Is Broader Better?
A Monetary Approach to Forecasting Economic Activity

Michael Ellington[a] & Marcin Michalski[a]

4th February 2021

Abstract

This paper investigates whether the use of broader Divisia monetary aggregates improves money's performance in forecasting economic activity within a time-varying parameter vector autoregressive (TVP-VAR) framework. We evaluate entire predictive densities from several alternative models of US output growth and inflation, each using eight different Divisia monetary aggregates. Using the broadest, M4 aggregate produces out-of-sample forecasts which consistently outperform those based on narrower measures of money, pooling of forecasts from several models, and a large-scale, 143-variable model. Our results show that TVP-VARs with Divisia M4 forecast economic activity more accurately than constant-parameter models with alternative or no measures of money.

JEL: E32, E47, E51, E52, E58.
Keywords: Density Forecasting, Quasi Bayesian Local Likelihood Methods, Vector Autoregression, Divisia money.

Acknowledgements
We thank Costas Milas and Pierre-Daniel Sarte for helpful comments. We received no financial support in conducting this research.

[a] University of Liverpool Management School, Chatham Street, Liverpool L69 7ZH, UK. M.Ellington@liverpool.ac.uk M.Michalski@liverpool.ac.uk.

1 Introduction

Accurate and reliable forecasts of future economic conditions are a cornerstone of formation and effective policy implementation by monetary and fiscal authorities, business decisions of commercial firms, and actions of individual economic agents. This puts the task of generating such predictions about the future state of the economy at the very heart of economics. Yet, several questions concerning which variables and classes of models ought to be used to accomplish that task remain.

In this paper, we investigate the empirical benefits of using the broadest, Divisia M4 monetary aggregate to forecast GDP growth and inflation within a time-varying parameter vector autoregressive framework. Monetary aggregates are often neglected in models used to predict key economic indicators such as output growth and inflation as a result of the emergence and dominance of Taylor Rules (Taylor, 1993) that guide monetary policy, and the allegations of a weak empirical link between money and economic activity (e.g. Friedman and Kuttner (1992), Bernanke and Blinder (1992)). As interest rates approach their zero-lower bound (ZLB), however, Taylor rules cease to be operational. Furthermore, the perceived lack of an empirical relationship between money and real activity is arguably a result of a measurement error (Barnett, 2012).

The construction of simple-sum measures of money by adding the component assets implies that they are perfect one-for-one substitutes. In reality, those component assets provide different services and therefore yield different returns. As these returns evolve over time they imply the existence of opportunity costs, in terms of foregone interest, which are also time-varying. The Divisia monetary aggregates derived by Barnett (1980) assign weights to component assets in accordance to their usefulness in making transactions: the more liquid the component, the higher the weight.[1]

Our study contributes to the literature on the use of money aggregates in out-of-sample forecasts of economic activity. The reduced-form evidence, in the form of correlation-type analysis, confirms a strong link between Divisia money and economic activity relative to atheoretical, simple-sum aggregates (Belongia and Ireland, 2016; Ellington, 2018). Furthermore, Barnett and Chauvet (2011) advocate the use of the broadest Divisia monetary aggregate, Divisia M4, since it can act as an early warning indicator for imminent financial crises and recession. In terms of forecasting, the existing evidence on using Divisia money to predict economic activity out-of-sample (e.g. Schunk (2001), Albuquerque et al. (2015)) is inconclusive. Our results, however, provide robust evidence that models featuring Divisia M4 aggregates produce more accurate forecasts of output growth and inflation

[1] The ensuing literature, such as Belongia (1996), Belongia and Chrystal (1991) and Hendrickson (2014), overturn the findings of Bernanke and Blinder (1992) and others when monetary aggregates have weights with a theoretical foundation.

than models without any measure of money.

We carry out a comprehensive out-of-sample analysis which documents significant improvements in forecasting performance of models using Divisia money stemming from the use of the broadest aggregate, Divisia M4. Our findings show that predictions based on Divisia M4 money exhibit superior accuracy relative to those based on narrower measures of money and on pooling forecasts from our alternative models. The conclusions about the greater utility of Divisia M4 reflect findings of Keating et al. (2014), who use that measure of money to identify monetary policy shocks that overcome the price puzzle of inflation rising as a result of a monetary contraction. They also relate to Keating et al. (2019), who deduce a model using Divisia M4 that yields responses that are free from price, output and liquidity puzzles. Furthermore, The Center for Financial Stability (CFS) echoes this recommendation by stating that Divisia M4 is the most appropriate monetary aggregate for almost all applications.

The majority of prior literature on forecasting economic activity using Divisia aggregates use constant-parameter models (see e.g. Belongia and Ireland, 2015, 2016). Our second contribution is in departing from this line of enquiry as we employ time-varying parameter models instead. Breaks in volatility are an important feature of macroeconomic data, induced for instance by the Great Inflation, the Great Moderation, and the Great Recession. Several studies account for such breaks by implementing time-varying parameter vector autoregressive models (TVP-VAR) of Primiceri (2005) for the analysis of monetary policy and other structural shocks (see Canova et al., 2007; Benati, 2008; Bianchi et al., 2009; Ellington, 2018). Such models are difficult to estimate, computationally costly, and usually constrained to no more than 6 variables. As a result, many studies focus on pseudo-forecasting exercises that stem from stochastic simulations of the model at each point in time (e.g. Benati and Mumtaz, 2007).

We use TVP-VARs that allow time-variation to evolve in a non-parametric manner (Petrova, 2019). This is different from conventional methods such as Primiceri (2005) that require a law of motion. If the true law of motion is misspecified then this can distort inference, and in turn affect forecasts. The methods we employ, however, allow us to estimate large-scale models that traditional TVP-VARs cannot handle. It also permits us to use parallel computing such that we achieve computational efficiency relative to traditional Markov Chain Monte Carlo (MCMC) methods (see e.g. Cogley and Sargent, 2005).

This study is the first of its kind to quantify the forecasting performance of theoretically consistent monetary aggregates using recursive TVP-VAR models. We estimate recursive forecasting TVP-VAR models of a range of different sizes over a forecast sample of 28 years. The largest of the model we estimate contains 147 variables, which we are able to do without having to rely on sparsity or shrinkage within an MCMC sampler typical for large-scale models (see e.g. Koop and Korobilis, 2013; Huber et al., 2020).

Finally, our results also provide strong out-of-sample evidence that TVP-VARs generate more accurate predictions of economic activity relative to constant parameter alternatives. Recent studies, such as Carriero et al. (2009, 2012); Cross et al. (2020), use large Bayesian VAR models with various shrinkage priors to successfully forecast exchange rates, government bond yields, and other macroeconomic indicators. We use such Bayesian VAR models with Divisia M4 to forecast economic conditions, compare the predictions with those obtained from our baseline TVP-VAR model, and find overwhelming evidence in support of time-varying, rather than constant, parameters.

2 Data and Methodology

2.1 Data

Our sample comprises data on US GDP, Consumer Price Index (CPI), wages, unemployment, oil prices, and corporate bond yields between 1967:Q2 and 2018:Q4, which we collect from the Federal Reserve Economic Database. We supplement this with information on the S&P500 composite price index and the house price index obtained from WRDS and Robert Shiller's website, respectively. Further, during periods when the official policy rate is constrained by the zero lower bound, we splice the federal funds rate from FRED with the shadow rate of Wu and Xia (2016) available from Jiang Cynthia Wu's website.

We use Divisia money aggregates obtained from the Center for Financial Stability as our measures of money. Specifically, we use eight measures ranging from the broadest, Divisia M4, which serves as our benchmark measure, to the narrowest, Divisa M1. The measures are observed at a monthly frequency, which we aggregate using 3-month averages to match the quarterly frequency of other variables in the sample. The definitions of component assets included in each measure of money are in Barnett et al. (2013). All variables enter each model as quarterly growth rates, with the exception of the spliced federal funds/shadow rate and the unemployment rate which enter in levels. All data series, their sources, and variable transformations are listed in Appendix A.

2.2 Econometric Methods: A Time-varying parameter VAR with nonparametric drift

Let y_t be an $N \times 1$ vector generated by a stable time-varying parameter (TVP) heteroskedastic VAR model with L lags:

$$y_t = \mathbf{B}_{0,t} + \sum_{p=1}^{L} \mathbf{B}_{p,t} y_{t-p} + \varepsilon_t, \;\; \varepsilon_t = \boldsymbol{\Xi}_t^{-\frac{1}{2}} \kappa_t, \;\; \kappa_t \backsim \text{NID}(0, \mathbf{I}_N) \tag{1}$$

where $\mathbf{B}_{0,t}, \mathbf{B}_{p,t}$ contain the time-varying intercepts and autoregressive matrices, respectively. Note that all roots of the polynomial, $\psi(z) = \det\left(\mathbf{I}_N - \sum_{p=1}^{L} z^p \mathbf{B}_{p,t}\right)$, lie outside the unit circle, and $\boldsymbol{\Xi}_t^{-1}$ is a positive definite time-varying covariance matrix. Stacking the time-varying intercepts and autoregressive matrices in the vector θ_t with $\mathbf{X}_t' = (\mathbf{I}_N \otimes x_t)$, where $x_t = \left(1, y_{t-1}', \ldots, y_{t-L}'\right)$ and $\otimes$ is the Kronecker product, the model can be written as:

$$y_t = \mathbf{X}_t' \theta_t + \boldsymbol{\Xi}_t^{-\frac{1}{2}} \kappa_t \tag{2}$$

The time-varying parameters of the model are estimated by employing Quasi-Bayesian Local Likelihood (QBLL) methods (Petrova, 2019).

Estimation of the model in Equation (2) requires re-weighting the likelihood function. Essentially, the weighting function gives higher proportions to observations surrounding the time period whose parameter values are of interest. The local likelihood function at time period k is given by:

$$\mathrm{L}_k \left(y | \theta_k, \boldsymbol{\Xi}_k, \mathbf{X}\right) \;\; \propto \;\; |\boldsymbol{\Xi}_k|^{\text{trace}(\mathbf{D}_k)/2} \exp\{-\frac{1}{2}(y - \mathbf{X}'\theta_k)' \left(\boldsymbol{\Xi}_k \otimes \mathbf{D}_k\right) \left(y - \mathbf{X}'\theta_k\right)\} \tag{3}$$

The $\mathbf{D}_k$ is a diagonal matrix whose elements hold the weights:

$$\mathbf{D}_k = \text{diag}(\vartheta_{k1}, \ldots, \vartheta_{kT}) \tag{4}$$

$$\vartheta_{kt} = \phi_{T,k} w_{kt} / \sum_{t=1}^{T} w_{kt} \tag{5}$$

$$w_{kt} = (1/\sqrt{2\pi}) \exp((-1/2)((k-t)/H)^2), \text{ for } k, t \in \{1, \ldots, T\} \tag{6}$$

$$\phi_{Tk} = \left(\left(\sum_{t=1}^{T} w_{kt}\right)^2\right)^{-1} \tag{7}$$

where ϑ_{kt} is a normalised kernel function. w_{kt} uses a normal kernel weighting function. ϕ_{Tk} gives the rate of convergence and behaves like the bandwidth parameter H in Equation (6), and it is the kernel function that provides greater weight to observations surrounding the parameter estimates at time k relative to more distant observations.

Using a Normal-Wishart prior distribution for $\theta_k | \, \Xi_k$ for $k \in \{1, \ldots, T\}$:

$$\theta_k | \Xi_k \backsim \mathcal{N}\left(\theta_{0k}, (\Xi_k \otimes \mathbf{\Omega}_{0k})^{-1}\right) \tag{8}$$

$$\Xi_k \backsim \mathcal{W}\left(\alpha_{0k}, \mathbf{\Gamma}_{0k}\right) \tag{9}$$

where θ_{0k} is a vector of prior means, $\mathbf{\Omega}_{0k}$ is a positive definite matrix, α_{0k} is a scale parameter of the Wishart distribution ($\mathcal{W}$), and $\mathbf{\Gamma}_{0k}$ is a positive definite matrix.

The prior and weighted likelihood function implies a Normal-Wishart quasi posterior distribution for $\theta_k | \, \Xi_k$ for $k = \{1, \ldots, T\}$. Formally let $\tilde{\mathbf{X}} = (x_1', \ldots, x_T')'$ and $\tilde{\mathbf{Y}} = (y1, \ldots, y_T)'$ then:

$$\theta_k | \Xi_k, \tilde{\mathbf{X}}, \tilde{\mathbf{Y}} \;\; \backsim \;\; \mathcal{N}\left(\tilde{\theta}_k, \left(\Xi_k \otimes \tilde{\mathbf{\Omega}}_k\right)^{-1}\right) \tag{10}$$

$$\Xi_k \;\; \backsim \;\; \mathcal{W}\left(\tilde{\alpha}_k, \tilde{\mathbf{\Gamma}}_k^{-1}\right) \tag{11}$$

with quasi posterior parameters

$$\tilde{\theta}_j = \left(\mathbf{I}_N \otimes \tilde{\mathbf{\Omega}}_k^{-1}\right)\left[\left(\mathbf{I}_N \otimes \tilde{\mathbf{X}}'\mathbf{D}_k\tilde{\mathbf{X}}\right)\hat{\theta}_k + \left(\mathbf{I}_N \otimes \mathbf{\Omega}_{0k}\right)\theta_{0k}\right] \tag{12}$$

$$\tilde{\mathbf{\Omega}}_k = \tilde{\mathbf{\Omega}}_{0k} + \tilde{\mathbf{X}}'\mathbf{D}_k\tilde{\mathbf{X}} \tag{13}$$

$$\tilde{\alpha}_k = \alpha_{0k} + \sum_{t=1}^{T} \vartheta_{kt} \tag{14}$$

$$\tilde{\mathbf{\Gamma}}_k = \mathbf{\Gamma}_{0k} + \tilde{\mathbf{Y}}'\mathbf{D}_k\tilde{\mathbf{Y}} + \mathbf{\Theta}_{0k}\mathbf{\Gamma}_{0k}\mathbf{\Theta}_{0k}' - \tilde{\mathbf{\Theta}}_k\tilde{\mathbf{\Gamma}}_k\tilde{\mathbf{\Theta}}_k' \tag{15}$$

where $\hat{\theta}_k = \left(\mathbf{I}_N \otimes \tilde{\mathbf{X}}'\mathbf{D}_k\tilde{\mathbf{X}}\right)^{-1}\left(\mathbf{I}_N \otimes \tilde{\mathbf{X}}'\mathbf{D}_k\right)y$ is the local likelihood estimator for θ_k. The matrices $\mathbf{\Theta}_{0k}$, $\tilde{\mathbf{\Theta}}_k$ are conformable matrices from the vector of prior means, θ_{0k}, and a draw from the quasi posterior distribution, $\tilde{\theta}_k$, respectively.

There are three benefits from applying this model. First, we can estimate large scale models with drifting coefficients that conventional methods cannot handle (Primiceri, 2005). This is due to the state-space representation of an N-dimensional TVP-VAR(L) model. In particular, these systems require an additional $N(3/2 + N(L + 1/2))$ state equations for every additional variable. Conventional Markov Chain Monte Carlo (MCMC) methods fail to estimate larger models, which in general confine one to (usually) fewer than 6 variables in the system. Second, the standard approach is fully parametric and requires one to specify a law of motion. This can distort inference if the true law of motion is misspecified (Petrova, 2019). Third, the methods used here permit direct estimation of the VAR's time-varying covariance matrix, which has an inverse-Wishart density and is symmetric positive definite at every point in time.

In estimating the model outlined, we use L=2 and a Minnesota Normal-Wishart prior with a shrinkage value $\varphi = 0.05$ and centre the coefficient on the first lag of each variable to 0.5 in each respective equation. The prior for the Wishart parameters are set following Kadiyala and Karlsson (1997). We experiment with various lag lengths, $L = \{2, 3, 4\}$; shrinkage values, $\varphi = \{0.05, 0.25, 0.5\}$; and values to centre the coefficient on the first lag of each variable, $\{0, 0.5, 0.7, 0.9\}$. Results provide similar conclusions to those we report below.

2.3 Forecasting Methodology

In order to assess the forecasting performance of Divisia M4 relative to its seven more narrow counterparts, we estimate a battery of TVP-VAR models outlined above in a recursive framework. We consider three different model sizes: i) small models with N=4 variables that contain GDP growth, inflation, the federal funds/shadow rate (Wu and Xia, 2016), and one of the eight measures of Divisia money; ii) medium models with N=6 variables which add wage growth and the unemployment rate to our N=4 variable models; and iii) large models with N=10 variables which further add oil price growth, the corporate bond spread, stock market returns and house price growth to the N=6 variable models. Consequently, for each model size we have $\mathcal{M} = 8$ forecasting models, resulting in 24 forecasting models.

We begin forecasting from 1990:Q2 using an initial sample of 91 observations. We run 5,000 simulations at each point in time at each recursion of the model. When the algorithm gets to the last time observation at each recursion, we forecast all variables in the system at horizon $h = \{1, 2, 3, 4\}$ quarters into the future for each of the 5,000 simulations. For $h > 1$, we obtain forecasts iteratively and analyse cumulative growth rates. We then move on to the next quarter and repeat until the end of the sample period in 2018:Q4, providing us with 28 years of out-of-sample pseudo-real-time forecasts.

We benchmarks forecasts from models using Divisia M4 in a pairwise manner against forecasts from corresponding models that use one of the more narrowly defined measures of money using the entire predictive density. We approximate predictive densities using kernel methods to account for the non-linearity in the data and our models (Alessandri and Mumtaz, 2017). To assess the accuracy of predictive densities, we rely on log-scores which measure the log-likelihood the model assigns to actual observations based on lagged data. In using log-scores, we are able to establish the forecasting performance of Divisia M4 using the entire predictive density as opposed to focusing on point forecasts and statistics such as root-mean-squared-errors.[2]

Typically, log-scores compare the average performance over a set of models, or model

[2]Note that we do compute these statistics using our models to assess the forecasting performance of Divisia M4 relative to its narrower counterparts using the mean of the respective predictive densities. These results, which are available upon request, show that one is unable to distinguish statistically or economically between Divisia M4 and any of its narrower counterparts.

specifications, over a given sample. In the context of our study, even if, on average, forecasts of GDP growth and inflation are more accurate using Divisia M4, two questions arise. First, can we distinguish whether Divisia M4 produces more accurate forecasts of GDP growth and inflation in particular states of the economy? Second, how robust is the evidence in favour of using the broadest monetary aggregate for forecasting GDP growth and inflation throughout the sample period? To answer these questions, we use log-predictive Bayes factors which summarise the difference between cumulative log-scores of models each period t (Geweke and Amisano, 2010). This metric uncovers points in time a set of two models compare and assesses the overall predictive performance of a model in comparison to another in real-time. We provide further details in Appendix B.

3 The Case for Time-varying Models

We first motivate the use of TVP-VAR models by assessing the forecasting accuracy of each TVP-VAR model relative to a linear Bayesian VAR model using the same variables.[3] In this exercise, we consider joint distribution of GDP growth and inflation using joint log-scores. This is justified economically, since individual log-scores place no emphasis on correlations and thus may be of limited use to forecasters wishing to use a specific set of variables for prediction. In the case of central banks which require information on predictions regarding the future state of the economy, a minimum requirement for policy-making purposes would be information on GDP and inflation.

Table 1 reports joint log-scores of GDP growth and inflation from TVP-VARs using each measure of money, relative to joint log-scores from linear Bayesian VAR models using the same measure of money at $h = \{1, 2, 3, 4\}$ quarter horizons over the forecasting sample. Panel A reports results from small models, while Panels B and C report results from medium and large models, respectively. Positive values indicate a more accurate joint predictive density of GDP growth and inflation from the TVP-VAR model and are in bold font.

In general, across all model sizes, the TVP-VARs generate more accurate joint log-predictive scores relative to simple linear Bayesian VAR models. The only cases where the Bayesian VAR produces more accurate joint log-scores are at four quarter horizons, particularly for models using M4X, M3, and M2A. Nevertheless, even considering average forecast performance provides evidence in favour of using TVP-VARs to forecasts economic activity. We therefore proceed in this manner when assessing the forecasting performance of Divisia M4 relative to the narrower definitions of money.

[3] The linear Bayesian VARs use Minnesota-Wishart priors similar to those we specify for the TVP-VAR models and use the same number of simulations to generate predictive densities.

Table 1: **Average Forecast Performance of TVP-VARs versus Bayesian VARs: Joint Log Predictive Scores for GDP growth and Inflation**

Notes: This table reports the joint log predictive scores of GDP growth, y_t and CPI inflation π_t for TVP-VAR models that each use a Divisia measure of money, at $h = \{1, 2, 3, 4\}$ quarter horizons relative to a Bayesian VAR that uses the same measure of Divisia. Positive values indicate that the TVP-VAR model produces a higher log-score relative to its corresponding Bayesian VAR model using the same measure of money. The forecast sample spans 1990:Q2–2018:Q4. Panel A reports results from N=4 models; Panels B and C report results from N=6 and N=10 variable models respectively. M4 refers to Divisia M4; M4X denotes Divisia M4 excluding Treasury Bills; M3 is Divisia M3; M2A is Divisia M2 All; M2 is Divisia M2; MZM is Divisia Money Zero-Maturity; M2M is Divisia M2M; and M1 is Divisia M1.

Panel A: Small Models, N=4 variables								
	M4	M4X	M3	M2A	M2	MZM	M2M	M1
h=1	**0.35**	**0.33**	**0.33**	**0.30**	**0.33**	**0.36**	**0.31**	**0.33**
2	**0.24**	**0.12**	**0.20**	**0.16**	**0.19**	**0.27**	**0.12**	**0.22**
3	**0.21**	**0.08**	**0.30**	**0.34**	**0.22**	**0.40**	**0.22**	**0.07**
4	-0.04	-0.29	-0.06	**0.01**	-0.15	0.00	**0.08**	-0.11
Panel B: Medium Models, N=6 variables								
	M4	M4X	M3	M2A	M2	MZM	M2M	M1
h=1	**0.31**	**0.31**	**0.28**	**0.33**	**0.33**	**0.34**	**0.29**	**0.30**
2	**0.27**	**0.19**	**0.32**	**0.10**	**0.23**	**0.21**	**0.06**	**0.13**
3	**0.29**	**0.17**	**0.32**	**0.16**	**0.42**	**0.34**	**0.23**	**0.39**
4	**0.19**	-0.10	**0.19**	-0.05	-0.12	**0.08**	0.00	**0.14**
Panel C: Large Models, N=10 variables								
	M4	M4X	M3	M2A	M2	MZM	M2M	M1
h=1	**0.32**	**0.34**	**0.31**	**0.35**	**0.32**	**0.34**	**0.36**	**0.34**
2	**0.10**	**0.13**	**0.29**	**0.33**	**0.14**	**0.21**	**0.33**	**0.33**
3	**0.31**	**0.22**	**0.29**	**0.36**	**0.33**	**0.22**	**0.52**	**0.22**
4	**0.03**	-0.19	-0.06	-0.05	**0.02**	-0.13	**0.30**	**0.27**

4 Results

We consider the forecasting performance of Divisia M4 relative to more narrow definitions of money in a pairwise manner. Table 2 assesses the average forecasting performance of models with Divisia M4 using individual log-scores of GDP growth and inflation. Panels A, B and C report results from small, medium and large models, respectively. For models using Divisia M4 we report raw log-score values. For the remaining seven measures of Divisia money, in each model size, we show log-score differences such that positive values indicate that the model using Divisia M4 outperforms its more narrow counterpart. In Table 3, read in a similar manner to Table 2, we show joint log-scores of GDP growth and inflation.

The individual log-scores in Table 2 provide mixed evidence around the forecasting performance of Divisia M4. There is no clear pattern that emerges from specific definitions of money across model size. However, medium models provide the strongest case where, across all forecast horizons for both GDP growth and inflation, Divisia M4 outperforms the narrower definitions of money.

Turning to joint log-scores, several observations emerge from Table 3. First, there is a stronger case for using Divisia M4 in general. Second, the N=4 variable models using the broadest monetary aggregate produce a more accurate joint distribution of GDP growth and inflation at h=1 and h=4 quarter horizons. Third, medium models utilising Divisia M4 provide a higher joint log-scores at short horizons, whereas large models exhibit stronger gains at longer horizons.

To investigate this question in real-time, we consider pairwise comparisons of models using Divisia M4 relative to a more narrow alternative using log predictive Bayes factors (Geweke and Amisano, 2010) focusing on a 4-quarter ahead horizon only.[4] For GDP growth and inflation, the period t Bayes factor is the cumulative difference between a model using Divisia M4 and a model using one of the narrower counterparts. Positive values indicate that the model using Divisia M4 outperforms the alternative. These statistics allow us to examine how evidence in favour of using Divisia M4 relative to one of its more narrow counterparts evolves over time.

Figure 1 shows the Bayes factors for GDP growth and inflation. Panel A reports statistics from N=4 variable models, and Panels B and C report results from N=6 and N=10 variable models, respectively. Overall, these results provide strong evidence in favour of using Divisia M4 from 2008 onwards. The N=4 and N=10 models suggest that Divisia M4 provides more accurate predictions of GDP growth for 4 out of the 7 narrower measures of money. The N=6 models indicate Divisia M4 is preferred for 5 out of the 7

[4]Results at shorter horizons are comparable to those we report in this paper and are available on request.

narrower measures of money from 2008 onwards. Turning to inflation, we can see that across all model sizes, Divisia M4 outperforms 5 out of the 7 narrower measures. Taken together, the results in Figure 1 demonstrate that Divisia M4 consistently outperforms M4X, M2, MZM, and M2M in terms of predicting GDP growth and inflation.

In Figure 2, we report the log predictive Bayes factors calculated on joint log-scores of GDP growth and inflation. Similar to Figure 1, Panels A, B, and C contain results on N=4, N=6, and N=10 variable models, respectively. On the whole, these plots show that models using Divisia M4 gradually increase from the burst of the dot-com bubble in 2000-2001. In particular, from the N=4 models Divisia M4 is preferred to all measures of money except M2A. Turning to Panels B and C, we can see that Divisia M4 is preferred relative to 5 out of 7 narrower measures. Overall, these plots provide an even stronger case for using Divisia M4 to predict output and inflation relative to the narrower alternatives.

These results provide evidence in favour of using the broadest measure of Divisia money when looking to forecast real activity. In particular, statistics showing forecast performance in real time uncover a stark shift in support of using Divisia M4 from 2008 until the end of the sample period. Furthermore, the log Bayes factors we compute on joint log-scores, which allow us to benchmark Divisia M4 against its more narrow counterparts in a pairwise manner, show a gradual increase from the beginning of our forecast sample with positive shifts occurring prior to the 2001 and the 2008 recessions.

Table 2: **Average Forecast Performance: Log Predictive Scores for GDP growth and Inflation**
Notes: This table reports the log predictive scores of GDP growth, y_t, and CPI inflation, π_t, for TVP-VAR models which use a Divisia measure of money, at $h = \{1, 2, 3, 4\}$ quarter horizons. The forecast sample spans 1990:Q2–2018:Q4. Panel A reports results from N=4 models; Panels B and C report results from N=6 and N=10 variable models, respectively. We report the log predictive scores at each horizon for the model using Divisia M4. All other columns report relative log-scores where we benchmark the model using Divisia M4 to each of its more narrow alternatives. M4X denotes Divisia M4 excluding Treasury Bills; M3 is Divisia M3; M2A is Divisia M2 All; M2 is Divisia M2; MZM is Divisia Money Zero-Maturity; M2M is Divisia M2M; and M1 is Divisia M1. Positive values indicate that the model using Divisia M4 produces a higher log-score relative to a model using a more narrow alternative.

Panel A: Small Models, N=4 variables

	M4		M4X		M3		M2A		M2		MZM		M2M		M1	
	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t
h=1	-1.02	-1.91	-0.12	-0.04	-0.09	-0.68	-0.13	-0.12	-0.12	0.21	-0.09	-0.25	-0.03	**0.09**	**0.04**	-0.11
2	-0.99	-1.61	**0.03**	-0.20	**0.09**	-0.34	-0.01	**0.14**	**0.03**	-0.09	**0.11**	-0.08	-0.17	-0.13	**0.14**	-0.56
3	-1.37	-0.80	-0.10	-0.05	-0.04	-0.19	-0.24	-0.01	-0.34	-0.19	**0.26**	**0.10**	-0.29	**0.17**	-0.13	-0.24
4	-1.39	-0.36	**0.08**	**0.02**	**0.41**	-0.06	-0.05	**0.06**	-0.12	**0.01**	**0.23**	**0.09**	-0.18	0.09	-0.14	-0.08

Panel B: Medium Models, N=6 variables

	M4		M4X		M3		M2A		M2		MZM		M2M		M1	
	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t
h=1	-0.89	-1.57	**0.10**	-0.13	0.00	-0.30	**0.26**	**0.07**	**0.01**	-0.22	**0.10**	-0.18	**0.01**	-0.08	0.00	-0.16
2	-1.09	-0.55	**0.15**	**0.35**	-0.25	**0.41**	**0.09**	**0.65**	**0.15**	**0.89**	**0.14**	**0.40**	**0.09**	0.76	**0.01**	**0.84**
3	-1.17	-0.39	**0.17**	**0.27**	**0.09**	**0.15**	**0.14**	**0.26**	**0.19**	**0.22**	-0.16	**0.23**	**0.11**	0.32	-0.19	**0.34**
4	-1.01	-0.34	**0.22**	**0.01**	-0.07	-0.01	**0.31**	**0.04**	**0.31**	-0.02	**0.26**	**0.08**	**0.26**	0.05	-0.03	**0.06**

Panel C: Large Models, N=10 variables

	M4		M4X		M3		M2A		M2		MZM		M2M		M1	
	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t
h=1	-0.96	-1.47	-0.12	**0.17**	-0.11	-0.38	-0.06	**0.35**	**0.02**	-0.40	-0.12	-0.06	-0.08	**0.55**	-0.11	-0.33
2	-0.80	-1.06	**0.07**	**0.68**	**0.45**	**0.12**	**0.17**	-0.26	**0.07**	-0.14	0.00	**0.49**	**0.13**	**0.46**	**0.20**	-0.54
3	-1.21	-0.81	-0.22	**0.02**	-0.16	-0.25	**0.05**	-0.25	-0.18	-0.41	-0.16	**0.13**	-0.46	-0.32	**0.18**	-0.45
4	-1.08	-0.38	-0.06	**0.10**	0.17	**0.02**	**0.14**	-0.06	-0.05	-0.04	**0.17**	**0.09**	-0.34	**0.02**	**0.03**	0.00

Table 3: **Average Forecast Performance: Joint Log Predictive Scores for GDP growth and Inflation**

Notes: This table reports the joint log predictive scores of GDP growth, y_t, and CPI inflation, π_t, for TVP-VAR models which use a Divisia measure of money, at $h = \{1, 2, 3, 4\}$ quarter horizons. The forecast sample spans 1990:Q2–2018:Q4. Panel A reports results from N=4 models; Panels B and C report results from N=6 and N=10 variable models, respectively. We report the log predictive scores at each horizon for the model using Divisia M4. All other columns report relative log-scores where we benchmark the model using Divisia M4 to each of its more narrow alternatives. M4X denotes Divisia M4 excluding Treasury Bills; M3 is Divisia M3; M2A is Divisia M2 All; M2 is Divisia M2; MZM is Divisia Money Zero-Maturity; M2M is Divisia M2M; and M1 is Divisia M1. Positive values indicate that the model using Divisia M4 produces a higher log-score relative to a model using a more narrow alternative.

	M4	M4X	M3	M2A	M2	MZM	M2M	M1
Panel A: Small Models, N=4 variables								
h=1	-1.55	**0.01**	**0.01**	**0.04**	**0.01**	0.00	**0.04**	**0.01**
2	-1.55	-0.02	**0.05**	**0.08**	-0.05	0.00	**0.03**	0.00
3	-1.56	-0.02	-0.08	-0.13	-0.07	-0.15	-0.15	**0.05**
4	-1.75	**0.07**	**0.01**	-0.01	**0.08**	**0.04**	**0.03**	**0.03**
Panel B: Medium Models, N=6 variables								
h=1	-1.59	-0.01	**0.02**	-0.01	-0.02	-0.02	**0.02**	**0.01**
2	-1.59	-0.02	-0.05	**0.04**	**0.01**	**0.03**	**0.09**	**0.04**
3	-1.59	**0.08**	-0.11	**0.08**	-0.12	-0.07	**0.02**	-0.10
4	-1.79	**0.03**	-0.05	-0.01	**0.11**	**0.03**	**0.06**	0.00
Panel C: Large Models, N=10 variables								
h=1	-1.58	-0.02	0.00	-0.02	**0.02**	-0.01	-0.02	-0.01
2	-1.68	-0.08	-0.16	-0.18	-0.06	-0.16	-0.14	-0.19
3	-1.47	**0.03**	**0.11**	**0.06**	**0.09**	-0.05	-0.11	-0.01
4	-1.66	**0.13**	**0.11**	**0.11**	**0.14**	**0.03**	-0.10	-0.19

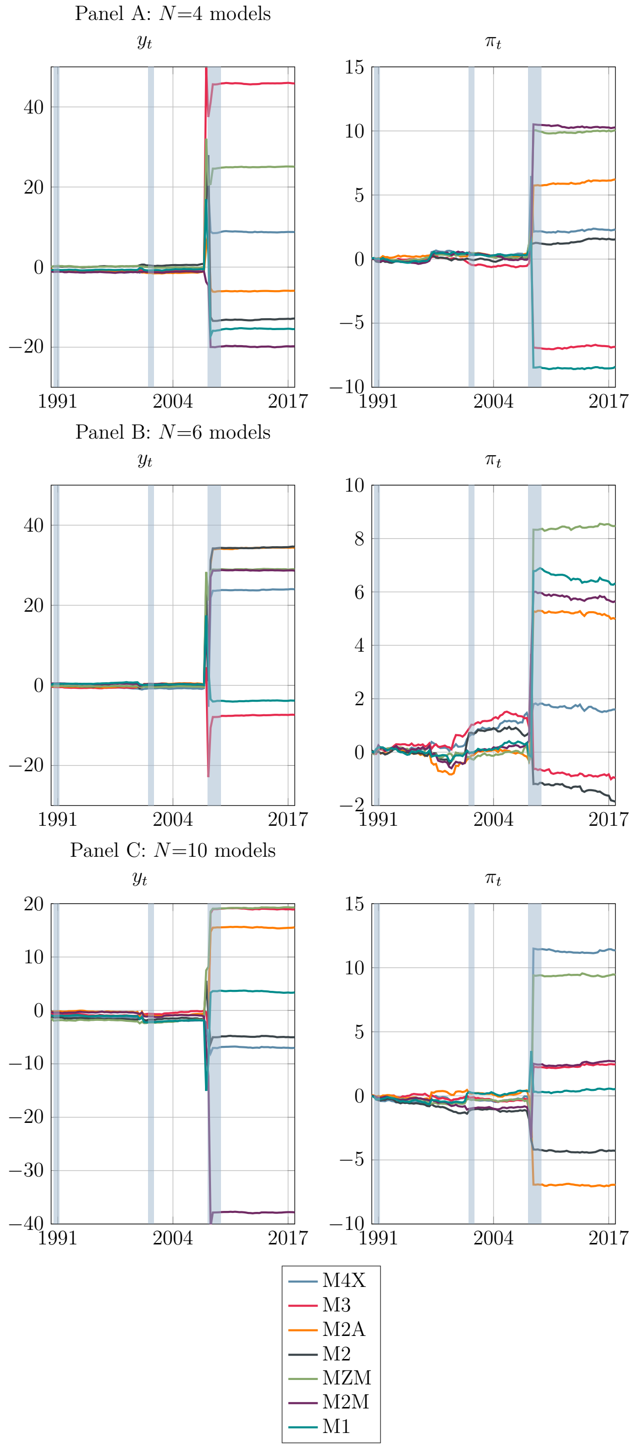

Figure 1: **Log Bayes Factors for GDP Growth and Inflation**
This figure reports the h=4 quarter ahead log Bayes factors in Geweke and Amisano (2010) for GDP growth and inflation from 1990:Q2 to 2017:Q4. y_t is GDP growth and π_t is inflation. These plots summarise the difference between cumulative log-scores of the TVP-VAR using Divisia M4 and a more narrow measure of Divisia money. Panel A reports results from N=4 variable TVP-VARs; Panels B and C report results for N=6 and N=10 TVP-VARs. Positive values indicate the overall predictive performance of the model using a Divisia M4 outperforms the narrower alternative in question. The shaded areas correspond to NBER recession dates.

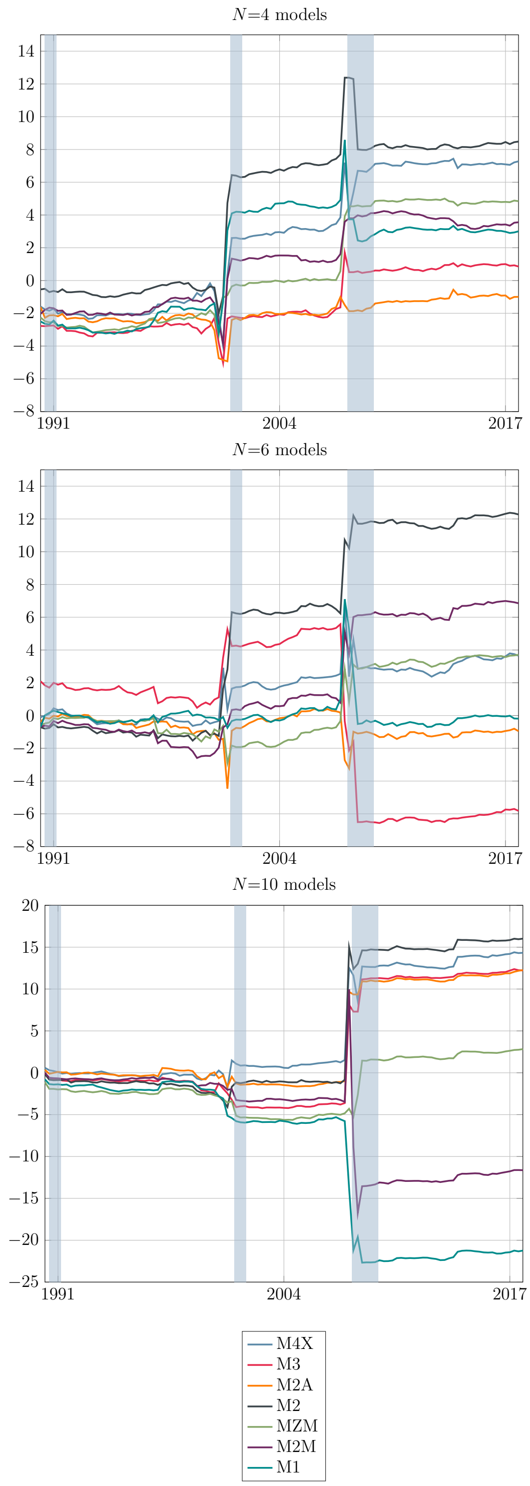

Figure 2: **Log Bayes Factors for Joint log-scores of GDP Growth and Inflation** This figure reports the h=4 quarter ahead log Bayes factors in Geweke and Amisano (2010) for joint log predictive scores of GDP growth and inflation from 1990:Q2 to 2017:Q4. These plots summarise the difference between cumulative joint log-scores of the TVP-VAR using Divisia M4 and a more narrow measure of Divisia money. Panel A reports results from N=4 variable TVP-VARs; Panels B and C report results for N=6 and N=10 TVP-VARs. Positive values indicate the overall predictive performance of the model using a Divisia M4 outperforms the narrower alternative in question. The shaded areas correspond to NBER recession dates.

4.1 Pooling Forecasting Models

A natural extension to our overall objective is to compare the forecasting performance of Divisia M4 against forecasts we pool from an array of $\mathcal{M}$ models. Specifically, for each model size, we evaluate the forecasts we obtain from the TVP-VARs using Divisia M4 relative forecasts that pool information from $\mathcal{M} = 7$ models using each of the narrower Divisia alternatives. Our second exercise benchmarks, for each model size, forecasts of GDP growth and inflation from the TVP-VARs using Divisia M4 against forecasts that pool information from all $\mathcal{M} = 8$ models. We consider two weighting schemes. The first is the optimal weighting scheme of Geweke and Amisano (2011) that updates weights recursively through time. The second is a naive equal weighting scheme.

Table 4 reports the average relative log-scores over the forecast sample at a 4-quarter horizon for GDP growth and inflation. Panels A and B show results using the optimal weighting scheme of Geweke and Amisano (2011) and equal weighting scheme, respectively. Results under heading 1 benchmark against pooled information from $\mathcal{M} = 7$ models, and those under heading 2 benchmark against $\mathcal{M} = 8$ models. Positive values indicate that the TVP-VAR using Divisia M4 outperforms the pooled forecasts. As shown in Panel A, across all model sizes, our benchmark TVP-VAR forecasting GDP growth outperforms pooled forecasts. Turning our attention to inflation, our benchmark model outperforms pooled information in small and medium models, and is on average equivalent for large models. Panel B reveals the same outcome, with results providing an even stronger case for our baseline model.

In Table 5, read in a similar manner to Table 4, we report sample averages of relative joint-log-scores of GDP growth and inflation at a 4-quarter horizon . Here, we can see that our benchmark TVP-VAR using Divisia M4 outperforms pooled forecasts using both an optimal and equal weighting scheme, and over all model sizes.

To assess the performance of our baseline model over time, we report log Bayes factors computed on joint log-scores at a 4-quarter horizon in Figure 3. We do so for both weighting schemes, as well as both model sets. Log Bayes factors gradually increase over the forecast sample. While we observe a steady increase over the forecast sample when pooling information using an equal weighting scheme, we see jumps in these statistics prior to the burst of the dot-com bubble under the optimal weights.

These results, which account for multiple models and measures of money, provide further evidence in favour of using the broadest measure of money when forecasting real activity. Additionally, our 'real time' statistics show surges in predictive performance prior to recessionary periods, consistent with Barnett and Chauvet (2011) who show that Divisia M4 acts as a useful leading indicator in predicting recessions, and Barnett et al. (2016) who document the benefits of using Divisia in nowcasting GDP.

Table 4: **Average Forecast Performance: Relative Log Predictive Scores for GDP growth and Inflation Benchmarked against Pooled Forecasts**
Notes: This table reports the log predictive scores of GDP growth, y_t, and CPI inflation, π_t, from TVP-VAR models using Divisia M4 at $h = 4$ quarter horizon relative to forecasts pooled from TVP-VARs using all narrower measures of Divisia. Panel A reports relative log predictive scores where the pooling information uses the optimal weighting scheme of Geweke and Amisano (2011), and Panel B assigns equal weights when pooling forecasts. Columns under 1, benchmark the forecasts from the TVP-VAR using Divisia M4 against pooled information from the remaining seven narrow measures of money. Columns under 2 benchmark Divisia M4 against pooled forecasts from models using all measures of money, including Divisia M4 itself. The narrower alternative measures are: Divisia M4X which is Divisia M4 excluding Treasury Bills; Divisia M3; Divisia M2 All; Divisia M2; Divisia Money Zero-Maturity; Divisia M2M; and Divisia M1. Positive values indicate that the forecast stemming from the model using Divisia M4 produces a higher log-score relative to pooled forecasts.

Panel A: Optimal weighting scheme				
	1		**2**	
	y_t	π_t	y_t	π_t
Small Models, N=4	**5.73**	**0.77**	**5.73**	**1.31**
Medium Models, N=6	**1.43**	**0.16**	**6.80**	**0.16**
Large Models, N=10	**2.24**	0.00	**2.24**	0.00

Panel B: Equal weighting scheme				
	1		**2**	
	y_t	π_t	y_t	π_t
Small Models, N=4	**19.03**	**18.85**	**18.72**	**23.25**
Medium Models, N=6	**18.78**	**18.14**	**23.59**	**25.82**
Large Models, N=10	**17.64**	**17.59**	**22.29**	**23.00**

Table 5: **Average Forecast Performance: Relative Joint Log Predictive Scores for GDP growth and Inflation Benchmarked against Pooled Forecasts**
Notes: This table reports the joint log predictive scores of GDP growth, y_t, and CPI inflation, π_t, from TVP-VAR models using Divisia M4 at $h = 4$ quarter horizon relative to forecasts pooled from TVP-VARs using all narrower measures of Divisia. Panel A reports relative log predictive scores where the pooling information uses the optimal weighting scheme of Geweke and Amisano (2011), and Panel B assigns equal weights when pooling forecasts. Columns under 1, benchmark the forecasts from the TVP-VAR using Divisia M4 against pooled information from the remaining seven narrow measures of money. Columns under 2 benchmark Divisia M4 against pooled forecasts from models using all measures of money, including Divisia M4 itself. The narrower alternative measures are: Divisia M4X which is Divisia M4 excluding Treasury Bills; Divisia M3; Divisia M2 All; Divisia M2; Divisia Money Zero-Maturity; Divisia M2M; and Divisia M1. Positive values indicate that the forecast stemming from the model using Divisia M4 produces a higher joint log-score relative to pooled forecasts.

Panel A: Optimal weighting scheme		
	1	**2**
Small Models, N=4	**0.82**	**1.24**
Medium Models N=6	**0.41**	**0.42**
Large Models, N=10	**0.41**	**0.41**
Panel B: Equal weighting scheme		
	1	**2**
Small Models, N=4	**8.89**	**10.34**
Medium Models N=6	**9.89**	**11.04**
Large Models, N=10	**9.69**	**10.99**

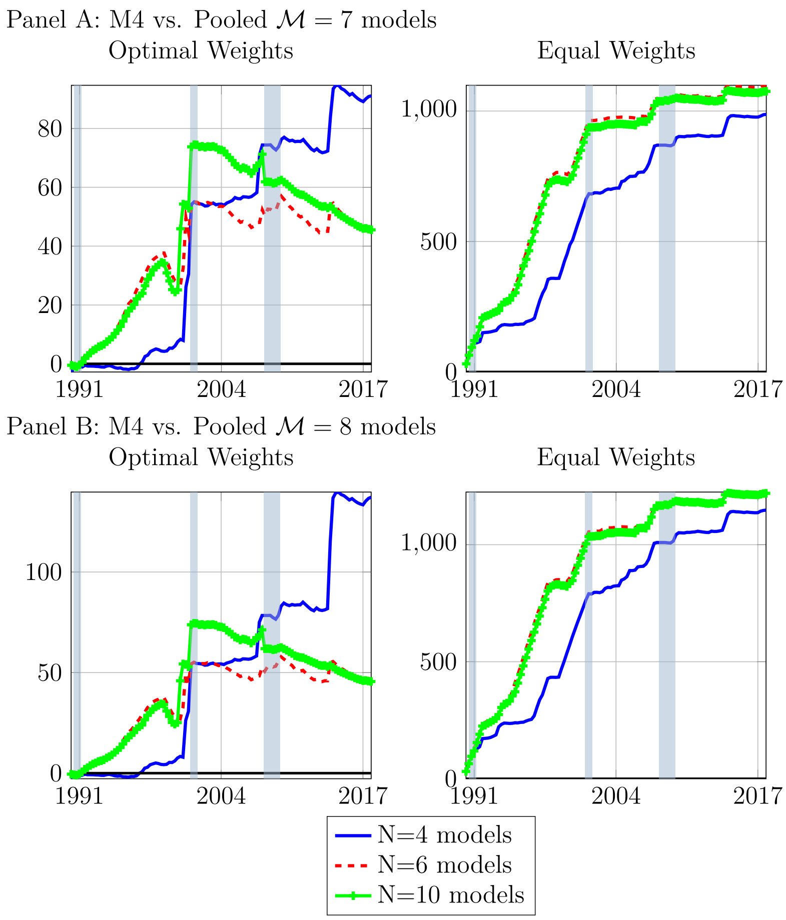

Figure 3: **Log Bayes Factors for Joint log-scores of GDP Growth and Inflation: Divisia M4 vs. Pooled Information**

This figure reports the h=4 quarter ahead log Bayes factors in Geweke and Amisano (2010) for joint log predictive scores of GDP growth and inflation from 1990:Q2 to 2017:Q4. These plots summarise the difference between cumulative joint log-scores of the TVP-VAR using Divisia M4 and log-scores pooled from a set of $\mathcal{M} = \{7, 8\}$ models. Panel A reports log Bayes factors based on joint log-scores from the model using Divisia M4 relative to pooling information from the models using the 7 narrower alternatives. Panel B reports the log Bayes factors based on joint log-scores from the model using Divisia M4 relative to pooling information from all 8 models that each contain a measure of Divisia money. Positive values indicate the overall predictive performance of the model using a Divisia M4 outperforms the pooled information. Solid lines refer to N=4 variable TVP-VARs; dashed lines refer to N=6 variable TVP-VARs; and solid dotted lines refer to N=10 variable models. The shaded areas correspond to NBER recession dates.

5 Extensions and Robustness Analysis

Having documented the superior performance of forecasts of economic activity generated from our preferred TVP-VAR model with Divisia M4, we now undertake several additional tests and extension in order to establish the robustness of our findings. We first investigate whether the inclusion of any Divisia monetary aggregate improves the forecasting performance of our models relative to comparable specifications with no measures of money. Next, we consider pairwise forecasting decision rules as in Giacomini and White (2006) which we base on left-tail weighted log-scores (Amisano and Giacomini, 2007). We then compare the forecasting performance of real activity for our benchmarks with a TVP-VAR model using 143 variables.Finally, we present results from two constant-parameter VAR models and from pooled models which use two alternative measures of inflation.

5.1 Do we need Divisia M4 at all?

In this subsection we assess the forecasting performance of Divisia M4 from our baseline models with comparable models that contain no monetary aggregate. For our N=4 variable models we compare these forecasts against an N=3 variable model. In a similar manner we remove Divisia M4 from our medium and large models models and compare these against our N=6 and N=10 variable models respectively.

Table 6 reports average forecast performance over the sample of models using Divisia M4 relative to analogous models with no measure of money. In Panel A we report relative log-scores of GDP growth and inflation, and in Panel B we report relative joint log-scores of GDP growth and inflation. Overall, it is clear that models containing Divisia M4 outperform those comparable models with no measure of money. Concerning GDP growth, individual log-scores show that as the model size increases, forecasts of GDP growth using Divisia M4 become increasingly more accurate relative to omitting money altogether. For inflation, models using Divisia M4 provide more accurate forecasts irrespective of size. Turning attention to joint log-scores, our results show increases in forecast performance at a 4-quarter horizon.

Figure 4 plots log Bayes factors that we compute on joint log-scores at a 4-quarter horizon to investigate gains over time. In general, across all model sizes, we realise gains in predictive ability prior to the burst of the Dot-com bubble with surges prior to and during the Great Recession for medium and large models. For small models, the benefits of using Divisia M4 relative to omitting money altogether appear from 2008 until the end of the sample period.

These results demonstrate that utilising a monetary approach, by including the broadest

Table 6: **Average Forecast Performance of Models using Divisia M4 Relative to Comparable Models using no Measure of Money**

Notes: Panel A of this table reports the log-scores of GDP growth, y_t, and inflation, π_t, of models using Divisia M4 relative to comparable models with no measure of money. Panel B of this table reports the joint log-scores of GDP growth and inflation using Divisia M4 relative to comparable models using no measure of money. Relative log-scores are averages over the forecast sample which spans 1990:Q2 to 2017:Q4 at a $h = \{1, 2, 3, 4\}$ horizon. Results under the heading: i) N=4 vs. N=3 refer to small models; ii) N=6 vs. N=5 refer to medium models; and iii) N=10 vs. N=9 refer to large models. Positive values indicate the model using Divisia M4 outperforms its analogous model with no measure of money and are in bold font.

Panel A: Relative log-scores of y_t, π_t

	N=4 vs. N=3		N=6 vs. N=5		N=10 vs. N=9	
	y_t	π_t	y_t	π_t	y_t	π_t
h=1	-0.14	-0.20	**0.07**	**0.31**	-0.04	**0.60**
2	-0.15	**0.05**	-0.18	**0.50**	**0.14**	**0.28**
3	-0.28	**0.13**	-0.14	**0.03**	**0.05**	-0.08
4	-0.10	**0.17**	**0.18**	-0.02	**0.51**	**0.12**

Panel B: Relative Joint log-scores of y_t, π_t

	N=4 vs. N=3	N=6 vs. N=5	N=10 vs. N=9
h=1	**0.02**	0.00	-0.02
2	-0.01	**0.06**	-0.13
3	**0.04**	**0.02**	-0.05
4	0.00	**0.08**	**0.02**

measure of money, provides substantial gains in predictive ability of models of economic activity. The results are similar to pairwise tests against narrow measures of money and therefore further highlight the advantages of using Divisia M4 in predicting the future state of the economy.

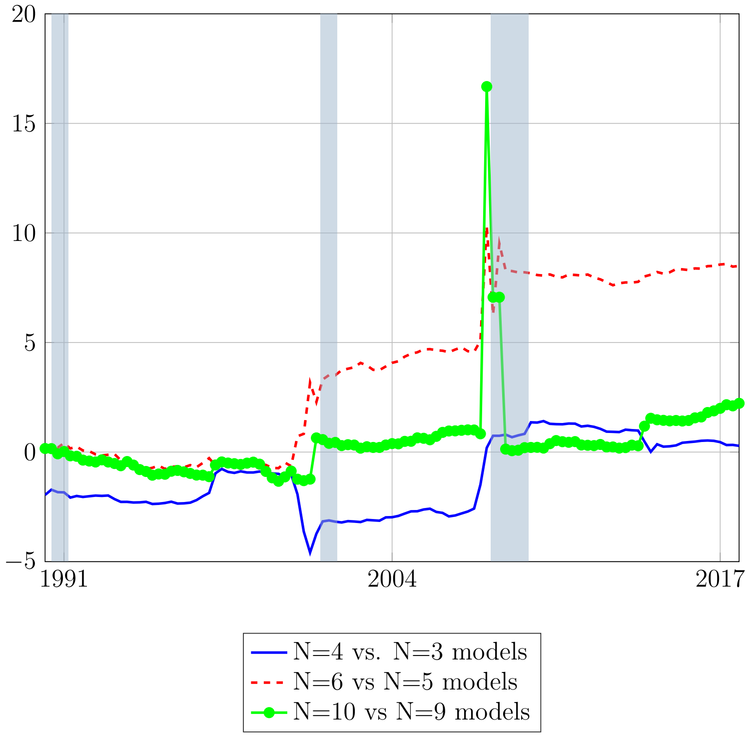

Figure 4: **Log Bayes Factors for Joint log-scores of GDP Growth and Inflation: Divisia M4 vs. Models with no Money**
This figure reports the h=4 quarter ahead log Bayes factors in Geweke and Amisano (2010) for joint log predictive scores of GDP growth and inflation from 1990:Q2 to 2017:Q4. These plots summarise the difference between cumulative joint log-scores of the TVP-VAR using Divisia M4 and log-scores a comparable model using no monetary aggregate. Solid lines refer to N=4 variable TVP-VARs; dashed lines refer to N=6 variable TVP-VARs; and solid dotted lines refer to N=10 variable models. The shaded areas correspond to NBER recession dates.

5.2 Forecasting Decision Rules from Weighted Log-Scores

To explore the forecasting performance of our benchmark TVP-VAR in a pairwise manner we construct forecast decision rules following Giacomini and White (2006). These statistics are out-of-sample and useful in assessing model performance in real time. Contrasting with log Bayes factors, these decision rules capture the short-run dynamics between two models (Alessandri and Mumtaz, 2017). Appendix B provides details of the strategy. Within this exercise, we base decision rules on left-tail weighted log-scores (Amisano and Giacomini, 2007). These metrics evaluate a model's accuracy in the, in our case, left tail of the distribution. In doing so we implicitly down-weight accurate forecasts around the mean of the target variables, which are GDP growth and inflation.

In Figure 5, we report the decision rules based on left-tail weighted log-scores of GDP growth and inflation in the respective left-hand and right-hand side plots at a 4-quarter horizon from 1993:Q2 to 2017:Q4. Positive values indicate that our benchmark model using Divisia M4 is favourable relative to one of its more narrow counterparts. First, considering GDP growth, there are negligible differences between our benchmark model and those using more narrow definitions of money until the 2008 recession. Then, on the whole, these decision rules indicate choosing our benchmark model until the end of our sample. Turning to inflation, the same conclusion holds. These results are largely consistent with the log Bayes factors we report earlier. However, in this case, the short-run dynamics of model discrepancies change abruptly in favour of using Divisia M4. Specifically, across all model sizes the Divisia M4-based model outperforms at least four of its more narrow counterparts.

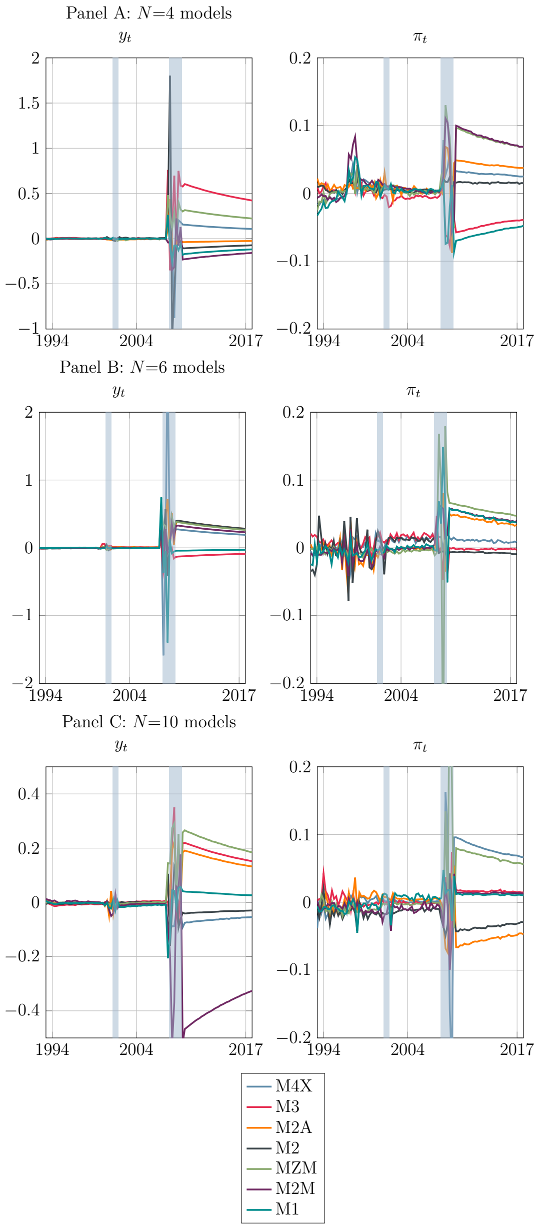

Figure 5: **Giacomini and White (2006) Decision Rules using left-tail weighted log-scores of GDP growth and inflation for log-scores of GDP Growth and Inflation**

This figure reports the h=4 quarter ahead Giacomini and White (2006) decision rules based on weighted log predictive scores (Amisano and Giacomini, 2007) of GDP growth and inflation from 1993:Q2 to 2017:Q4. These plots summarise the predictive accuracy of Divisia M4 relative a narrower alternative using TVP-VAR models. Panel A reports results from N=4 variable TVP-VARs; Panels B and C report results for N=6 and N=10 TVP-VARs. Positive values indicate the forecaster should choose the model using Divisia M4 as it outperforms the narrower alternative in question. The shaded areas correspond to NBER recession dates.

5.3 Assessing the Forecasting Performance of Divisia M4 using Big Data

We now assess the performance Divisia M4 using an even larger information set. Specifically, we use a subset of the McCracken and Ng (2020) FRED-QD database that compiles 248 macroeconomic variables. We remove all series relating to National Income and Product accounts and prices, apart from GDP and CPI inflation, and those with missing data when at the beginning of our sample period in 1967:Q2.

We estimate three extra-large TVP-VAR models in order to gauge the forecasting performance of Divisia M4 for economic activity. Our first model is a N=141 variable TVP-VAR that replaces the four monetary aggregates from our subset of the FRED-QD database with Divisia M4. Our second model is a N=143 variable TVP-VAR with no measures of Divisia money, but with time series relating to money and credit, specifically the monetary base, M1, M2, and MZM reported in the FRED-QD database. Our final model removes the monetary aggregates from the FRED-QD database and adds the seven narrower measures of Divisia, producing in a N=147 variable model. The forecast horizon, number of simulations to generate predictive densities, and lag length are identical to those we use in our main analysis.[5]

Table 7 presents log-scores of each of our three high dimensional models. Panel A reports individual log-scores for GDP growth and inflation, and Panel B shows joint log-scores of GDP growth and inflation. The N=143 and N=147 results are relative to our benchmark high dimensional model that contains Divisia M4 as the monetary aggregate. First, concerning the N=143 variable model, in Panel A, we can see that our benchmark model produces more accurate forecasts of both GDP growth and inflation relative to the N=143 variable model; the same holds true when looking at joint log-scores in Panel B. Turning to the N=147 variable model, the results are less strong. However, we do see benefits from using Divisia M4 from individual and joint log-scores.

Overall, these results show that using the broadest measure of money results in gains in predictive accuracy within a big data application. The gains are strongest at longer horizons and suggest that using the broadest monetary aggregate with theoretically-founded weights outperforms a battery of simple sum measures within the FRED-QD database. When we benchmark Divisia M4 against its seven more narrow alternatives in this big data context, gains are somewhat smaller. However, note that at longer horizons the model using Divisia M4 provides at least as accurate predictions of GDP growth and inflation as the N=147 variable model throughout the forecast sample.

[5] We use a high performance server to run the recursive forecasting exercise for TVP-VARs with $N = \{141, 143, 147\}$ variables. It takes approximately four days to obtain the forecasts from each respective model.

Table 7: **Average Forecast Performance of High Dimensional Models using Divisia M4 versus Different Alternatives**

Notes: Panel A of this table reports the log-scores of GDP growth, y_t, and inflation, π_t, from high dimensional models. Panel B of this table reports the joint log-scores of GDP growth and inflation from high dimensional models. The benchmark model contain N=141 variables and includes the Divisia M4 monetary aggregate. We compare these forecasts against an N=143 variable model that uses the monetary aggregates in the FRED-QD database (McCracken and Ng, 2020) and an N=147 variable model that contains the seven narrow Divisia aggregates as well as a subset of the FRED-QD database. We express the log-scores from these alternative models relative to our benchmark. Positive values indicate indicate the model using Divisia M4 outperforms the alternative comparable model at the $h = \{1, 2, 3, 4\}$ quarter horizons and are in bold font.

Panel A: log-scores of y_t, π_t						
	N=141		N=143 vs. N=141		N=147 vs. N=141	
	y_t	π_t	y_t	π_t	y_t	π_t
h=1	-0.819	-1.194	**0.008**	-0.174	-0.012	-0.155
2	-0.645	-0.590	-0.006	**0.173**	-0.012	**0.040**
3	-0.560	-0.329	**0.001**	**0.044**	-0.002	**0.025**
4	-0.516	-0.307	**0.001**	-0.006	**0.007**	-0.017

Panel B: Joint log-scores of y_t, π_t			
	N=141	N=143 vs. N=141	N=147 vs. N=141
h=1	-1.559	-0.007	-0.013
2	-1.414	**0.032**	-0.086
3	-1.227	**0.045**	**0.099**
4	-0.964	**0.079**	-0.007

5.4 Alternative VAR Models

Here we benchmark the performance of Divisia M4 against narrower measures using alternative VAR models, namely: i) linear Bayesian VAR models with a Minnesota-Normal Wishart Prior, and ii) VAR models following the algorithm in Korobilis and Pettenuzzo (2019) using a Normal Gamma prior with Minnesota shrinkage.[6] The forecast horizon and number of simulations to generate predictive densities are consistent with our main results, as is the number of lags. Tables 8 and 9 report the log-scores for GDP growth and inflation from Bayesian VAR models and VAR models in the spirit of Korobilis and Pettenuzzo (2019) and are read in the same manner as Table 2.

First, considering Table 8, observe that in general Divisia M4 produces at least as, if not more accurate predictions (values of 0.00) of both GDP growth and inflation across all horizons. This result becomes stronger as the dimension of the model increases. Turning to the results in Table 9, the same interpretation holds true. Note that relative gains of Divisia M4 are particularly prominent for inflation across all model sizes. For GDP, there is still evidence that Divisia M4 generates at least as accurate predictions, although these differences in log-scores are on average smaller than those from our main analysis, models in Table 8, and other robustness checks. That notwithstanding, these results indicate that our conclusions are not specific to model choice.

[6] We adapt the code made available on Dimitris Korobilis' website: https://sites.google.com/site/dimitriskorobilis/matlab/ssbcvar.

Table 8: **Average Forecast Performance: Log Predictive Scores from Bayesian VAR models**

Notes: This table reports the log predictive scores of GDP growth, y_t, and CPI inflation, π_t, for VAR models that each use a Divisia measure of money, at $h = \{1, 2, 3, 4\}$ quarter horizons. The forecast sample spans 1990:Q2–2018:Q4. Panel A reports results from N=4 models; Panels B and C report results from N=6 and N=10 variable models respectively. We report the log predictive scores at each horizon for the model using Divisia M4. All other columns report relative log-scores where we benchmark the model using Divisia M4 to each of its more narrow alternatives. M4X denotes Divisia M4 excluding Treasury Bills; M3 is Divisia M3; M2A is Divisia M2 All; M2 is Divisia M2; MZM is Divisia Money Zero-Maturity; M2M is Divisia M2M; and M1 is Divisia M1. Positive values indicate that the model using Divisia M4 produces a higher log-score relative to a model using a more narrow alternative.

Panel A: Small Models, N=4 variables

	M4		M4X		M3		M2A		M2		MZM		M2M		M1	
	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t
h=1	-1.03	-1.27	-0.02	**0.40**	-0.02	**0.12**	0.00	**0.23**	-0.01	**0.51**	**0.01**	**0.32**	0.00	-0.09	**0.03**	**0.11**
2	-0.91	-0.87	-0.05	**0.02**	-0.04	-0.14	-0.01	-0.11	-0.03	-0.07	0.00	**0.16**	-0.01	**0.04**	-0.03	**0.08**
3	-0.79	-0.54	-0.01	**0.04**	-0.02	-0.01	**0.04**	0.00	**0.02**	**0.03**	**0.05**	**0.01**	**0.08**	-0.01	**0.01**	**0.02**
4	-0.78	-0.52	-0.02	**0.01**	-0.03	0.00	**0.04**	0.00	-0.01	0.00	-0.02	0.00	-0.03	**0.01**	**0.04**	**0.01**

Panel B: Medium Models, N=6 variables

	M4		M4X		M3		M2A		M2		MZM		M2M		M1	
	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t
h=1	-1.05	-0.94	-0.02	**0.84**	-0.02	**1.04**	**0.04**	**0.84**	-0.01	**0.45**	-0.01	**0.89**	**0.03**	**0.61**	0.00	**0.87**
2	-0.97	-0.83	**0.11**	0.12	**0.02**	**0.26**	**0.07**	**0.27**	-0.04	**0.09**	-0.01	**0.08**	-0.02	**0.28**	**0.12**	**0.09**
3	-0.89	-0.59	**0.10**	-0.01	**0.08**	**0.03**	**0.09**	**0.02**	**0.16**	-0.02	**0.24**	-0.02	**0.19**	-0.01	**0.22**	**0.01**
4	-0.94	-0.56	**0.02**	0.00	**0.05**	0.00	**0.10**	**0.01**	-0.04	**0.01**	-0.03	0.00	**0.07**	0.00	-0.05	0.00

Panel C: Large Models, N=10 variables

	M4		M4X		M3		M2A		M2		MZM		M2M		M1	
	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t
h=1	-1.01	-1.10	0.00	**0.38**	0.00	**0.35**	**0.02**	**0.59**	**0.01**	**0.36**	**0.01**	**0.63**	**0.03**	**0.29**	**0.02**	**0.40**
2	-0.95	-0.83	-0.10	-0.07	-0.02	**0.18**	-0.07	**0.09**	-0.04	**0.11**	-0.03	**0.15**	0.00	**0.03**	-0.05	**0.01**
3	-0.85	-0.63	-0.03	**0.03**	**0.07**	**0.02**	**0.02**	**0.02**	**0.06**	-0.01	**0.05**	**0.01**	**0.02**	0.00	**0.02**	0.00
4	-0.79	-0.62	0.00	0.00	**0.07**	0.00	**0.07**	0.00	**0.16**	0.01	**0.02**	0.00	**0.02**	**0.01**	**0.07**	0.00

Table 9: **Average Forecast Performance: Log Predictive Scores from Korobilis and Pettenuzzo (2019) VAR models**

Notes: This table reports the log predictive scores of GDP growth, y_t, and CPI inflation, π_t, for VAR models that each use a Divisia measure of money, at $h = \{1, 2, 3, 4\}$ quarter horizons. The forecast sample spans 1990:Q2–2018:Q4. Panel A reports results from N=4 models; Panels B and C report results from N=6 and N=10 variable models respectively. We report the log predictive scores at each horizon for the model using Divisia M4. All other columns report relative log-scores where we benchmark the model using Divisia M4 to each of its more narrow alternatives. M4X denotes Divisia M4 excluding Treasury Bills; M3 is Divisia M3; M2A is Divisia M2 All; M2 is Divisia M2; MZM is Divisia Money Zero-Maturity; M2M is Divisia M2M; and M1 is Divisia M1. Positive values indicate that the model using Divisia M4 produces a higher log-score relative to a model using a more narrow alternative.

Panel A: Small Models, N=4 variables

	M4		M4X		M3		M2A		M2		MZM		M2M		M1	
	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t
h=1	3.397	3.465	**0.013**	**0.008**	-0.002	**0.006**	-0.031	**0.003**	-0.032	**0.013**	-0.027	**0.016**	-0.026	0.000	**0.009**	**0.004**
2	3.529	3.519	**0.014**	**0.006**	-0.009	-0.001	-0.037	**0.003**	-0.030	**0.001**	-0.031	**0.001**	-0.036	**0.002**	**0.012**	**0.001**
3	3.582	3.557	**0.008**	**0.001**	-0.012	-0.003	-0.042	**0.001**	-0.040	-0.001	-0.037	**0.003**	-0.039	-0.001	**0.001**	**0.007**
4	3.616	3.581	**0.014**	-0.001	-0.008	-0.004	-0.035	-0.002	-0.045	-0.004	-0.040	**0.001**	-0.039	0.000	**0.006**	**0.005**

Panel B: Medium Models, N=6 variables

	M4		M4X		M3		M2A		M2		MZM		M2M		M1	
	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t
h=1	3.418	3.469	**0.011**	**0.005**	-0.002	**0.016**	-0.031	**0.012**	-0.028	**0.018**	-0.021	**0.015**	-0.019	**0.005**	**0.014**	**0.010**
2	3.555	3.522	**0.013**	**0.001**	-0.006	0.004	-0.033	**0.008**	-0.028	**0.003**	-0.027	**0.007**	-0.024	**0.005**	**0.009**	**0.006**
3	3.617	3.557	**0.014**	-0.004	-0.005	**0.007**	-0.038	**0.007**	-0.028	**0.001**	-0.027	0.000	-0.027	**0.008**	**0.009**	**0.002**
4	3.656	3.581	**0.013**	-0.005	-0.002	**0.007**	-0.037	**0.007**	-0.028	-0.001	-0.031	-0.002	-0.028	**0.008**	**0.008**	**0.004**

Panel C: Large Models, N=10 variables

	M4		M4X		M3		M2A		M2		MZM		M2M		M1	
	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t	y_t	π_t
h=1	3.422	3.839	**0.012**	**0.001**	-0.002	**0.005**	-0.022	**0.003**	-0.031	**0.004**	-0.027	**0.001**	-0.028	**0.004**	**0.015**	**0.008**
2	3.557	3.863	**0.010**	-0.001	-0.005	**0.003**	-0.029	**0.004**	-0.033	**0.006**	-0.034	**0.001**	-0.027	**0.007**	**0.009**	**0.005**
3	3.623	3.874	**0.013**	-0.002	-0.006	-0.001	-0.026	**0.002**	-0.026	**0.003**	-0.026	0.000	-0.023	**0.003**	**0.014**	**0.003**
4	3.656	3.882	**0.014**	-0.003	-0.003	-0.001	-0.020	**0.001**	-0.032	0.000	-0.029	**0.003**	-0.023	**0.001**	**0.007**	**0.002**

5.5 Pooling Forecasting Models: Alternative Inflation Measures

We now assess the robustness of our main findings by considering two alternative measures of inflation, namely GDP deflator inflation and personal consumption expenditures (PCE) inflation. Our focus within this exercise is on joint log-scores from models using Divisia M4 relative to pooled information.[7]

Table 10 reports average relative joint log-scores over the forecast sample at a 4-quarter horizon for GDP growth and inflation and is read in a similar way to Table 5. Panels A and B report results using optimal and equal weighting schemes when pooling forecast information, respectively. We report these statistics using GDP deflator inflation and PCE inflation under the headings GDP Def. and PCE, respectively. Subheadings 1 and 2 refer to our benchmark model in comparison to $\mathcal{M} = 7$ models, and $\mathcal{M} = 8$ models, respectively. Overall, our benchmark models using Divisia M4 outperform both cases of pooled forecasts across all model sizes.

Next, we consider log Bayes factors based on joint log-scores at a 4-quarter horizon from models using GDP deflator inflation and PCE inflation in Figures 6 and 7, respectively. When compared with Figure 3, it is clear that these results are consistent with our baseline analysis. We observe a gradual increase in these statistics with surges prior to the 2001 and the 2008 recessions, particularly when pooling information using optimal weighting schemes. Turning to results using equal weighting schemes, as with our baseline results, there is a gradual increase throughout the forecast sample. Overall, these extensions indicate no differences in the forecasting performance of Divisia M4 relative to different inflation measures. Also, our conclusions regarding GDP growth do not change in light of alternative inflation measures.

[7] All other results are are consistent with our main findings and are available upon request.

Table 10: **Average Forecast Performance: Relative Joint Log Predictive Scores for GDP growth and Inflation Benchmarked against Pooled forecasts using Alternative Measures on Inflation**

Notes: This table reports the joint log predictive scores of GDP growth, y_t and CPI inflation π_t from TVP-VAR models using Divisia M4 at $h = 4$ quarter horizon relative to forecasts pooled from TVP-VARs using all narrower measures of Divisia. Panel A reports relative log predictive scores where the pooling information uses the optimal weighting scheme of Geweke and Amisano (2011), and Panel B assigns equal weights when pooling forecasts. Results under the headings GDP def. and PCE are from models using GDP deflator inflation, and Personal Consumption Expenditure inflation respectively. Results under subheading 1, benchmark the forecasts from the TVP-VAR using Divisia M4 against pooled information from the remaining seven narrow measures of money. Results under subheading 2 benchmark Divisia M4 against pooled forecasts from models using all measures of money, including Divisia M4 itself. The narrower alternative measures are: Divisia M4X which is Divisia M4 excluding Treasury Bills; Divisia M3; Divisia M2 All; Divisia M2; Divisia Money Zero-Maturity; Divisia M2M; and Divisia M1. Positive values indicate that the forecast stemming from the model using Divisia M4 produces a higher joint log-score relative to pooled forecasts.

Panel A: Optimal weighting scheme				
	GDP Def.		**PCE**	
	1	**2**	**1**	**2**
Small Models, N=4	**1.73**	**1.73**	**7.07**	**7.72**
Medium Models N=6	**2.09**	**2.31**	**9.11**	**9.10**
Large Models, N=10	**2.48**	**1.46**	**9.36**	**9.28**

Panel B: Equal weighting scheme				
	GDP Def.		**PCE**	
	1	**2**	**1**	**2**
Small Models, N=4	**13.60**	**14.86**	**25.79**	**27.32**
Medium Models N=6	**13.87**	**14.99**	**27.38**	**28.19**
Large Models, N=10	**13.32**	**14.48**	**27.09**	**27.77**

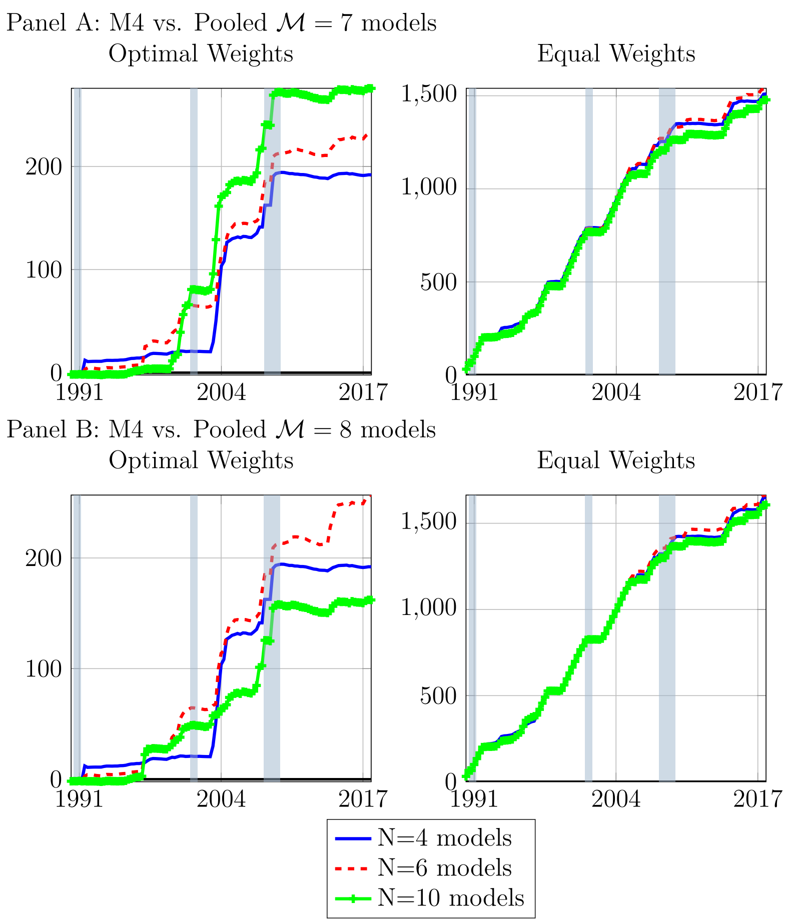

Figure 6: **Log Bayes Factors for Joint log-scores of GDP Growth and Inflation: Divisia M4 vs. Pooled Information: Models using GDP Deflator Inflation**
This figure reports the h=4 quarter ahead log Bayes factors in Geweke and Amisano (2010) for joint log predictive scores of GDP growth and inflation from 1990:Q2 to 2017:Q4. These plots summarise the difference between cumulative joint log-scores of the TVP-VAR using Divisia M4 and log-scores pooled from a set of $\mathcal{M} = \{7, 8\}$ models. Panel A reports log Bayes factors based on joint log-scores from the model using Divisia M4 relative to pooling information from the models using the 7 narrower alternatives. Panel B reports the log Bayes factors based on joint log-scores from the model using Divisia M4 relative to pooling information from all 8 models that each contain a measure of Divisia money. Positive values indicate the overall predictive performance of the model using a Divisia M4 outperforms the pooled information. Solid lines refer to N=4 variable TVP-VARs; dashed lines refer to N=6 variable TVP-VARs; and solid dotted lines refer to N=10 variable models. The shaded areas correspond to NBER recession dates.

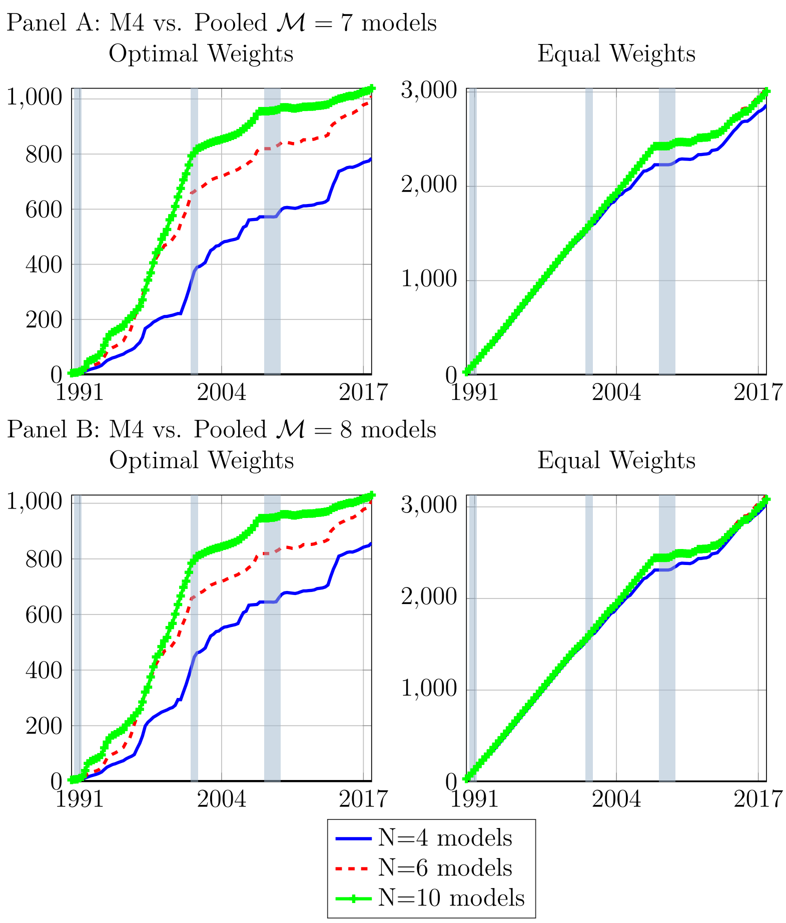

Figure 7: **Log Bayes Factors for Joint log-scores of GDP Growth and Inflation: Divisia M4 vs. Pooled Information: Models using PCE Inflation**
This figure reports the h=4 quarter ahead log Bayes factors in Geweke and Amisano (2010) for joint log predictive scores of GDP growth and inflation from 1990:Q2 to 2017:Q4. These plots summarise the difference between cumulative joint log-scores of the TVP-VAR using Divisia M4 and log-scores pooled from a set of $\mathcal{M} = \{7, 8\}$ models. Panel A reports log Bayes factors based on joint log-scores from the model using Divisia M4 relative to pooling information from the models using the 7 narrower alternatives. Panel B reports the log Bayes factors based on joint log-scores from the model using Divisia M4 relative to pooling information from all 8 models that each contain a measure of Divisia money. Positive values indicate the overall predictive performance of the model using a Divisia M4 outperforms the pooled information. Solid lines refer to N=4 variable TVP-VARs; dashed lines refer to N=6 variable TVP-VARs; and solid dotted lines refer to N=10 variable models. The shaded areas correspond to NBER recession dates.

6 Conclusions

Based on robust theoretical foundations, Divisia monetary aggregates overcome many of the shortcomings of simple-sum measures of money and are shown to be an empirically more useful measure of money supply. In this paper, we examine whether the use of the broadest, M4 Divisia monetary aggregate improves the predictive power of forecasts of US GDP growth and inflation relative to a number of alternative, narrower measures of money. To that end, we assess entire predictive densities obtained from a number of time-varying parameter vector autoregressive models, estimated using the non-parametric methods outlined in (Petrova, 2019) which do not require an ex-ante specified law of motion and are thus less susceptible to misspecification errors.

Our key results provide robust evidence of superior out-of-sample forecasting performance of models using Divisia M4 relative to alternative specifications using seven different, narrower Divisia measures of money, as well as to forecasts obtained from pooling predictions from those alternative models.

The reliability of our findings is further confirmed in a battery of robustness tests and extensions. Our results demonstrate that models featuring Divisia M4 perform better than comparable models which do not feature any measure of money. We also document the improvements in accuracy of forecasts of economic activity stemming from the use of time-varying parameter models relative to constant-parameter models. Finally, we assess Divisia M4's forecasting performance in a large-scale exercise utilising three TVP-VAR models with more than 140 macroeconomic and financial variables, and confirm its usefulness, especially at longer forecast horizons.

Our findings lend strong empirical support in favour of a monetary approach to forecasting economic activity, which ought to utilise the superior performance of the broad Divisia M4 monetary aggregate over its narrower counterparts. As such, the results we present in this study are of significant importance to both fiscal and monetary policymakers, commercial organisations, and individual economic agents whose actions depend on accurate and reliable predictions about the future state of the economy.

References

Albuquerque, B., Baumann, U., Seitz, F., 2015. The Information Content of Money and Credit for US Activity. ECB Working Paper No. 1803 .

Alessandri, P., Mumtaz, H., 2017. Financial Conditions and Density Forecasts for US Output and Inflation. Review of Economic Dynamics 24, 66–78.

Amisano, G., Giacomini, R., 2007. Comparing Density Forecasts via Weighted Likelihood Ratio Tests. Journal of Business & Economic Statistics 25, 177–190.

Barnett, W.A., 1980. Economic Monetary Aggregates an Application of Index Number and Aggregation Theory. Journal of Econometrics 14, 11–48.

Barnett, W.A., 2012. Getting It Wrong: How Faulty Monetary Statistics Undermine the Fed, the Financial System, and the Economy. MIT press.

Barnett, W.A., Chauvet, M., 2011. How Better Monetary Statistics could have Signaled the Financial Crisis. Journal of Econometrics 161, 6–23.

Barnett, W.A., Chauvet, M., Leiva-Leon, D., 2016. Real-time Nowcasting of Nominal GDP with Structural Breaks. Journal of Econometrics 191, 312–324.

Barnett, W.A., Liu, J., Mattson, R.S., van den Noort, J., 2013. The new CFS Divisia monetary aggregates: Design, construction, and data sources. Open Economies Review 24, 101–124.

Belongia, M.T., 1996. Measurement Matters: Recent Results from Monetary Economics Reexamined. Journal of Political Economy 104, 1065–1083.

Belongia, M.T., Chrystal, K.A., 1991. An Admissible Monetary Aggregate for the United Kingdom. Review of Economics and Statistics 73, 497–503.

Belongia, M.T., Ireland, P.N., 2015. Interest Rates and Money in the Measurement of Monetary Policy. Journal of Business & Economic Statistics 33, 255–269.

Belongia, M.T., Ireland, P.N., 2016. Money and Output: Friedman and Schwartz Revisited. Journal of Money, Credit and Banking 48, 1223–1266.

Benati, L., 2008. The "Great Moderation" in the United Kingdom. Journal of Money, Credit and Banking 40, 121–147.

Benati, L., Mumtaz, H., 2007. US Evolving Macroeconomic Dynamics: A Structural Investigation. ECB Working Paper No. 746 .

Bernanke, B.S., Blinder, A.S., 1992. The Federal Funds Rate and the Channels of Monetary Transmission. American Economic Review 82, 901–921.

Bianchi, F., Mumtaz, H., Surico, P., 2009. The Great Moderation of the Term Structure of UK Interest Rates. Journal of Monetary Economics 56, 856–871.

Canova, F., Gambetti, L., Pappa, E., 2007. The Structural Dynamics of Output Growth and Inflation: Some International Evidence. The Economic Journal 117.

Carriero, A., Kapetanios, G., Marcellino, M., 2009. Forecasting exchange rates with a large Bayesian VAR. International Journal of Forecasting 25, 400 – 417.

Carriero, A., Kapetanios, G., Marcellino, M., 2012. Forecasting government bond yields with large Bayesian vector autoregressions. Journal of Banking & Finance 36, 2026 – 2047.

Cogley, T., Sargent, T.J., 2005. Drifts and Volatilities: Monetary Policies and Outcomes in the post WWII US. Review of Economic Dynamics 8, 262–302.

Cross, J.L., Hou, C., Poon, A., 2020. Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. International Journal of Forecasting 36, 899 – 915.

Ellington, M., 2018. The Case for Divisia Monetary Statistics: A Bayesian Time-varying Approach. Journal of Economic Dynamics and Control 96, 26–41.

Friedman, B.M., Kuttner, K.N., 1992. Money, Income, Prices, and Interest Rates. American Economic Review 82, 472–492.

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

Geweke, J., Amisano, G., 2011. Optimal Prediction Pools. Journal of Econometrics 164, 130–141.

Giacomini, R., White, H., 2006. Tests of Conditional Predictive Ability. Econometrica 74, 1545–1578.

Hendrickson, J.R., 2014. Redundancy or Mismeasurement? A Reappraisal of Money. Macroeconomic Dynamics 18, 1437–1465.

Huber, F., Koop, G., Onorante, L., 2020. Inducing Sparsity and Shrinkage in Time-Varying Parameter Models. Journal of Business & Economic Statistics In press.

Kadiyala, K.R., Karlsson, S., 1997. Numerical methods for estimation and inference in Bayesian VAR-models. Journal of Applied Econometrics 12, 99–132.

Keating, J.W., Kelly, L.J., Smith, A.L., Valcarcel, V.J., 2019. A Model of Monetary Policy Shocks for Financial Crises and Normal Conditions. Journal of Money, Credit and Banking 51, 227–259.

Keating, J.W., Kelly, L.J., Valcarcel, V.J., 2014. Solving the Price Puzzle with an Alternative Indicator of Monetary Policy. Economics Letters 124, 188–194.

Koop, G., Korobilis, D., 2013. Large time-varying parameter VARs. Journal of Econometrics 177, 185 – 198.

Korobilis, D., Pettenuzzo, D., 2019. Adaptive hierarchical priors for high-dimensional vector autoregressions. Journal of Econometrics 212, 241–271.

McCracken, M., Ng, S., 2020. FRED-QD: A Quarterly Database for Macroeconomic Research. NBER Working Paper, No 26872. Available: https://www.nber.org/papers/w26872 .

Petrova, K., 2019. A Quasi-Bayesian Local Likelihood Approach to Time Varying Parameter VAR Models. Journal of Econometrics 212, 286–306.

Primiceri, G.E., 2005. Time Varying Structural Vector Autoregressions and Monetary Policy. Review of Economic Studies 72, 821–852.

Schunk, D.L., 2001. The Relative Forecasting Performance of the Divisia and Simple Sum Monetary Aggregates. Journal of Money, Credit and Banking 33, 272–283.

Taylor, J.B., 1993. Discretion Versus Policy Rules in Practice, in: Carnegie-Rochester conference series on public policy, Elsevier. pp. 195–214.

Wu, J.C., Xia, F.D., 2016. Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. Journal of Money, Credit and Banking 48, 253–291.

Appendix A: Data Definitions and Sources

Variables, Sources, and Definitions

Variable	**Source/Code**	**Definition**
GDP	FRED, GDPC1	Quarterly growth rate
CPI inflation	FRED, CPIAUSCL	Quarterly growth rate
GDP deflator inflation	FRED, GDPDEF	Quarterly growth rate
PCE inflation	FRED, PCEPI	Quarterly growth rate
Federal Funds/shadow rate	FRED, DFF and Jing Cynthia Wu's website	Spliced Federal Funds rate with Jing Cynthia Wu's shadow rate during ZLB period.
Wages	FRED, A576RC1	Quarterly growth rate
Unemployment	FRED, UNRATE	Raw, rate
Oil price	FRED, WTISPLC	Quarterly growth rate
Corporate Bond Spread	FRED, DBAA and DAAA/Moody's	Moody's Baa corporate bond yield minus Moody's Aaa corporate bond yield.
S&P500 composite stock index	WRDS	Quarterly growth rate
House prices	FRED/Robert Shiller's website	Quarterly growth rate
Divisia M1	Center for Financial Stability	Quarterly growth rate
Divisia M2M	Center for Financial Stability	Quarterly growth rate
Divisia MZM	Center for Financial Stability	Quarterly growth rate
Divisia M2	Center for Financial Stability	Quarterly growth rate
Divisia M2A	Center for Financial Stability	Quarterly growth rate
Divisia M3	Center for Financial Stability	Quarterly growth rate
Divisia M4X	Center for Financial Stability	Quarterly growth rate
Divisia M4X	Center for Financial Stability	Quarterly growth rate

Appendix B: Forecast Evaluation Strategies

Log Bayes Factors

Given a pair of models (A, B) and a variable x_t, the log Bayes factors at period t for forecast horizon h is

$$\text{BF}_{t,h}^{A,B} = \sum_{\tau=1}^{t} \left[\ln(LS_{\tau,h}^{A}) - \ln(LS_{\tau,h}^{B})\right]$$

where $LS_{\tau,h}^{i} = p(x_{\tau}^{o}|Y_{\tau-1})$ is the predictive density from model i for x_t that we evaluate at the actual observation x_t^o. For log Bayes factors that we calculate on joint log-scores, we replace $LS_{\tau,h}^{i}$ with the corresponding joint log-score $JLS_{\tau,h}^{i}$.

Giacomini and White (2006) Decision Rules

Using Giacomini and White (2006), and following Alessandri and Mumtaz (2017) we construct decision rules in a pairwise manner. Given two forecasts for y_{t+h}, f_t and g_t, as well as a loss function L_{t+h} we can write

$$H_0 : \mathbb{E}_t \left[L_{t+h}(Y_{t+h}, f_t) - L_{t+h}(Y_{t+h}, g_t)\right] \equiv \mathbb{E}_t \left[\Delta L_{m,t+h}|I_t\right] = 0$$

where I_t is information set available at time t. In our case, we calculate decision rules based on left-tail weighted log-scores (Amisano and Giacomini, 2007). We exploit any persistence in $\Delta L_{m,t}$ to set up a decision rule that tells us for every t the expectation of which model should perform better in period $t+h$. For each window $[1, ..., T_0 + k - h]$, where T_0=1993Q2 and $k = h, ..., T - h$, we estimate the following regression

$$\Delta L_{m,t+h} = \delta^{'}[1 \;\; \Delta L_{m,t}] + \epsilon_t$$

where δ is a coefficient vector that we use to calculate the decision criterion $\mathcal{C}_t = \hat{\delta}_k^{'}[1 \;\; \Delta L_{m,t}]$ at each of the k windows. The decision criteria is as follows: if $\mathcal{C}_t < 0$ ($\mathcal{C}_t > 0$) choose model f_t (g_t).