# Mixing Mixed Frequency and Diffusion Indices in Good Times and in Bad: An Assessment Based on Historical Data Around the Great Recession of 2008*

Kihwan Kim[1], Hyun Hak Kim[2], and Norman R. Swanson[3]

[1]Korea Energy Economics Institute, [2]Kookmin Univeristy, and [3]Rutgers University

November 2021

In the field of economics, recent advances in the areas of machine learning, shrinkage, and variable selection have been spectacularly successful. In one key area of study, advances in both modelling and estimation have enabled empirical practitioners to show the usefulness of latent factors designed to efficiently extract common information from interesting new datasets. At the center of this "big data" success are diffusion and mixed frequency indices, which have proven useful time and time again in forecasting contexts. This paper lends further support to recent claims of the usefulness of these sorts of indices, albeit with a twist. We focus on a historical dataset than contains the Great Recession of 2008, and show that the usefulness of said indices is pronounced during "low" GDP growth periods, while simple autoregressive models are adequate during "high growth" periods. This finding stems from the introduction of very simple "hybrid" models that employ dynamic recursive (rolling) thresholding in order to switch between benchmark linear models and more complex index driven models, depending on GDP growth conditions. In the context of predicting both quarterly real GDP growth and CPI inflation, these hybrid models are found to be superior, for all forecast horizons. When comparing the hybrid models against a host of alternatives, mean square forecast error gains reach as high as 35%, during the Great Recession, and remain significant throughout our entire prediction period. Additionally, the very best short-term GDP forecasting models contain variants of the Aruoba et al. (2009) business conditions index, although these models are most useful when diffusion indices are also incorporated. Thus, mixing mixed frequency and diffusion indices matters. Finally, across all experiments, we find strong new evidence of the usefulness of survey predictions, including those from the Survey and Professional Forecasters, and those from the Livingston Survey.

*JEL classification:* C22, C51.

*Keywords*: Forecasting, Diffusion index, Mixed frequency data, Factor model, Recursive estimation, Kalman filter.

---

* Kihwan Kim (paradiso4@gmail.com), 405-11, Jongga-Ro, Jung-Gu, Ulsan, Korea, 44543. Hyun Hak Kim (Corresponding author) (hyunhak.kim@kookmin.ac.kr), Department of Economics, Kookmin University, Seoul, Korea., Norman R. Swanson (nswanson@econ.rutgers.edu), Department of Economics, Rutgers University, New Brunswick, NJ, USA. The authors owe many thanks to Nii Ayi Armah, Valentina Corradi, John Landon-Lane, Bruce Mizrach, and Xiye Yang for providing comments on earlier versions of this work. The views expressed herein are those of the authors and do not necessarily reflect those of the Korea Energy Economics Institute.

# 1 Introduction

In the field of economics, recent advances in the areas of machine learning, shrinkage, and variable selection have been spectacularly successful. In one key area of study, theoretical advances, both in modelling and estimation, have enabled empirical practitioners to show the usefulness of latent factors designed to efficiently extract common information from interesting new datasets. At the center of this "big data" success are diffusion and mixed frequency indices, which have proven useful time and time again in forecasting contexts. A very incomplete list of key contributions in this area include Forni et al. (2000, 2005), Bai and Ng (2002), Bai and Ng (2002, 2013), Stock and Watson (2002a,b, 2006), Ghysels et al. (2007), and Aruoba, Diebold and Scotti (2009, henceforth ADS).

In this paper, we explore the usefulness of diffusion and mixed frequency indices of the variety discussed in the above papers, in the context of predicting U.S. GDP and CPI inflation around the time of the Great Recession of 2008.[1] In particular, we present the results of an extensive series of experiments wherein standard linear specifications are compared with models utilizing: diffusion indices extracted from largescale monthly macroeconomic datasets, mixed frequency indices extracted from carefully selected small mixed frequency datasets, and survey predictions. More importantly, we introduce very simple "hybrid" models that employ recursive, rolling, and fixed thresholding in order to "switch" between benchmark linear models and more complex mixed frequency and diffusion index models (i.e., "big data" models) that may contain survey predictions. Thresholds are time-varying, and are determined in real-time by examining extant measures of GDP growth, using various windowing techniques. Simply put, we wish to assess whether very intuitive and simple thresholding techniques lead to notable improvement in predictive model performance; and we find this to be the case. Moreover, our thresholding experiments indicate that the usefulness of "big data" indices is dependent upon whether GDP growth is low (under 1%, say) or high. When it is low, they are useful. When it is high, prediction of GDP and CPI requires little more than autoregressive type models. The notion behind the thresholding used in this paper can be easily motivated using a standard example. Namely, consider the case of correlation in the stock market. When in good times, financial instrument returns may be (nearly) independent, allowing for simple risk diversification and asset allocation based on Sharpe type factor-regression analysis, say. This is often done in the hedge fund industry, for

[1]For further discussion of recent advances in mixed frequency modelling, see Ghysels et al. (2006, 2007), and the 2016 special issue of the *Journal of Econometrics* on the topic, with editorial comment by Ghysels and Marcelino (2016).

instance. However, when the market "goes south", correlations that were hithertofore zero become non-zero, causing a previously diversified portfolio to simply follow the market. Thus, the dynamic behavior of the financial instrument returns in this example change markedly depending on business conditions. This in turn begs the question as to whether "markedly" different models should be used in the two "regimes", when the objective is prediction. Our prediction experiments use two such markedly different classes of models. One class involves standard benchmark linear specifications. The other class involves utilization of "big data".

Sargent and Sims (1977) found that a small number of common factors explain much of the variation in various macroeconomic variables. A multitude of theoretical and empirical advances associated with constructing latent factor indices have occurred since the publication of this paper.[2,3] For example, in the forecasting literature, prediction models that utilize estimates of latent factors have been extensively studied (e.g., see Stock and Watson (1999, 2002a, 2006), Boivin and Ng (2006), Bai and Ng (2009), Armah and Swanson (2010), D'Agostino and Giannone (2012),Kim and Swanson (2014, 2018b), and the references cited therein). In D'Agostino and Giannone (2012), the authors show that different model specifications involving latent factors are useful at different times in the economic cycle. This is not surprising, given the wealth of research stressing the importance of regime switching that is related to phases of the business cycle. More generally, model instability, regardless of whether it is driven by the business cycle, is an important topic of research in the area of forecasting. In the context of the construction of diffusion indices, Stock and Watson (2008) show that independent and mild factor loading instability may not appreciably affect factor estimation. Along the same lines, Carrasco and Rossi (2016) derive rates of convergence for approximate (or misspecified) factor models. Our approach is to be agnostic concerning misspecification, and to propose very simple dynamic thresholding techniques for "switching" between the use of standard "small data" linear forecasting models, and "big data" models that utilize diffusion indices and mixed frequency factors. We find that use of even the very simplest GDP-based dynamic thresholds which utilize rolling windows of data (called $\tau_t^{rol}$ thresholding) and recursive windows of data (called $\tau_t^{rec}$ thresholding) for "switching" yield impressive gains when predicting real GDP growth and CPI inflation.

[2] The idea of utilizing statistical estimates of common factors dates back to Spearman (1904). Refer to Swanson (2016) for further discussion of the broad history of machine learning, shrinkage, and variable selection in the context of factor modelling.

[3] A few important papers include Stock and Watson (1988, 1989, 2002b, 2006), Forni and Reichlin (1998), Ding and Hwang (1999), Forni et al. (2000, 2005), Bai and Ng (2002, 2006a), Bai and Ng (2007, 2013), Bai (2003), Mariano and Murasawa (2003), Boivin and Ng (2005, 2006), Hallin and Liska (2007), Aruoba et al. (2009), Aruoba and Diebold (2010), Onatski (2009), Corradi and Swanson (2014), and Marcellino et al. (2015).

In our prediction experiments, we examine not only the "hybrid" models discussed above, but we also examine a large set of linear models containing combinations of autoregressive and dynamic distributed lag terms, mixed frequency indices, and diffusion indices. More specifically, we examine 37 different sets mixed frequency indicators and associated indices; as well as a number of diffusion indices extracted from a large-scale macroeconomic dataset containing 143 variables. One feature of our setup is that it allows us to carry out a systematic examination of the usefulness of the mixed frequency factor model in ADS(2009). We document when and how the important ADS business conditions index is useful for constructing predictions of both real GDP growth and CPI inflation.[4] We additionally examine the usefulness of survey information available from the Survey of Professional Forecasters and from the Livingston Survey.

Our main results can be summarized as follows. First, for 1-quarter ahead GDP prediction, under $\tau_t^{rol}$ thresholding, the ADS index is not only useful, but yields the very best prediction model, in terms of mean square forecast error. However, this "top" model involves combining the ADS index with diffusion indices, so that it is a combination mixed frequency and diffusion index model that dominates all other specifications. This result lends strong support to the notion that the daily ADS index produced by the FRBP is not only useful as a business conditions index, but is also highly useful for short-term GDP forecasting.

Second, regardless of forecast horizon and variable being predicted, the very best models always involve dynamic thresholding, and model combination never yields a "top-5" model. The only exception to this rule is for annual 1-year ahead GDP growth prediction, in which case setting $\tau_t = \tau = 0$ yields the mean-square forecast error "best" ($MSFE$-best) model. This result implies that not only is simple thresholding useful in contexts where mixtures of mixed frequency and diffusion indices are included in forecasting models, but that forecast combination, long held out to be virtually unbeatable in numerous aggregate macroeconomic forecasting contexts, is dominated under simple thresholding rules. However, it is interesting to note that model combination does actually yield superior predictions at all forecast horizons, for the case of CPI inflation, but only when no hybrid models of any sort are considered.

Third, $MSFE$ percentage gains associated with use of our hybrid models for forecasting GDP vary from around 10% to as much as 35%, depending on the sample period analyzed. Notably, the hybrid models perform particularly well during the Great Recession, where $MSFE$ gains are 3 times as high as those based on the entire forecast period from 1987 -

[4]Interesting related research is contained in Balke et al. (2017), who construct a daily real activity index using Beige Book information.

2012. This suggests that our simple thresholds are serving, roughly speaking, to differentiate between "high growth" and "low growth" epsiodes, and that during "low growth" episodes it pays to utilize models with mixed frequency and diffusion indices, while during "high growth" episodes, it suffices to utilize simple autoregressive models. Moreover, findings are qualitatively the same when forecasting CPI inflation using the same dynamic thresholding mechanism as that used in the case of GDP. Various possible reasons for this are discussed later in the paper, and our arguments are not unrelated to the story of correlation in the stock market discussed above.

Fourth, a final key element of the our results concerns the usefulness of survey variables. Note that a variety of the indicator sets used in the construction of our mixed frequency indices include either Livingston or SPF survey predictions of GDP growth. Findings regarding indices containing these variables depend upon whether one is predicting GDP growth or CPI inflation. For GDP, in hybrid cases, a subset of top five performing models contain Livingston GDP predictions in their mixed frequency indices. No survey variables are contained in the top performing non-hybrid models. In stark contrast, for CPI, both Livingston and SPF forecasts appear in *all* 1st ranked specifications, from amongst all non-hybrid models. Moreover, when hybrid specifications are considered, at least one top 5 model includes either SPF or Livingston survey variables, regardless of forecast horizon, thresholding method, or version of inflation being forecasted. We thus have strong new evidence of the usefulness of these surveys, at multiple prediction horizons.

The rest of the paper is organized as follows. In Section 2, we present the modelling framework, including discussions of diffusion and mixed frequency indices, as well as the specification of (hybrid) prediction models. In Section 3, we outline the dataset used in our analysis. Section 4 contains the results of our prediction experiments. Finally, in Section 5, we provide concluding remarks. Various technical details and additional empirical results are contained in two appendices.

# 2 The Modelling Framework

## 2.1 Diffusion Indices

The diffusion indices or common factors examined in our empirical analysis are based on the following setup. Suppose that a multidimensional normalized random variable, $X_t$, is generated according to the following dynamic factor model (henceforth DFM):

$$X_{it} = \lambda_i(L)' f_t + e_{it}, \tag{1}$$

for $i = 1, 2, ..., N$, and $t = 1, 2, ..., T$, where $X_{it}$ is a single datum, $f_t$ is a $q \times 1$ vector of latent common factors, $\lambda_i(L)$are $q \times 1$ vector lag polynomials in nonnegative powers of $L$, and $e_{it}$ is an idiosyncratic shock. That is, $N$ series of data are assumed to be composed of two parts, common components, $\lambda_i(L)f_t$, and idiosyncratic errors $e_{it}$, for each $i$. As discussed in Stock and Watson (2006), a standard assumption is that the factors and idiosyncratic errors are uncorrelated and that the idiosyncratic error terms are mutually uncorrelated, at all leads and lags. One can also allow for some degree of serial correlation in this model. Under the assumption that the lag polynomials have finite dimension, $p$, we can write the model in equation (1) in static form, as follows:

$$X_t = \Lambda F_t + e_t \ , \tag{2}$$

where $F_t = (f_t' f_{t-1}' ... f_{t-p+1}')'$ is an $r \times 1$ vector,with $r \leq pq$, $r$ is the number of static factors, and $\Lambda$ is the factor loading matrix. The static factors in equation (2) are estimated using principal component analysis.

In particular, following Stock and Watson (2006), let $k$ ($k < \min\{N, T\}$) be an arbitrary number of factors, assume that $N < T$, let $\Lambda$ be the $N \times k$ matrix of factor loadings, $(\Lambda_1, \Lambda_2, ..., \Lambda_N)'$, and let $F$ be a $k \times T$ matrix of factors $(F_1, F_2, ..., F_T)$. From equation (2), estimates of $\Lambda$ and $F_t$ are obtained by solving the following optimization problem :

$$V = \min_{F,\Lambda} \frac{1}{T} \sum_{t=1}^{T} (X_t - \Lambda F_t)' (X_t - \Lambda F_t), \text{ subject to } \Lambda'\Lambda = I_k \ . \tag{3}$$

We treat $F_1, ..., F_T$ as fixed parameters to be estimated after normalizing $\Lambda$. Given $\widehat{\Lambda}$, the solution to equation (3) satisfies $\widehat{F}_t = (\widehat{\Lambda}'\widehat{\Lambda})^{-1}\widehat{\Lambda}' X_t$. Substituting this into equation (3) yields

$$\begin{aligned} V &= \min \frac{1}{T} \sum_{t=1}^{T} X_t' (I - \Lambda(\Lambda'\Lambda)^{-1}\Lambda') X_t, \text{ subject to } \Lambda'\Lambda = I_k & (4) \\ &= \max tr((\Lambda'\Lambda)^{-\frac{1}{2}} \Lambda' \sum\nolimits_{XX} \Lambda(\Lambda'\Lambda)^{-\frac{1}{2}}, \text{ subject to } \Lambda'\Lambda = I_k & (5) \\ &= \max \ \Lambda' \sum\nolimits_{XX} \Lambda, \text{ subject to } \Lambda'\Lambda = I_k, & (6) \end{aligned}$$

where $\sum_{XX} = T^{-1} \sum_{t=1}^{T} X_t X_t'$. This optimization is solved by setting $\widehat{\Lambda}$ equal to the eigenvectors of $X'X$ corresponding to its $r$ largest eigenvalues. Then, construct $\widehat{F}_t = \widehat{\Lambda}' X_t$.

Following Bai and Ng (2002), after estimating $\widehat{\Lambda}$ and $\widehat{F}_t$, let $\hat{V}(k) = T^{-1}\sum_{t=1}^{T}(X_t - \widehat{\Lambda}\widehat{F}_t)'(X_t - \widehat{\Lambda}\widehat{F}_t)$ be the sum of squared residuals from regressions of $X_t$ on the $k$ factors and $IC(k) = \log(\hat{V}(k)) + k(\frac{N+T}{NT})\log(C_{NT}^2)$ be the information criterion, where $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. A consistent estimator of the true number of factors is $r = \arg\min_{0\leq k\leq \bar{k}} IC(k)$, where $\overline{k}$ is the maximum number of factors. Since this important paper by Bai and Ng, many additional estimators of $r$ have been proposed (see e.g. Carrasco and Rossi (2016), Kim and Swanson (2018b), and the references cited therein). Examination of their performance in our prediction experiments is left to future research.

## 2.2 Mixed Frequency Indices

Stock and Watson (1989) construct business condition indices from four monthly variables (industrial production, real manufacturing, trade and sales, number of employees on nonagricultural payroll, and personal income less transfer payments). Mariano and Murasawa (2003) add quarterly real GDP to the single index model of Stock and Watson, and develop a mixed frequency model with latent factors. ADS (2009) take all of this research one important step further, and construct mixed frequency business conditions indices at a daily frequency. Finally, Camacho et al. (2014, 2018) include one-time Markov switching, and Marcellino et al. (2015) incorporate stochastic volatility into the standard latent factor model used in mixed frequency modelling.

In this paper, we utilize the mixed frequency dynamic factor model presented in ADS, wherein it is assumed that the latent dynamics of an index, say $mf_t$, follows a zero-mean $AR(p)$ process, and is generated daily, so that the index $t$ denote daily increments. Thus, our mixed frequency (MF) index evolves according to:

$$mf_t = \rho_1 mf_{t-1} + \cdots + \rho_p mf_{t-p} + e_t, \tag{7}$$

where $e_t$ is white noise with unit variance. Suppose that we have $J$ indicator variables that we wish to model, and let $y_t^i$, denote a single datum at time $t$, for variable $i$, $i = 1, ..., J$. Now, assume that this variable depends depends *linearly* on $mf_t$, and possibly also on various exogenous variables $w_t^1, \ldots, w_t^k$ and/or n lags of $y_t^i$. This leads to the following measurement equation:

$$y_t^i = c_i + \beta_i mf_t + \delta_{i1} w_t^1 + \cdots + \delta_{ik} w_t^k + \gamma_{i1} y_{t-1}^i + \cdots + \gamma_{in} y_{t-n}^i + u_t^i, \tag{8}$$

where the $u_t^i$ are contemporaneously and serially uncorrelated innovations. At time $t$, the $i$th indicator, $y^i$, may be missing. For example, if $y_t^i$ is quarterly real GDP, then it is "missing" on many days within the quarter. To handle the missing data problem, ADS distinguish between stock and flow variables, and between observed data and missing data. Suppose that $\widetilde{y}_t^i$ denotes a stock variable, recorded at a low frequency. At any time, $t$, if $y_t^i$ is observed, then $\widetilde{y}_t^i = y_t^i$. If it is not observed, then $\widetilde{y}_t^i = NA$. Thus, the stock variable at time $t$ is:

$$\widetilde{y}_t^i = \begin{cases} y_t^i & \text{, if } y_t^i \text{ is observed} \\ NA & \text{, otherwise} \end{cases} . \tag{9}$$

Combining equations (8) and (9), the measurement equation for a stock variable is:

$$\widetilde{y}_t^i = \begin{cases} c_i + \beta_i mf_{t-i} + \gamma_{i1}\widetilde{y}_{t-i}^i + \cdots + \gamma_{ip}\widetilde{y}_{t-n}^i + u_t^i & \text{, if } y_t^i \text{ is observed} \\ NA & \text{, otherwise} \end{cases} . \tag{10}$$

Unlike a stock variable, a flow variable is assumed to exist at a higher frequency, but is only recorded at lower frequencies, and can thus can be interpreted as an intra-period sum of daily values, so that a flow variable is defined as:

$$\widetilde{y}_t^i = \begin{cases} \sum_{j=1}^{D_i} y_{t-j+1}^i & \text{, if } y_t^i \text{ is observed} \\ NA & \text{, otherwise} \end{cases} , \tag{11}$$

where $D_i$ denotes the number of days in a period. Combining equations (8) and (11), the measurement equation for a flow variable is:

$$\widetilde{y}_t^i = \begin{cases} \sum_{i=1}^{D_i} c_i + \beta_i \sum_{i=1}^{D_i} mf_{t-i+1} + \gamma_{i1}\widetilde{y}_{t-D_i}^i + \cdots + \gamma_{in}\widetilde{y}_{t-nD_i}^i + u_t^{*i} & \text{, if } y_t^i \text{ is observed} \\ NA & \text{, otherwise} \end{cases} . \tag{12}$$

We use the sum of state variables for the period (i.e., $\sum_{i=1}^{D_i} mf_{t-i+1}$) in the measurement equation in the case of flow variable, as in ADS. For example, in a quarterly real GDP growth measurement equation, all daily factors from the first day to the last day of the quarter are summed and plugged into the measurement equation. Note that different temporal aggregation schemes between lower frequency flow variables and daily state variables may be considered. For further discussion, refer to Mariano and Murasawa (2003).

Here, equation (7) is the state equation and equations (9) and (11) are the measurement equations. Together, these equations constitute a state-space model, for which both smoothed and unsmoothed estimation algorithms are discussed in Appendix A. Broadly speaking, under the assumption that errors in state and measurement equations are normally distributed, it is straightforward to estimate this model using the Kalman filtering and prediction error decomposition.

## 2.3 Mixed Frequency and Diffusion Index Models (with Thresholding)

The diffusion and mixed frequency indices discussed above are used in a variety of forecasting experiments in the sequel, as discussed in the introduction to this paper. In particular, adhering to the approach used in Stock and Watson (2002a,b), Bai and Ng (2006b) and Kim and Swanson (2014), we examine the following prediction models:

$$\widehat{y}_{t+h} = \widehat{c}^h + \sum_{ki=1}^{r} \sum_{j=1}^{\widehat{p}_k^h} \widehat{\psi}_j^{h,ki} DI_{t-j}^{ki} + \sum_{i=1}^{\widehat{q}} \widehat{\phi}_i^h y_{t-i} + \varepsilon_{t+h}, \tag{13}$$

$$\widehat{y}_{t+h} = \widehat{c}^h + \sum_{j=1}^{\widehat{p}^h} \widehat{\psi}_j^h MF_{t-j} + \sum_{i=1}^{\widehat{q}^h} \widehat{\phi}_i^h y_{t-i} + \varepsilon_{t+h}, \tag{14}$$

and

$$\widehat{y}_{t+h} = \widehat{c}^h + \sum_{ki=1}^{r} \sum_{j=1}^{\widehat{p}_k^h} \widehat{\psi}_j^{h,ki} DI_{t-j}^{ki} + \sum_{j=1}^{\widehat{p}^h} \widehat{\psi}_j^h MF_{t-j} + \sum_{i=1}^{\widehat{q}^h} \widehat{\phi}_i^h y_{t-i} + \varepsilon_{t+h}, \tag{15}$$

where $y_t$ is a scalar target variable being predicted, $\widehat{y}_{t+h}$ are predictions thereof, $\varepsilon_{t+h}$ is a disturbance term, $h$ denotes forecast horizon, $\widehat{c}^h, \widehat{\psi}_j^{h,k}$ and $\widehat{\phi}_i^h$ are estimated using least squares (LS), and $\widehat{p}_k^h$, $\widehat{q}_k^h$, $\widehat{p}^h$, and $\widehat{q}^h$ are selected via use of the Schwarz Information Criterion (SIC). In these models, the diffusion indices, (i.e., $DI_t^{ki}$, for $ki = 1, ..., r$) are estimated recursively using a large-scale monthly dataset. This is done prior to recursive estimation of our prediction models. When predicting the growth of quarterly real GDP, we set $r = 1$, while for monthly prediction, we set $r = 2$. This choice is consistent with empirical findings in the literature (see e.g. Stock and Watson (2002b), D'Agostino and Giannone (2012), and Kim and Swanson (2018a)). In a set of experiments not reported here, $r$ was recursively estimated, and prediction results were found to be inferior to those based on our simpler strategy of fixing $r$. (Complete results are available upon request.) Mixed frequency indices (i.e., $MF_t$) are also estimated prior to recursive estimation of our prediction models. A variety of different mixed frequency datasets are utilized for this step, as discussed below. Finally, all forecasting equations are estimated both with an without autoregressive terms. For a detailed examination of the usefulness of including autoregressive terms in models such as those examined here, see Clements and Galvao (2008).

In our prediction experiments, we also consider a very simple class of "hybrid" forecasting models that combine purely autoregressive models with more complex models that include diffusion and mixed frequency indices, via simple thresholding rules. The specification of these models is predicated on the fact that various machine learning, shrinkage and variable selection techniques that involve choices concerning which loss functions and tuning parameters to use are often utilized when the practitioner is faced with large-dimensional and multiple frequency datasets. However, the choice of loss functions and tuning parameters is complicated, and we wish to evaluate diffusion and mixed frequency indices through a very different lens. Namely, we propose an extremely simple alternative class of forecasting models based on so-called "thresholding", and assess whether this simple variety of models yields improved predictions, relative to the models outlined in equations (13), (14), and (15).[5] In particular, we examine the following hybrid prediction model:

$$\widehat{y}_{t+h}^{Hybrid} = I\{GDP_t > \tau_t\} \times \widehat{y}_{t+h}^{AR} + (1 - I\{GDP_t > \tau_t\}) \times \widehat{y}_{t+h}, \tag{16}$$

where $\widehat{y}_{t+h}^{AR}$ is the prediction from a purely autoregressive (AR) model, with lags selected via the SIC (this is called our AR(SIC) model), $\widehat{y}_{t+h}$ is a prediction from one of the models defined in equations (13), (14), and (15), and $GDP_t$ is the historical variable, in growth rates, that is used in this triggering mechanism. The thresholding parameter, $\tau_t$, is defined using a number of simple schemes. Namely, we consider a recursively estimated threshold (i.e., $\tau_t \equiv \tau_t^{rec} = \frac{1}{t}\sum_{j=1}^{t} GDP_j$, $t = R, ..., R + P - h$, with $R + P = T$), and a threshold estimated using rolling windows of data (i.e., $\tau_t \equiv \tau_t^{rol} = \frac{1}{R}\sum_{j=t-R+1}^{t} GDP_j$, $t = R, ..., R+P-h$). In this model, thus, a linear $AR(SIC)$ model is combined with an alternative model that includes diffusion and/or mixed frequency indices. In the prediction experiments reported on below, we also examine two other alternative models. These include multivariate distributed lag and multivariate autoregressive distributed lag models (see below for further discussion). These two models are "stand alone" prediction models, but are also combined with our benchmark $AR(SIC)$ model in order to specify two alternative variants of the hybrid model in equation (16). Finally, we also consider the case where $\tau_t \equiv \tau = 0$. All threshold parameters are

[5]See Chudik et al. (2016) for a related discussion in the context of variable selection using large-dimensional datasets.

estimated in a real-time fashion, in keeping with the real-time setup in this paper. However, in order to shed further light on the characteristics of $\tau_t$, we additionally calculate the fixed value of $\tau_t \equiv \tau^{post}$ that leads to "$MSFE$-best" predictions, ex post. This "cherry-picking" exercise allows us to assess the potential gains to our simple thresholding, were a fixed in-feasible threshold known in advance. We do not consider the case where a time varying in-feasible threshold is known in advance.

Consider GDP. One way to view our threshold is that we are eschewing the usual practice of defining thresholds based on signals concerning whether we are in expansion or recession, and are instead simply assessing, in real-time, whether GDP is growing above or below its average, as calculated using either a rolling or recursive data sample. When GDP is below its average, we construct predictions using our mode complex model. One way of thinking about this setup is that in times of lower growth, GDP is more difficult to predict, in the sense that the informational content of additional variables becomes relevant. At the same time, when growth is above average, there is "smooth sailing", and the informational content of other variables is subsumed in the autoregressive component of the model. This argument is not dissimilar to the threshold and "banding" arguments made in many papers, where, for example, a variable is assumed to follow a random walk in a certain "band", and is assumed to follow another process outside the "band". It is also not dissimilar to the observation that correlations amongst variables become pronounced in "bad times", while the same correlations are small in "good times".[6] Of course, the case where $\tau_t = 0$ is "closer" to the usual expansion/recession definitions. Our objective, then, is to assess whether extremely simple thresholding rules lead to improved prediction, and if so, which variable(s) do these thresholding rules work for.

In closing this section, it is worth stressing that switching between models depends solely upon the value of real GDP growth. This is true regardless of whether we are forecasting real GDP growth or CPI inflation. Interestingly, when CPI inflation is instead used to trigger switching between CPI models, hybrid model performance is actually worse than that based on a GDP trigger mechanism. This lends support to our claim discussed below that the triggering mechanism based on GDP, roughly speaking, acts as a signal of a "low-growth" state, in which case our more complicated models become more accurate, predictively. Moreover,

[6] A standard example is the case of correlation in the stock market. When in good times, financial instruments may be (nearly) independent, allowing for simple risk diversification and asset allocation based on Sharpe type regression analysis. This is often done in the hedge fund industry, for instance. However, when the market "goes south", correlations that were hithertofore zero become non-zero, causing a previously diversified portfolio to simply follow the market. Thus, the dynamic behavior of the financial instruments in this example change markedly depending on business conditions.

the “best” variable for signalling this state is GDP in our experiments. This makes sense, given the NBER statement that: “*A recession is a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales*”[7]; and given that CPI inflation does not appear in this NBER list of recession indicators.

# 3 Data

We utilize two distinct U.S. datasets. The first dataset is used for constructing mixed frequency indices. The frequencies of variables in this dataset range from daily to semi-annual. This dataset includes the variables used in the construction of the Federal Reserve Bank of Philadelphia (FRBP) business conditions index due to ADS (2009), hereafter called the ADS, as well as two GDP growth rate predictions tracked by the SPF and Livingston surveys. Non-survey series were obtained from FRED, the exhaustive online dataset maintained by the Federal Reserve Bank of St. Louis, while survey data were obtained from the online Real-Time Data Research Center website of the FRBP. The lowest frequency variable in the dataset is a GDP growth prediction from the Livingston Survey, which is available every six months and has been extensively studied in the literature. In our implementation, we include the mean and median of two-step ahead (two-quarter ahead) forecasts of real GDP growth from first half year of 1971 to second half year of 2012. Also included in our dataset is a quarterly SPF variable defined as projected real GDP growth. This survey variable is included for a time period spanning the fourth quarter of 1968 to the fourth quarter of 2012, and as in the case of the Livingston data, we examine mean and median variants, although in this case they correspond to one-quarter ahead forecasts, rather than bi-annual forecasts. The real gross domestic product variable (GDP) in our dataset is seasonally adjusted, and spans 1960.1 to 2013.2. Monthly indicators included in the dataset are the index of industrial production (1960.2-2013.6), total non-agricultural employees on payroll (1960.2-2013.6), real manufacturing, trade and sales (1967.1-2013.6), real personal income less transfer payment (1960.2-2013.6), and the consumer price index (CPI) for all items (1960.2-2013.6). All series are seasonally adjusted, and mnemonics used in the sequel when discussing these variables are given in panel (a) of Table 3. Finally, our weekly variable is initial claims for unemployment insurance for the period January 7, 1967 to June 29, 2013, and our daily variable is a government bond spread (i.e., the difference between the 10-year Treasury-bond yield and

[7]see http://www.nber.org/cycles/jan08bcdc_memo.html.

the 3-month Treasury-bill yield), for the period January 2, 1962 to June 28, 2013.

The second dataset that we utilize is an existing largescale macroeconomic dataset, and is used when constructing our diffusion indices. This dataset contains 143 monthly U.S. variables for the period 1960.1 - 2012.12, and is the dataset examined by Kim and Swanson (2014). It is this dataset that determines the length of our out-of-sample prediction period. Series in this dataset are contained in the following categories: industrial productions, employment, manufacturing, trade, sales, housing starts, inventories, orders or unfilled orders, stock price indices, exchange rates, interest rates and spreads, money and credit related quantities, and price related indices such as the consumer price index and personal consumption expenditures. This dataset also has a group of survey variables. One such variable is the Michigan consumer expectations index, and six others comprise a group known as the purchasing manager's indices (or national association of purchasing manager's indices). As mentioned elsewhere, all series are transformed to stationarity. A complete listing of the variables in this dataset, as well as transformations used, is available in the online appendix to Kim and Swanson (2018b).

# 4 Empirical Results

## 4.1 Experimental Setup

The benchmark model considered in this paper is the AR(SIC) model. This model is combined in a number of ways with diffusion and mixed frequency indices, as discussed above. We also estimate the following linear multivariate distributed lag (DL) models, and multivariate autoregressive distributed lag (DLAR) models. In particular, we construct forecasting models that are specified as follows:

$$\widehat{y}_{t+h} = \widehat{c}^h + \sum_{j=1}^{\overline{J}} \sum_{i=1}^{\widehat{p}^{h,j}} \widehat{\psi}_i^{h,j} x_{t-i}^j + \sum_{i=1}^{\widehat{q}^h} \widehat{\phi}_i^h y_{t-i}, \tag{17}$$

where the exogenous variables, $x_t$, in the above expression are the same variables used to extract our mixed frequency indices, and lags are selected via use of the SIC, as discussed above. The target variables in our experiments are quarterly real GDP growth and CPI inflation (see below for further details), and all prediction models are summarized in Table 1. The different sets of exogenous variables used in this context are described in Table 2. The forecast frequency in all cases is the same as the frequency of $y_t$. For a discussion of pastcasting, nowcasting, and prediction at lower frequencies than those at which the data are recorded, see Kim and Swanson (2018a).

When predicting lower frequency variables, such as GDP, using higher frequency variables, such as our mixed frequency indices (which are estimated at a daily frequency), the last available observation on the lower frequency variable is utilized in our forecasting models. In general, we ensure that predictions use all available information, in real-time. The frequency of the different variables used in our experiments is given in panel (a) of Table 2. Notice that in addition to examining the use of standard daily, weekly, and monthly indicators, we also include real GDP predictions from the well known SPF and Livingston surveys. This allows us to assess whether these surveys are useful, when utilized for the estimation of our mixed frequency indices, and when utilized in the $x_t$ variable given in equation (17). Finally, it should be stressed that our mixed frequency indicators are constructed using a small number of indicator variables (i.e., those listed in panel (a) of Table 3). panel (b) of the same table defines 37 different sets of mixed frequency indicators used to construct 37 alternative indices. The indicator set named "C" contains the set of variables used in the construction of the ADS business conditions index updated regularly by the FRBP. This set includes IC1, Pay, IP, RM, PI, and GDP, using the variable mnemonics contained in panel (a) of the table. This set subsumes the four variables used in the original ADS (2009) paper, and used in our above discussion of mixed frequency modeling. Our diffusion indices, on the other hand, are constructed using the large-scale monthly dataset discussed above, and do not utilize the same variables as those used in the construction of our mixed frequency indices. In all experiments, all variables are transformed to stationarity. For a listing of the variables used in diffusion index construction, see the online appendix to Kim and Swanson (2018b). For a discussion of the timing of our datasets, see the above section entitled "Data".

All mixed frequency and diffusion indices, as well as all parameters in our forecasting models, are estimated recursively, so that predictions are made in pseudo real-time. The reason that we use the phrase "pseudo real-time" is that we do not use real-time datasets in our analysis, and indeed there is no real-time dataset currently available that would allow us to do so. Using standard notation, our first estimations are carried out using the first $R$ observations in our datasets, and predictions for period $R+h$ are constructed. Then, models are re-estimated using $R+1$ observations, and predictions for period $R+h$ are constructed. This procedure is carried out until the sample is exhausted, yielding sets of $P-h+1$, predictions and prediction errors, where $h$ is the forecast horizon, set alternatively to 1, 2, 4, and 8 quarters ahead for predictions of real GDP growth, and 1, 3, 6, and 12 months ahead for predictions of our monthly CPI target variable. The prediction periods are 1987.1 to 2012.4 for quarterly real GDP growth, and 1987.1 to 2012.12 for monthly CPI inflation. We construct predictions for two varieties of these two target variables. Namely,

$$y^A_{t+h} = 100 \cdot \ln(Y_{t+h}/Y_t) \text{ and } y^I_{t+h} = 100 \cdot \ln(Y_{t+h}/Y_{t+h-1}),$$

where $Y_t$ denotes the "levels" variable, and the superscripts $A$ and $I$ denote whether cumulative or one-period ahead growth is targeted. Thus, two variants of each variable are predicted. Models are evaluated using the $MSFE$, defined as:

$$MSFE = \frac{1}{P-h+1} \sum_{t=R+h}^{T} (y_{t+h} - \widehat{y}_{t+h})^2,$$

and the ratios of $MSFE$s of model $i$, say, and the benchmark $AR(SIC)$ model are given by $RMSFE = MSFE_i/MSFE_{AR(SIC)}$. Inference on prediction sequences of model $i$, when compared with the $AR(SIC)$ model, is carried out using the pair-wise Diebold-Mariano (DM: 1995) predictive accuracy test. The null hypothesis of DM test statistic is:

$$H_0 : E\left(f(\varepsilon^h_{AR(SIC),t}) - f(\varepsilon^h_{i,t})\right) = 0,$$

where $\varepsilon^h_{AR(SIC),t}$ and $\varepsilon^h_{i,t}$ are the true prediction errors of model $i$ and the AR(SIC) model, respectively, and $f(\cdot)$ is the loss function, assumed to be quadratic. The DM test statistic is $DM = \sqrt{P}\frac{\overline{d}}{\sqrt{\widehat{\sigma}^2_d}}$, where $\overline{d} = \frac{1}{P}\sum_{t=R+h}^{T} \widehat{d}_t$, $\widehat{d}_t = (\widehat{\varepsilon}^h_{AR(SIC),t})^2 - (\widehat{\varepsilon}^h_{i,t})^2$. Here, $\widehat{\varepsilon}^h_{AR(SIC),t}$ and $\widehat{\varepsilon}^h_{i,t}$ are estimates of $\varepsilon^h_{AR(SIC),t}$ and $\varepsilon^h_{i,t}$. Additionally, $\widehat{\sigma}^2_d$ is a HAC standard error of $\widehat{d}_t$. The DM test statistic has a limiting standard normal distribution, under the assumption that parameter estimation error vanishes as $P, R \to \infty$, and that the two prediction models are non-nested. For a discussion of testing in nested contexts, see McCracken (2000) and Clark and McCracken (2001, 2005).

As a final check of our results, we note that in the forecasting literature, it is well known that simple forecast combinations often outperform forecasts based on individual models (e.g., see Timmermann (2006)). Given this fact, we consider 10 different equal weighted model combinations. The different model combinations are detailed in panel (a) of Table 2. Note that both smoothed and non-smoothed mixed frequency indices are utilized in all models and model combinations. Smoothing is discussed above.

## 4.2 Findings

Table 4 contains numerical summary statistics based on our prediction experiments. In particular, $RMSFE$s are tabulated for the "best" 5 models in each of 4 categories. The categories include: (1) the class of all models (i.e., benchmark linear, index, and combination models), excluding hybrid models of any kind; (2) the class of all models, but only including hybrid models with $\tau_t^{rec}$ thresholding; (3) the class of all models, but only including hybrid models with $\tau_t^{rol}$ thresholding; and (4) the class of all models, but only including hybrid models with $\tau = 0$ thresholding. Results of DM tests comparing each "winning" model to our benchmark AR(SIC) model are given by one, two, or three stars, denoting rejection of a one-sided DM test in favor or the non-AR(SIC) model, at the 1%, 5%, and 10% levels, respectively. Bold $RMSFE$s in the 2nd, 3rd, and 4th columns of numerical entries in this table denote cases where threshold-type models ranked from 1 through 5 have lower point $MSFE$s than similarly ranked models involving no thresholding. In these same columns, entries superscripted with an "A" denote models that have lower $MSFE$s than the very best non-threshold-type model, for each forecast horizon. Similarly, entries superscripted with an "B" denote models that have lower $MSFE$s than the very best non-threshold-type models, for each forecast horizon (excluding non-threshold combination models). Finally, entries superscripted with an "C" denote models that have lower $MSFE$s than the very best non-threshold-type models, for each forecast horizon (including only combination models).

The results contained in Table 4 point to a number of clear conclusions for the case of GDP. First, for 1-quarter ahead GDP prediction (see the first block of entries in panel (a) of Table 4), the ADS index is not only useful, but yields the very best prediction model, in terms of $RMSFE$. This occurs when $\tau_t^{rol}$ thresholding is utilized, yielding a $RMSFE$ of 0.705. Moreover, all forms of thresholding lead to improved predictive accuracy. However, in this "top" model, the ADS index is combined with diffusion indices, so that it is a hybrid model that utilizes both mixed frequency (MF) indices and diffusion (DI) indices that dominates all other specifications, when $h = 1$. Indeed, cursory examination of the GDP results across all forecast horizons in panel (a) indicates that many models ranked among the top 5 include a mixture of MF and DI indices. This result lends strong support to the notion that the daily ADS index produced by the FRBP is not only useful as a business conditions index, but is also highly useful for short-term GDP forecasting.

Drilling down a bit further yields additional interesting findings. In particular, if the $RMSFE$ of our best non-hybrid model is compared with the $RMSFE$ of our best hybrid model (i.e., the model with $\tau_t^{rol}$ thresholding), but only for the period of 2007.4 - 2009.2

(i.e., the period during the Great Recession), then the hybrid model yields a 34.3% $MSFE$ improvement. Notably, the same hybrid model still dominates across our entire forecast period (i.e., 1987.1 - 2012.1), but the improvement in $MSFE$ is only 9.4%.[8] In this sense, our mixed frequency and diffusion index models are clearly performing "best" during periods of low growth (e.g. the Great Recession), while during other epsidodes, the AR(SIC) model sometimes dominates, accounting for the relative reduction (from 34% to 9%) in $MSFE$ gains associated with use of our hybrid model when evaluating performance during the entire forecasting period.[9] These findings are mirrored when CPI inflation is examined (see Table 4, panel (b)), although it is recursive thresholding that yields the biggest $MSFE$ percentage gains, when utilizing our hybrid model, and gains during the Great Recession are only 8%, and decline to 3% when evaluating performance during the entire forecasting period. See Figures 2 and 3 for graphical depictions of these findings. In particular, to see how our thresholding approach fared during the recent Great Recession (GR), refer to Figure 2 for GDP and Figure 3 for CPI. Consider Figure 2. panel (b) of this figure blows up the period containing the GR, and the solid line (i.e., the line that reaches the lowest point over the entire time span plotted) in this panel is that of actual real GDP growth. Interestingly, it is *only* our hybrid model with $\tau_t^{rol}$ thresholding that achieves predictions near actual values at the very lowest point in the recession (see the shaded area in the plot). This is taken as evidence confirming our notion above finding that "low-growth" states lead to the superior performance of models that utilize $\tau_t^{rol}$ thresholding. Moreover, inspection of Panel C of the same figure reveals that the fact that $\tau_t^{rol}$ thresholding yields the "$MSFE$-best" hybrid model across our entire sample period (see the following paragraph for a complete discussion of this finding) is highly dependent upon the inclusion of the Great Recession in our forecasting period. Indeed, inspection of the plots in Panel C of Figure 3 reveals that $\tau_t^{rec}$ thresholding is almost always preferred to $\tau_t^{rol}$ thresholding, with the exception of the GR period, when comparing all predictions between 1987 and 2012.[10]

Second, the finding discussed above, concerning the fact that the very best model for $h = 1$ involves thresholding, carries over to all forecast horizons, regardless of whether the target forecast variable is $GDP_{t+h}^{A}$ or $GDP_{t+h}^{I}$. Namely, for $h = 1, 2, 4$, and 8, and for incremental or cumulative GDP growth, the 1st ranked "$MSFE$-best" models utilize $\tau_t^{rol}$ thresholding.

[8] This figure can be readily calculated by comparing the $RMSFE$s of 0.778 and 0.705 given in the first row of entries in Table 4, panel (a). For a graphical depiction of these findings, refer to Figure 2.

[9] For discussion of the changing performance over time of GDP prediction models, see Rossi and Sekhposyan (2010).

[10] As shall be subsequently discussed, $\tau_t^{rec}$ thresholding is always preferred to $\tau_t^{rol}$ thresholding for CPI inflation prediction, so that it is actually $\tau_t^{rec}$ thresholding is largely preferred in all of our experiments.

The only exception to this rule is the case of $GDP^{I}_{t+4}$ , where $\tau = 0$ thresholding yields the "$MSFE$-best" model. This result implies that not only is simple thresholding useful in contexts where mixtures of MF and DI indices are included, but that forecast combination, which is often unbeatable in a variety of aggregate macroeconomic forecasting contexts, is dominated under simple thresholding rules. Figure 1 plots $\tau^{rol}_t$ and $\tau^{rec}_t$ used in our prediction experiments. As expected, $\tau^{rol}_t$ is more volatile than $\tau^{rec}_t$ , and has decreased considerably in recent years, after remaining surprisingly stable for around 15 years. Interestingly, even $\tau^{rec}_t$ has decreased to all time lows in recent years, suggesting that use of a constant threshold is not optimal, perhaps due to regime shifts and other varieties of model instability. As confirmation of this conclusion, turn to Table B2 in Appendix B, in which $RMSFE$s that correspond to those in Table 4 are tabulated, except that the threshold is fixed, and is calculated ex-post (i.e., it is "cherry picked").[11] The interesting take-away from this table is that the best $RMSFE$s are often not appreciably lower than those calculated using our truly ex-ante, but time varying, thresholds. Turning back to Figure 1, it is noteworthy that while $\tau^{rol}_t$ does vary over time, it always lies between approximately 0.7% and 0.9%. This real GDP growth rate range is quite "tight", and while not zero, is rather close to zero. One way of viewing this feature is that prediction becomes "more difficult", hence requiring more complex models in certain regions of the range of GDP growth. However, unlike many papers that utilize notions of recession to determine thresholds, we use a very much simpler approach. Based on our findings, perhaps the economy should be viewed as being in a "low growth" state when quarterly real GDP growth falls below 1%, say, and not when standard recession dating metrics signal a recession.

Third, regardless of forecast horizon, and for all 4 categories, the top ranked model (i.e., the model denoted by "$Ranking = 1$" in the first column of the table, for each forecast horizon) utilizes mixed frequency indices constructed via use of the smoothed Kalman filter, as denoted by the inclusion of "SL" in model names. The incidence of models with and without autoregressive terms, as denoted by models that contain "AR" in their names, is spotty, and many top ranked models do not utilize AR terms. This speaks to the ability of MF/DI technology to "mop up" autoregressive information required for predicting real GDP growth.

Fourth, almost every hybrid model under $\tau^{rol}_t$ thresholding yields a $MSFE$ that is significantly lower than that of the $AR(SIC)$ model, and involves combining MD and DI indices. Additionally, our other benchmark linear models (i.e., the linear multivariate DL and mul-

[11] These "cherry picked" thresholds are given in Table B1.

tivariate autoregressive DL models) are never in the top 5 models, regardless of forecast horizon or whether we model $GDP^A_{t+h}$ or $GDP^I_{t+h}$. These findings lend further support to our earlier conclusions that $\tau^{rol}_t$ thresholding is useful, and that combination mixed frequency (MF) / diffusion index (DI) model that dominates all other specifications considered in this paper.

Now, consider CPI inflation prediction. Turning to panel (b) of Table 4, the first thing to note is that the findings discussed above in the context of predicting real GDP growth apply to CPI inflation, with a couple of notable exceptions. First, while thresholding remains very important, and indeed our hybrid models are "$MSFE$-best" in all cases except for $CPI^I_{t+12}$, it is largely $\tau^{rec}_t$ thresholding that yields the best models. Thus, it remains the case that when we are in a state of the economy involving "low growth", we benefit from using our more heavily parameterized hybrid models for predicitng CPI inflation. Interestingly, in the case of CPI, the threshold appears slightly more stable than in the case of GDP, in the sense that $\tau^{rec}_t$ thresholding is peferred to $\tau^{rol}_t$ thresholding. This is a somewhat surprising finding. Second, if only non-hybrid models are compared, model combination yields the "$MSFE$-best" model in 6 of 7 forecast horizon / target variable variant permutations. (Recall that model combination never "wins" under GDP prediction.) Still, in the truly "$MSFE$-best" models, which are always hybrid models, model combination never plays a role. Instead, it is always models utilizing mixtures of MF and DI indices that dominate, just as when predicting GDP.

Finally, note that percentage gains associated with use of our hybrid models vary from around 10% to as much as 40%. It remains to see whether these gains can be bested using other simple forms of thresholding, or by more complicated prediction models. Overall, though, our experiments are surprisingly robust in their support of the use of hybrid threshold models.

A final key element of the results contained in Table 4 pertains to the use of survey variables. Recall first that a variety of the indicator sets used in the construction of our mixed frequency indices include either Livingston or SPF survey predictions of GDP growth. In panel (b) of Table 3, it is noted that the indicator sets including these variables are sets "G" through "N". Thus, if top performing models in Table 4 include these letters in their names, we have direct evidence of the usefulness of these surveys. Findings vary greatly depending upon whether one is predicting GDP growth or CPI inflation. For GDP, in hybrid cases, a subset of top 5 performing models for $GDP^I_{t+2}$ and $GDP^I_{t+8}$ contain Livingston GDP predictions in their mixed frequency indices. No survey variables are contained in the top performing non-hybrid models. In stark contrast, for CPI, both Livingston and SPF forecasts appear in *all* 1st ranked specifications, from amongst all non-hybrid models. Moreover, when hybrid specifications are considered, at least one top 5 model includes either SPF or Livingston survey variables, regardless of forecast horizon or thresholding method; and regardless of whether $CPI^A_{t+h}$ or $CPI^I_{t+h}$ is being predicted. Moreover, in most cases it is a Livingston survey variable that is in the indicator set used for MF index construction. We conclude that this finding constitutes strong new evidence of the usefulness of these surveys, at multiple prediction horizons.[12]

Please now refer to Figures 4 and 5 in the paper. These figures contain plots of $RMSFE$s, cumulated over time as the forecast period increases. The plots in these figures aid us in discovering whether the gains associated with the use of our hybrid models are sample-period specific, for example. Consider Figure 3. The panels denoted by "Average" refer to predictions of $GDP^A_{t+h}$, while those denoted by "Increments" refer to predictions of $GDP^I_{t+h}$. It is apparent from inspection of these plots that our "$MSFE$-best" hybrid models with $\tau_t^{rol}$ thresholding are almost everywhere superior to "$MSFE$-best" non-thresholding (called "Linear Model") and combination models, regardless of forecast sample period. This story is less clear for CPI, as is evident upon inspection of Figure 4. In this figure, the various plots indicate a close race between $\tau_t^{rec}$ and $\tau_t^{rol}$ thresholding, although we know from Table 4 that $\tau_t^{rec}$ thresholding prevails. Moreover, model combination also fares well in the case of CPI prediction; a finding supported by noting that for non-hybrid models, combination always yields the 1st ranked model, regardless of forecast horizon, and regardless of whether $CPI^A_{t+h}$ or $CPI^I_{t+h}$ is being predicted.

# 5 Concluding Remarks

We present the results of a set of prediction experiments wherein standard linear specifications, including autoregressive and autoregressive distributed lag models, are compared with models utilizing: diffusion indices extracted from largescale monthly macroeconomic datasets, mixed frequency indices extracted from small mixed frequency datasets, and survey predictions. Additionally, we employ very simple recursive, rolling, and fixed thresholding in order to construct a class of "hybrid" models which "switch" between benchmark linear

[12]Note also that survey variables that differ from our SPF and Livingston survey variables are included in our diffusion index dataset, as discussed above. Thus, all diffusion indices contain survey variables, although we have not examined factor loadings or constructed diffusion indices with and without these variables in order to assess the relevance of the inclusion of these variables. Such an assessment is left to future research.

models and more complex models that also include diffusion indices, mixed frequency indices, and survey predictions. Thresholds are time-varying, and are determined in real-time by examining extant measures of GDP growth, using various windowing techniques. We find that thresholding is very useful for prediction, in the sense that hybrid GDP growth and CPI inflation prediction models are always preferred to a variety of alternative non-hybrid models. Additionally, the hybrid models perform particularly well during the Great Recession, suggesting that our simple thresholds are serving, roughly speaking, to differentiate between "good times" and "bad times", and that during "bad times" (or periods of "low growth"), a case can be made for specifying much more complicated prediction models than those useful during "good times". We further find that the daily ADS index produced by the FRBP is not only useful as a business conditions index, but is also highly useful for short-term GDP forecasting. Finally, we present strong new evidence of the predictive usefulness of GDP survey predictions from the Survey of Professional Forecasters and the Livingston Survey.

**Compliance with Ethical Standards:**

Funding: There is no funding source.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: This article does not contain any studies with human participants performed by any of the authors.

Table 1: Prediction Model Specifications*

| Model | Description |
|---|---|
| Group 1: *Autoregressive model (lags selected using the SIC)* | |
| AR(SIC) | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h y_{T-i+1}$ |
| Group 1: *Distribute Lag model (lags selected using the SIC)* | |
| DL | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{j=1}^{K} \sum_{i=1}^{p} \widehat{\beta}_i^{j,h} X_{T-i+1}^j$ |
| Group 1: *Autoregressive Distributed Lag model (lags selected using the SIC)* | |
| DLAR | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^{k,h} y_{T-i+1} + \sum_{j=1}^{k} \sum_{i=1}^{p} \widehat{\beta}_i^{j,h} X_{T-i+1}^j$ |
| Group 1: *Diffusion Index model* | |
| DI | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{j=1}^{K} \sum_{i=1}^{p} \widehat{\alpha}_i^{j,h} DI_{T-i+1}^j$ |
| Group 1: *Diffusion Index model with AR component* | |
| DIAR | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h y_{T-i+1} + \sum_{k=1}^{K} \sum_{i=1}^{p} \widehat{\beta}_i^{j,h} DI_{T-i+1}^j$ |
| Group 2: *Mixed Frequency model (MF)* | |
| MF_X_NSF | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h X\_NSF_{T-i+1}$ |
| MF_X_SF | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h y_{T-i+1} + \sum_{i=1}^{q} \widehat{\beta}_i^h X\_SF_{T-i+1}$ |
| MF_X_NSL | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h X\_NSL_{T-i+1}$ |
| MF_X_SL | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h y_{T-i+1} + \sum_{i=1}^{q} \widehat{\beta}_i^h X\_SL_{T-i+1}$ |
| Group 3: *Mixed Frequency model with AR component (MFAR)* | |
| MFAR_X_NSF | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i y_{T-i+1} + \sum_{i=1}^{q} \widehat{\beta}_i^h X\_NSF_{T-i+1}$ |
| MFAR_X_SF | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i X\_SF_{T-i+1}$ |
| MFAR_X_NSL | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h y_{T-i+1} + \sum_{i=1}^{q} \widehat{\beta}_i^h \mathrm{X\_NSL}_{T-i+1}$ |
| MFAR_X_SL | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h y_{T-i+1} + \sum_{i=1}^{q} \widehat{\beta}_i^h \mathrm{X\_SL}_{T-i+1}$ |
| Group 4: *Mixed Frequency model with DI component (MFDI)* | |
| MFDI_X_NSF | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h X\_NSF_{T-i+1} + \sum_{j=1}^{k} \sum_{i=1}^{q} \widehat{\beta}_i^{j,h} \mathrm{DI}_{T-i+1}^j$ |
| MFDI_X_SF | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h X\_NSF_{T-i+1} + \sum_{j=1}^{k} \sum_{i=1}^{q} \widehat{\beta}_i^{j,h} \mathrm{DI}_{T-i+1}^j$ |
| MFDI_X_NSL | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h X\_NSF_{T-i+1} + \sum_{j=1}^{k} \sum_{i=1}^{q} \widehat{\beta}_i^{j,h} \mathrm{DI}_{T-i+1}^j$ |
| MFDI_X_SL | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h X\_NSF_{T-i+1} + \sum_{j=1}^{k} \sum_{i=1}^{q} \widehat{\beta}_i^{j,h} \mathrm{DI}_{T-i+1}^j$ |
| Group 5: *Mixed Frequency model with DI and AR components (MFDIAR)* | |
| MFDIAR_X_NSF | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h y_{T-i+1} + \sum_{i=1}^{p} \widehat{\beta}_i^h \mathrm{X\_NSF}_{T-i+1} + \sum_{j=1}^{k} \sum_{i=1}^{q} \widehat{\gamma}_i^j \mathrm{DI}_{T-i+1}^j$ |
| MFDIAR_X_SF | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h y_{T-i+1} + \sum_{i=1}^{p} \widehat{\beta}_i^h \mathrm{X\_SF}_{T-i+1} + \sum_{j=1}^{k} \sum_{i=1}^{q} \widehat{\gamma}_i^j \mathrm{DI}_{T-i+1}^j$ |
| MFDIAR_X_NSL | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h y_{T-i+1} + \sum_{i=1}^{p} \widehat{\beta}_i^h \mathrm{X\_NSL}_{T-i+1} + \sum_{j=1}^{k} \sum_{i=1}^{q} \widehat{\gamma}_i^j \mathrm{DI}_{T-i+1}^j$ |
| MFDIAR_X_SL | $y_{T+h\|T} = \widehat{\alpha}_0^h + \sum_{i=1}^{p} \widehat{\alpha}_i^h y_{T-i+1} + \sum_{i=1}^{p} \widehat{\beta}_i^h \mathrm{X\_SL}_{T-i+1} + \sum_{j=1}^{k} \sum_{i=1}^{q} \widehat{\gamma}_i^j \mathrm{DI}_{T-i+1}^j$ |

* Notes: Prediction models are grouped into five groups. The first group contains benchmark models, i.e, AR, DL, DLAR, DI and DIAR models. The second group contains mixed frequency indices, while the third group is the same, but with added autoregressive terms. Finally, the fourth group of models combines mixed frequency and diffusion indices, while fifth group is the same, but with added autoregressive terms. The mnemonics introduced in this table are utilized in Table 4, wherein prediction experiment results are gathered.

## Table 2: Prediction Model Specifications - Additional Details*

### (a) Combination Models and Thresholds

| *Model* | *Description* |
|---|---|
| Group 6: *Forecast Combinations* | |
| $\widehat{y}_{T+h\|T} = \frac{1}{I}\sum_{i=1}^{I}\widehat{y}^{i}_{T+h\|T}, i = 1, .., I,$ *where* $I =$ *number of models used in combination* | |
| $CMA1$ | *All benchmark models i.e,* $AR, DL, DLAR, DI$ *and* $DIAR$ |
| $CMA2$ | *All nonsmoothed* $MF$ *models* |
| $CMA3$ | *All smoothed* $MF$ *models* |
| $CMA4$ | *All nonsmoothed* $MFAR$ *models* |
| $CMA5$ | *All smoothed* $MFAR$ *models* |
| $CMA6$ | *All nonsmoothed* $MFDI$ *models* |
| $CMA7$ | *All smoothed* $MFDI$ *models* |
| $CMA8$ | *All nonsmoothed* $MFDIAR$ *models* |
| $CMA9$ | *All smoothed* $MFDIAR$ *models* |
| $CMA10$ | *All models* |
| *Hybrid Models* | |
| | *If* real GDP *growth at time* $t > \tau_t$, *then* *use AR(SIC) model for forecasting,* *else use one selected alternative model from Groups 1-5* |
| *Threshold Mechanisms (recursive, rolling, and zero) used in hybrid models* | |
| *Recursive* | $\tau_t^{rec} = \frac{1}{t}\sum_{j=1}^{t} GDP_j$, for $t = R, ..., R+P-h$. |
| *Rolling* | $\tau_t^{rol} = \frac{1}{R}\sum_{j=t-R+1}^{t} GDP_j$, for $t = R, ..., R+P-h$. |
| *Zero* | $\tau_t = \tau = 0$ (real GDP growth rate equals 0). |

### (b) Smoothing and Matching Timing

| | First Day | Last Day |
|---|---|---|
| Smoothed | SF | SL |
| Nonsmoothed | NSF | NSL |

* Notes: See notes to Table 1. A 6th group of prediction models containing various forecasts combinations is given in panel (a), along with definitions of thesholding mechanisms used in hybrid models that comibine an AR(SIC) model with one of the other models listed in Table 1. Notation denoting the use of smoothed and unsmoothed mixied frequency indices is given in panel (b) of the table. In this notation, "S" and "NS" denote smoothed and unsmoothed indices, respectively. Additionally, in the notation introduced in panel (b), and in the context of mixed frequency modelling, "F" denotes models for which monthly (quarterly) indicators utilize observations from the first day of the month (or quarter), while "L" denotes models for which monthly (quarterly) indicators utilize observations from the last day of the month (or quarter). Thus, for forecasting quarterly variables, the "F" indicator sets which utilize both quarterly and monthly data, use "first day" observations, while for forecasting monthly variables, only monthly indicators utilize observations from the first day of the month. "F" and "L" experiments are performed using indicator sets A to F (see panel (b) of Table 3 for a list of indicator sets). All other indicator sets utilize "L" dating. For further details, refer to Section 4 and Appendix 1.

Table 3: Variable Definitions and Indicator Sets Used for Constructing Mixed Frequency Indices*

(a) Variable Definitions

| Frequency | Variables | Abbreviation | Transformation |
|---|---|---|---|
| Daily | Government Bond Spread | SPR | $X_t - X_{t-1}$ |
| Weekly | Initial Claims for Unemployment Insurance | | |
| | Level | IC1 | $X_t$ |
| | Growth Rate | IC2 | $\ln(X_t) - \ln(X_{t-1})$ |
| Monthly | Payroll | Pay | $\ln(X_t) - \ln(X_{t-1})$ |
| | Industrial Production | IP | $\ln(X_t) - \ln(X_{t-1})$ |
| | Real Manufacturing Trade & Sales | RM | $\ln(X_t) - \ln(X_{t-1})$ |
| | Real Personal Income less Transfer Payments | PI | $\ln(X_t) - \ln(X_{t-1})$ |
| | Consumer Price Index | CPI | $\ln(X_t) - \ln(X_{t-1})$ |
| Quarterly | Real GDP growth, historical data | GDP | $\ln(X_t) - \ln(X_{t-1})$ |
| | Real GDP growth, mean, SPF | SPF1 | |
| | Real GDP growth, median, SPF | SPF2 | |
| | Real GDP growth, mean, Livingston | LIV1 | |
| | Real GDP growth, median, Livingston | LIV2 | |

(b) Indicator Sets Used for Constructing Mixed Frequency Indices

| Set | Variables used in Index Construction | Set | Variables used in Index Construction |
|---|---|---|---|
| D1 | SPR | | |
| W1 | IC1 | WM4 | IC1,Pay, IP, RM, PI |
| M1 | Pay | Q1 | GDP |
| M2 | Pay, IP | MQ1 | Pay, GDP |
| M3 | Pay, IP, RM | MQ2 | Pay, IP, GDP |
| M4 | Pay, IP, RM, PI | MQ3 | Pay, IP, RM, GDP |
| WM1 | IC1,Pay | MQ4 | Pay, IP, SPF1 |
| WM2 | IC1,Pay,IP | MQ5 | Pay, IP, RM, PI, GDP |
| WM3 | IC1,Pay, IP, RM | INF | IC1,Pay, IP, RM, PI, CPI, GDP |
| A | IC1, Pay, IP, GDP | K | SPR, IC1, Pay, IP, RM, PI, GDP, SPF1 |
| B | IC1, Pay, IP, RM, GDP | L | SPR, IC1, Pay, IP, RM, PI, GDP, SPF2 |
| C | IC1, Pay, IP, RM, PI, GDP | M | SPR, IC1, Pay, IP, RM, PI, GDP, LIV1 |
| D | SPR, IC1, Pay, IP, GDP | N | SPR, IC1, Pay, IP, RM, PI, GDP, LIV2 |
| E | SPR, IC1, Pay, IP, RM, GDP | O | IC2, Pay, IP, GDP |
| F | SPR, IC1, Pay, IP, RM, PI, GDP | P | IC2, Pay, IP, RM, GDP |
| G | IC1, Pay, IP, RM, PI, GDP, SPF1 | Q | IC2, Pay, IP, RM, PI, GDP |
| H | IC1, Pay, IP, RM, PI, GDP, SPF2 | R | SPR, IC2, Pay, IP, GDP |
| I | IC1, Pay, IP, RM, PI, GDP, LIV1 | S | SPR, IC2, Pay, IP, RM, GDP |
| J | IC1, Pay, IP, RM, PI, GDP, LIV2 | T | SPR, IC2, Pay, IP, RM, PI, GDP |

* Notes: panel (a) contains variable definitions. The government bond spread is difference between the 10 year Treasury constant maturity yield and 3 month Treasury-Bill yield. See Section 3 for complete details concerning the sample periods of the datsets used in prediction experiments. Our levels initial claims variable corresponds to that used in ADS (2009), although a growth variant of this variable is also included in various experiments. Two varieties of GDP survey predictions are utilized in our experiments, including 1-quarter ahead predictions from the Survey of Professional Forecasters (SPF), and 2-quarter ahead predictions from the Livingston Survey (LIV).Various mixed frequency indices are constructed as discussed in Section 2.2, using the indicator sets summarized in panel (b) of the table. All such indices are constructed at a daily frequency.

Table 4: Top-5 MSFE-Best Models, by Forecast Horizon and Variable Transformation*

(a) Real GDP growth

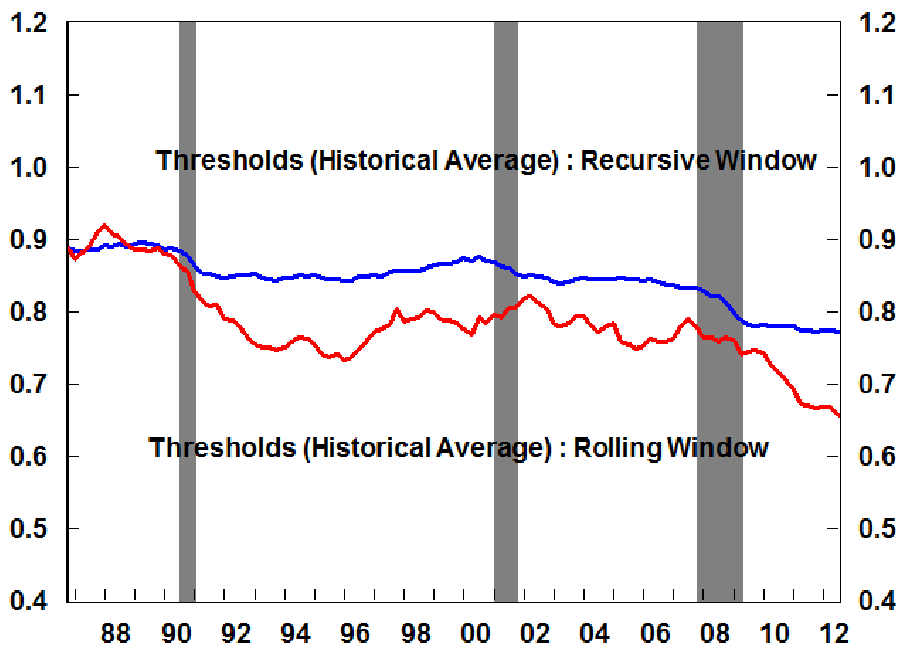
| *Ranking* | *Linear & Average* | *RMSFE* | *Hybrid Model with* $\tau^{\text{rec}}$ | *RMSFE* | *Hybrid Model with* $\tau^{\text{rol}}$ | *RMSFE* | *Hybrid Model with* $\tau$=0 | *RMSFE* |
|---|---|---|---|---|---|---|---|---|
| **GDP**$_{t+1}$ | | | | | | | | |
| 1 | MF_B_NSL | 0.778 | MFAR_MQ2_SL | $\mathbf{0.727}^{*A}$ | MFDI_C_SF | $\mathbf{0.705}^{*A}$ | MFDIAR_B_SF | $\mathbf{0.766}^{A}$ |
| 2 | MFAR_O_NSL | 0.779* | MFDI_C_SF | $\mathbf{0.728}^{A}$ | MFAR_S_NSL | $\mathbf{0.711}^{*A}$ | MFDIAR_M4_SL | $\mathbf{0.766}^{A}$ |
| 3 | MFAR_B_NSL | 0.780 | MFAR_MQ4_SL | $\mathbf{0.728}^{*A}$ | MFDIAR_C_SF | $\mathbf{0.713}^{*A}$ | MFDI_M3_SL | $\mathbf{0.767}^{A}$ |
| 4 | MF_O_NSL | 0.782* | MFDIAR_MQ4_SL | $\mathbf{0.730}^{*A}$ | MFAR_MQ2_SL | $\mathbf{0.716}^{*A}$ | MFDIAR_M3_SL | $\mathbf{0.767}^{A}$ |
| 5 | MFAR_MQ2_NSL | 0.783* | MF_MQ4_SL | $\mathbf{0.731}^{*A}$ | MFAR_T_NSL | $\mathbf{0.718}^{*A}$ | DLAR | $\mathbf{0.768}^{A}$ |
| **GDP**$^{A}_{t+2}$ | | | | | | | | |
| 1 | MFDI_MQ1_SL | 0.775 | MFDI_MQ1_SL | $\mathbf{0.718}^{*A}$ | MFDI_MQ1_SL | $\mathbf{0.710}^{*A}$ | MFDIAR_M3_SL | $0.782^{C}$ |
| 2 | MFDI_MQ2_SL | 0.780 | MFDI_MQ2_SL | $\mathbf{0.741}^{A}$ | MFDI_MQ2_SL | $\mathbf{0.729}^{*A}$ | MFDI_MQ3_SL | $0.782^{C}$ |
| 3 | MF_A_NSL | 0.791 | MF_MQ2_SL | $\mathbf{0.747}^{A}$ | MF_MQ2_SL | $\mathbf{0.732}^{A}$ | MFDIAR_MQ3_SL | $0.784^{C}$ |
| 4 | MFDIAR_MQ2_SL | 0.791 | MFDIAR_MQ1_SL | $\mathbf{0.749}^{*A}$ | MFDIAR_MQ4_SL | $\mathbf{0.745}^{A}$ | MF_MQ3_SL | $0.784^{C}$ |
| 5 | MFDIAR_MQ1_SL | 0.793 | MFDIAR_MQ4_SL | $\mathbf{0.760}^{A}$ | MFDIAR_MQ1_SL | $\mathbf{0.746}^{*A}$ | MFAR_MQ3_SL | $0.785^{C}$ |
| **GDP**$^{A}_{t+4}$ | | | | | | | | |
| 1 | MFDIAR_Q_SL | 0.839 | MFDIAR_Q_SL | $\mathbf{0.812}^{A}$ | MFDIAR_Q_SL | $\mathbf{0.804}^{A}$ | MF_E_SF | $0.868^{*C}$ |
| 2 | MFAR_Q_SL | 0.841 | MFDIAR_T_SL | $\mathbf{0.817}^{A}$ | MFAR_Q_SL | $\mathbf{0.809}^{A}$ | MF_F_SF | $0.868^{*C}$ |
| 3 | MFDI_Q_SL | 0.848 | MFAR_Q_SL | $\mathbf{0.819}^{A}$ | MFDIAR_T_SL | $\mathbf{0.809}^{A}$ | MFAR_MQ3_SL | $0.870^{*C}$ |
| 4 | MF_Q_SL | 0.851 | MFDI_Q_SL | $\mathbf{0.824}^{A}$ | MFDI_Q_SL | $\mathbf{0.815}^{A}$ | MFAR_E_SF | $0.870^{*C}$ |
| 5 | MFDI_O_SL | 0.858 | MF_Q_SL | $\mathbf{0.828}^{A}$ | MF_Q_SL | $\mathbf{0.818}^{A}$ | MFAR_F_SF | $0.870^{*C}$ |
| **GDP**$^{A}_{t+8}$ | | | | | | | | |
| 1 | MF_Q_SL | 0.866** | MF_Q_SL | $\mathbf{0.863}^{**A}$ | MF_Q_SL | $\mathbf{0.858}^{**A}$ | MF_E_SF | $0.921^{**C}$ |
| 2 | MFAR_Q_SL | 0.883** | MF_T_SL | $0.877^{**C}$ | MF_T_SL | $0.871^{**C}$ | MF_F_SF | $0.921^{**C}$ |
| 3 | MF_T_SL | 0.884** | MFDIAR_T_SL | $0.882^{**C}$ | MFDIAR_T_SL | $0.877^{**C}$ | MF_Q_SL | $0.923^{**C}$ |
| 4 | MFDI_T_SL | 0.890** | MFDI_T_SL | $0.884^{**C}$ | MFAR_Q_SL | $0.879^{**C}$ | MFAR_Q_SL | $0.927^{**C}$ |
| 5 | MFDIAR_T_SL | 0.890** | MFAR_T_SL | $0.887^{**C}$ | MFDI_T_SL | $0.880^{**C}$ | MF_T_SL | $0.927^{**C}$ |
| **GDP**$^{I}_{t+2}$ | | | | | | | | |
| 1 | MFDIAR_MQ1_SL | 0.877* | MFDIAR_MQ1_SL | $\mathbf{0.857}^{*A}$ | MFDIAR_MQ1_SL | $\mathbf{0.858}^{*A}$ | MFAR_M4_SL | $0.906^{C}$ |
| 2 | MFDI_MQ1_SL | 0.879* | MFDI_MQ1_SL | $\mathbf{0.858}^{**A}$ | MFDI_MQ1_SL | $\mathbf{0.858}^{**A}$ | MF_M4_SL | $0.924^{C}$ |
| 3 | MFDI_O_SL | 0.908 | MF_MQ1_SL | $0.895^{**C}$ | MF_MQ2_SL | $0.890^{C}$ | MFDIAR_M_SL | $0.926^{C}$ |
| 4 | MFDIAR_O_SL | 0.910 | MFAR_MQ1_SL | $0.899^{*C}$ | MFDIAR_M1_SL | $0.892^{*C}$ | MFAR_M_SL | $0.929^{C}$ |
| 5 | MFDI_MQ2_SL | 0.911 | MFDIAR_M1_SL | $0.899^{*C}$ | MFDI_M1_SL | $0.892^{*C}$ | MFDIAR_M4_SL | $0.933^{C}$ |
| **GDP**$^{I}_{t+4}$ | | | | | | | | |
| 1 | MFAR_INF_SL | 0.989 | MFAR_INF_SL | $0.993^{C}$ | MFAR_INF_SL | $0.993^{C}$ | MFAR_D_SF | $\mathbf{0.986}^{*A}$ |
| 2 | MFDIAR_T_SL | 0.993 | MFAR_T_SL | $0.995^{C}$ | MFAR_T_SL | $0.996^{C}$ | MFAR_INF_SL | $0.993^{**C}$ |
| 3 | MFAR_Q1_SL | 0.993* | MFDIAR_T_SL | $0.995^{C}$ | MFAR_WM4_SL | $0.996^{C}$ | MFDIAR_WM3_SL | $0.994^{C}$ |
| 4 | MFAR_C_SF | 0.994 | MFAR_D_SF | $0.996^{C}$ | MFAR_WM3_SL | $0.997^{C}$ | MFDIAR_MQ5_SL | $0.994^{C}$ |
| 5 | MFAR_T_SL | 0.994 | MFAR_C_SF | $0.996^{C}$ | MFAR_C_SF | $0.998^{C}$ | MF_INF_SL | $0.995^{C}$ |
| **GDP**$^{I}_{t+8}$ | | | | | | | | |
| 1 | MF_F_SF | 0.964* | MF_F_SF | $\mathbf{0.951}^{**A}$ | MF_J_SL | $\mathbf{0.934}^{***A}$ | MF_J_SL | $0.961^{**C}$ |
| 2 | MF_D_SF | 0.965* | MF_D_SF | $\mathbf{0.952}^{***A}$ | MF_F_SF | $\mathbf{0.952}^{**A}$ | MF_J_NSL | $0.972^{**C}$ |
| 3 | MF_D1_NSL | 0.967* | MF_INF_SL | $\mathbf{0.953}^{**A}$ | MF_D_SF | $\mathbf{0.954}^{***A}$ | MF_F_SF | $0.975^{*C}$ |
| 4 | MF_D1_SL | 0.968* | MF_E_SF | $\mathbf{0.955}^{**A}$ | MF_INF_SL | $\mathbf{0.956}^{**A}$ | MF_INF_SL | $0.976^{C}$ |
| 5 | MF_E_SF | 0.968* | MF_INF_NSL | $\mathbf{0.956}^{**A}$ | MF_E_SF | $\mathbf{0.956}^{**A}$ | MF_D_SF | $0.977^{*C}$ |

(b) CPI Inflation

| *Ranking* | *Linear & Average* | *RMSFE* | *Hybrid Model with* $\tau^{rec}$ | *RMSFE* | *Hybrid Model with* $\tau^{rol}$ | *RMSFE* | *Hybrid Model with* $\tau$=0 | *RMSFE* |
|---|---|---|---|---|---|---|---|---|
| | | | | | $\mathbf{CPI}_{t+1}$ | | | |
| 1 | MFDIAR_S_NSL | 0.972 | MFAR_D_SL | **0.964**$^{A}$ | MFAR_MQ5_SL | **0.971**$^{A}$ | MFAR_D_SL | **0.971**$^{A}$ |
| 2 | MFAR_MQ5_SL | 0.972 | MFAR_N_SL | **0.966**$^{A}$ | MFAR_D_SL | **0.971**$^{A}$ | MFAR_MQ5_SL | 0.975 |
| 3 | CMA10 | 0.974 | MFAR_MQ5_SL | **0.969**$^{A}$ | MFAR_N_SL | **0.972**$^{A}$ | MFAR_N_SL | 0.978 |
| 4 | MFAR_I_SL | 0.975 | MFAR_I_SL | **0.970**$^{A}$ | MFAR_W1_NSL | 0.974 | MFAR_F_SL | 0.982 |
| 5 | MFAR_D_SL | 0.976 | MFAR_F_SL | **0.973**$^{A}$ | MFAR_W1_SL | 0.975 | MFAR_E_SL | 0.983 |
| | | | | | $\mathbf{CPI}^{A}_{t+3}$ | | | |
| 1 | CMA10 | 0.913** | MFDI_WM3_SL | **0.896**$^{A}$ | MFDI_WM3_SL | **0.922**$^{*B}$ | MFDI_WM1_SL | **0.910**$^{**A}$ |
| 2 | MFDI_WM3_SL | 0.948 | MFDI_E_NSL | **0.903**$^{A}$ | MFDI_E_NSL | **0.928**$^{B}$ | MFDI_MQ1_NSL | **0.911**$^{**A}$ |
| 3 | MFDI_WM4_SL | 0.959 | MFDI_WM4_SL | **0.907**$^{A}$ | MFDI_N_SL | **0.931**$^{*B}$ | MFDI_R_NSL | **0.911**$^{**A}$ |
| 4 | MFDI_MQ1_SL | 0.960 | MFDI_N_SL | **0.908**$^{A}$ | MFDI_N_NSL | **0.933**$^{*B}$ | MFDI_WM2_NSL | **0.912**$^{**A}$ |
| 5 | MFDI_MQ1_NSL | 0.962 | MFDI_N_NSL | **0.908**$^{A}$ | MFDI_WM4_SL | **0.933**$^{B}$ | MFDI_WM1_NSL | **0.912**$^{**A}$ |
| | | | | | $\mathbf{CPI}^{A}_{t+6}$ | | | |
| 1 | CMA10 | 0.831*** | MFDI_MQ5_SL | **0.744**$^{**A}$ | MFDI_MQ5_SL | **0.769**$^{**A}$ | MF_M_NSL | **0.821**$^{**A}$ |
| 2 | MFDI_W1_NSL | 0.854 | MFDI_WM3_SL | **0.755**$^{**A}$ | MFDI_WM3_SL | **0.774**$^{**A}$ | MFDI_WM1_SL | **0.824**$^{**A}$ |
| 3 | MFDI_WM3_SL | 0.856 | MFDI_D_SL | **0.757**$^{**A}$ | MFDI_N_SL | **0.775**$^{**A}$ | MFDI_WM1_NSL | **0.824**$^{**A}$ |
| 4 | MFDI_MQ5_SL | 0.856 | MFDI_N_SL | **0.757**$^{**A}$ | MFDI_D_SL | **0.776**$^{**A}$ | MFDI_D_SL | **0.826**$^{**A}$ |
| 5 | MFDI_MQ1_SL | 0.859 | MFDI_INF_NSL | **0.763**$^{**A}$ | MFDI_E_SL | **0.780**$^{**A}$ | MFDI_R_NSL | **0.826**$^{**A}$ |
| | | | | | $\mathbf{CPI}^{A}_{t+12}$ | | | |
| 1 | CMA10 | 0.750*** | MFDI_MQ5_SL | **0.681**$^{***A}$ | MFDI_N_SL | **0.709**$^{***A}$ | MFDI_N_SL | **0.813**$^{***A}$ |
| 2 | MFDI_WM3_SL | 0.801** | MFDI_N_SL | **0.682**$^{***A}$ | MFDI_MQ5_SL | **0.716**$^{***A}$ | MFDI_E_SL | **0.813**$^{***A}$ |
| 3 | MFDI_MQ5_SL | 0.802** | MFDI_D_SL | **0.689**$^{***A}$ | MFDI_D_SL | **0.717**$^{***A}$ | MFDI_D_SL | **0.813**$^{***A}$ |
| 4 | MFDI_WM4_SL | 0.810* | MFDI_WM3_SL | **0.689**$^{***A}$ | MFDI_E_SL | **0.717**$^{***A}$ | MFDI_D1_SL | **0.815**$^{***A}$ |
| 5 | MFDI_D1_SL | 0.811* | MFDI_E_SL | **0.691**$^{***A}$ | MFDI_WM3_SL | **0.722**$^{***A}$ | MFDI_D1_NSL | **0.815**$^{***A}$ |
| | | | | | $\mathbf{CPI}^{I}_{t+3}$ | | | |
| 1 | CMA10 | 0.901** | MFDI_F_NSF | **0.880**$^{**A}$ | MFDI_WM3_SL | **0.889**$^{A}$ | MFDI_F_NSF | **0.884**$^{**A}$ |
| 2 | MFDI_WM3_SL | 0.915 | MFDI_WM3_SL | **0.881**$^{A}$ | MFDI_F_NSF | **0.890**$^{*A}$ | MFDI_D_SF | **0.888**$^{**A}$ |
| 3 | MFDI_WM4_SL | 0.915 | MFDI_WM4_SL | **0.883**$^{A}$ | MFDI_D1_NSL | **0.893**$^{*A}$ | MFDI_Q1_NSL | **0.889**$^{**A}$ |
| 4 | MFDI_F_NSF | 0.916 | MFDI_D1_NSL | **0.886**$^{A}$ | MFDI_WM4_SL | **0.893**$^{*A}$ | MFDI_D_NSF | **0.889**$^{**A}$ |
| 5 | MFDI_D1_NSL | 0.917 | MFDI_MQ5_SL | **0.887**$^{A}$ | MFDI_D_NSF | **0.896**$^{*A}$ | MFDI_D1_NSL | **0.890**$^{**A}$ |
| | | | | | $\mathbf{CPI}^{I}_{t+6}$ | | | |
| 1 | CMA10 | 0.946** | MFDI_W1_SL | **0.887**$^{***A}$ | MFDI_K_SL | **0.875**$^{***A}$ | MF_M1_NSL | **0.885**$^{**A}$ |
| 2 | CMA1 | 0.968 | MFDI_N_SL | **0.888**$^{**A}$ | MFDI_F_SF | **0.877**$^{***A}$ | MF_M_NSL | **0.885**$^{**A}$ |
| 3 | MFDI_W1_SL | 0.970 | MFDI_K_SL | **0.889**$^{**A}$ | MFDI_W1_SL | **0.880**$^{***A}$ | MF_WM1_SL | **0.886**$^{**A}$ |
| 4 | MFDI_MQ5_SL | 0.970 | MFDI_E_SL | **0.890**$^{**A}$ | MFDI_N_NSL | **0.880**$^{***A}$ | MF_MQ1_NSL | **0.889**$^{**A}$ |
| 5 | MFDI_D1_SL | 0.973 | MFDI_F_SF | **0.890**$^{**A}$ | MFDI_N_SL | **0.880**$^{***A}$ | MF_M2_NSL | **0.892**$^{**A}$ |
| | | | | | $\mathbf{CPI}^{I}_{t+12}$ | | | |
| 1 | CMA10 | 0.928*** | MFDI_I_SL | **0.943**$^{**B}$ | MFDI_D1_SL | 0.965 | MFDI_F_NSF | 0.959*** |
| 2 | MFDI_K_SL | 0.944* | MFDI_M1_SL | 0.944** | MFDI_D1_NSL | 0.965 | MFDI_I_NSL | 0.960*** |
| 3 | MFDI_F_SF | 0.945* | MFDI_F_NSF | 0.944** | MFDI_N_SL | 0.969 | MFDI_F_NSL | 0.960*** |
| 4 | MFDI_S_SL | 0.947* | MFDI_MQ1_NSL | 0.944** | MFDI_I_SL | 0.969 | MFDI_F_SF | 0.960*** |
| 5 | MFDI_E_SF | 0.948* | MFDI_D1_SL | 0.944** | MFDI_MQ1_NSL | 0.970 | MF_M1_SL | 0.960*** |

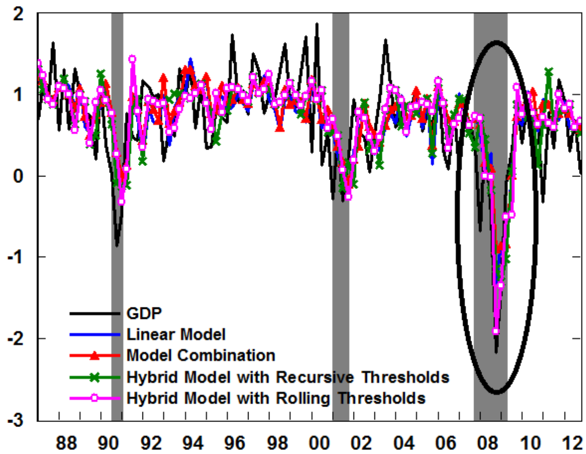
(*) Notes: This table summarizes the top 5 models for 4 groups of models. The first group, denoted "Linear & Average", contains all models listed in Table 1. In this group, no hybrid models are included, but every other variety of model is, including those with mixed frequency indices and diffusion indices. Combination models are also included in this group. The remaining three groups of models are all hybrid type variants of the models given in Table 1, whereby the thresholding summarized in panel (a) of Table 2 is implemented. Numerical entries are relative mean square forecast errors (RMSFEs) where the MSFE of the listed model is divided by the MSFE of the benchmark AR(SIC) model. *, **, and *** denote rejection, at the 10%, 5% and 1% levels, respectively, of a one-sided DM predictive accuracy test comparing the listed model with the AR(SIC) model. Rejections are indicative of the AR(SIC) being "MSFE-inferior". Mnemonics used to denote the different models are given in Tables 1-3. Entries superscripted with an "A" denote cases where the top ranked "Linear & Average' model is "MSFE-inferior" to a particular hybrid model. Entries superscripted with an "B" denote cases where the top ranked "Linear & Average' model, excluding forecast combination models, is "MSFE-inferior" to a particular hybrid model. Entries superscripted with an "C" denote cases where the top ranked "Linear & Average' model, including only forecast combination models, is "MSFE-inferior" to a particular hybrid model. Dynamic thresholds used in hybrid models are plotted in Figure 1. See Sections 4 for complete details.

Figure 1: Rolling and Recursive Real GDP Growth Thresholds*

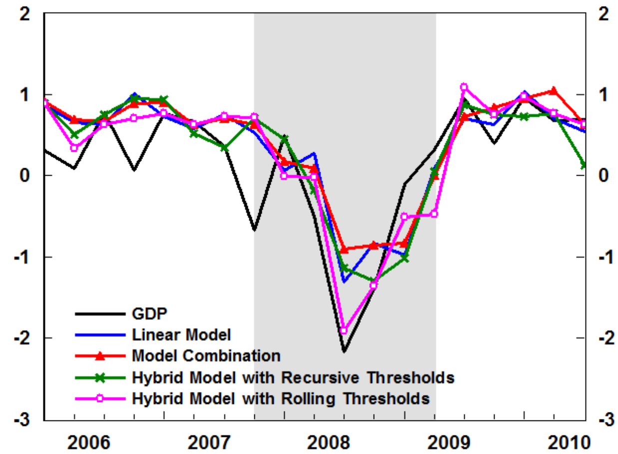



* Notes: The figure displays recursive and rolling dynamic thresholds defined in panel (a) of Table 2, and used in construction forecasts based on the hybrid models reported on in Table 4. NBER recession periods are shaded.

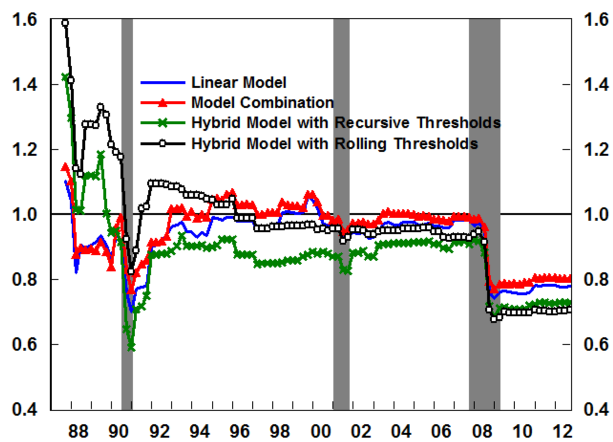
Figure 2: One-Quarter Ahead Real GDP Growth Forecasts*




(a) Whole Forecasting Period



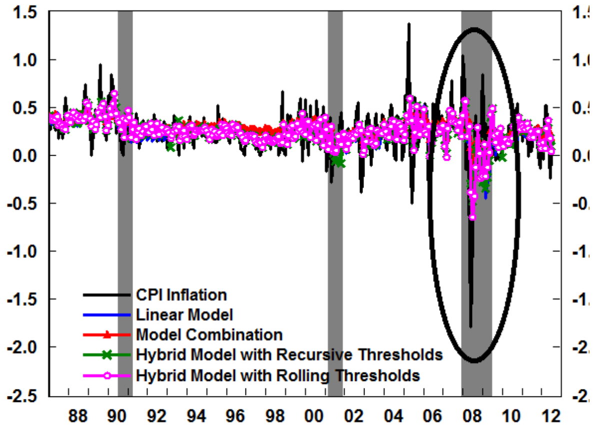

(b) During Great Recession



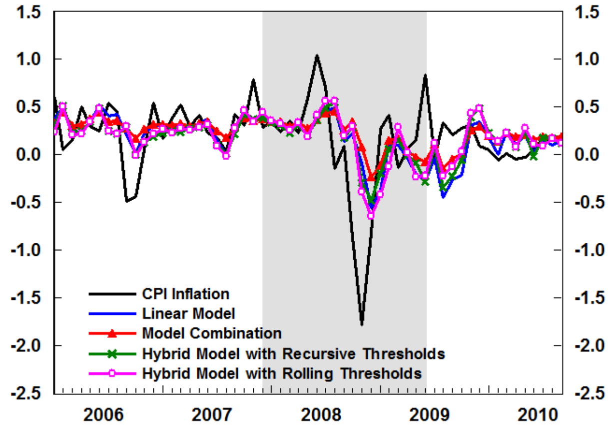

(c) Cumulative RMSFEs

* Notes: Panel (a) and (b) plot actual GDP against 1-quarter ahead GDP forecasts from various prediction models. The shaded areas indicate NBER recessions. Panel C plots corresponding cumulative RMSFEs, where the divisor MSFE used in the construction of the RMSFEs is the benchmark AR(SIC) model. See Section 4 and the notes to Tables 1-4 for further details.

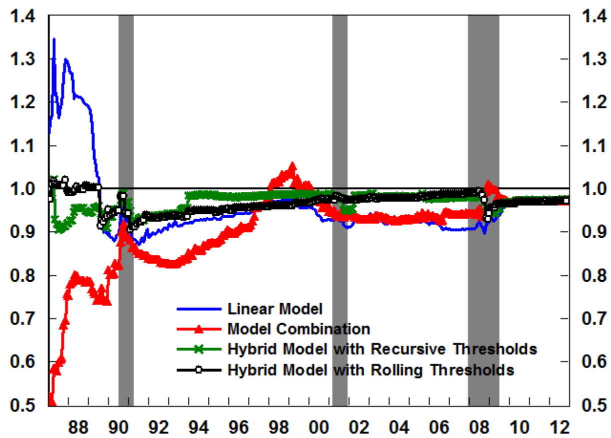
Figure 3: One-Month Ahead CPI Inflation Forecasts*




(a) Whole Forecasting Period



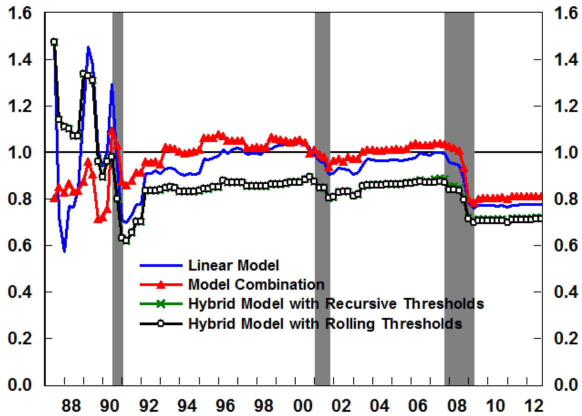

(b) During Great Recession



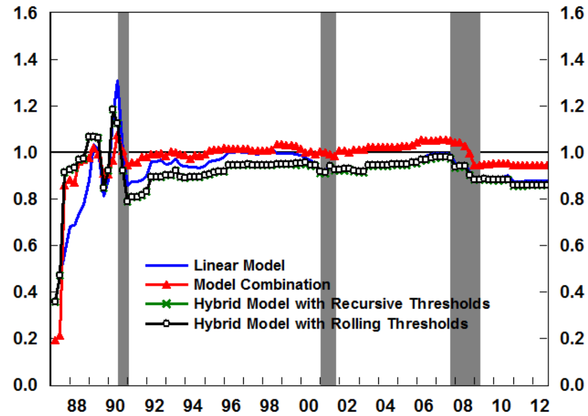

(c) Cumulative RMSFEs

* Notes: See notes to Figure 2.

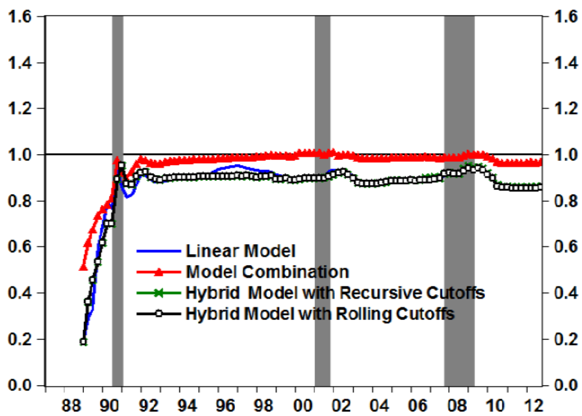
Figure 4: Cumulative RMSFEs of GDP Forecasts*



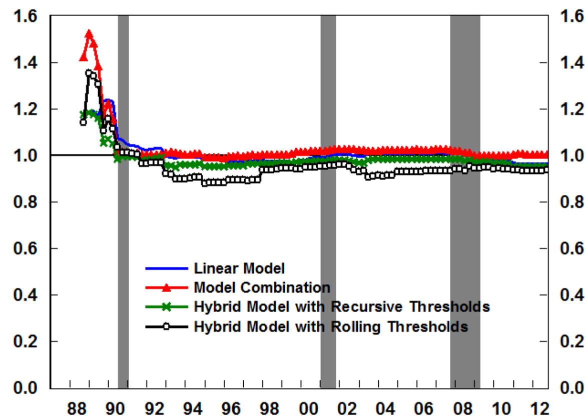

(a) 2-Quarter, Average




(b) 2-Quarter, Increments



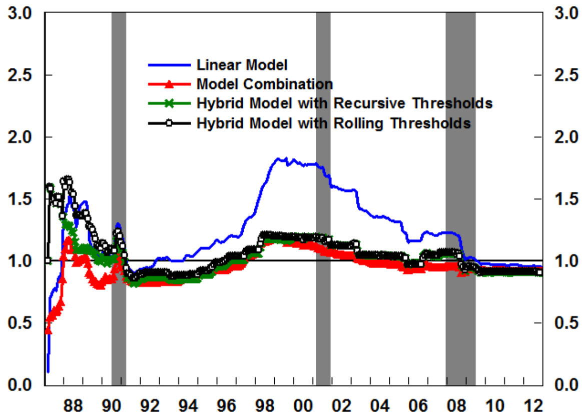

(c) 8-Quarter, Average



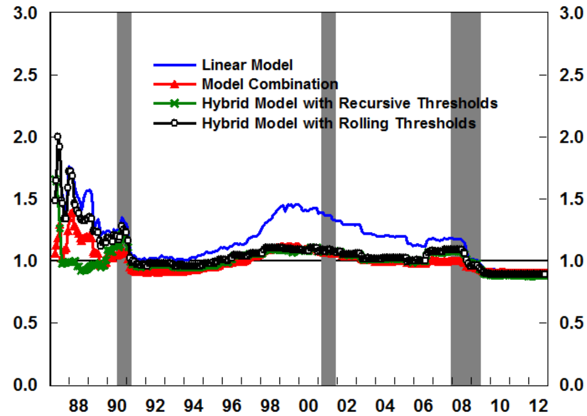

(d) 8-Quarter, Increments

* Notes: See notes to Figure 2.

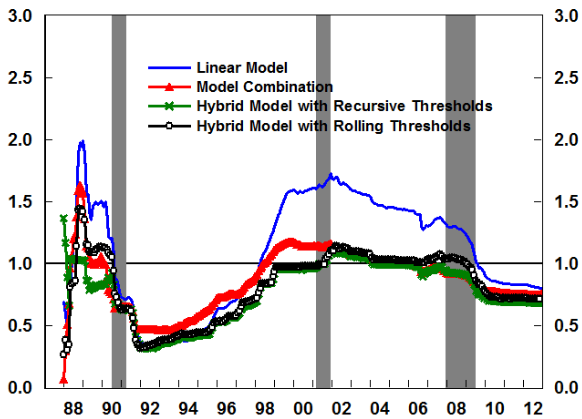
Figure 5: Cumulative RMSFEs of CPI Inflation Forecasts*



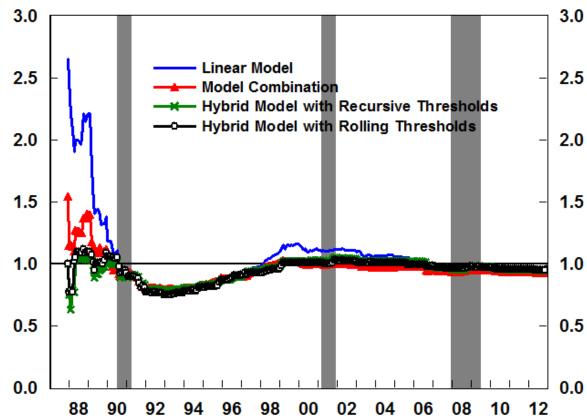

(a) 3-Month, Average




(b) 3-Month, Increments




(c) 12-Month, Average




(d) 12-Month, Increments

* Notes: See notes to Figure 2.

# Appendices

## A State Space Estimation of Mixed Frequency Indices: An Example

As in ADS (2009), assume that there are four indicators, including real GDP (a quarterly flow indicator), the number of employees on nonagricultural payrolls (a monthly stock indicator), initial claims for unemployment insurance (a weekly flow indicator), and the yield spread between the 10-year U.S. Treasury bond and the corresponding 3-month T-bill ( a daily stock variable). The measurement equation for the daily spread, which is missing for weekends and holidays, with an autoregressive term and no exogenous variables, is:

$$\widetilde{y}_t^{Spread} = \begin{cases} c_1 + \beta_1 mf_t + \gamma_1 \widetilde{y}_{t-1}^{i} + u_t^1 & , \text{ if } y_t^{Spread} \text{ is observed} \\ NA & , \text{ otherwise.} \end{cases}.$$

The measurement equations for weekly initial claims, which is flow variable and is missing for six days per week, with an autoregressive term, is:

$$\widetilde{y}_t^{IC} = \begin{cases} \sum_{i=1}^{7} c_2 + \alpha_2 \sum_{i=1}^{7} mf_{t-i+1} + \gamma_2 \widetilde{y}_{t-W}^{IC} + u_t^{*2} & , \text{ if } y_t^{IC} \text{ is observed} \\ NA & , \text{ otherwise.} \end{cases}.$$

The measurement equation for monthly nonagricultural payroll employees, which is stock variable and is observed on one day each month, with an autoregressive term, is:

$$\widetilde{y}_t^{Pay} = \begin{cases} c_3 + \alpha_3 mf_t + \gamma_3 \widetilde{y}_{t-M}^{Pay} + u_t^3 & , \text{ if } y_t^{Pay} \text{ is observed} \\ NA & , \text{ otherwise.} \end{cases}.$$

The measurement equations for quarterly real GDP, which is flow variable and observed at one day during a quarter is

$$\widetilde{y}_t^{GDP} = \begin{cases} \sum_{i=1}^{Q} c_4 + \alpha_4 \sum_{i=1}^{Q} mf_{t-i+1} + \gamma_4 \widetilde{y}_{t-Q}^{GDP} + u_t^{*4} & , \text{ if } y_t^{GDP} \text{ is observed} \\ NA & , \text{ otherwise.} \end{cases}.$$

The state equation is assumed to follow a zero mean AR(1) process. That is,

$$mf_t = \rho mf_{t-1} + e_t.$$

The above measurement equations and state equation can be summarized in vector form (assume normally distributed errors).

Measurement Equation

$$\begin{bmatrix} \widetilde{y}_t^1 \\ \widetilde{y}_t^2 \\ \widetilde{y}_t^3 \\ \widetilde{y}_t^4 \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ 0 & \alpha_2 & 0 & \alpha_4 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \alpha_4 \ or \ 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}' \begin{bmatrix} mf_t \\ \vdots \\ mf_{t-\bar{q}} \\ u_t^1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ \gamma_2 & 0 & 0 \\ 0 & \gamma_3 & 0 \\ 0 & 0 & \gamma_4 \end{bmatrix} \begin{bmatrix} \widetilde{y}_{t-W}^2 \\ \widetilde{y}_{t-M}^3 \\ \widetilde{y}_{t-Q}^4 \end{bmatrix} + \begin{bmatrix} 0 \\ u_{2t}^* \\ u_t^3 \\ u_{4t}^* \end{bmatrix}$$

$$\mathbf{y}_t = \mathbf{Z}_t \beta_t + \mathbf{\Gamma} \mathbf{w}_t + \varepsilon_t.$$

State Equation

$$\begin{bmatrix} mf_{t+1} \\ mf_t \\ mf_{t-1} \\ \vdots \\ mf_{t-\bar{q}+1} \\ u_{t+1}^1 \end{bmatrix} = \begin{bmatrix} \rho & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & \gamma_1 \end{bmatrix} \begin{bmatrix} mf_t \\ mf_{t-1} \\ mf_{t-2} \\ \vdots \\ mf_{t-\bar{q}} \\ u_t^1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e_t \\ \zeta_t \end{bmatrix},$$

$$\beta_{t+1} = \mathbf{T} \beta_t + \mathbf{R} \eta_t .$$

$$\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{0}_{4\times 1} \\ \mathbf{0}_{2\times 1} \end{bmatrix}, \begin{bmatrix} H_t & 0 \\ 0 & Q \end{bmatrix} \right),$$

and

$$H_t = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma_{2t}^{*2} & 0 & 0 \\ 0 & 0 & \sigma_{3t}^2 & 0 \\ 0 & 0 & 0 & \sigma_{4t}^{*2} \end{bmatrix}, \ Q = \begin{bmatrix} 1 & 0 \\ 0 & \sigma_1^2 \end{bmatrix},$$

where the $u_{jt}^*$ and $\sigma_{jt}^{*2}$ signify the measurement error and the variance thereof, respectively, in case of flow variable $j$. The sum of daily MF indexes for the week or the quarter is plugged into the measurement equation. In the case of stock variables such as monthly payroll and daily spread, one MF index on the day at which data is available is plugged into the measurement equation. The coefficient matrix, $Z_t$, in the measurement equation, is time-varying because the MF index is evolving on a daily basis, and the number of days in the quarter is time-varying. The large dimension of state vector, $\beta_t$, makes estimation difficult, especially in forecasting applications. ADS adopt the so-called Harvey cumulator, which is common in the mixed frequency literature. Define the $C_t^{D_i}$ cumulator variable for period $D_i$, e.g., quarter $(Q)$,month $(M)$,week $(W)$ as:

$$\begin{aligned} C_t^{D_i} &= I_t C_{t-1}^{D_i} + mf_t \\ &= I_t C_{t-1}^{D_i} + \rho_1 mf_{t-1} + \rho_2 mf_{t-2} + \cdots + \varepsilon_t, \end{aligned}$$

where $I_t$ is an indicator variable defined as

$$I_t^{D_i} = \begin{cases} 0, & \text{if } t \text{ is the first day of the period } D_i \\ 1, & \text{otherwise.} \end{cases} .$$

Then, for the quarterly cumulator $C_t^Q$,which is the sum of daily MF factors during the quarter (i.e., $\sum_{j=1}^{Q_t} mf_{t-j+1}$), the measurement equation for real GDP, for example, is:

$$\widetilde{y}_t^{GDP} = \begin{cases} \sum_{t=1}^{Q} c_4 + \alpha_4 C_t^Q + \gamma_4 \widetilde{y}_{t-Q}^{GDP} + u_t^{*4} & \text{, if } \mathrm{y}_t^{GDP} \text{ is observed} \\ NA & \text{, otherwise.} \end{cases} ,$$

Here, $Q_t$ is the number of days in a quarter and is time-varying. Now, the state space system can be represented using a 4x1 state vector, which has much smaller dimension, compared to the 93x1 dimension used in the previous state-space representation of the model. Namely:

$$\begin{array}{ccccc} \begin{bmatrix} mf_{t+1} \\ C_{t+1}^W \\ C_{t+1}^M \\ C_{t+1}^Q \end{bmatrix} & = & \begin{bmatrix} \rho & 0 & 0 & 0 \\ \rho & I_t^W & 0 & 0 \\ \rho & 0 & I_t^M & 0 \\ \rho & 0 & 0 & I_t^Q \end{bmatrix} \begin{bmatrix} mf_t \\ C_t^W \\ C_t^M \\ C_t^Q \end{bmatrix} & + & \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [e_t], \\ \beta_{t+1} & = & \mathbf{T}_t \beta_t & + & \mathbf{R}\eta_t, \end{array}$$

$$\begin{array}{l} \begin{bmatrix} \widetilde{y}_t^1 \\ \widetilde{y}_t^2 \\ \widetilde{y}_t^3 \\ \widetilde{y}_t^4 \end{bmatrix} = \begin{bmatrix} c_1 \\ \sum_{i=1}^{W} c_2 \\ c_3 \\ \sum_{i=1}^{Q} c_4 \end{bmatrix} + \begin{bmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 0 & 0 \\ \alpha_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_4 \end{bmatrix} \begin{bmatrix} mf_t \\ C_t^W \\ C_t^M \\ C_t^Q \end{bmatrix} \\ \qquad + \begin{bmatrix} \gamma_1 & 0 & 0 & 0 \\ 0 & \gamma_2 & 0 & 0 \\ 0 & 0 & \gamma_3 & 0 \\ 0 & 0 & 0 & \gamma_4 \end{bmatrix} \begin{bmatrix} \widetilde{y}_{t-D}^1 \\ \widetilde{y}_{t-W}^2 \\ \widetilde{y}_{t-M}^3 \\ \widetilde{y}_{t-Q}^4 \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t}^* \\ u_t^3 \\ u_{4t}^* \end{bmatrix}, \\ \mathbf{Y}_t = \mathbf{C} + \mathbf{Z}\beta_t + \mathbf{\Gamma}\mathbf{w}_t + \varepsilon_t \end{array}$$

This state space model can readily be estimated. For illustrative purposes, and following our above discussion, consider :

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{C} + \mathbf{Z}\beta_t + \mathbf{\Gamma}\mathbf{w}_t + \varepsilon_t \\ \beta_t &= \mathbf{T}_t \beta_{t-1} + \mathbf{R}\eta_t, \end{aligned}$$

where $\varepsilon_t \sim \mathbf{N}(0, Q)$ and $\eta_t \sim \mathbf{N}(0, H)$, $\mathbf{Y}_t$ is a vector of indicators, possibly having missing observations, and $\beta_t$ is the latent state vector. This model contains the mixed frequency factor $mf_t$, and the weekly, monthly, and quarterly cumulator. In the vector of exogenous variables $\mathbf{w}_t$, one autoregressive term of each indicator is included as in ADS. With missing data, and assuming normality, the Kalman filter can be used to estimate this system (see Kim and Nelson (KN: 1999) for details). Following KN, let $\mathbf{Y}_t \equiv [y_{t,}^1 y_{t,}^2 ..., y_t^N]$, $\mathbf{Y}_{t|t-1} = E[\mathbf{Y}_t|\mathbf{Y}_{t-1}]$, $\eta_{t|t-1} = \mathbf{Y}_t - \mathbf{Y}_{t|t-1}$, $\mathbf{F}_{t|t-1} = cov[\eta_{t|t-1}]$, $\beta_{t|t} = E(\beta_t|\mathbf{Y}_t)$, $\mathbf{P}_{t|t} = cov(\beta_t|\mathbf{Y}_t)$, $\beta_{t|t-1} \equiv E(\beta_t|\mathbf{Y}_{t-1})$, and $\mathbf{P}_{t|t-1} = cov(\beta_t|\mathbf{Y}_{t-1})$. The Kalman filter consists of following six equations: four prediction equations and two updating equations. For any $t$, with no missing observations,

$$\begin{aligned}
\beta_{t|t-1} &= \mathbf{T}_t\beta_{t-1|t-1}, && (18)\\
\mathbf{P}_{t|t-1} &= \mathbf{T}_t\mathbf{P}_{t-1|t-1}\mathbf{T}_t' + \mathbf{RHR}', && (19)\\
\eta_{t|t-1} &= \mathbf{Y}_t - \mathbf{Y}_{t|t-1} = \mathbf{Y}_t - \mathbf{C} - \mathbf{Z}\beta_{t|t-1} - \mathbf{\Gamma w}_t, && (20)\\
\mathbf{F}_{t|t-1} &= \mathbf{ZP}_{t|t-1}\mathbf{Z}' + \mathbf{Q}, && (21)\\
\beta_{t|t} &= \beta_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}'\mathbf{F}_{t|t-1}^{-1}\eta_{t|t-1}, && (22)\\
\mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{Z}'\mathbf{F}_{t|t-1}^{-1}\mathbf{ZP}_{t|t-1}. && (23)
\end{aligned}$$

Two prediction steps are associated with the state equation and the two more prediction steps are performed using the measurement equations. Given initial choices of state vector, $\beta_{0|0}$ and its covariance matrix, project the its future value of state vector and its covariance matrix using (18) and (19). In (20) and (21), the vector of prediction errors and associated covariance matrix are obtained after comparing the realized observations with predictions of them. Finally, update the state information (and associated covariance matrix) using (22) and (23).

If observations are missing, the measurement equation in vector form is modified as the observed number of observations changes. That is, measurement equations associated with missing data are removed from the measurement equation yielding:

$$\begin{aligned}
\mathbf{Y}_t^* &= \mathbf{C}^* + \mathbf{Z}^*\beta_t + \mathbf{\Gamma}_t^*\mathbf{w}_t + \mathbf{u}_t^*, && (24)\\
& \boldsymbol{u}_t^* \sim N\ (0, Q^*).
\end{aligned}$$

In actual applications, prediction steps are performed using the modified measurement equation (i.e., equation (24)). For any $t$, with missing observations, one thus utilizes the following system:

$$\begin{aligned}
\beta_{t|t-1} &= \mathbf{T}_t\beta_{t-1|t-1},\\
\mathbf{P}_{t|t-1} &= \mathbf{T}_t\mathbf{P}_{t-1|t-1}\mathbf{T}_t' + \mathbf{RHR}',\\
\eta_{t|t-1}^* &= \mathbf{Y}_t^* - \mathbf{Y}_{t|t-1}^* = \mathbf{Y}_t^* - \mathbf{C}^* - \mathbf{Z}^*\beta_{t|t-1} - \mathbf{\Gamma}^*\mathbf{w}_t, && (25)\\
\mathbf{F}_{t|t-1}^* &= \mathbf{Z}^*\mathbf{P}_{t|t-1}\mathbf{Z}^{*\prime} + \mathbf{Q}^*, && (26)\\
\beta_{t|t} &= \beta_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}^{*\prime}\mathbf{F}_{t|t-1}^{*-1}\eta_{t|t-1}^*,
\end{aligned}$$

$$\mathbf{P}_{t|t} \quad = \quad \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{Z}^{*\prime}\mathbf{F}^{*-1}_{t|t-1}\mathbf{Z}^{*}\mathbf{P}_{t|t-1}.$$

Finally, if all data are missing at period $t$, only prediction steps based on the state equation are required, yielding:

$$\begin{aligned}
\beta_{t|t-1} &= \mathbf{T}_t\beta_{t-1|t-1}, \\
\mathbf{P}_{t|t-1} &= \mathbf{T}_t\mathbf{P}_{t-1|t-1}\mathbf{T}_t' + \mathbf{RHR}'.
\end{aligned}$$

Assuming i.i.d. normal errors in the measurement and state equations, maximum likelihood estimation (MLE) can be applied using the prediction error decomposition to the linear state space model. Specifically, when the all $N$ variables are observed at time $t$, the log-likelihood is incrementally increased as follows:

$$\log L = -\frac{1}{2}[N\log 2\pi + (\log|\mathbf{F}_{t|t-1}| \ + \eta'_{t|t-1}\mathbf{F}^{-1}_{t|t-1}\eta_{t|t-1})].$$

For missing data, the incremental log-likelihood at time $t$ is

$$\log L = -\frac{1}{2}[N^*\log 2\pi + (\log|\mathbf{F}^*_{t|t-1}| \ + \eta^{*\prime}_{t|t-1}\mathbf{F}^{*-1}_{t|t-1}\eta^*_{t|t-1})],$$

where $N^* < N$ is the number of available observations at time $t$, and $\eta^*_{t|t-1}$ and $\mathbf{F}^*_{t|t-1}$ are defined above. Finally, if all indicators are missing, the incremental change of the likelihood is zero. In extracting an MF index, two steps are needed. The first step involves estimating the parameters in the models. Using MLE, given initial parameter, state vector, and covariance matrix choices, find the estimates of the parameters which is maximizing the log-likelihood. In a second step, given parameter estimates, one can extract an MF index in the state vector by running the Kalman filter. For the initial choices of state vector, $\beta_{0|0}$, and of its covariance matrix, $\mathbf{P}_{0|0}$, under stationarity, the unconditional mean of state vector, $E(\beta_t)$, and its covariance matrix, $E(\mathbf{P}_t)$ are used. Following Durbin and Koopman (2001), and in addition to the above "non-smoothed" approach, we also consider use of a fixed interval smoothing algorithm for this step. If all data are observed at period $t$, and with $r_T = 0$,

$$\begin{aligned}
\beta_{t|T} &= \beta_{t|t} \ + \ \mathbf{P}_t\mathbf{r}_{t-1}, \\
\mathbf{P}_{t|T} &= \mathbf{P}_{t|t} \ - \ \mathbf{P}_{t|t}\mathbf{N}_{t-1}\mathbf{P}_{t|t}, \\
\mathbf{r}_{t-1} &= \mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{v}_t \ + \ \mathbf{L}_t'\mathbf{r}_t, \\
\mathbf{N}_{t-1} &= \mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{Z}_t \ + \ \mathbf{L}_t'\mathbf{N}_t
\end{aligned}$$

where

$$
\begin{aligned}
\mathbf{L}_t &= \mathbf{T}_t - \mathbf{K}_t\mathbf{Z}_t, \\
\mathbf{v}_t &= \mathbf{Y}_t - E(\mathbf{Y}_t|\mathbf{Y}_{t-1}).
\end{aligned}
$$

If some observations are missing at period $t$, set:

$$
\begin{aligned}
\beta_{t|T} &= \beta_{t|t} + \mathbf{P}_t\mathbf{r}^*_{t-1}, \\
\mathbf{P}_{t|T} &= \mathbf{P}_{t|t} - \mathbf{P}_{t|t}\mathbf{N}^*_{t-1}\mathbf{P}_{t|t}, \\
\mathbf{r}^*_{t-1} &= \mathbf{Z}^{*\prime}_t\mathbf{F}^{*-1}_t\mathbf{v}^*_t + \mathbf{L}^{*\prime}_t\mathbf{r}^*_t, \\
\mathbf{N}^*_{t-1} &= \mathbf{Z}^{*\prime}_t\mathbf{F}^{*-1}_t\mathbf{Z}^*_t + \mathbf{L}^{*\prime}_t\mathbf{N}^*_t\text{Ł}^*_t,
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{L}^*_t &= \mathbf{T}^*_t - \mathbf{K}^*_t\mathbf{Z}^*_t, \\
\mathbf{v}^*_t &= \mathbf{Y}^*_t - E(\mathbf{Y}^*_t|\mathbf{Y}_{t-1}).
\end{aligned}
$$

If all data are missing at period $t$,

$$
\begin{aligned}
\beta_{t|T} &= \beta_{t|t} + \mathbf{P}_t\mathbf{r}_{t-1}, \\
\mathbf{P}_{t|T} &= \mathbf{P}_{t|t} - \mathbf{P}_{t|t}\mathbf{N}_{t-1}\mathbf{P}_{t|t}, \\
\mathbf{r}_{t-1} &= \mathbf{T}'_t\mathbf{r}_t, \\
\mathbf{N}_{t-1} &= \mathbf{T}'_t\mathbf{N}_t\mathbf{T}_t.
\end{aligned}
$$

# B Additional Empirical Results Based on “Cherry Picking ”a Fixed Threshold for Use in Hybrid Models

**Table B1: Thresholds ($\tau^{Post}$) of Real GDP Growth For Each Target Variable***

**(a) $\tau^{Post}$ in GDP Growth Hybrid Models**

| h | 1 | 2, A | 4, A | 8, A | 2, I | 4, I | 8, I |
|---|---|---|---|---|---|---|---|
| Cutoffs | 0.758 | 0.684 | 0.560 | 0.838 | 0.692 | 0.671 | 0.478 |

**(b) $\tau^{Post}$ in CPI Inflation Hybrid Models**

| h | 1 | 3, A | 6, A | 12, A | 3, I | 6, I | 12, I |
|---|---|---|---|---|---|---|---|
| Cutoffs | -1.399 | 0.560 | 0.956 | 0.517 | 0.245 | 0.025 | 1.206 |

* Notes: This table presents fixed “cherry picked” GDP growth thresholds. These thresholds are those leading to the “best” MSFEs, ex post, and are separately calculated for each forecast horizon, denoted by 1,2,4, and 8 (quarters ahead) for GDP prediction, and 1,3,6, and 12 (months ahead) for CPI prediction. All thresholoding is done via evaluation of GDP growth, as outlined in panel (a) of Table 2. Finally, “A” and “I” refer to whether “average” of “incremental” versions of GDP growth and inflation are predicted, as outlined in Section 4.1. See Section 2.3 for further details on the thresholding mechansim used.

**Table B2: Summary of Top-5 "MSFE-Best" models by Ex-Post Experiments***

**(a) Real GDP Growth**

| *Ranking* | *Linear Model* | *RMSFE* | *Average Model* | *RMSFE* | *Hybrid (Cherry Picked Threshold)* | *RMSFE* |
|---|---|---|---|---|---|---|
| | | | | | **$GDP^{A}_{t+1}$** | |
| 1 | *MF_B_NSL* | 0.778 | *CMA5* | 0.801 | *MFDIAR_C_SF* | **0.659**$^{*A}$ |
| 2 | *MFAR_O_NSL* | 0.779* | *CMA3* | 0.809 | *MFDI_C_SF* | **0.663**$^{*A}$ |
| 3 | *MFAR_B_NSL* | 0.780 | *CMA10* | 0.817 | *MF_C_SF* | **0.694**$^{*A}$ |
| 4 | *MF_O_NSL* | 0.782* | *CMA4* | 0.817 | *MFDIAR_M4_SL* | **0.694**$^{*A}$ |
| 5 | *MFAR_MQ2_NSL* | 0.783* | *CMA2* | 0.819 | *MFAR_S_NSL* | **0.695**$^{*A}$ |
| | | | | | **$GDP^{A}_{t+2}$** | |
| 1 | *MFDI_MQ1_SL* | 0.775 | *CMA5* | 0.810 | *MFDI_MQ1_SL* | **0.705**$^{**A}$ |
| 2 | *MFDI_MQ2_SL* | 0.780 | *CMA9* | 0.821 | *MF_MQ2_SL* | **0.724**$^{*A}$ |
| 3 | *MF_A_NSL* | 0.791 | *CMA3* | 0.829 | *MFDI_MQ2_SL* | **0.731**$^{*A}$ |
| 4 | *MFDIAR_MQ2_SL* | 0.791 | *CMA7* | 0.830 | *MFDI_M1_SL* | **0.735**$^{*A}$ |
| 5 | *MFDIAR_MQ1_SL* | 0.793 | *CMA4* | 0.834 | *MFDIAR_MQ4_SL* | **0.737**$^{*A}$ |
| | | | | | **$GDP^{A}_{t+4}$** | |
| 1 | *MFDIAR_Q_SL* | 0.839 | *CMA9* | 0.888 | *MFDIAR_Q_SL* | **0.801**$^{A}$ |
| 2 | *MFAR_Q_SL* | 0.841 | *CMA5* | 0.891* | *MFDIAR_T_SL* | **0.807**$^{A}$ |
| 3 | *MFDI_Q_SL* | 0.848 | *CMA7* | 0.900 | *MFAR_Q_SL* | **0.809**$^{A}$ |
| 4 | *MF_Q_SL* | 0.851 | *CMA3* | 0.913 | *MFDI_Q_SL* | **0.813**$^{A}$ |
| 5 | *MFDI_O_SL* | 0.858 | *CMA10* | 0.917 | *MF_Q_SL* | **0.819**$^{A}$ |
| | | | | | **$GDP^{A}_{t+8}$** | |
| 1 | *MF_Q_SL* | 0.866** | *CMA5* | 0.966 | *MF_Q_SL* | **0.859**$^{**A}$ |
| 2 | *MFAR_Q_SL* | 0.883** | *CMA3* | 0.967 | *MF_T_SL* | 0.873$^{**C}$ |
| 3 | *MF_T_SL* | 0.884** | *CMA9* | 1.005 | *MFDIAR_T_SL* | 0.877$^{**C}$ |
| 4 | *MFDI_T_SL* | 0.890** | *CMA7* | 1.007 | *MFDI_T_SL* | 0.879$^{**C}$ |
| 5 | *MFDIAR_T_SL* | 0.890** | *CMA10* | 1.016 | *MFAR_Q_SL* | 0.882$^{**C}$ |
| | | | | | **$GDP^{I}_{t+2}$** | |
| 1 | *MFDIAR_MQ1_SL* | 0.877* | *CMA5* | 0.944 | *MFDI_MQ1_SL* | **0.857**$^{**A}$ |
| 2 | *MFDI_MQ1_SL* | 0.879* | *CMA9* | 0.945 | *MFDIAR_MQ1_SL* | **0.860**$^{*A}$ |
| 3 | *MFDI_O_SL* | 0.908 | *CMA4* | 0.966 | *MFAR_MQ4_SL* | 0.881$^{*C}$ |
| 4 | *MFDIAR_O_SL* | 0.910 | *CMA7* | 0.968 | *MFDIAR_MQ4_SL* | 0.885$^{*C}$ |
| 5 | *MFDI_MQ2_SL* | 0.911 | *CMA10* | 0.971 | *MFDIAR_M1_SL* | 0.886$^{*C}$ |
| | | | | | **$GDP^{I}_{t+4}$** | |
| 1 | *MFAR_INF_SL* | 0.989 | *CMA1* | 0.998 | *MFDIAR_T_SL* | **0.985**$^{*A}$ |
| 2 | *MFDIAR_T_SL* | 0.993 | *CMA5* | 1.012 | *MFDIAR_Q_SL* | 0.990$^{C}$ |
| 3 | *MFAR_Q1_SL* | 0.993* | *CMA9* | 1.026 | *MFDIAR_WM3_SL* | 0.991$^{C}$ |
| 4 | *MFAR_C_SF* | 0.994 | *CMA3* | 1.029 | *MFAR_D_SF* | 0.992$^{C}$ |
| 5 | *MFAR_T_SL* | 0.994 | *CMA10* | 1.031 | *MFDIAR_WM4_SL* | 0.994$^{C}$ |
| | | | | | **$GDP^{I}_{t+8}$** | |
| 1 | *MF_F_SF* | 0.964* | *CMA3* | 1.003 | *MF_J_SL* | **0.922**$^{***A}$ |
| 2 | *MF_D_SF* | 0.965* | *CMA1* | 1.011 | *MF_F_SF* | **0.957**$^{**A}$ |
| 3 | *MF_D1_NSL* | 0.967* | *CMA2* | 1.016 | *MF_J_NSL* | **0.957**$^{***A}$ |
| 4 | *MF_D1_SL* | 0.968* | *CMA5* | 1.017 | *MF_INF_SL* | **0.958**$^{**A}$ |
| 5 | *MF_E_SF* | 0.968* | *CMA10* | 1.024 | *MF_D_SF* | **0.958**$^{**A}$ |

## Table B2: (Cont.)
## (b): CPI Inflation

| *Ranking* | *Linear Model* | *RMSFE* | *Average Model* | *RMSFE* | *Hybrid (Cherry Picked Threshold)* | *RMSFE* |
|---|---|---|---|---|---|---|
| | | | | | $\mathbf{CPI}^{A}_{t+1}$ | |
| 1 | *MFDIAR_S_NSL* | 0.972 | *CMA*10 | 0.974 | *MF_WM1_SL* | **0.941**$^{*A}$ |
| 2 | *MFAR_MQ5_SL* | 0.972 | *CMA*1 | 0.988 | *MF_K_SL* | **0.942**$^{*A}$ |
| 3 | *MFAR_I_SL* | 0.975 | *CMA*5 | 0.990 | *MF_K_NSL* | **0.942**$^{*A}$ |
| 4 | *MFAR_D_SL* | 0.976 | *CMA*4 | 0.991 | *MF_D_NSF* | **0.942**$^{*A}$ |
| 5 | *MFAR_L_SL* | 0.976 | *CMA*8 | 0.993 | *MF_N_SL* | **0.942**$^{*A}$ |
| | | | | | $\mathbf{CPI}^{A}_{t+3}$ | |
| 1 | *MFDI_WM3_SL* | 0.948 | *CMA*10 | 0.913** | *MFDI_WM3_SL* | **0.886**$^{*A}$ |
| 2 | *MFDI_WM4_SL* | 0.959 | *CMA*6 | 0.964 | *MFDI_E_SL* | **0.890**$^{*A}$ |
| 3 | *MFDI_MQ1_SL* | 0.960 | *CMA*7 | 0.972 | *MFDI_E_NSL* | **0.890**$^{*A}$ |
| 4 | *MFDI_MQ1_NSL* | 0.962 | *CMA*1 | 0.977 | *MFDI_N_SL* | **0.893**$^{*A}$ |
| 5 | *MFDI_W1_NSL* | 0.963 | *CMA*5 | 1.022 | *MFDI_N_NSL* | **0.893**$^{*A}$ |
| | | | | | $\mathbf{CPI}^{A}_{t+6}$ | |
| 1 | *MFDI_W1_NSL* | 0.854 | *CMA*10 | 0.831*** | *MFDI_MQ5_SL* | **0.726**$^{***A}$ |
| 2 | *MFDI_WM3_SL* | 0.856 | *CMA*7 | 0.869 | *MFDI_WM3_SL* | **0.735**$^{***A}$ |
| 3 | *MFDI_MQ5_SL* | 0.856 | *CMA*6 | 0.869 | *MFDI_W1_SL* | **0.739**$^{***A}$ |
| 4 | *MFDI_MQ1_SL* | 0.859 | *CMA*1 | 0.909* | *MFDI_W1_NSL* | **0.741**$^{***A}$ |
| 5 | *MFDI_W1_SL* | 0.863 | *CMA*4 | 1.061 | *MFDI_N_SL* | **0.742**$^{**A}$ |
| | | | | | $\mathbf{CPI}^{A}_{t+12}$ | |
| 1 | *MFDI_WM3_SL* | 0.801** | *CMA*10 | 0.750*** | *MFDI_N_SL* | **0.654**$^{***A}$ |
| 2 | *MFDI_MQ5_SL* | 0.802** | *CMA*7 | 0.829* | *MFDI_E_SL* | **0.661**$^{***A}$ |
| 3 | *MFDI_WM4_SL* | 0.810* | *CMA*6 | 0.836* | *MFDI_MQ5_SL* | **0.665**$^{***A}$ |
| 4 | *MFDI_D1_SL* | 0.811* | *CMA*1 | 0.867* | *MFDI_D_SL* | **0.666**$^{***A}$ |
| 5 | *MFDI_D1_NSL* | 0.811* | *CMA*4 | 0.980 | *MFDI_K_SL* | **0.667**$^{***A}$ |
| | | | | | $\mathbf{CPI}^{I}_{t+3}$ | |
| 1 | *MFDI_WM3_SL* | 0.915 | *CMA*10 | 0.901** | *MFDI_F_NSF* | **0.871**$^{**A}$ |
| 2 | *MFDI_WM4_SL* | 0.915 | *CMA*7 | 0.925 | *MFDI_D_SF* | **0.874**$^{***A}$ |
| 3 | *MFDI_F_NSF* | 0.916 | *CMA*6 | 0.926 | *MFDI_D_NSF* | **0.875**$^{**A}$ |
| 4 | *MFDI_D1_NSL* | 0.917 | *CMA*1 | 0.928** | *MFDI_T_NSL* | **0.881**$^{**A}$ |
| 5 | *MFDI_I_SL* | 0.919 | *CMA*4 | 0.997 | *MFDI_E_NSF* | **0.881**$^{**A}$ |
| | | | | | $\mathbf{CPI}^{I}_{t+6}$ | |
| 1 | *MFDI_W1_SL* | 0.970 | *CMA*10 | 0.946** | *MFDI_K_SL* | **0.844**$^{***A}$ |
| 2 | *MFDI_MQ5_SL* | 0.970 | *CMA*1 | 0.968 | *MFDI_F_SF* | **0.846**$^{***A}$ |
| 3 | *MFDI_D1_SL* | 0.973 | *CMA*7 | 0.976 | *MFDI_K_NSL* | **0.847**$^{***A}$ |
| 4 | *MFDI_MQ1_SL* | 0.974 | *CMA*6 | 0.979 | *MFDI_N_SL* | **0.848**$^{***A}$ |
| 5 | *MFDI_N_NSL* | 0.974 | *CMA*4 | 1.032 | *MFDI_N_NSL* | **0.849**$^{***A}$ |
| | | | | | $\mathbf{CPI}^{I}_{t+12}$ | |
| 1 | *MFDI_K_SL* | 0.944* | *CMA*10 | 0.928*** | *MFDI_K_SL* | **0.920**$^{**A}$ |
| 2 | *MFDI_F_SF* | 0.945* | *CMA*7 | 0.951* | *MFDI_F_SF* | **0.921**$^{**A}$ |
| 3 | *MFDI_S_SL* | 0.947* | *CMA*6 | 0.956* | *MFDI_E_SF* | **0.925**$^{**A}$ |
| 4 | *MFDI_E_SF* | 0.948* | *CMA*1 | 0.975 | *MFDI_M1_SL* | **0.930**$^{**B}$ |
| 5 | *MFDI_M1_SL* | 0.948* | *CMA*8 | 0.985 | *MFDI_M1_NSL* | **0.931**$^{**B}$ |

(*) Notes: See notes to Table 4 and Table A1. In these results, the group denoted as "Hybrid Model with Cherry Picked Threshold" corresponds to models chosen based on use of the cherry-picked thresholds tabulated in Table A1.

# References


Armah, N. A. and Swanson, N. R. (2010). Diffusion index models and index proxies: Recent results and new direction. *European Journal of Pure and Applied Mathematics*, 3:478–501.

Aruoba, S. and Diebold, F. (2010). Real-time macroeconomic monitoring: Real activity, inflation, and interactions. *American Economic Review*, 100:20–24.

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

Bai, J. (2003). Testing parametric conditional distributions of dynamic models. *The Review of Economics and Statistics*, 85(3):531–549.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2006a). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Ng, S. (2006b). Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics*, 131(1-2):507–537.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics*, 25:52–60.

Bai, J. and Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, 24(4):607–629.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176:18–29.

Balke, N. S., Fulmer, M., and Zhang, R. (2017). Incorporating the beige book into a quantitative index of economic activity. *Journal of Forecasting*, 36(5):497–514.

Boivin, J. and Ng, S. (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking*, 1(3):117–152.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Camacho, M., Pérez-Quirós, G., and Poncela, P. (2014). Green shoots and double dips in the euro area: A real time measure. *International Journal of Forecasting*, 30:520–535.

Camacho, M., Pérez-Quirós, G., and Poncela, P. (2018). Markov-switching dynamic factor models in real time. *International Journal of Forecasting*, 34(4):598–611.

Carrasco, M. and Rossi, B. (2016). In-sample inference and forecasting in misspecified factor models. *Journal of Business and Economic Statistics*, 34:313–338.

Chudik, A., Kapetanios, G., and Pesaran, M. (2016). Big data analytics: A new perspective. Working paper no. 268, Federal Reserve Bank of Dallas Working Paper No. 268.

Clark, T. and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, 105:85–110.

Clark, T. and McCracken, M. W. (2005). Evaluating direct multi-step forecasts. *Econometric Reviews*, 24:369–404.

Clements, M. P. and Galvao, A. B. (2008). Macroeconomic forecasting with mixed frequency data. *Journal of Business & Economic Statistics*, 26:546–554.

Corradi, V. and Swanson, N. R. (2014). Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics*, 182:100–118.

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

Ding, A. A. and Hwang, J. T. G. (1999). Prediction intervals, factor analysis models, and high-dimensional empirical linear prediction. *Journal of the American Statistical Association*, 94(446):446–455.

Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, 100:830–840.

Forni, M. and Reichlin, L. (1998). Let's get real: A factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies*, 65(3):453–73.

Ghysels, E. and Marcelino, M. (2016). The econometric analysis of mixed frequency data sampling. *Journal of Econometrics*, 193:291–293.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006). Predicting volatility: Getting the most out of return data sampled at different frequencie. *Journal of Econometrics*, 131:59–95.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). Midas regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.

Hallin, M. and Liska, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102:603–617.

Kim, H. H. and Swanson, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, 178(2):352–367.

Kim, H. H. and Swanson, N. R. (2018a). Methods for backcasting, nowcasting and forecasting using factor-midas: With an application to korean gdp. *Journal of Forecasting*, 37(3):281–302.

Kim, H. H. and Swanson, N. R. (2018b). Mining big data using parsimonious factor and shrinkage methods. *International Journal of Forecasting*, 34(2):339–354.

Marcellino, M., Porqueddu, M., and Venditti, F. (2015). Short-term gdp forecasting with a mixed frequency dynamic factor model with stochastic volatility. *Journal of Business and Economic Statistics*, 34(1):118–127.

Mariano, R. and Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18:427–443.

McCracken, M. W. (2000). Robust out-of-sample inference. *Journal of Econometrics*, 99:195–223.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77:1447–1479.

Rossi, B. and Sekhposyan, T. (2010). Have economic models' forecasting performance for us output growth and inflation changed over time, and when? *International Journal of Forecasting*, 26:808–835.

Sargent, T. and Sims, C. (1977). Business cycle modeling without pretending to have too much a priori economic theory. Working Papers 55, Federal Reserve Bank of Minneapolis.

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology*, 15:201–293.

Stock, J. and Watson, M. (1988). A probability model of the coincident economic indicators. In *Leading Economic Indicators: New Approaches and Forecasting Records*, number 2772. Cambridge University Press.

Stock, J. and Watson, M. (1989). New indexes of coincident and leading economic indicators. In *NBER Macroeconomics Annual 1989, Volume 4*, pages 351–409. National Bureau of Economic Research, Inc.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44(2):293–335.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–62.

Stock, J. H. and Watson, M. W. (2006). Forecasting with many predictors. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, chapter 10, pages 515–554. Elsevier.

Stock, J. H. and Watson, M. W. (2008). Forecasting in dynamic factor models subject to structural instability. In Castle, J. and Shephard, N., editors, *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*. Oxford University Press.

Swanson, N. (2016). Comment on: In sample inference and forecasting in misspecified factor models. *Journal of Business and Economic Statistics*, 34:348–353.

Timmermann, A. G. (2006). Forecast combinations. In Elliott, G., C., G., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, chapter 4, pages 135–196. Elsevier.