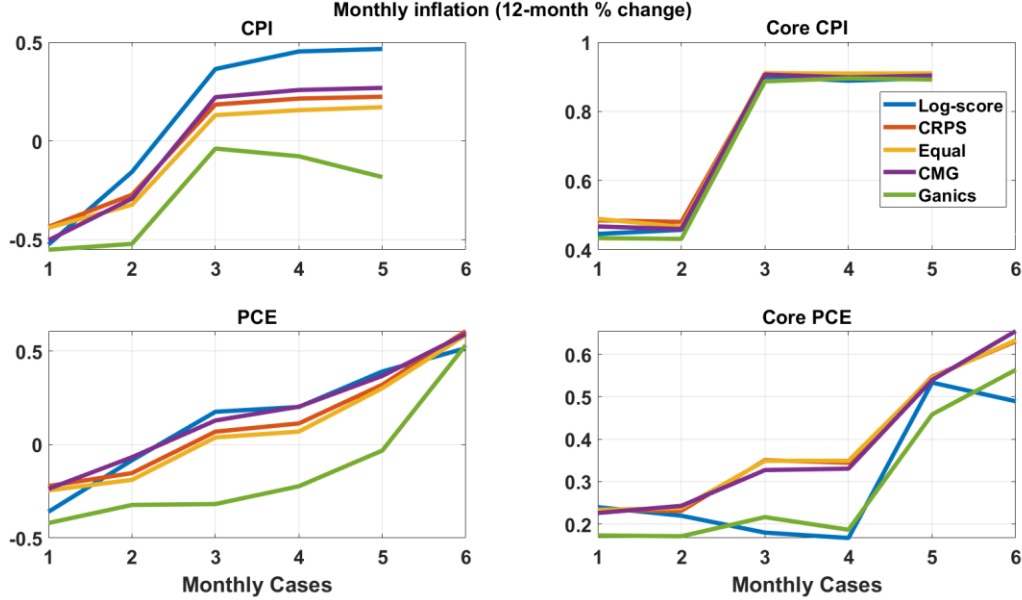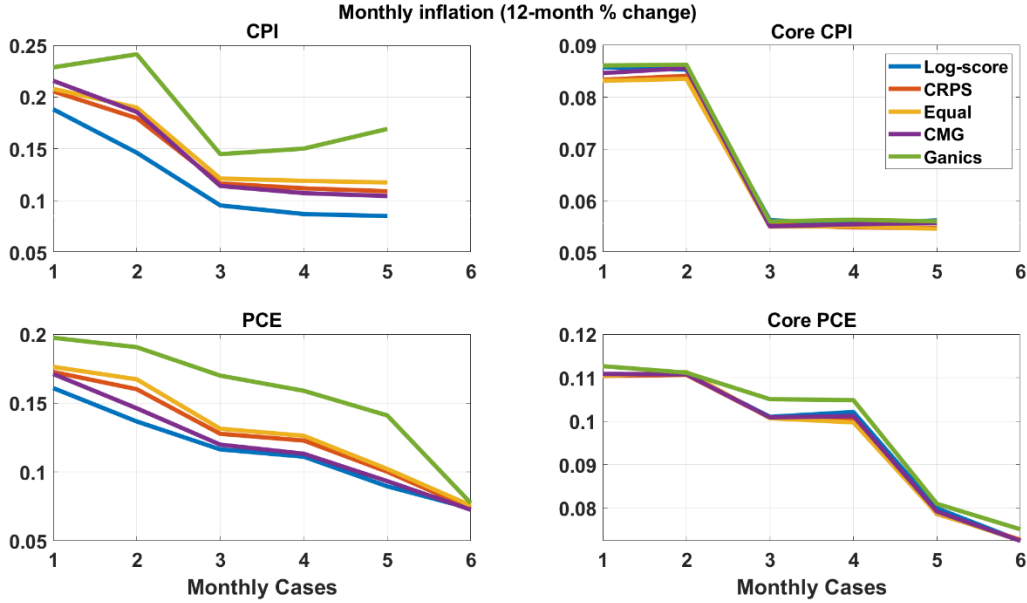
**Monthly inflation (12-month % change)**



CPI — Core CPI — PCE — Core PCE

Legend: Log-score, CRPS, Equal, CMG, Ganics

### (b) Relative accuracy based on CRPS

**Monthly inflation (12-month % change)**



CPI — Core CPI — PCE — Core PCE

Legend: Log-score, CRPS, Equal, CMG, Ganics

Notes: The top panel plots the average log score and the bottom panel plots the average CRPS for grand combinations based on log score, CRPS, equal, CMG, and Ganics weighting schemes. The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

## Figure 5: Point Nowcasting Performance across Grand Combinations
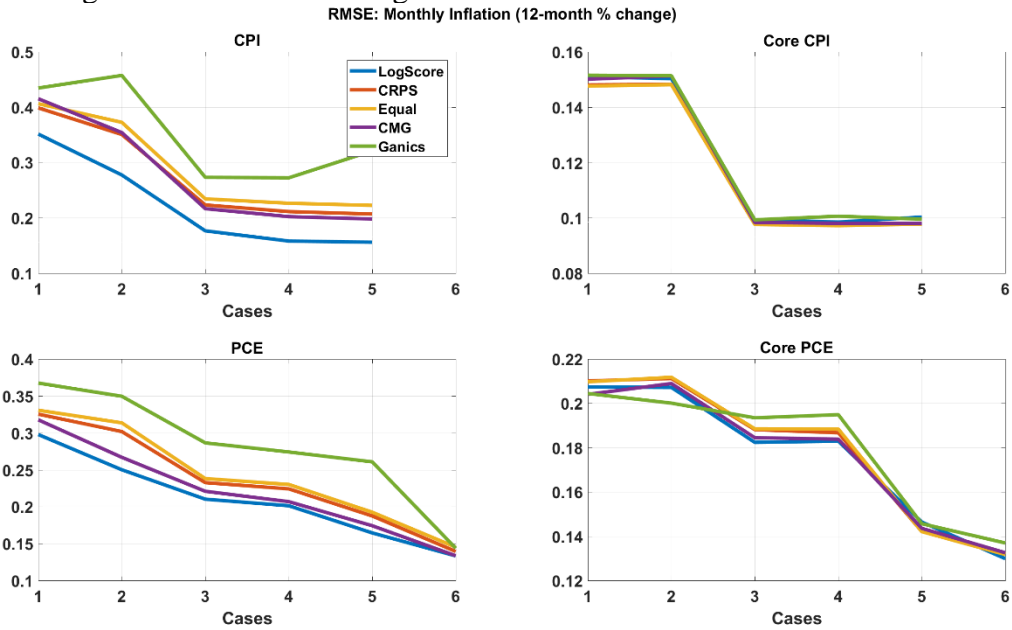


Notes: The figure plots the RMSEs for grand combinations based on log score, CRPS, equal, CMG, and Ganics weighting schemes and using the flexible aggregation strategy. The cases reflect the point in time when each nowcast was made relative to the target nowcast month; see Table 2. The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure 6: Density Performance of Grand Combination vs. Its Components
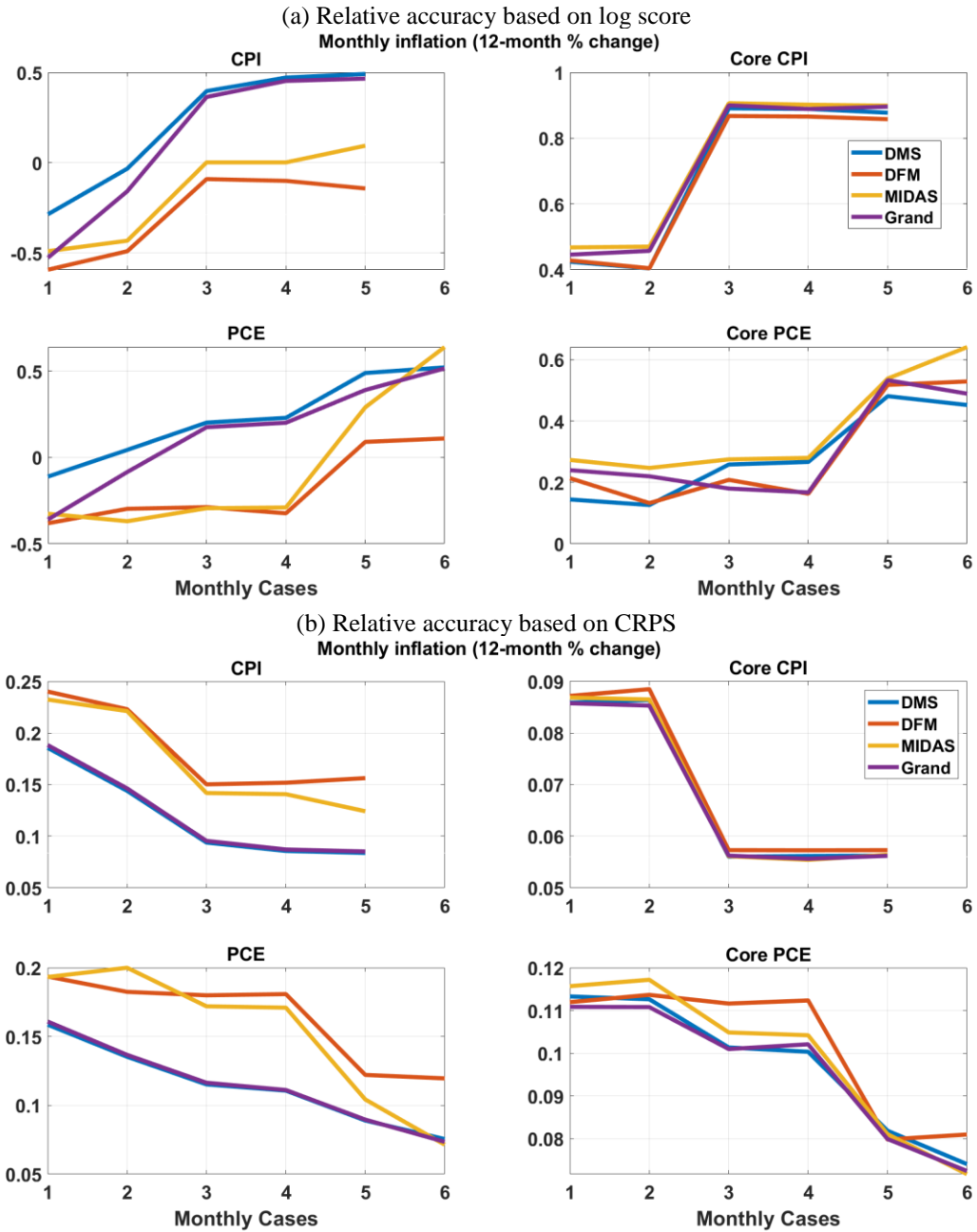
(a) Relative accuracy based on log score

(b) Relative accuracy based on CRPS

Notes: The top panel plots the average log score and the bottom panel plots the average CRPS for the grand combination based on the log score weighting scheme and combinations based on the DMS model class, MIDAS model class, and DFM model class, where each individual model class uses the log score weighting scheme. The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

## Figure 7: Point Nowcasting Performance, Grand Combination vs. DMS
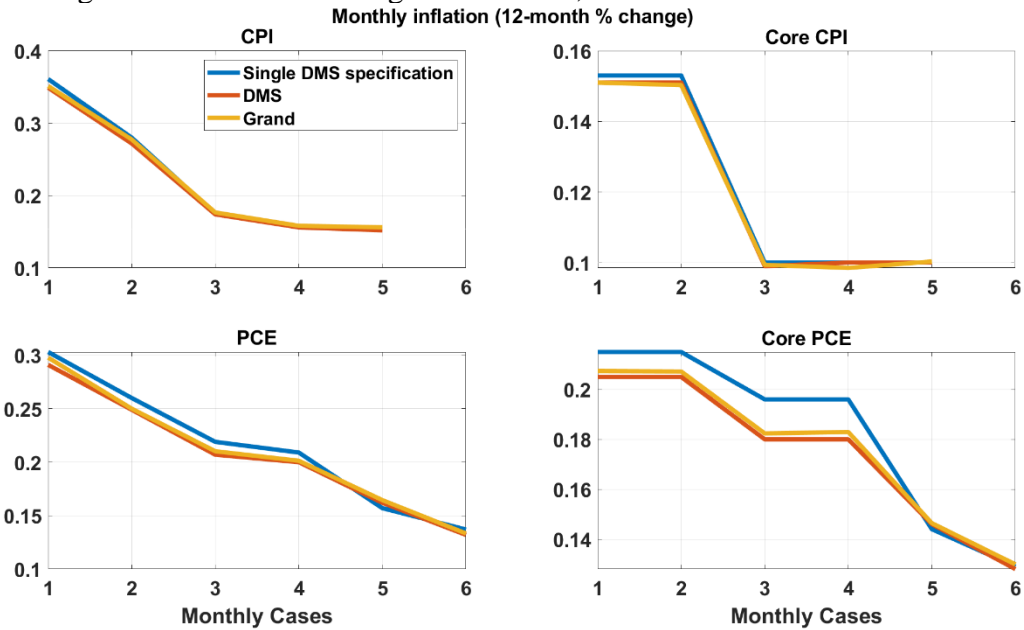


Notes: The figure plots the RMSE for the grand combination based on log score and using the flexible aggregation strategy; the stage 1 combination from the DMS model class; and a single specification from the DMS model class based on Knotek and Zaman (2017). The cases reflect the point in time when each nowcast was made relative to the target nowcast month; see Table 2. The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

## Figure 8: Weights and Higher-Order Moments
### (a) CPI inflation
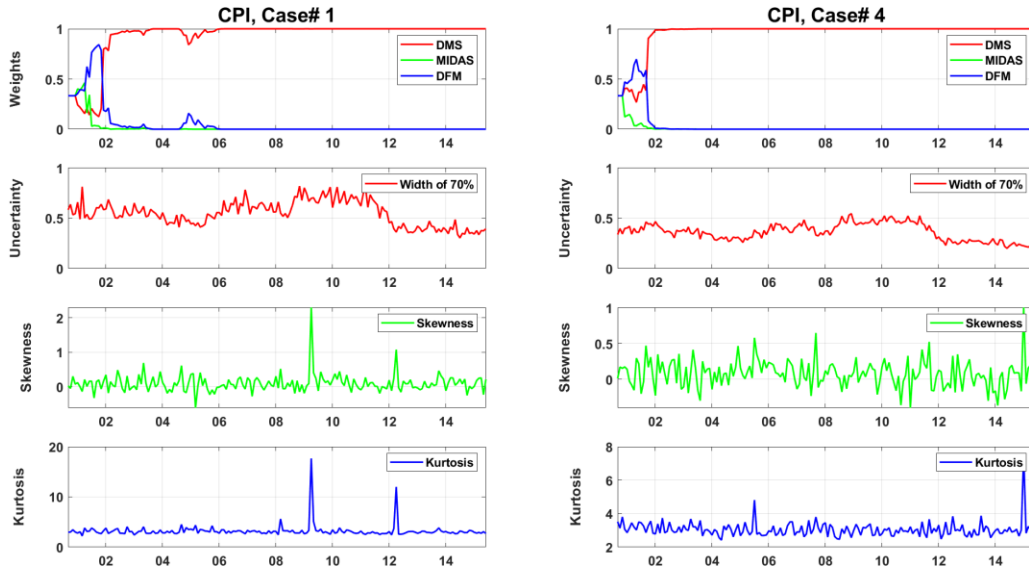


### (b) Core CPI inflation
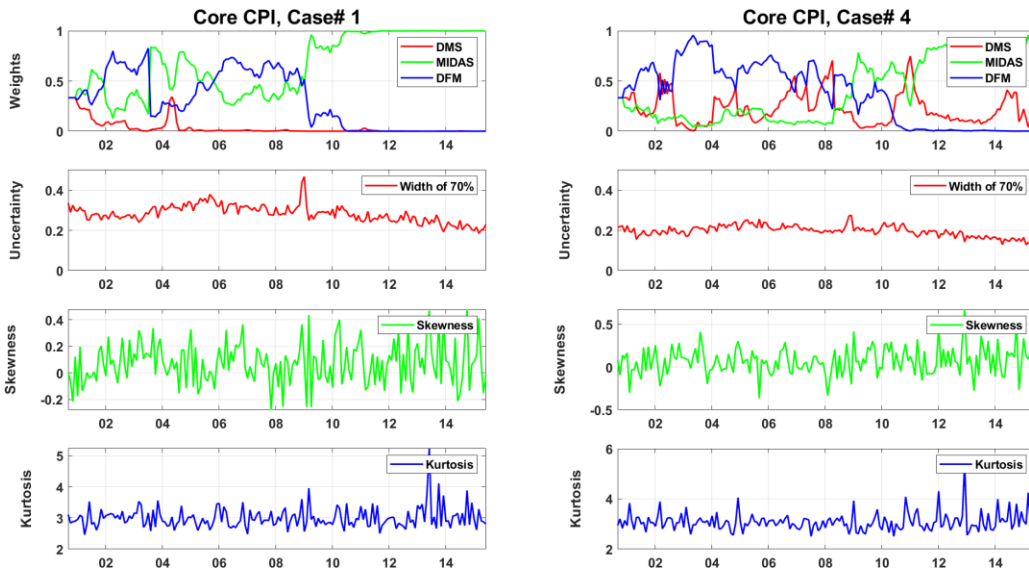
## Figure 8: Weights and Higher-Order Moments (continued)
### (c) PCE inflation
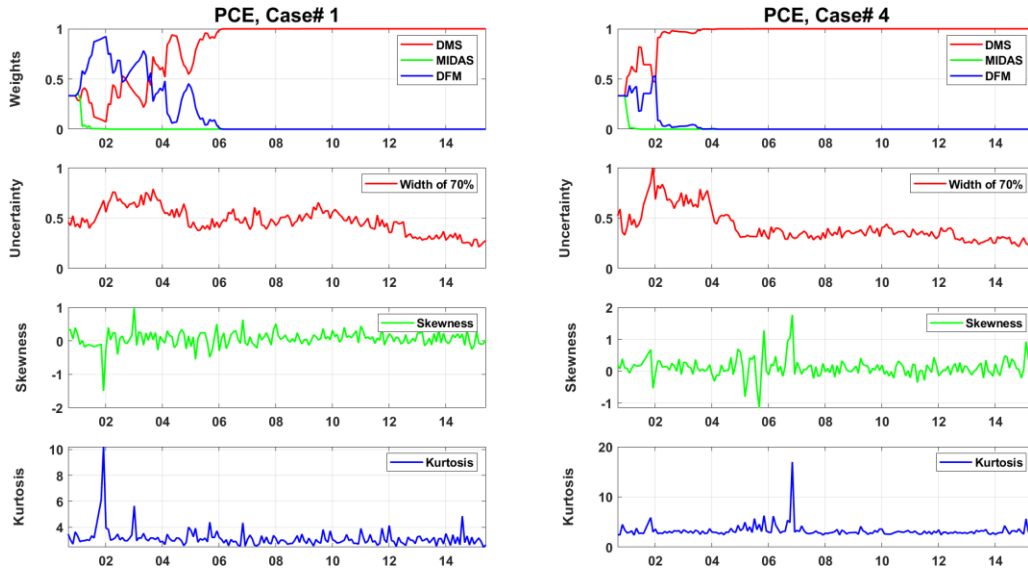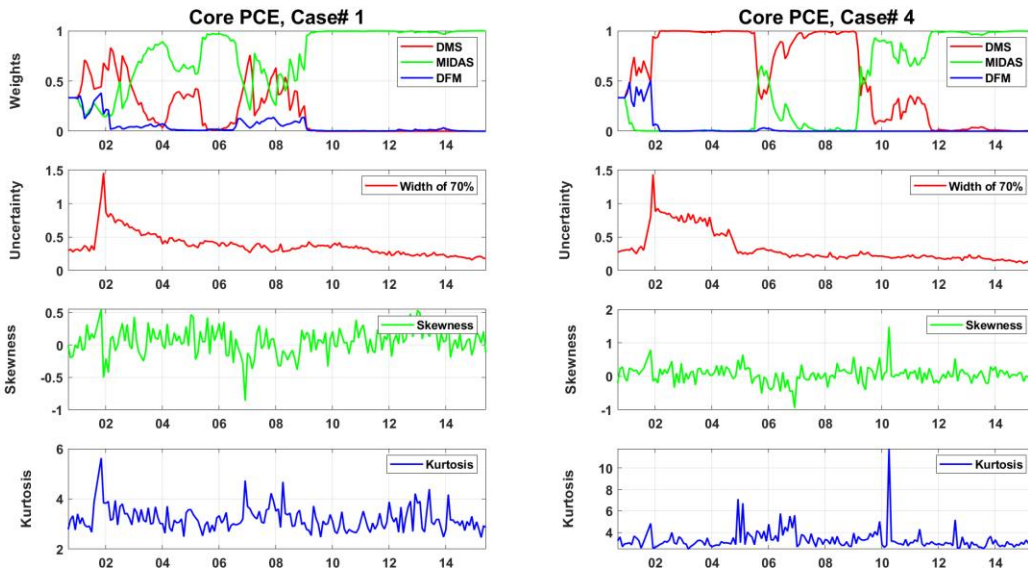


### (d) Core PCE inflation



Notes: The first row of each panel plots the evolution of the weights for the three model classes underlying the grand combination, based on the flexible aggregation strategy and log-score weighting scheme. (Each model class is a combination of multiple model specifications.) The second row plots estimates of dynamic uncertainty, defined as the width of the 70% prediction intervals. The last two rows plot time-varying estimates of skewness and kurtosis. The sample period spans September 2000 through June 2015. **Supplementary Appendix to Real-Time Density Nowcasts of U.S. Inflation: A Model-Combination Approach**[*]

Edward S. Knotek II
Federal Reserve Bank of Cleveland

Saeed Zaman
Federal Reserve Bank of Cleveland
University of Strathclyde, UK

October 20, 2020

## A.1. Description of Mixed-Frequency Models and Simulation Procedures

### A.1.1. MIDAS Model

Following Knotek and Zaman (2017, KZ), a general representation of an ADL-MIDAS model with leads takes the following form,

$$\pi_{t+h} = \alpha_{(h)} + \sum_{J=0}^{P(M)-1} \chi_{j+1,(h)} \, \pi_{t-j} + \sum_{J=0}^{P(M)-1} \gamma_{j+1,(h)} \, Z_{t-j} +$$

$$\beta_h \sum_{J=0}^{P(HF)-1} \omega_{P(HF)-j} \left( \theta_{(h)}^{HF} \right) X_{P(HF)-j,t+1}^{HF} + e_{t+h} \qquad (1)$$

where $Z$ refers to other monthly variables; $P(M)$ refers to the number of lags of the monthly regressors (we set to 1); and $P(HF)$ refers to the number of high-frequency observations, $X_{1,t+1}^{HF}, \ldots, X_{P(HF),t+1}^{HF}$ in month t+1 (i.e., the target nowcast month). The notation (h) indicates that the coefficients are independently estimated for each forecast horizon (h). In nowcasting monthly inflation, $h$ will range from 1 to 2, whereas in nowcasting quarterly inflation, $h$ will range from 1 to 4. An assumption of $\sum_{J=0}^{P(HF)-1} \omega_{P(HF)-j} \left( \theta_{(h)}^{HF} \right) = 1$ helps identify $\beta_h$.

Density construction: Drawing errors from the normal distribution

Let $T$ be the total number of observations (i.e., the length of the estimation window).

1. For $h=1,..,4$

2. Estimate the model specified in equation (1) using nonlinear least squares to obtain the parameter estimates $\hat{\alpha}_{(h)}, \hat{\chi}_{(h)}, \hat{\gamma}_{(h)}, \hat{\beta}_{(h)}(\hat{\theta}_{(h)})$

3. Based on the estimates in the previous step, compute the sequence of residuals $\hat{e}_{t+h}$

4. For $d=1,\ldots,D$

a. Sample $e^*_{t+h}$ from the empirical distribution of $e_{t+h} \sim N(0, var(\ddot{e}_{t+h}))$, where $\ddot{e}_{t+h} =$

   $\left(\frac{T}{T-k}\right)^{0.5} \hat{e}_{t+h}$ and $k$ is the number of regressors in eq. (1).

b. Generate a simulated series $\pi^*_{t+h}$ using

   $$\pi^{*}_{t+h}{}^{(d)} = \hat{\alpha}_{(h)} + \sum_{J=0}^{P(M)-1} \hat{\chi}_{j+1,(h)}\, \pi_{t-j} + \sum_{J=0}^{P(M)-1} \hat{\gamma}_{j+1,(h)}\, Z_{t-j} +$$

   $$\hat{\beta}_{(h)} \sum_{J=0}^{P(HF)-1} \omega_{P(HF)-j}\left(\hat{\theta}^{HF}_{(h)}\right) X^{HF}_{P(HF)-j,t+1} + e^*_{t+h}$$

c. REPEAT

5. The empirical distribution $\{\pi^*_{t+h}\}^{D}_{d=1}$ constitutes the estimate of the density nowcast corresponding to the forecast horizon, $h$

Note that, in step 4a above, the draws are obtained from a distribution of modified residuals because the variance of the modified residuals is a better estimate of the true variance of the least squares estimate of the error term $e_{t+h}$ in equation (1). To further explain why this is the case, recall that the variance of the residuals $\hat{e}_{t+h}$ is the sum of the squared residuals divided by $T$, whereas the variance of the least squares estimate should be divided by $T-k$, where $k$ is the number of regressors in the regression. Therefore, the original series of residuals are rescaled to correct the variance (see Davidson and MacKinnon, 2006).

This simple procedure accounts for shock uncertainty only; i.e., it does not account for the parameter uncertainty. However, in preliminary exercises, the difference in the density accuracy between this procedure and a bootstrapping procedure that also takes into account parameter uncertainty was very small.

**A.1.2. DFM Model**

Our implementation of the mixed-frequency DFM follows Modugno (2013) and KZ. The dynamic factor model takes the general form:

$$y_t = Cf_t + \varepsilon_t, \ \ \varepsilon_t \sim N(0, \Sigma) \tag{2}$$

where $t$ refers to the trading-day frequency, $y_t$ is a vector of observations, $C$ is a block diagonal matrix of factor loadings, $\varepsilon_t$ is a vector of idiosyncratic components, and $f_t$ is a vector of latent common factors following VAR dynamics:

$$Bf_t = A(L)f_{t-1} + u_t, \ u_t \sim N(0, Q), \tag{3}$$

49

where $B$ and $A(L)$ are matrices governing factor dynamics, some of which may be time-varying, and $u_t$ is a vector of residuals.

With monthly, weekly, and daily data, $y_t = [y_t^M, y_t^W, y_t^D]'$, we have three corresponding factors, $f_t = [f_t^M, f_t^W, f_t^D]'$, each of dimension $r \times 1$. The monthly factor(s) $f_t^M$ and the weekly factor(s) $f_t^W$ are a function of the daily factor(s) $f_t^D$. Thus equations (2) and (3) can be written as:

$$\begin{bmatrix} y_t^M \\ y_t^W \\ y_t^D \end{bmatrix} = \begin{bmatrix} C_M & 0 & 0 \\ 0 & C_W & 0 \\ 0 & 0 & C_D \end{bmatrix} \begin{bmatrix} f_t^M \\ f_t^W \\ f_t^D \end{bmatrix} + \begin{bmatrix} \varepsilon_t^M \\ \varepsilon_t^W \\ \varepsilon_t^D \end{bmatrix} \tag{4}$$

and

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_t^M \\ f_t^W \\ f_t^D \end{bmatrix} = \begin{bmatrix} \Theta_t^M & 0 & 0 \\ 0 & \Theta_t^W & 0 \\ 0 & 0 & A_D \end{bmatrix} \begin{bmatrix} f_{t-1}^M \\ f_{t-1}^W \\ f_{t-1}^D \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ u_t^D \end{bmatrix} \tag{5}$$

The matrices $C_M$, $C_W$, and $C_D$ are the loadings for the monthly, weekly, and daily variables. $\Theta_t^M$ and $\Theta_t^W$ are time-varying coefficients: $\Theta_t^M$ is equal to zero the day after the release of the monthly data and is equal to one elsewhere; similarly, $\Theta_t^W$ is equal to zero the day after the release of the weekly data and is equal to one elsewhere.

Assuming that the monthly variables and weekly variables in our system at any time $t$ represent a stock (i.e., a snapshot), accordingly the monthly first difference (or growth rate) and weekly first difference (or growth rate) of those variables can be formed by summing up their respective daily first differences (or growth rates).

To produce forecasts far into the future, the daily factors are forecast via the transition equation (5) and are translated to daily nowcasts and aggregated to weekly and monthly nowcasts via equation (4). Following Modugno (2013), we estimate the model with the expectation-maximization (EM) algorithm as detailed in Bańbura and Modugno (2014).

Density construction: Standard bootstrapping procedure

Our procedure closely follows the factor model bootstrapping procedure detailed in Aastveit et al. (2014).

Let $T$ be the number of observations (i.e., the length of the estimation window).

1. Estimate the model specified in equations (2) and (3) to obtain parameter estimates $\hat{A}^{(0)}$, $\hat{B}^{(0)}$, $\hat{C}^{(0)}$, $\hat{Q}^{(0)}$, $\hat{\Sigma}^{(0)}$, $\hat{f}^{(0)}$. Let $\hat{A} = \hat{A}^{(0)}$, $\hat{B} = \hat{B}^{(0)}$, $\hat{C} = \hat{C}^{(0)}$, $\hat{Q} = \hat{Q}^{(0)}$, and $\hat{\Sigma} = \hat{\Sigma}^{(0)}$.

2. For $d=1,\ldots,D$, do the following

   a. Simulate draws $u_t^*$ from the empirical distribution of $u_t \sim N(0, \hat{Q})$

   b. Generate bootstrap series $f_t^*$ using $\hat{B} f_t^* = \hat{A}(L) f_{t-1}^* + u_t^*$ where $u_t^*$ is obtained in the previous step

   c. Simulate draws $\varepsilon_t^*$ from the empirical distribution of $\varepsilon_t \sim N(0, \hat{\Sigma})$

   d. Generate bootstrap series $y_t^*$ using $y_t^* = \hat{C} f_t^* + \varepsilon_t^*$ where $\varepsilon_t^*$ and $f_t^*$ are obtained in the previous two steps.

   e. Using $y_t^*$ re-estimate the model in equations (2) and (3) to obtain an updated set of parameter and factor estimates, $\hat{B}^{(d)}$, $\hat{C}^{(d)}$, $\hat{Q}^{(d)}$, $\hat{\Sigma}^{(d)}$, $\hat{f}^{(d)}$. Set $\hat{A} = \hat{A}^{(d)}$, $\hat{B} = \hat{B}^{(d)}$, $\hat{C} = \hat{C}^{(d)}$, $\hat{Q} = \hat{Q}^{(d)}$, and $\hat{\Sigma} = \hat{\Sigma}^{(d)}$

   f. Based on the parameter and factor estimates obtained in the previous step construct forecasts of factors via equation (3), which are then aggregated up to produce nowcasts (and forecasts) for monthly inflation, $\pi_{t+h}^{*}{}^{(d)}$ via equation (2).

   g. REPEAT

3. The empirical distribution $\{\pi_{t+h}^*\}_{d=1}^{D}$ constitutes the estimate of the density nowcast corresponding to the forecast horizon, $h$

### A.1.3. DMS Model

As discussed in the body of the paper, the DMS model is essentially a collection of univariate and multivariate regressions applied to disaggregate components and aggregate inflation. To appropriately account for uncertainty, we devise two separate bootstrapping algorithms for univariate and multivariate formulations. The difference between these two algorithms is only slight but it helps improve the density accuracy of monthly inflation.

We first describe the general-purpose bootstrap algorithm for the multivariate regression followed by the description for the univariate regression.

A general representation for a multivariate regression can be written as follows,

$$y_t = \beta_0 + \alpha X_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2) \tag{6}$$

Assume that $\hat{\beta}_0, \hat{\alpha}, \hat{\sigma}^2$ are the OLS estimates obtained through the estimation of equation (6) over the sample 1,…,$T$. $\hat{\varepsilon}_t$ are the least squares residuals with mean 0 and variance $\hat{\sigma}^2$.

Density construction, algorithm 1: Wild block bootstrap for density forecasts

For $d$=1,….,$D$ do the following.

1. Construct a transformed series of residuals $\{\ddot{\varepsilon}_t\}_{t=1}^T$ from the OLS residuals $\{\hat{\varepsilon}_t\}_{t=1}^T$, where $\ddot{\varepsilon}_t = h(\hat{\varepsilon}_t)u_t$ and $u_t \sim N(0,1)$. $h$ is a transformation function that modifies the original least squares residuals to correct them for possible heteroscedasticity. Various choices for $h$ have been suggested in the literature. Following Chernick and LaBudde (2011, Ch. 6, Section 6.6), we set

$$h(\hat{\varepsilon}_t) = \frac{\hat{\varepsilon}_t}{1-H} \quad \text{where } H = X(X'X)^{-1}X'$$

   We also tried $h(\hat{\varepsilon}_t) = \frac{\hat{\varepsilon}_t}{(1-H)^{1/2}}$, another widely used transformation.

2. Sampling from $\ddot{\varepsilon}$:
   a. To correct for possible serial correlation (following Aastveit et al., 2014), we draw blocks of consecutive errors from $\ddot{\varepsilon}$. We define the block size, $b_{size} = 4$; it is common to set it greater than or equal to the forecast horizon; $T$ is the number of observations; and $b_{number} = ceil(\frac{T}{b_{size}})$, is an integer that denotes the number of non-overlapping blocks of consecutive errors.
   b. For $l$=1,…,$b_{size}$ and $j$=1,…,$b_{number}$ construct the bootstrap sample for $y^*$

   $$y^*_{(j-1)b_{size}+l} = \hat{\beta}_0 + \hat{\alpha}X_{(j-1)b_{size}+l} + \varepsilon^*_{(j-1)b_{size}+l}$$

   where $\varepsilon^*_{(j-1)b_{size}+l} = \ddot{\varepsilon}_{(j-1)b_{size}+l} \cdot \delta_j$, and $\delta_j$ is set as a Rademacher variable, following Davidson and Flachaire (2008) and Aastveit et al (2014):

   $$\delta_j = \begin{cases} +1, & with\ probability\ 0.5 \\ -1, & with\ probability\ 0.5 \end{cases}$$

   We also experimented with $\delta_j \sim N(0,1)$, but doing so slightly worsened the accuracy of the density forecasts.

3. Based on the bootstrap sample $y^*$ (constructed in the previous step), re-estimate the model in equation (6) to obtain updated estimates $\hat{\beta}_0^{(d)}, \hat{\alpha}^{(d)}, \hat{\sigma}^{2(d)}$.

4. Use $\hat{\beta}_0^{(d)}$ and $\hat{\alpha}^{(d)}$ in equation (6) to generate iterative forecasts, $\hat{y}_{t+h}^{(d)}$ up to $h$ periods ahead. (We also experimented with a modified step 4: when generating iterative forecasts $\hat{y}_{t+h}^{(d)}$ we drew from $\varepsilon^* \sim N(0, var(\varepsilon^*))$ for each $h$. This alternative made no difference to the overall results.)

5. REPEAT

6. The empirical distribution of $\{\hat{y}_{t+h}\}_{d=1}^{D}$ constitutes our estimate of the $h$-step-ahead density.


Next, we describe the algorithm that we apply to the univariate AR regressions. A general representation for a univariate AR regression is:

$$y_t = \beta_0 + \sum_{j=1}^{P} \alpha_j y_{t-j} + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2) \tag{7}$$

Assume that $\hat{\beta}_0$, $[\hat{\alpha}_j]_{j=1}^{P}$, $\hat{\sigma}^2$ are the OLS estimates obtained through the estimation of equation (7) over the sample consisting of $1, \ldots, T$ observations. $\hat{\varepsilon}_t$ are the least squares residuals with mean 0 and variance $\hat{\sigma}^2$.


Density construction, algorithm 2: Parametric bootstrap for density forecasts

For $d=1, \ldots, D$ do the following.

1. Construct a transformed series of residuals $\{\ddot{\varepsilon}_t\}_{t=1}^{T}$ from the residuals $\{\hat{\varepsilon}_t\}_{t=1}^{T}$, where $\ddot{\varepsilon}_t = \left(\frac{T}{T-k}\right)^{0.5} \hat{\varepsilon}_t$ and $k$ is the number of regressors, in this case $k = P + 1$; P is the number of lags of the dependent variable. We also experimented with $\ddot{\varepsilon}_t = h(\hat{\varepsilon}_t)u_t$ and $u_t \sim N(0,1)$ but this produced inferior nowcasts.

2. Sample a sequence of $\{\varepsilon^*\}_{t=1}^{T}$ from $\ddot{\varepsilon} \sim N(0, var(\ddot{\varepsilon}))$ and then construct a bootstrap sample of $\{y^*\}_{t=1}^{T}$ using

$$y_t^* = \hat{\beta}_0 + \sum_{j=1}^{P} \hat{\alpha}_j y_{t-j}^* + \varepsilon_t^*$$

3. Based on the bootstrap sample $y^*$ re-estimate the model in equation (7) to obtain updated estimates $\hat{\beta}_0^{(d)}$, $[\hat{\alpha}_j^{(d)}]_{j=1}^{P}$

53

4. Use $\hat{\beta}_0^{(d)}, \left[\hat{\alpha}_j^{(d)}\right]_{j=1}^P$ in equation (7) to iteratively generate forecasts, $\hat{y}_{t+h}^{(d)}$ up to $h$ periods ahead. (We also experimented with a modified step 4: when generating iterative forecasts $\hat{y}_{t+h}^{(d)}$ we draw from $\ddot{\varepsilon} \sim N(0, var(\ddot{\varepsilon}))$ for each $h$. This alternative made no difference to the overall results.)

5. REPEAT

6. The empirical distribution of $\{\hat{y}_{t+h}\}_{d=1}^D$ constitutes our estimate of the $h$-step-ahead density.

Using the same notation as in KZ, the general representation of the DMS model for monthly headline (or core) inflation is

$$A_{s(t)} \mathbf{Z_t} = B_{s(t)} + C_{s(t)} \mathbf{X_t} + \sum_{j=1}^J D_{j,s(t)} \mathbf{Z_{t-j}} + \varepsilon_{s(t)} \qquad (8)$$

where $\mathbf{Z}_t$ is an $n \times 1$ vector of aggregates, $\mathbf{X}_t$ is an $m \times 1$ vector of disaggregates that are informative over $\mathbf{Z}_t$, and $\varepsilon_{s(t)} \sim N(\mathbf{0}, \mathbf{\Sigma})$. The coefficient matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}_j$ are $n \times n$, $n \times 1$, $n \times m$, and $n \times n$, respectively, and are allowed to vary over time depending on the available information set, denoted $s(t)$; in particular, $\mathbf{C}$ and $\mathbf{D}_j$ measure the weights put on the disaggregates and lagged aggregates, respectively.

Nowcasting core inflation

Let $\mathbf{Z}_t = [\pi_t^{\text{Core CPI}}, \pi_t^{\text{Core PCE}}]'$ and $\mathbf{X}_t = \mathbf{0}$ in equation (8). We specify two possible regression specifications for core inflation. The first one is a univariate AR, and the second is a bridge equation (i.e., multivariate regression), which regresses core CPI on core PCE and a constant. Conditional on the available information, equation (8) reduces to either a univariate AR or a combination of a univariate AR and bridge equation.

Univariate AR: $\pi_t^{Core} = \beta_0 + \sum_{j=1}^P \alpha_j \pi_{t-j}^{Core} + \varepsilon_t$.

Bridge equation: $\pi_t^{CorePCE} = \gamma_0 + \theta \pi_t^{CoreCPI} + u_t$.

In cases where we have an additional monthly release of core CPI compared with core PCE, and only core PCE remains to be nowcasted: (1) The forecasts of core CPI are produced using a univariate AR, and algorithm 2 is used to produce density forecasts. (2) The nowcast of core PCE is produced using a bridge regression. The forecasts up to $h$ steps ahead are produced using a univariate regression that treats the nowcast from a bridge regression as an initial value.

54

To produce density estimates (nowcasts and forecasts), algorithm 2 is used. In all other cases, both core CPI and core PCE are nowcasted (and forecasted) using a univariate AR model. The density estimates are computed based on algorithm 2.

Nowcasting food inflation

Nowcasts for food inflation are produced and used to nowcast headline inflation in all cases except: (1) when we are unable to produce a nowcast for gasoline inflation, and (2) when we have an additional reading for PCE inflation ($\pi^{PCE}$) compared to CPI inflation ($\pi^{CPI}$). Similar to core PCE, we adopt a parsimonious approach to produce nowcasts of food inflation by simply estimating a univariate AR,

$$\pi_t^{food} = \beta_0 + \sum_{j=1}^{P} \alpha_j \pi_{t-j}^{food} + \varepsilon_t$$

Density nowcasts (and forecasts) are produced using algorithm 2.

Nowcasting gasoline inflation

Following KZ, we generate nowcasts (and forecasts) for gasoline inflation based on the availability of weekly gasoline prices and daily oil prices. If weekly gasoline prices are available in the current month, these form the basis for that month's gasoline inflation nowcast. We use a daily random walk in oil prices to extend (i.e., forecast) the oil price series by one additional month. If oil price data or a forecast for oil prices is available for a month but gasoline prices are not available from within that month, then we produce nowcasts or forecasts for gasoline inflation ($\hat{\pi}^{gasoline}$) via a two-stage regression procedure (see KZ for details). In the first stage, a longer-run relationship between monthly gasoline prices and monthly oil prices is assumed via the following regression:

$$P_{t-1}^{\text{Gasoline (NSA)}} = \alpha + \beta P_{t-1}^{\text{Oil}} + e_{1,t-1} \tag{9}$$

Denote $\tilde{P}_{t-1}^{Gasoline(NSA)}$ as the fitted monthly gasoline prices obtained by estimating equation (9). In the second stage, we estimate an error correction model that uses the lagged gap between gasoline prices and their predicted (longer-run) values obtained in the first stage via the following regression:

$$\Delta P_{t-1}^{\text{Gasoline (NSA)}} = b\Delta P_{t-1}^{\text{Oil}} + c\left(P_{t-2}^{\text{Gasoline (NSA)}} - \tilde{P}_{t-2}^{\text{Gasoline (NSA)}}\right) + e_{2,t-1} \qquad (10)$$

Using the estimated coefficients in equations (9) and (10) and iterating forward equations (9) and (10) we generate $\hat{P}_{t-1+h}^{Gasoline(NSA)}$ and $\widehat{\Delta P}_{t-1+h}^{Gasoline(NSA)}$ and in turn estimates of $\hat{\pi}_{t-1+h}^{Gasoline(NSA)}$. The estimates are seasonally adjusted to produce $\hat{\pi}_{t-1+h}^{Gasoline}$. The density forecasts are produced by applying algorithm 1 sequentially to equations (9) and (10). For each simulation $d$, $\hat{\pi}_{t-1+h}^{Gasoline(NSA),d}$ is seasonally adjusted to obtain the corresponding $\hat{\pi}_{t-1+h}^{Gasoline,d}$.

Nowcasting headline inflation

Let $\mathbf{Z}_t = [\pi_t^{\text{CPI}}, \pi_t^{\text{PCE}}]'$ and $\mathbf{X}_t = \left[\pi_t^{\text{Core CPI}}, \pi_t^{\text{Core PCE}}, \pi_t^{\text{Food}}, \pi_t^{\text{Gasoline}}\right]'$. In cases where we have an additional release of $\pi_t^{CPI}$, equation (8) reduces to a bridge equation for $\pi_t^{PCE}$ and a univariate AR for $\pi_t^{CPI}$.

Univariate AR: $\pi_t^{CPI} = \beta_0 + \sum_{j=1}^{P} \alpha_j \pi_{t-j}^{CPI} + \varepsilon_t$.

Bridge equation: $\pi_t^{PCE} = \gamma_0 + \theta \pi_t^{CPI} + u_t$.

Density estimates are constructed using algorithm 2. In cases where we have nowcasts of $\hat{\pi}_t^{Gasoline}$, equation (8) reduces to a multivariate regression,

$$\pi_t^{CPI} = b_1 + c_{11}\pi_t^{CoreCPI} + c_{13}\pi_t^{Food} + c_{14}\pi_t^{Gasoline} + e_t^{CPI} \qquad (11)$$

$$\pi_t^{PCE} = b_2 + c_{22}\pi_t^{CorePCE} + c_{23}\pi_t^{Food} + c_{24}\pi_t^{Gasoline} + e_t^{PCE} \qquad (12)$$

The density nowcasts (and forecasts) for CPI and PCE inflation are produced by separately applying algorithm 1 to equations (11) and (12). In very few cases, where we lack estimates of $\hat{\pi}_t^{Gasoline}$ and do not have an additional reading for $\pi_t^{CPI}$, equation (8) reduces to univariate AR,

$$\pi_t^{CPI} = \beta_1 + \sum_{j=1}^{P} \alpha_j^{CPI} \pi_{t-j}^{CPI} + \varepsilon_t^{CPI} \qquad (13)$$

$$\pi_t^{PCE} = \beta_2 + \sum_{j=1}^{P} \alpha_j^{PCE} \pi_{t-j}^{PCE} + \varepsilon_t^{PCE} \qquad (14)$$

The density nowcasts (and forecasts) are generated by separately applying algorithm 2 on equations (13) and (14).

In all of our simulation procedures, $D$=500. Early experimentation suggested that we would normally obtain similar results if we instead set $D$=1000.

## A.2. Mechanics of Density Combination and Graphical illustration

Assume at time $t$, we have $i = 1, ..., M$ (potentially different) empirical distributions $f_{i,t}(y_t)$ for a variable $y_t$. We wish to combine them using a given set of $M$ weights, $w_{i,t}$.

Step 1: Looking across all the $M$ empirical distributions, $f_{i,t}(y_t)$, determine the (global) minimum value and (global) maximum value of $y_t$. Denote $x_t^{min}$ as the minimum value and $x_t^{max}$ as the maximum value.

Step 2: Define a grid $x_t \in \{x_t^{min}, ..., x_t^{max}\}$ of $S$ equally spaced intervals such that $x_{k-1} < x_k$.

Step 3: Transform each of the $i = 1, ..., M$ empirical distributions $f_{i,t}(y_t)$ to a probability density function (pdf), $p_{i,t}(y_t)$ using the grid $x_t$ as the domain. The Gaussian kernel function (Matlab: ksdensity function) is applied to construct a smoothed $p_{i,t}(y_t)$. Using the same grid $x_t$ to construct each of the $M$ pdfs will guarantee that all the pdfs that are to be combined together at time $t$ have the same domain; that is, they are all positioned over the same grid.

Step 4: With all pdfs positioned over the same domain (grid), the combination can be achieved by simply adding up the $M$ different densities using the corresponding weights $w_{i,t}$ (for linear combination) or raised to a power of $w_{i,t}$ for a log pool combination. The combined density $g_t(y_t)$ will also be positioned over the same grid (domain) $x_t$ as the $M$ individual densities.

We set $S$=500. Early experimentation suggested that the results were very similar if we set $S$=1000.

Note that our procedure dynamically adjusts the grid $x_t$ at each time $t$. Alternatively, we could just set it to a predefined interval but then the interval has to be wide enough to encompass all the individual empirical distributions for all $t$=1,…,$T$ (i.e., over the evaluation sample). Given the breadth of our analysis, including the number of variables considered and both monthly and quarterly rates, having a grid that adjusts dynamically was more efficient for our application.

In implementing our algorithm, we have benefitted from and are grateful for the PROFOR Matlab toolbox (developed by researchers at the Norges Bank, Bank of England, and Warwick Business School). We have modified some of the functions of the toolbox to fit our needs.

### A.3. Comparing Properties of Grand Combinations across Weighting Schemes

Figure 8 in the body of the paper shows the weights and higher-order moments from using the log score weighting scheme to generate the stage 1 and stage 2 combinations. Figures A15, A16, and A17 show the weights and higher-order moments from the CMG, Ganics, and CRPS weighting schemes, respectively. We summarize six key results from this comparison.

First, for CPI inflation and PCE inflation, the DMS combination gets the highest weight in all weighting schemes with the exception of Ganics. Furthermore, the DMS maintains its ranking with incoming information over the course of the month.

Second, the CMG and Ganics grand combinations for CPI inflation and PCE inflation provide stronger evidence of both kurtosis and skewness than the combination based on log score weights. This finding is associated with the grand combination being composed of more diverse components in these cases; that is, the DMS combination, the DFM combination, and the MIDAS combination are all assigned nonzero weights in the grand combination. Different weighting schemes can lead to combinations with very different compositions, as is evident by very different profiles of the weights assigned to the three model classes over time. In general, the greater the diversity in the composition of the grand combination, the greater is the evidence of skewness and kurtosis.[38] But greater flexibility in terms of accommodating skewness and kurtosis does not necessarily translate into improved accuracy. We say this because for CPI inflation the grand combination based on the log-score weighting scheme is more accurate than grand combinations based on other schemes, yet it displays less evidence of skewness and kurtosis on average compared with other grand combinations. This improved accuracy is mainly coming from the significantly more accurate mean of the density nowcast constructed from the log-score scheme, which puts high weight on the stage 1 DMS combination, compared with grand combinations based on other weighting schemes.

Third, in the case of core inflation, the patterns observed in the properties of the grand combination are generally comparable across the various weighting schemes, even though the

---

[38] We highlight a result in regard to grand combinations for CPI inflation (case 4) produced using the log score weighting scheme (see Figure 8) and the CMG weighting scheme (Figure A15). Both schemes assign a weight of 100% to DMS at least in the last few years of the evaluation sample, yet the profiles of the kurtosis property of the grand combinations across the two schemes are very different for this period. This finding arises in part because the underlying composition of the two respective stage 1 DMS combinations is quite different; see Figure A18.

weights assigned to the densities of the three modeling classes differ. This result stems from the fact that the estimates of density nowcasts for core inflation are generally similar across the different modeling classes; so irrespective of the approach used to combine the component density nowcasts, the resulting estimates of the combined density nowcasts are similar. This latter pattern also explains the comparable accuracy results for core inflation shown in Figures 4 and 5 (especially in the case of core CPI). Relatedly, the weight profiles across different weighting schemes (for core inflation) indicate a high incidence of fast switching across the three combinations. The evidence of time-varying switching across density combinations highlights the importance of combining density estimates from a range of models to circumvent the instability issues of using a single model.

Fourth, the CRPS weighting scheme assigns positive weights to the three combinations across all inflation measures and at all representative dates (shown for cases 1 and 4), reflecting the generous assessment of the CRPS metric. In the case of core inflation, the weights are pretty evenly distributed across the DMS, DFM, and MIDAS combinations.

Fifth, in our application, the two optimal combination weighting schemes (CMG and Ganics) yield weight profiles that are remarkably different, especially in the case of CPI inflation and PCE inflation. However, the different profiles are not unexpected, given the earlier results that showed MIDAS and DFM combinations producing well-calibrated densities compared with DMS, which tends to do quite well in relative accuracy scoring. The weights produced from the Ganics approach display quite a bit of variability early in the sample. This variability is also present to a degree in the results reported in Ganics (2017) using industrial production data.

## A.4. Additional References

Chernick, Michael R., and Robert A. LaBudde. 2011. *An Introduction to Bootstrap Methods with Applications to R.* John Wiley and Sons.

Davidson, Russell, and Emmanuel Flachaire. 2008. "The Wild Bootstrap, Tamed at Last." *Journal of Econometrics* 146(1): 162-69. https://doi.org/10.1016/j.jeconom.2008.08.003.

Davidson, Russell, and James G. MacKinnon. 2006. "Bootstrap Methods in Econometrics." In Kerry Patterson and Terence C. Mills, eds. *Palgrave Handbook of Econometrics: Vol 1, Econometric Theory*.

Figure A1: Illustration of Combining Densities with Linear and Log Opinion Pools
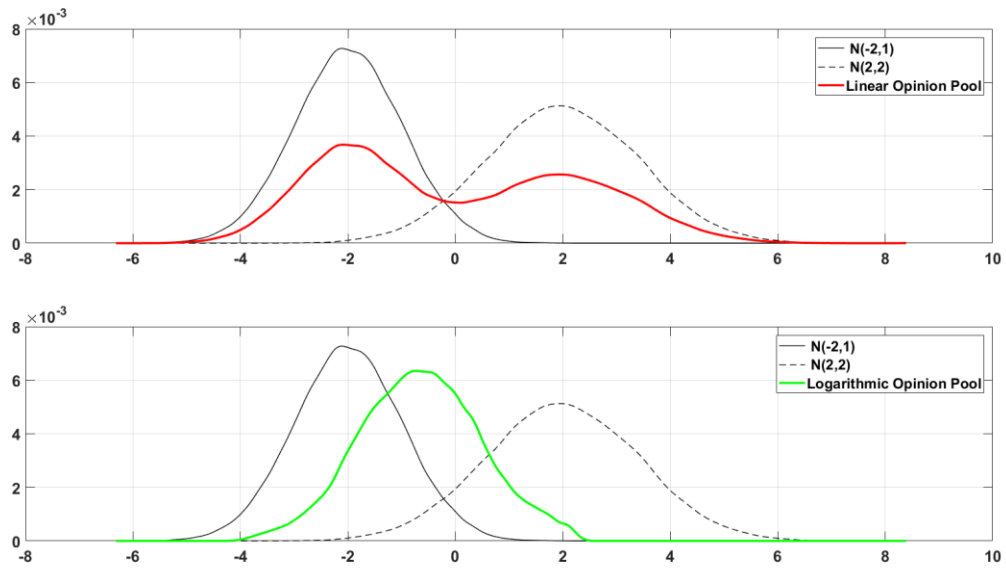


Notes: A simple example (motivated by Kascha and Ravazzolo, 2010) on combining two densities with very different mean and variances via two different functional forms.

## Figure A2: Example Stage 1 DMS Combination
### Month-over-month inflation (%)
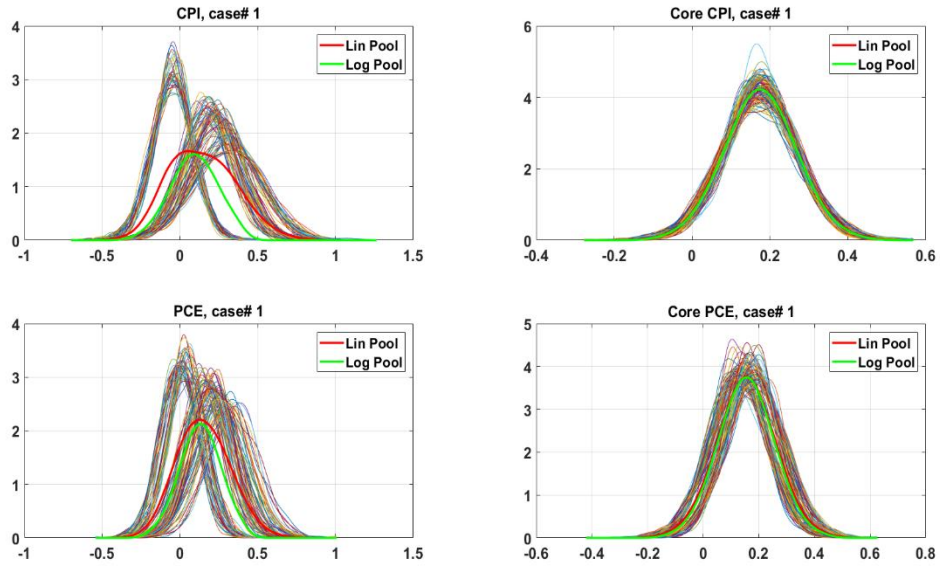


Notes: Single specification density nowcasts (thin lines) underlying the stage 1 DMS combination, linear pool nowcasts (thick red lines), and log pool nowcasts (thick green lines) for case 1 for the month of September 2000.

Figure A3: Comparisons between Single Specifications vs. Stage 1 Combinations



**Density Accuracy: Monthly Inflation (12-month % change)**
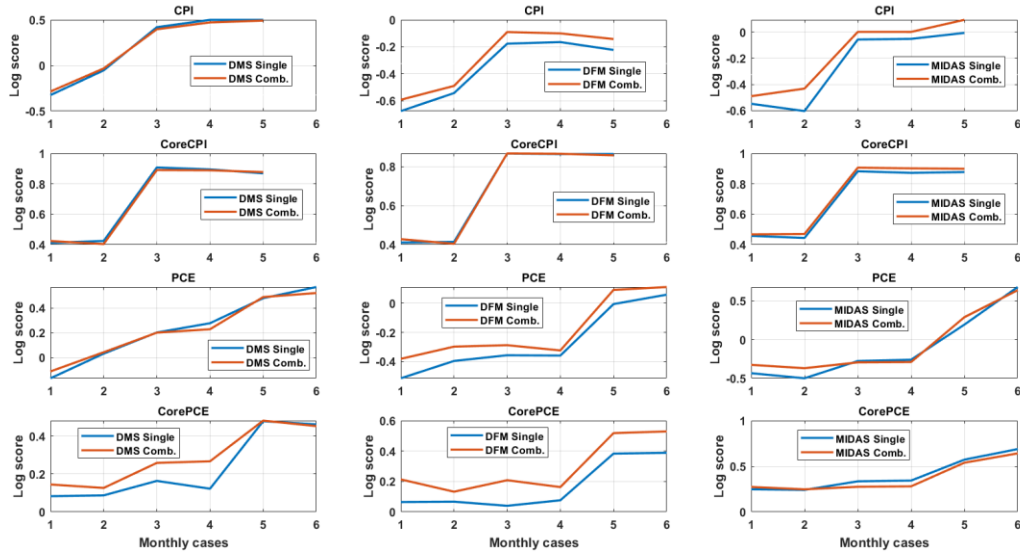
Notes: Average log scores at different nowcast origins for single specifications and stage 1 combinations within model classes. The evaluation sample is September 2000 through June 2015. We exclude September 2001 and October 2001 from the average log score calculations for PCE inflation and core PCE inflation.

Figure A4: Real-Time Density Nowcasts

(a) CPI inflation



**CPI, Case# 1**

**CPI, Case# 4**

(b) Core CPI inflation



**Core CPI, Case# 1**

**Core CPI, Case# 4**
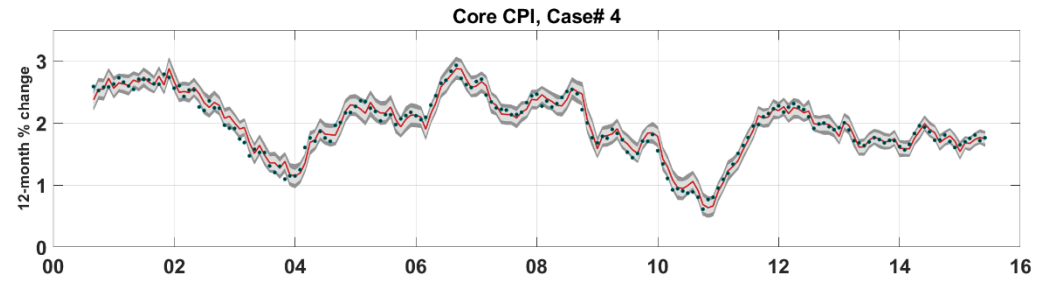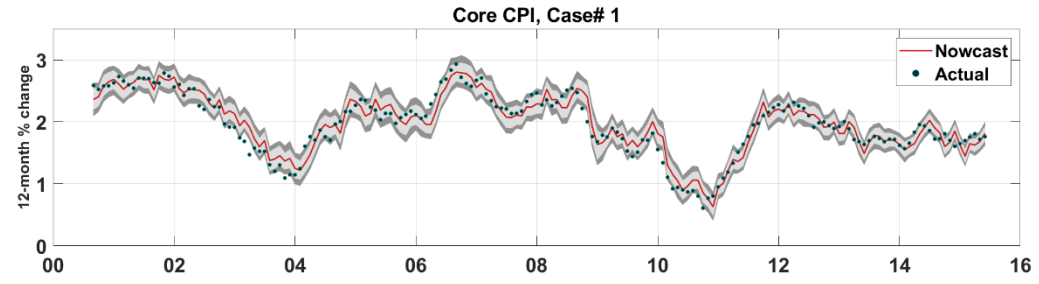
# Figure A4: Real-Time Density Nowcasts (continued)
## (c) PCE inflation

**PCE, Case# 1**



**PCE, Case# 4**



## (d) Core PCE inflation

**Core PCE, Case# 1**



**Core PCE, Case# 4**
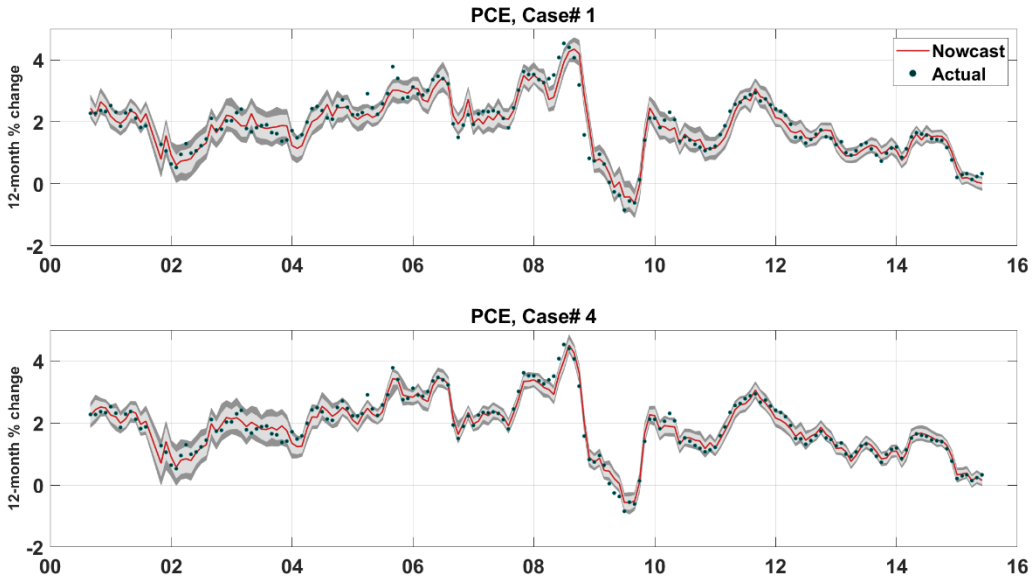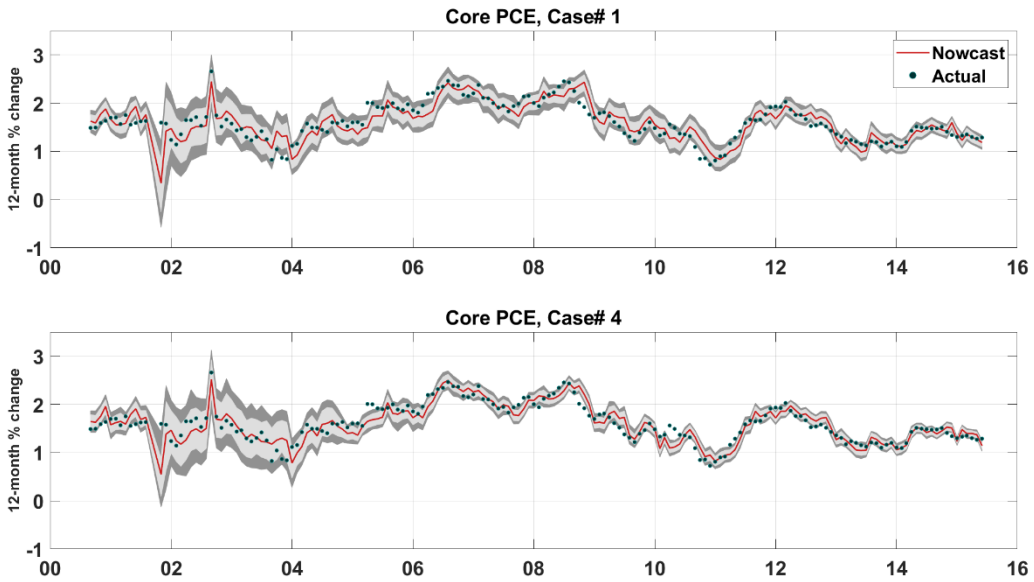


Notes: The figure shows the out-of-sample nowcasts generated using real-time data from the grand combination with the log score weighting scheme and the flexible aggregation strategy at two different points in each month (case 1 and case 4) for the 12-month trailing inflation rate. The shaded areas represent 70% and 90% prediction intervals. The sample period spans September 2000 through June 2015.

Figure A5: Weights Underlying Grand Combination based on Ganics Weighting Scheme
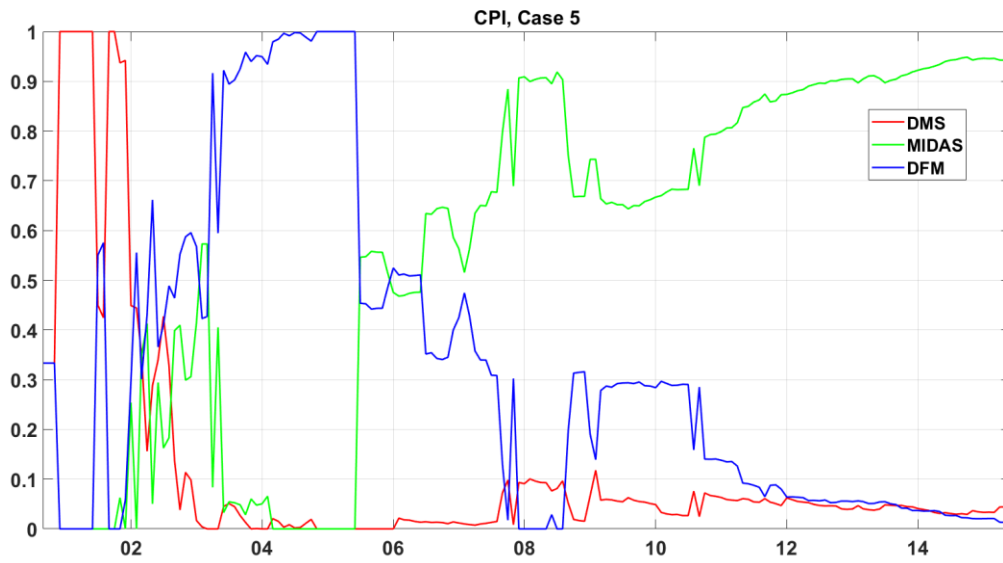


Notes: The figure plots the evolution of the weights applied to each of the stage 1 density combinations from the DMS, MIDAS, and DFM model classes to form the stage 2 combination, based on nowcasts generated for monthly (year-over-year) inflation at case 5 for nowcasting CPI inflation. The sample period spans September 2000 through June 2015.

Figure A6: Density Performance of Grand Combination vs. Its Components: Month-Over-Month Inflation
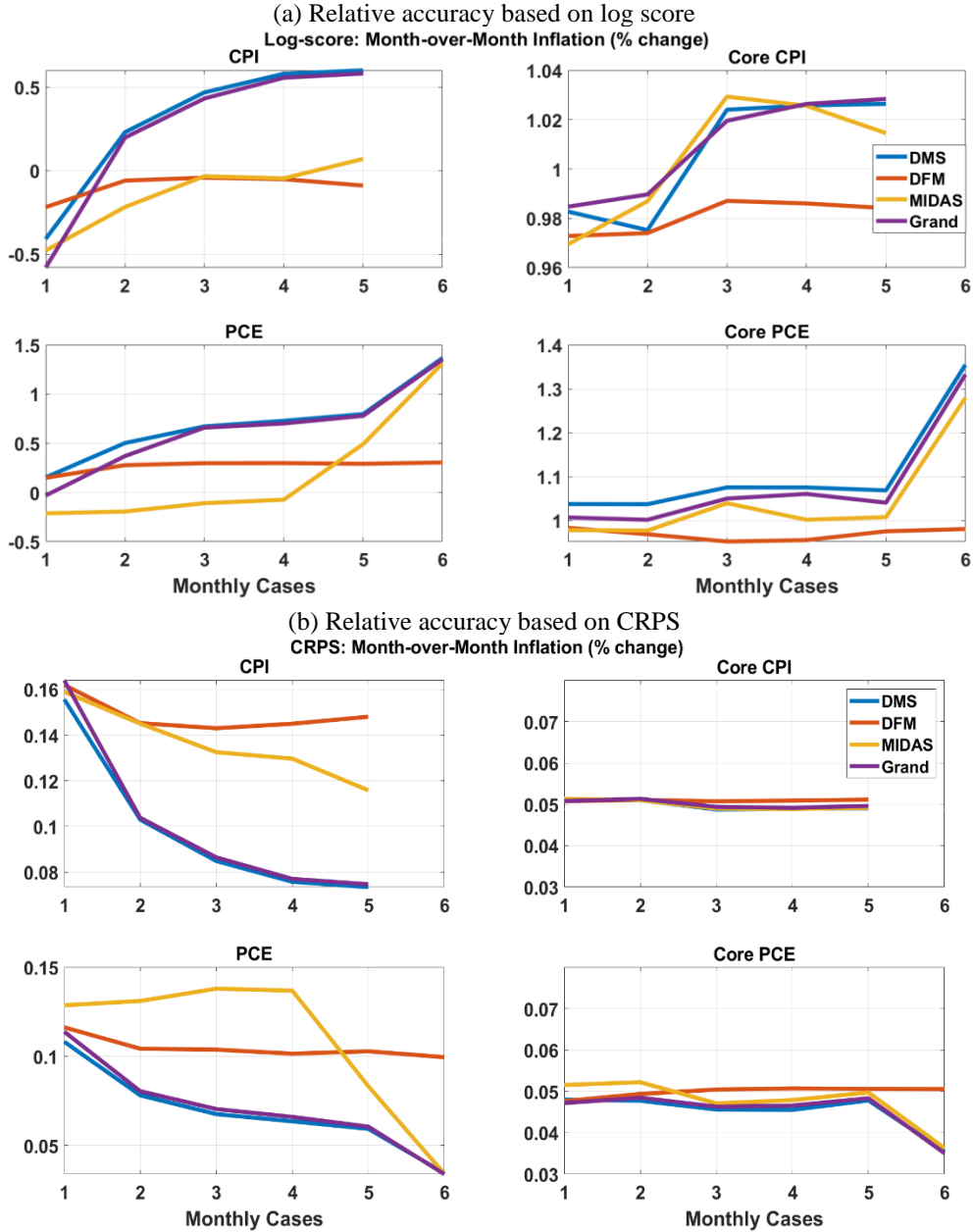
(a) Relative accuracy based on log score



(b) Relative accuracy based on CRPS



Notes: The top panel plots the average log score and the bottom panel plots the average CRPS for the grand combination based on the log score weighting scheme and combinations based on the DMS model class, MIDAS model class, and DFM model class, where each individual model class uses the log score weighting scheme. The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure A7: Point Nowcasting Performance, Grand Combination vs. DMS: Month-Over-Month Inflation
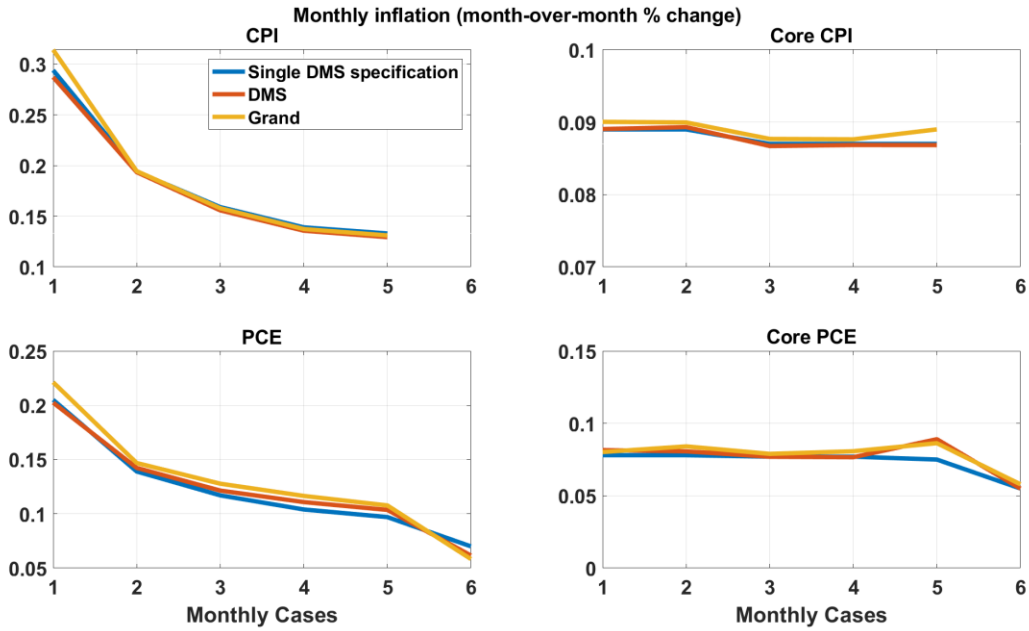


Notes: The figure plots the RMSE for the grand combination based on log score and using the flexible aggregation strategy; the stage 1 combination from the DMS model class; and a single specification from the DMS model class based on Knotek and Zaman (2017). The cases reflect the point in time when each nowcast was made relative to the target nowcast month; see Table 2. The evaluation sample runs from September 2000 through June 2015; we omit September 2001 and October 2001 for PCE inflation and core PCE inflation calculations.

Figure A8: Density Performance of Grand Combination vs. Its Components: Quarterly Inflation
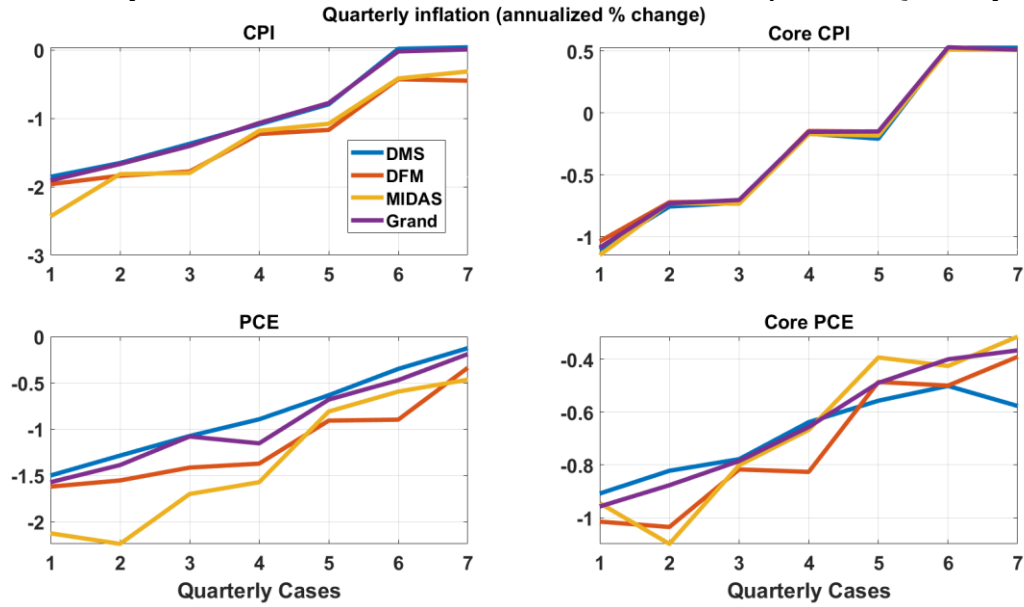


Notes: Average log score for the grand combination based on the log score weighting scheme and combinations based on the DMS model class, MIDAS model class, and DFM model class, where each individual model class uses the log score weighting scheme. The evaluation sample runs from 2000Q4 through 2015Q2; we omit 2001Q3 and 2001Q4 for PCE inflation and core PCE inflation calculations.

Figure A9: Point Nowcasting Performance, Grand Combination vs. Other Combinations and Single DMS Specification: Quarterly Inflation
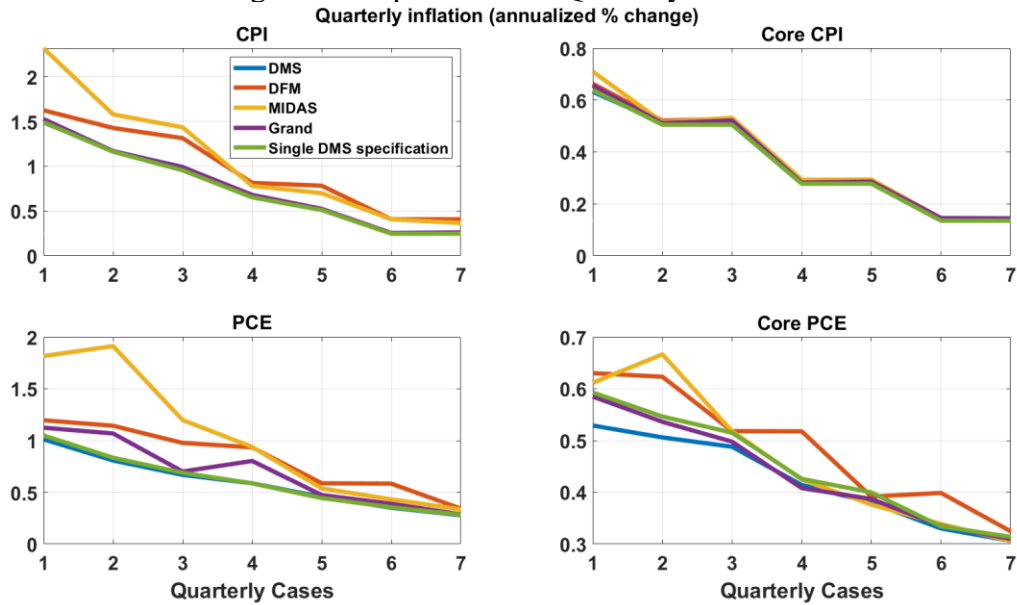


Notes: The figure plots the RMSE for the grand combination based on log score and using the flexible aggregation strategy; the stage 1 combinations from the DMS model class, DFM model class, and MIDAS model class; and a single specification from the DMS model class based on Knotek and Zaman (2017). The cases reflect the point in time when each nowcast was made relative to the target nowcast quarter; see Table 2. The evaluation sample runs from 2000Q4 through 2015Q2; we omit 2001Q3 and 2001Q4 for PCE inflation and core PCE inflation calculations.

Figure A10: Weights for Stage 1 DMS Combinations, Log Score Weighting Scheme
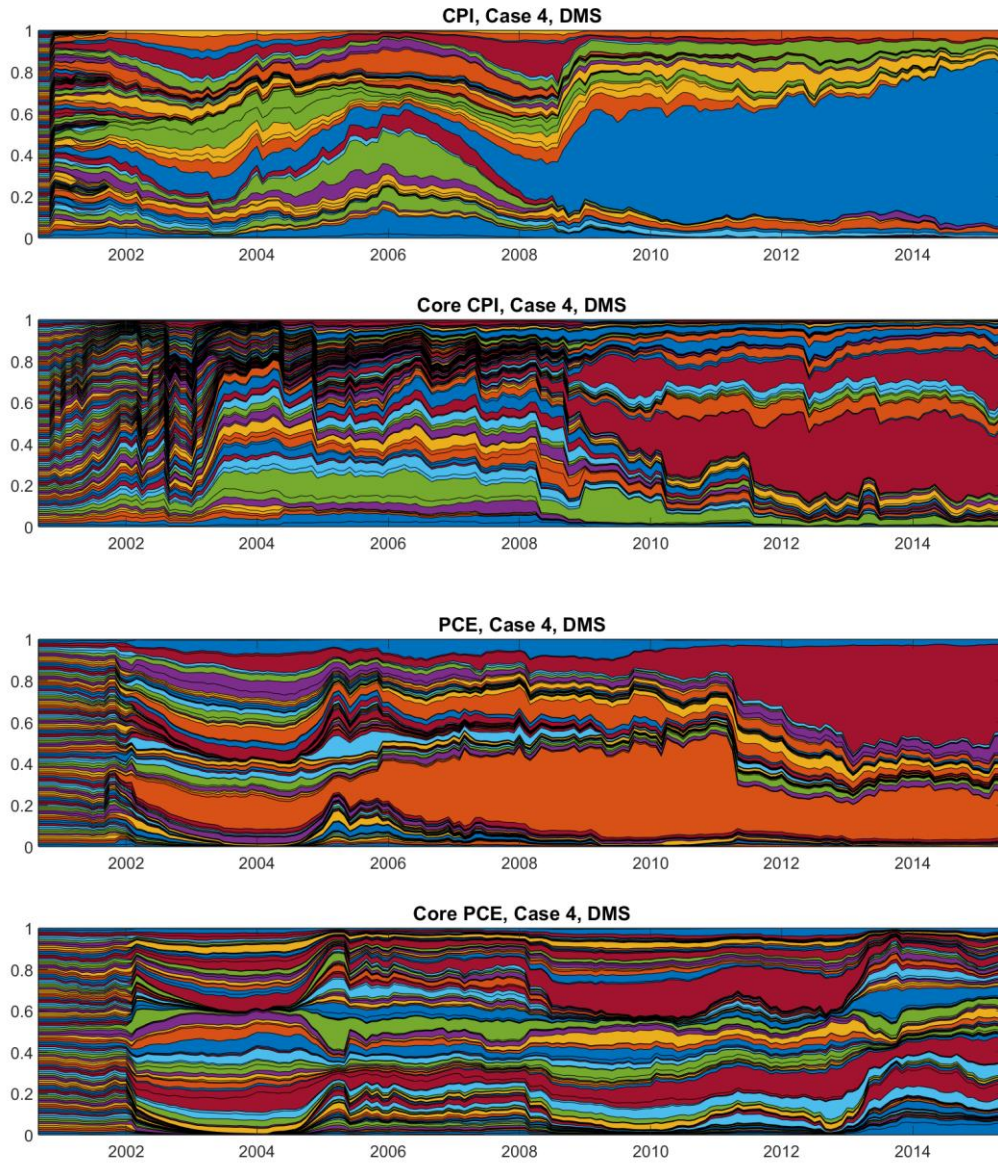


Notes: The figure plots the evolution of the weights for underlying individual candidate densities for the stage 1 DMS combination at case 4. Each color shade represents a particular individual candidate density. There are 108 candidate densities. The sample period spans September 2000 through June 2015.

Figure A11: Weights for Stage 1 DFM Combinations, Log Score Weighting Scheme
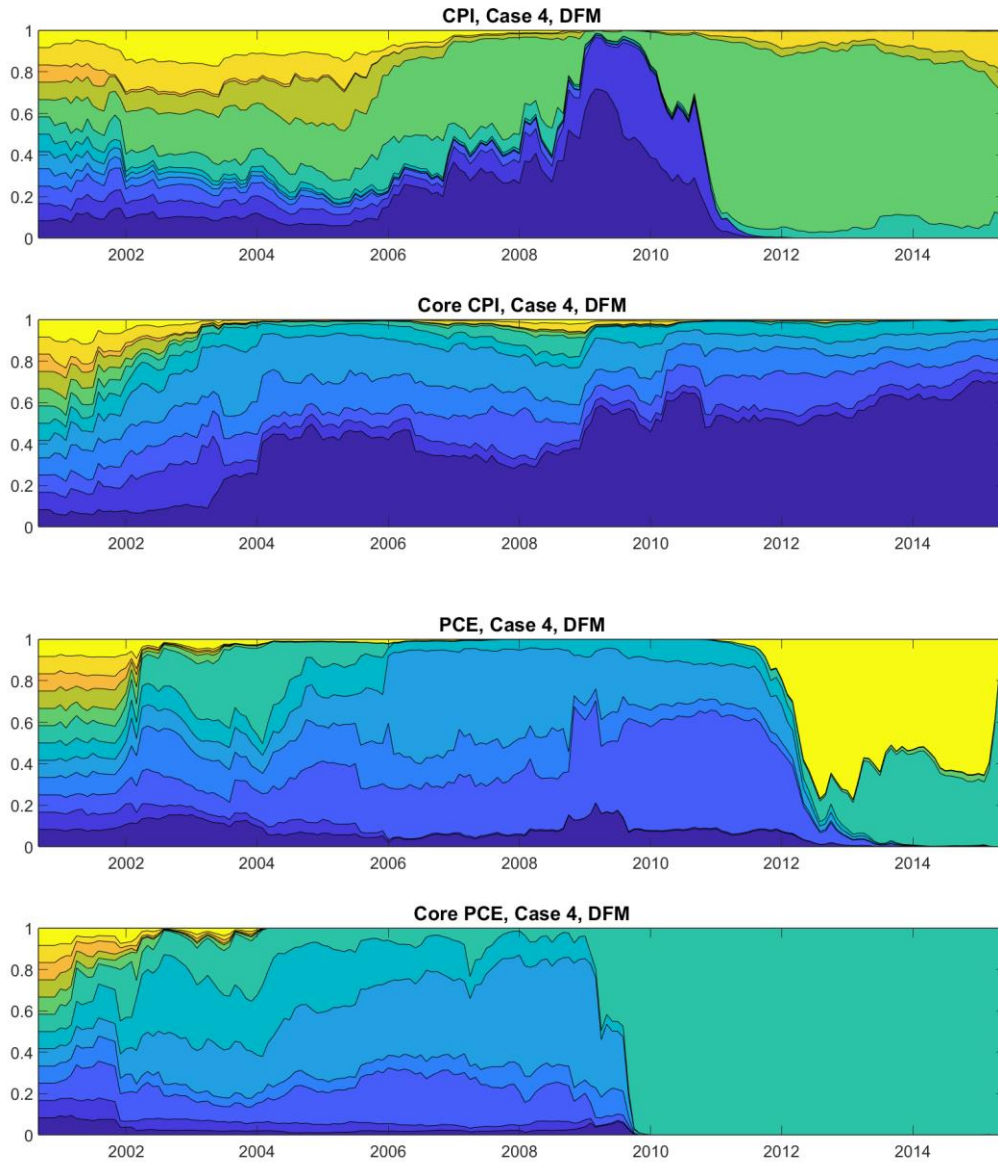


Notes: The figure plots the evolution of the weights for of underlying individual candidate densities for the stage 1 DFM combination at case 4. Each color shade represents a particular individual candidate density. There are 12 candidate densities. The sample period spans September 2000 through June 2015.

Figure A12: Weights for Stage 1 MIDAS Combinations, Log Score Weighting Scheme
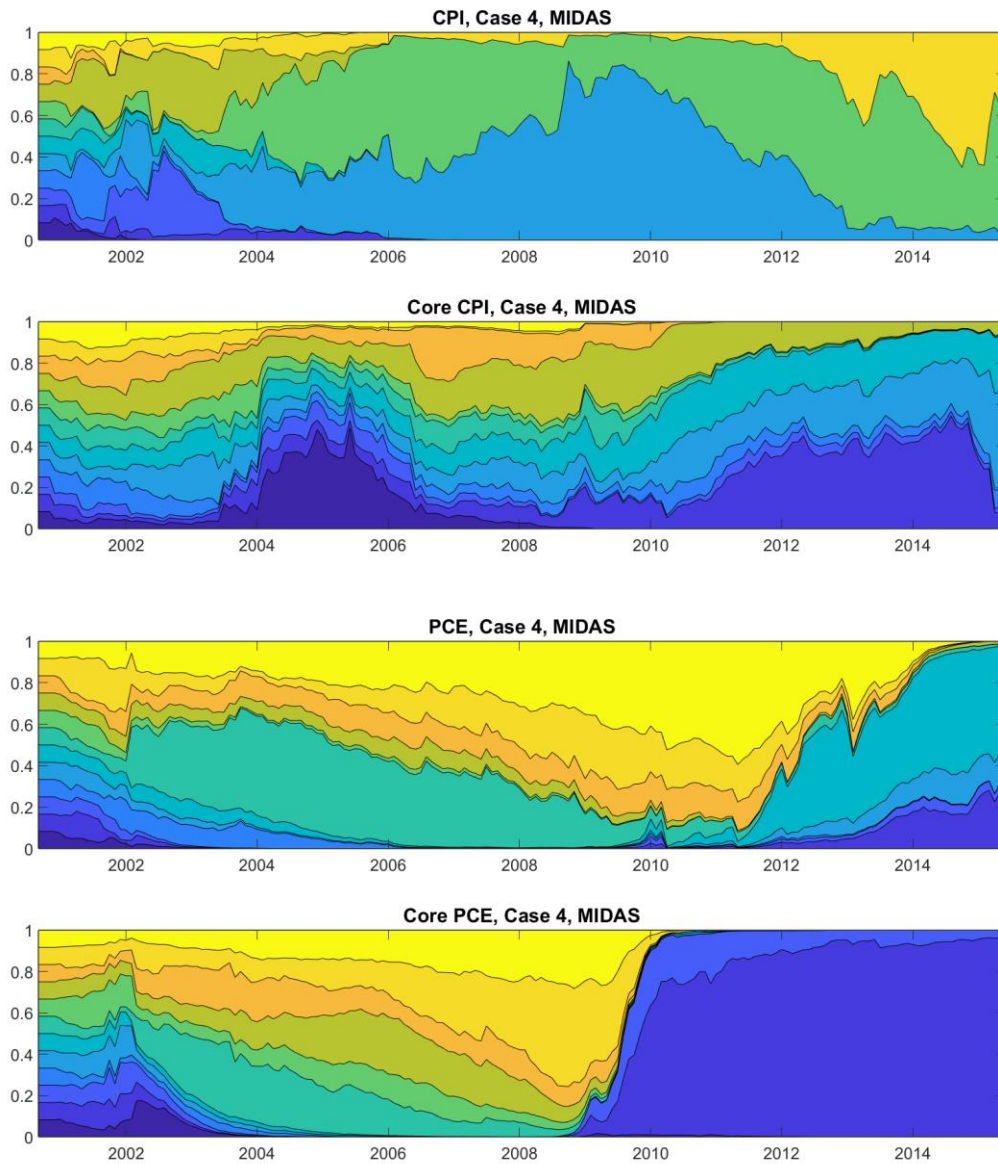
**CPI, Case 4, MIDAS**

**Core CPI, Case 4, MIDAS**

**PCE, Case 4, MIDAS**

**Core PCE, Case 4, MIDAS**

Notes: The figure plots the evolution of the weights for underlying individual candidate densities for the stage 1 MIDAS combination at case 4.  Each color shade represents a particular individual candidate density.  There are 12 candidate densities.  The sample period spans September 2000 through June 2015.

Figure A13: Time-Varying Uncertainty Estimates for Density Nowcasts of Quarterly Inflation
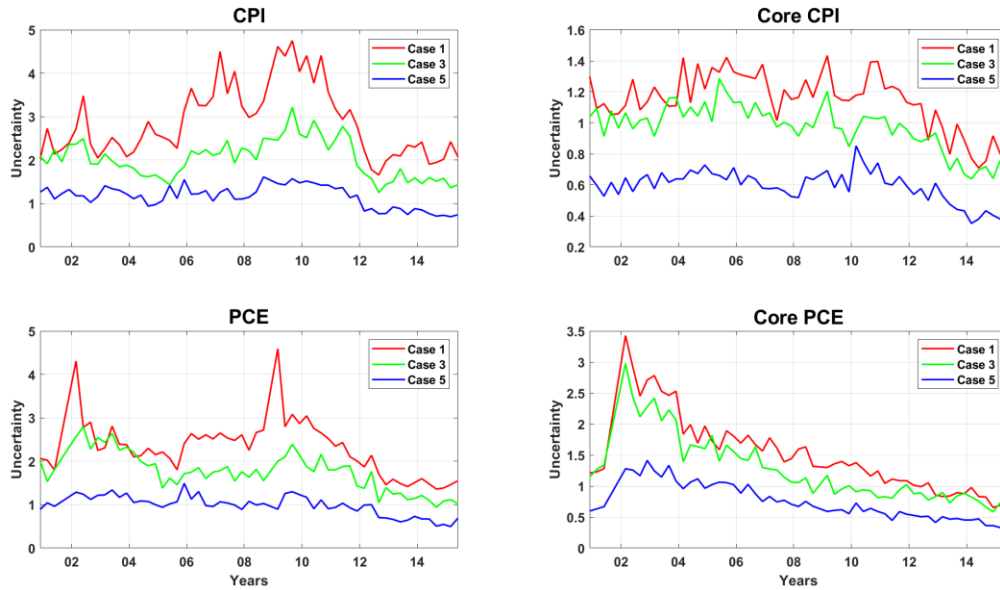


Notes: Uncertainty is measured as the width of the 70% prediction intervals. Estimates are for the grand combination based on the flexible aggregation strategy and log score weighting scheme for case 1 (last day of the preceding quarter), case 3 (last day of the first month of the quarter), and case 5 (last day of the second month of the quarter); see Table A1. The sample period spans 2000Q4 through 2015Q2.

Figure A14: Stage 2 Grand Combination of DMS, DFM, and MIDAS Combinations
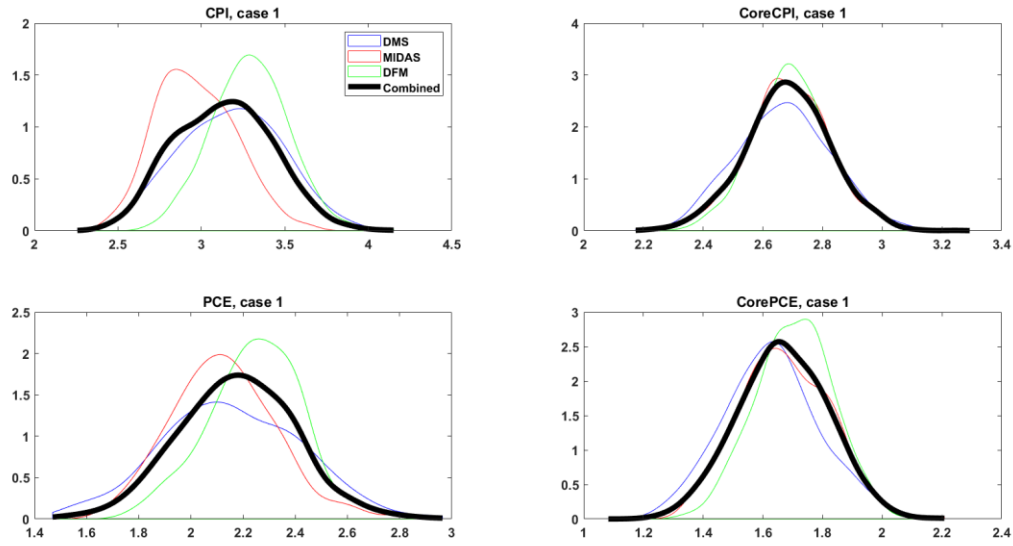


Notes: The figure illustrates a grand combination for 12-month inflation rates as of case 1 (the last day of the previous month) for nowcasting the target month of January 2001 and the three stage 1 combinations from the DMS, MIDAS, and DFM model classes that are used to construct the grand combination.

Figure A15: Weights and Higher-Order Moments, CMG Weighting Scheme
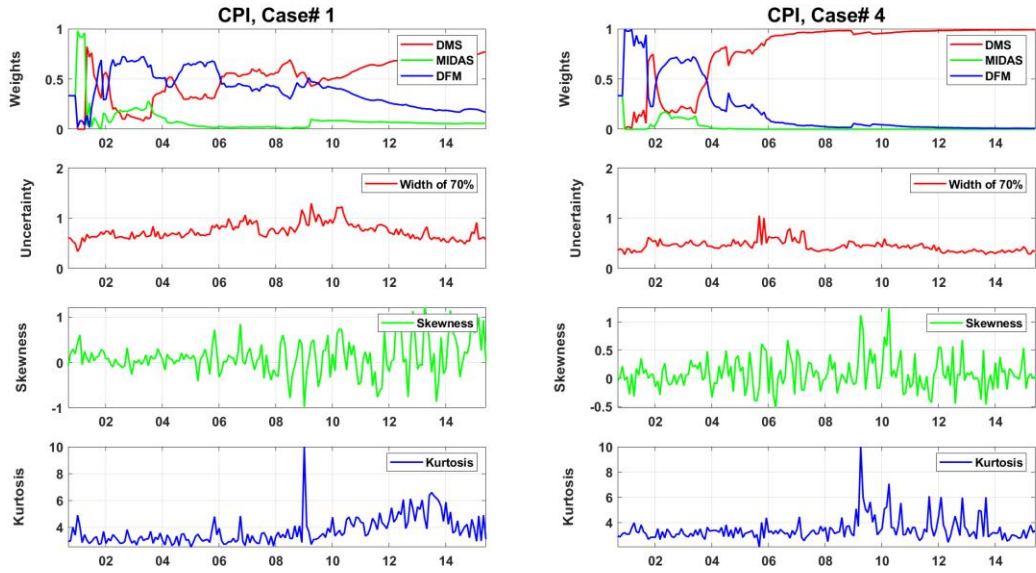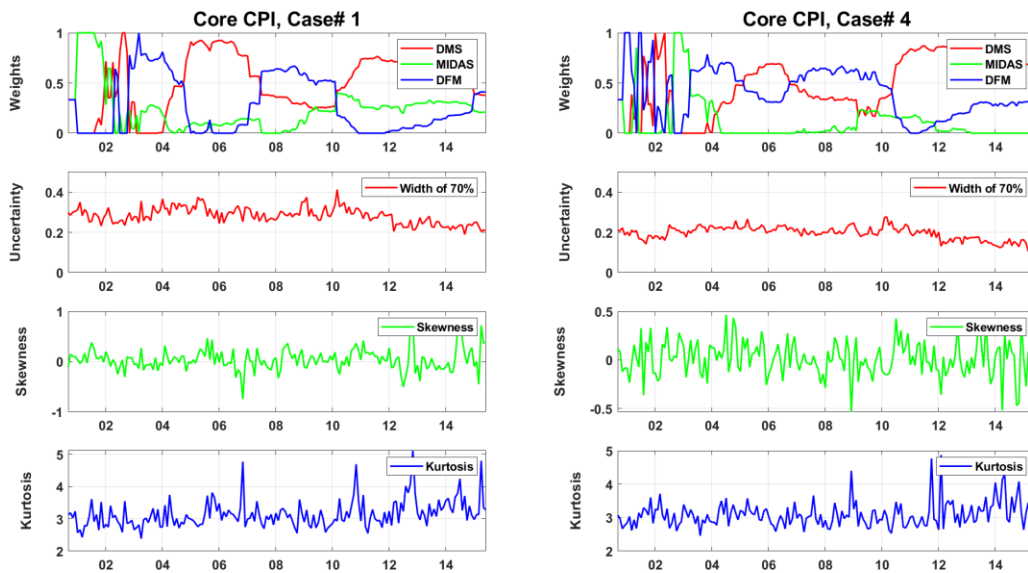
(a) CPI inflation
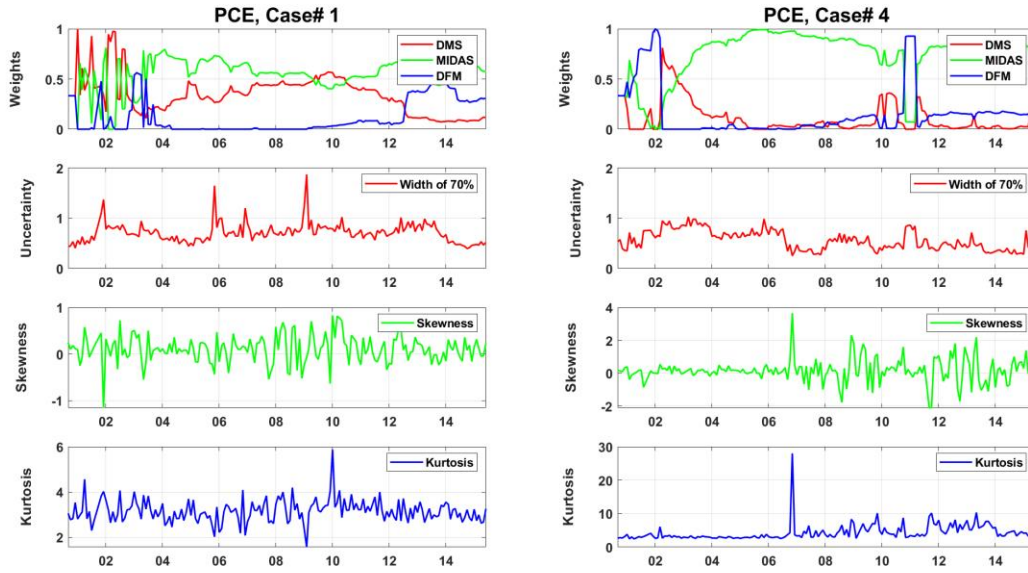


(b) Core CPI inflation
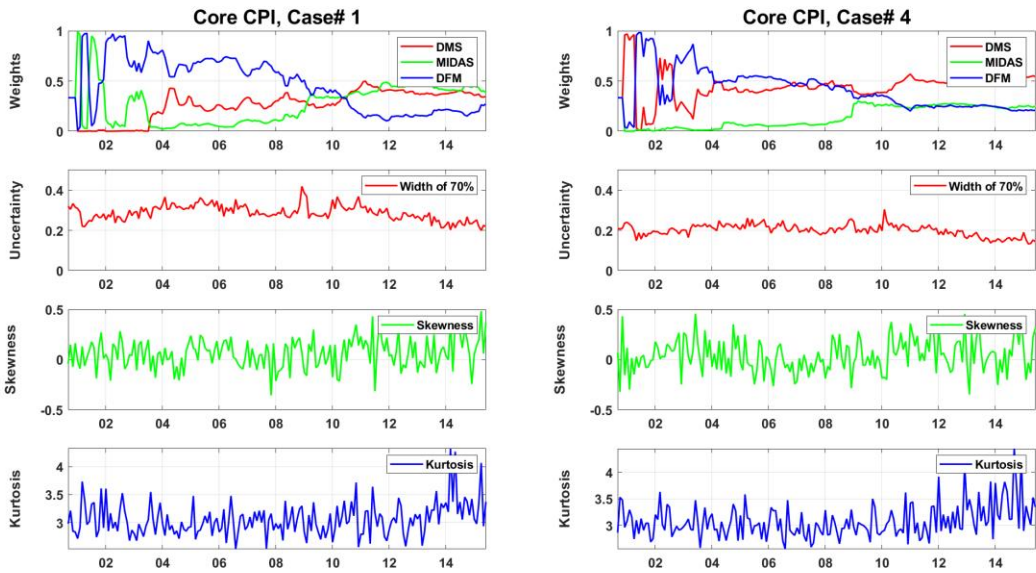
## Figure A15: Weights and Higher-Order Moments, CMG Weighting Scheme (continued)
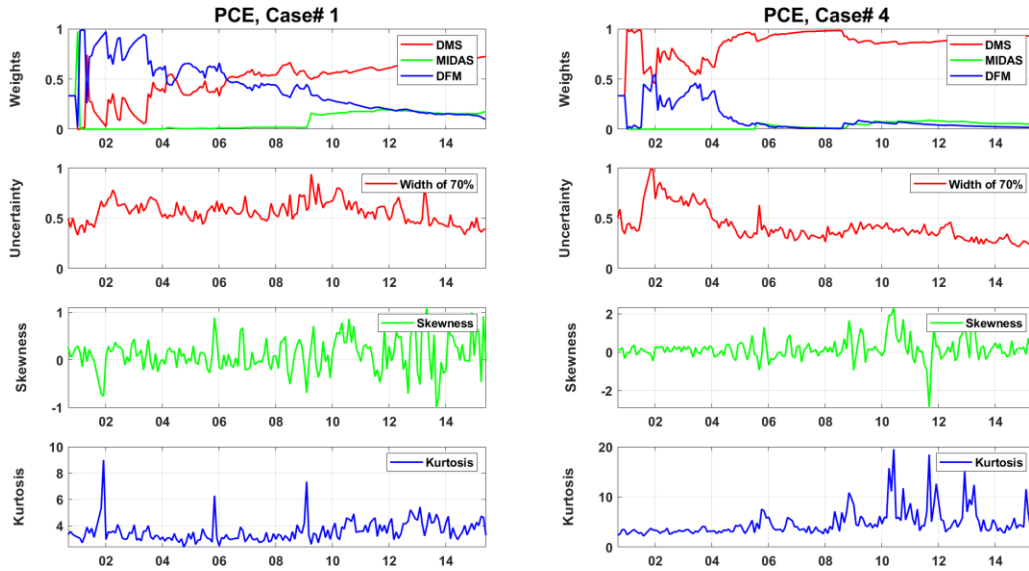
### (c) PCE inflation



### (d) Core PCE inflation
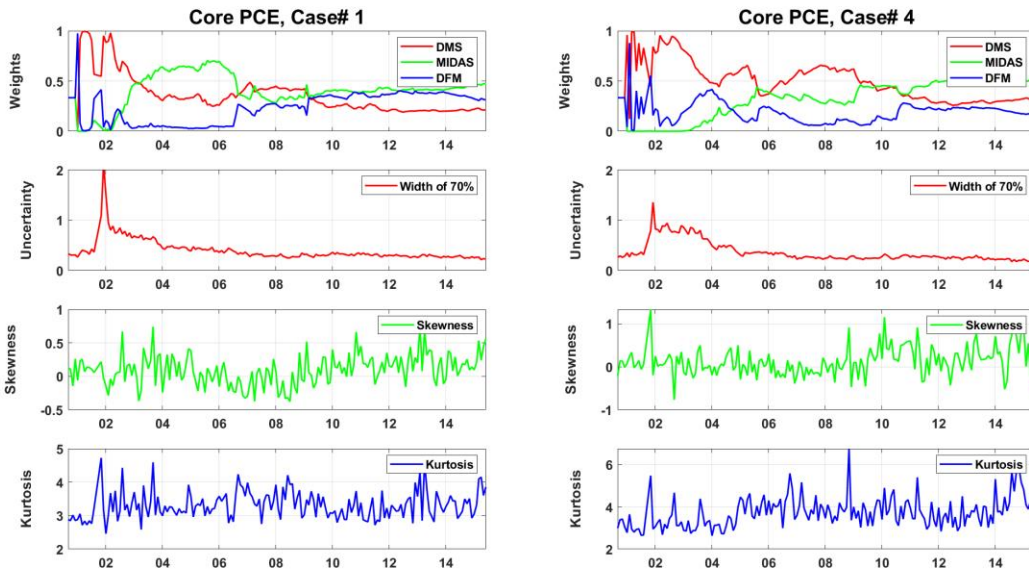


Notes: The first row of each panel plots the evolution of the weights for the three model classes underlying the grand combination, based on the flexible aggregation strategy and CMG weighting scheme. (Each model class is a combination of multiple model specifications.) The second row plots estimates of dynamic uncertainty, defined as the width of the 70% prediction intervals. The last two rows plot time-varying estimates of skewness and kurtosis. The sample period spans September 2000 through June 2015.

Figure A16: Weights and Higher-Order Moments, Ganics Weighting Scheme
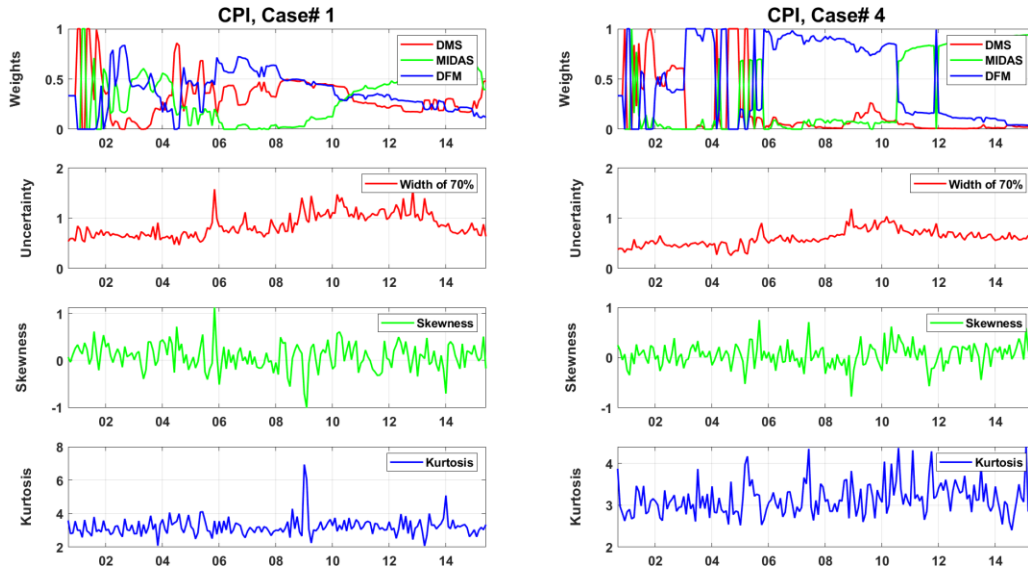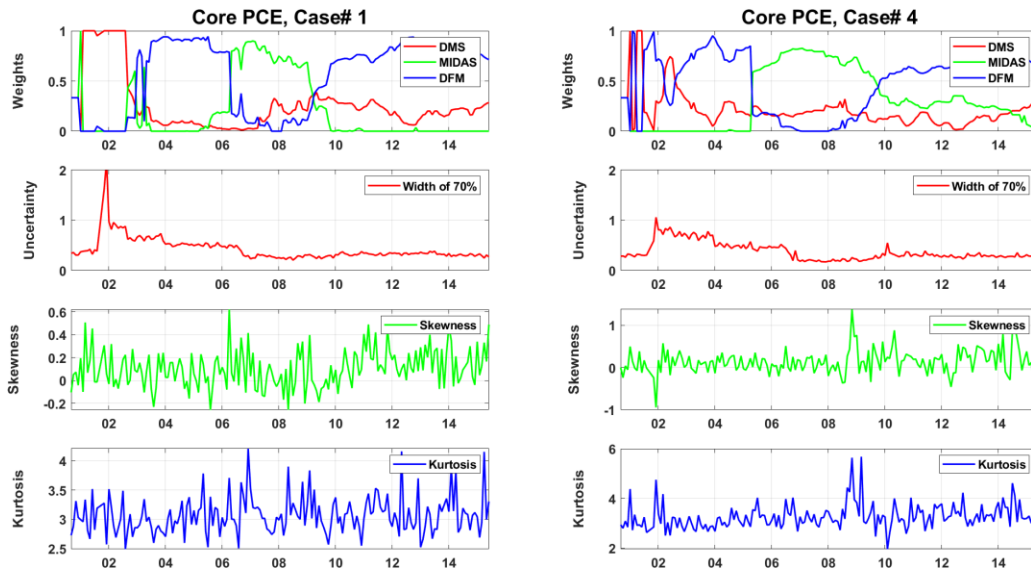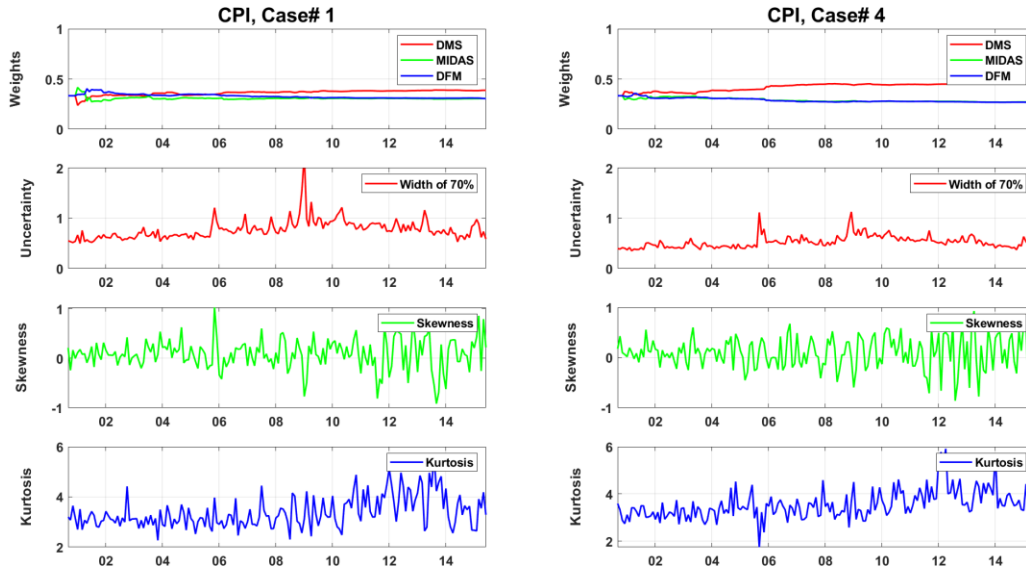
(a) CPI inflation

(b) Core CPI inflation

# Figure A16: Weights and Higher-Order Moments, Ganics Weighting Scheme (continued)

## (c) PCE inflation



## (d) Core PCE inflation



Notes: The first row of each panel plots the evolution of the weights for the three model classes underlying the grand combination, based on the flexible aggregation strategy and the Ganics weighting scheme. (Each model class is a combination of multiple model specifications.) The second row plots estimates of dynamic uncertainty, defined as the width of the 70% prediction intervals. The last two rows plot time-varying estimates of skewness and kurtosis. The sample period spans September 2000 through June 2015.

# Figure A17: Weights and Higher-Order Moments, CRPS Weighting Scheme

## (a) CPI inflation
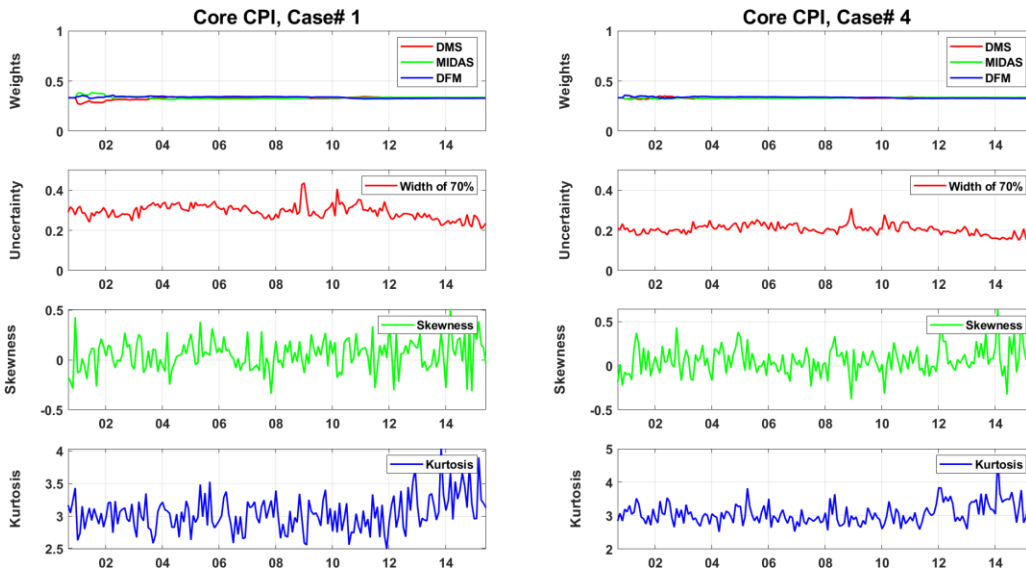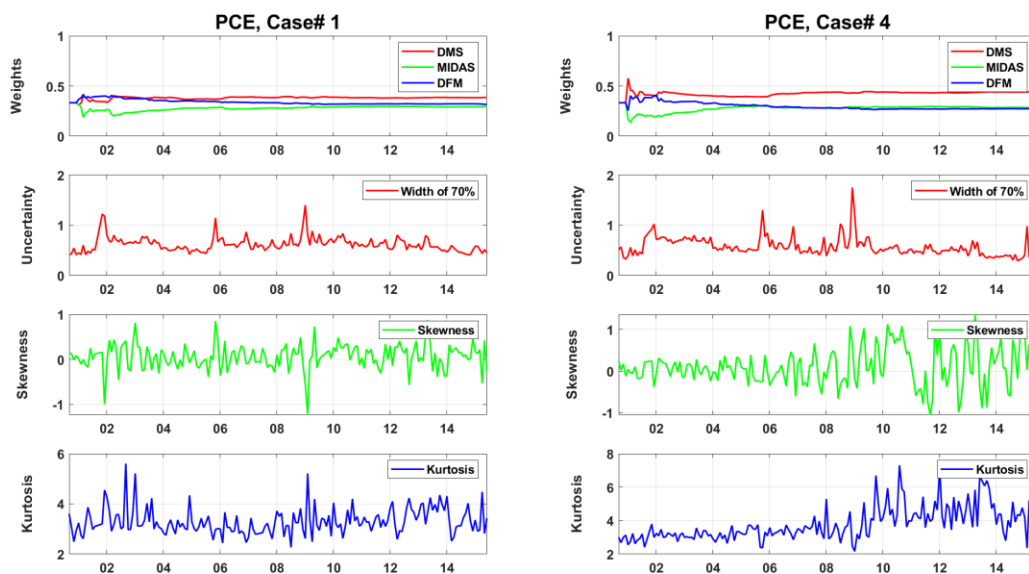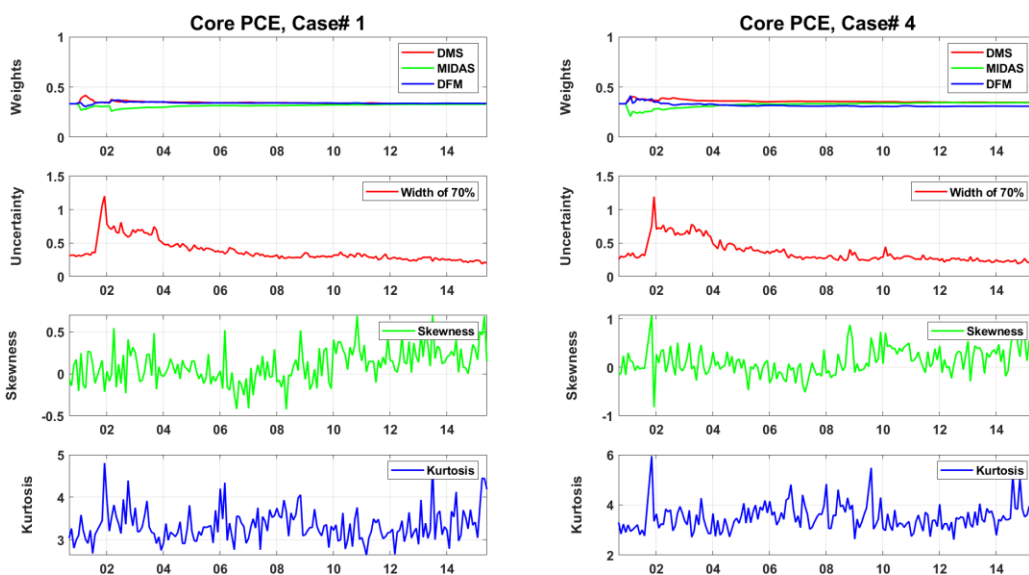


## (b) Core CPI inflation

# Figure A17: Weights and Higher-Order Moments, CRPS Weighting Scheme (continued)

## (c) PCE inflation



## (d) Core PCE inflation



Notes: The first row of each panel plots the evolution of the weights for the three model classes underlying the grand combination, based on the flexible aggregation strategy and the CRPS weighting scheme. (Each model class is a combination of multiple model specifications.) The second row plots estimates of dynamic uncertainty, defined as the width of the 70% prediction intervals. The last two rows plot time-varying estimates of skewness and kurtosis. The sample period spans September 2000 through June 2015.

Figure A18: Comparison of Weights within the DMS Model Class, Log Score Weighting Scheme vs. CMG Weighting Scheme
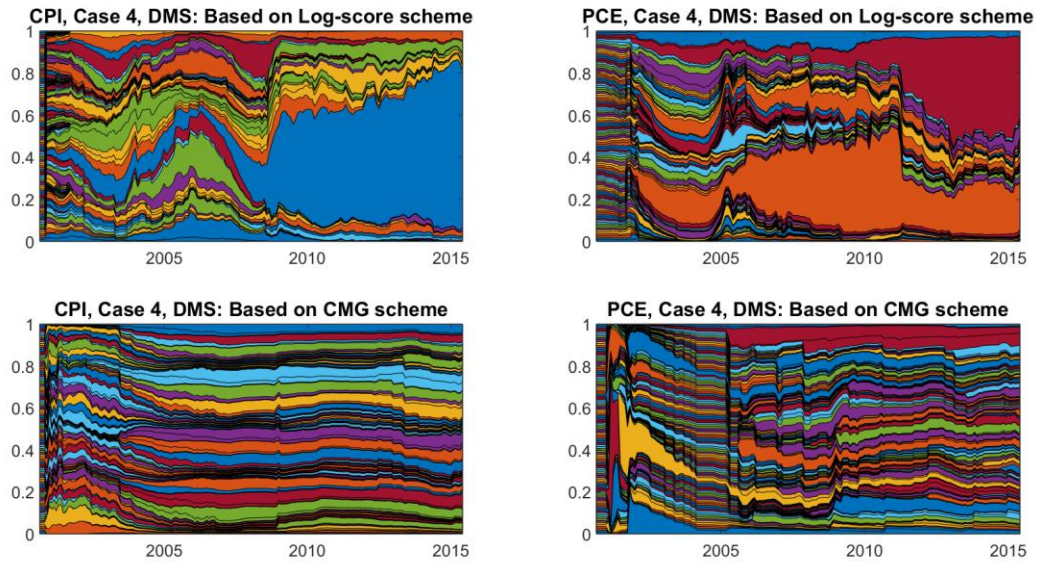


Notes: The figure plots the evolution of weights of the underlying individual candidate densities. Each color shade represents a particular individual candidate density. There are 108 candidate densities. The richness in the color variation indicates that no single candidate density dominates others. The left panel displays the weights for the stage 1 DMS combination constructed using the log score weighting scheme, and the right panel displays weights for the stage 1 DMS combination constructed using the CMG weighting scheme. The flexible aggregation method is used in both cases. The sample period spans September 2000 through June 2015.

**Table A1 Representative Dates for Quarterly Nowcasting Performance**

| Case | Date | Information Set (Example: Nowcasting target quarter is Q1) | Months to Forecast |
|------|------|------------------------------------------------------------|--------------------|
| 1 | Last day of the previous quarter | December 31: Have CPI and PCE through November; high-frequency information through December 31 | CPI: h=4 (Dec., Jan., Feb., Mar.)<br>PCE: h=4 (Dec., Jan., Feb., Mar.) |
| 2 | Day 15 of month 1 of the target quarter | January 15: Receive CPI for December and have PCE through November; high-frequency information through end of second week of January, which includes two weekly retail gasoline readings from January | CPI: h=3 (Jan., Feb., Mar.)<br>PCE: h=4 (Dec., Jan., Feb., Mar.) |
| 3 | Last day of month 1 of the target quarter | January 31: Have CPI for December and receive PCE for December; high-frequency information for all of January, which includes all four weekly retail gasoline readings from January | CPI: h=3 (Jan., Feb., Mar.)<br>PCE: h=3 (Jan., Feb., Mar.) |
| 4 | Day 15 of month 2 of the target quarter | February 15: Receive CPI for January and have PCE through December; high-frequency information through end of second week of February, which includes two weekly retail gasoline readings from February | CPI: h=2 (Feb., Mar.)<br>PCE: h=3 (Jan., Feb., Mar.) |
| 5 | Last day of month 2 of the target quarter | February 28: Have CPI for January and receive PCE for January; high-frequency information for all of February, which includes all four weekly retail gasoline readings from February | CPI: h=2 (Feb., Mar.)<br>PCE: h=2 (Feb., Mar.) |
| 6 | Day 15 of month 3 of the target quarter | March 15: Receive CPI for February and have PCE through January; high-frequency information through end of second week of March, which includes two weekly retail gasoline readings from March | CPI: h=1 (Mar.)<br>PCE: h=2 (Feb., Mar.) |
| 7 | Last day of month 3 of the target quarter | March 31: Have CPI for February and receive PCE for February; high-frequency information for all of March, which includes all four weekly retail gasoline readings from March | CPI: h=1 (Mar.)<br>PCE: h=1 (Mar.) |