Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Department Socioeconomics

Forecasting growth of U.S. aggregate and household-sector M2 after 2000 using economic uncertainty measures

Artur Tarassow

DEP (Socioeconomics) Discussion Papers
Macroeconomics and Finance Series
2/2017

Hamburg, 2017

Forecasting growth of U.S. aggregate and household-sector M2 after 2000 using economic uncertainty measures*

Artur Tarassow
Faculty of Economics and Social Sciences
University of Hamburg
artur.tarassow@wiso.uni-hamburg.de
+49-40-42838-8683

August 29, 2017

Abstract

This paper evaluates the predictive out-of-sample forecasting properties of six different economic uncertainty variables for both growth in aggregate M2 and growth in household-sector M2 in the U.S. using data between 1971m1 and 2014m12. The core contention is that economic uncertainty improves both forecast accuracy as well as direction-of-change forecasts of real money stock growth. We estimate linear ARDL models using the iterated rolling-window forecasting scheme combined with two different indicator selection procedures. Forecast accuracy is evaluated by RMSE and the Diebold-Mariano test. Direction-of-change forecasts are assessed by means of the Kuipers Score and the Pesaran-Timmermann test. The results indicate an increased relevance of certain economic uncertainty measures for forecasting growth in both real aggregate as well as real household-sector M2 since 2000.

JEL Classifications: C22; E41; E47

Key Words: money demand, uncertainty, risk, multi-step forecasts, forecast comparison

*Address for correspondence: Faculty of Economics and Social Sciences, University of Hamburg, Welckerstraße 8, Hamburg, 20354, Germany. Email: artur.tarassow@wiso.uni-hamburg.de. Tel: +49-40-42838-8683. We are grateful to Matthew Greenwood-Nimmo, Thomas Theobald, Efrem Castelnuovo, Tomasz Woźniak, Barry Rafferty and Patrick Harms for helpful comments. Any remaining errors or omissions are strictly our own.

1 Introduction

In this article we seek to analyse whether economic uncertainty provides predictive power to out-of-sample forecast growth of the U.S. real monetary aggregate M2 as well as real U.S. household-sector M2. Even though monetary aggregates have not played a central role in the formulation of U.S. monetary policy since the 1980s, policymakers continue to use monetary data as a source of information about the state of the economy. For instance, money demand models are still estimated by policymakers to predict long-run inflation trends (Bernanke, 2006). Furthermore, within the New Keynesian camp, money can serve the purpose of an additional indicator of the monetary conditions prevailing in the market as it may provide timely information about variables that are measured imperfectly (Coenen et al., 2005; Beck and Wieland, 2007). Others have shown that targeting money growth leads to improved economic outcomes in standard New Keynesian models when monetary policy acts under discretion (Söderström, 2005). Hence, there are good reasons to think that policy makers have a deep interest in accurate money forecasts.

The reason why monetary aggregates have not played a more prominent role in U.S. monetary policy is the recurrent instability in the (long-run) relationship between various monetary aggregates and other nominal variables. Among the potential factors leading to this instability, economists see, for instance, institutional deregulation and financial innovation. However, others stress the relevance of different types of economic uncertainty or risk as relevant determinants of money demand (Carpenter and Lange, 2003). This latter aspect is central to our study.

The recent great financial crisis was accompanied by a regained interest in private actors' liquidity preferences. Studies on money demand indicate that U.S. households are risk-averse against volatility. Risk-aversion implies that the demand for liquid and safe assets moves procyclically with changes in, for instance, capital market risk, inflation risk or macroeconomic risk. These types of risk or uncertainty[1] have increased—at least temporarily—as a consequence of the recent great financial crisis (GFC henceforth) as well as unconventional monetary policy which has led to money growth rates seen as incompatible with stable price inflation and increased stock market volatility (Baker et al., 2013; Jurado et al., 2015; Ludvigson et al., 2015).

The literature discusses some channels through which economic uncertainty may affect households' money holdings. For instance, as the cost of investing in stocks and bonds has declined and households hold broader sets of monetary assets, there are reasons to believe that money holdings may have become more sensitive to financial as well as inflation risk (Cook and Choi, 2007). Assuming that the money-growth-to-inflation nexus remains intact, an inflation-targeting central bank needs to monitor financial and inflation risk to future inflation. This provides another argument for the inclusion of financial stability measures into a central bank's objective function, as the stabilization of financial markets can be seen as an additional pillar for ensuring

[1] In the literature one typically distinguishes between risk and uncertainty. *Risk* typically refers to the odds of an outcome when the probability distribution is known, the term *uncertainty* describes the case when the probability distribution of the data generating process is unknown. Hence, uncertainty is defined as the conditional volatility of a disturbance that is unforecastable from the perspective of economic agents (Ludvigson et al., 2015). However, in this paper we will explicitly not distinguish between risk and uncertainty as this distinction is not of importance for our forecast analysis.

price stability (Cronin et al., 2011). Another line of argument emphasizes the increased risk of long periods of secular stagnation in advanced countries accompanied by a high preference for liquidity as a major cause (Caballero and Farhi, 2013; Bossone, 2014). Lastly, Benhabib et al. (2001) and Christiano and Rostagno (2001) argue that the stabilizing Taylor principle may break down if money growth rates exceed some target rate. In such a case, the central bank needs to adjust money supply for re-stabilisation which requires a close monitoring of actual and expected money growth rates. For a central bank which has a particular obligation to increased employment and sustain price stability, these concerns require the need to forecast growth of monetary aggregates.

While the empirical money demand literature has been mainly concerned with the question whether the dynamics in economic uncertainty measures help to explain the breakdown of the standard long-run money demand relationship consisting of a monetary aggregate, an income and opportunity cost measure (see e.g. Carpenter and Lange, 2003), there has been no extensive work on the out-of-sample predictive power of economic uncertainty for forecasting the dynamics of monetary aggregates.

This paper attempts to study the out-of-sample predictive content of various economic uncertainty measures for growth of both real U.S. aggregate and household-sector demand for M2 using data between 1971m1 and 2014m12. The focus of our analysis, however, will be on more recent forecasting dynamics since the year 2000 including the New Economy Bust as well as the recent GFC episode. To do so, we evaluate the pseudo out-of-sample forecasting power of six different uncertainty measures using a linear ARDL model. Our study makes three main contributions to the literature:

(i) Even though many papers were published on the relevance of various uncertainty variables for re-establishing a stable long-run money demand relationship, we are not aware of any article studying systematically the out-of-sample forecasting properties of economic uncertainty variables for growth of a monetary aggregate in the U.S.

(ii) Instead of studying the effect of a single uncertainty variable, we will evaluate the out-of-sample forecasting properties of the following six uncertainty variables among which some have been used in the money demand literature before while others are rather known from recent business cycle studies. To be more concrete, we evaluate the forecasting power of inflation uncertainty (Stock and Watson, 2007; Wright, 2011), the stock market risk premium (Fama and French, 1988), the VXO implied volatility measure for the S&P 100 (Bloom, 2009), the financial market uncertainty index (Ludvigson et al., 2015), the macroeconomic uncertainty index (Jurado et al., 2015) and the economic policy uncertainty index (Baker et al., 2013).

(iii) Apart from out-of-sample forecast accuracy, we also evaluate out-of-sample direction-of-change forecasts at different horizons by means of the Kuipers score and Pesaran-Timmermann test. Furthermore, we run rolling-window forecast schemes combined with two different recursive indicator selection procedures.

Our main findings support the claim that the consideration of certain economic uncertainty variables helps to improve forecasting monetary aggregates:

(i) The benchmark ARDL(12,12,12) model of money demand including lags of growth of a real monetary measure, growth of a real income measure and an opportunity cost measure combined with recursive indicator selection provides good out-of-sample forecast accuracy (measured by the root mean squared forecast error) at the short- and medium-term horizon for growth of *aggregate* M2 for the period between 2000m1 and 2014m12. However, the implied volatility VXO financial market uncertainty series as well as the macroeconomic uncertainty measure provide additional and relevant forecasting information. These results are also confirmed for the U.S. aggregate Divisia M4 money stock measure.

(ii) With regard to forecasting growth of *household-sector* M2, the benchmark ARDL(12,12,12) model combined with automatic indicator selection provides reasonable out-of-sample forecast accuracy at the short- and medium-term horizon for the period between 2000m1 and 2014m12. However, inflation uncertainty series, economic policy uncertainty and simple equally-weighted model combination helps to improve the point forecast accuracy.

(iii) The standard money demand model yields poor direction-of-change forecasts for growth of *aggregate* M2 as measured by the Kuipers Score and the Pesaran-Timmermann test on market timing between 2000m1 and 2014m12. Inflation uncertainty, financial market uncertainty and economic policy uncertainty provide statistically relevant information for short- and medium-term direction-of-change forecasts, respectively. Overall, direction-of-change forecasting monetary growth of *aggregate* M2 has become much harder since 2000.

(iv) Direction-of-change forecasting growth of *household-sector* M2 has become much harder since 2000. While the standard money demand model yields good short-term out-of-sample direction-of-change forecasts, considering the inflation uncertainty measure, the economic policy measure, the stock market risk premium or the financial market uncertainty series, respectively, helps to improve this type of forecast considerably—especially for the period since 2007m1 covering the recent GFC period.

Especially the effects of inflation risk, income risk and capital market risk on money holdings can be understood against the background of money as a store of value in intertemporal household optimization models (Größl and Fritsche, 2006; Größl and Tarassow, 2015). While macroeconomic uncertainty as well as economic policy uncertainty is correlated with future and expected nominal income, inflation uncertainty is linked to expected real income as well as expected real returns on investment. This latter argument plays a crucial role for households' pension plans and their portfolio allocations. Also financial market uncertainties are expected to have repercussions on allocating resources for pension plans.

As mentioned before, the empirical money demand literature is mainly concerned with the in-sample fit of otherwise standard money demand models which are augmented by some uncertainty variable. For instance, Carpenter and Lange (2003) and Cook and Choi (2007) use the implied volatility VXO measure from the S&P 100 as an additional regressor in their money demand model for the U.S. Cook and Choi (2007) use the stock market risk premium (see also Größl and Tarassow, 2015), a measure of liquidity risk and the corporate spread, respectively, as a potential measure of uncertainty in a cointegrating framework. Among others, Carstensen (2006) finds a significant long-run impact of stock market volatility on euro area money demand. Greiber and Lemke (2005) show that the consideration of uncertainty factors improves the statistical fit of the model in both the euro area as well as U.S. Also de Bondt (2009) presents evidence for the relevance of precautionary motives, stemming from the labour market, for money demand holdings in the euro area (see Seitz and von Landesberger, 2014, in a similar vein). The effect of macroeconomic as well as monetary uncertainty on U.S. money demand dynamics was studied by Cronin et al. (2011). Lastly, Higgins and Majin (2009) and Größl and Tarassow (2015) examined the role of inflation uncertainty for money demand in the U.S.

The topic of uncertainty has recently gained great interest since the seminal work of Bloom (2009) for business cycle analysis. This large growing body of literature on uncertainty can be divided into three branches. One branch of the literature has been concerned with the construction of uncertainty series (Jurado et al., 2015; Ludvigson et al., 2015; Rossi et al., 2016). Another strand has discussed the issue of whether uncertainty causes an economic outcome (e.g. a recession) or is actually a consequence of this outcome (Bachmann et al., 2013; Ludvigson et al., 2015). Several studies on the repercussions of uncertainty shocks on various economic variables were published during the last couple of years. A common finding is that uncertainty proxies are strongly countercyclical even though the macroeconomic impact of the various uncertainty measures can be very different from each other (Rossi et al., 2016). For instance, a countercyclical relationship between real activity and stock market volatility was found by Bloom (2009). The dispersion in firm-level earnings, industry-level earnings, total factor productivity, and the predictions of forecasters have been used as other proxies for uncertainty (Bloom et al., 2013). Others have used disagreement from the Survey of Professional Forecasters and analyst uncertainty (D'Amico and Orphanides, 2008). Wright (2011) emphasized the role of inflation uncertainty for predicting the term structure, and finds a positive relationship between long-term inflation uncertainty and bond term premium in a large cross-section of countries. Stock and Watson (2012) argue that uncertainty shocks and liquidity/risk shocks are highly correlated. They find that the largest negative shock contributing to the recession period 2007q4–2009q2 are due to both liquidity/risk and uncertainty shocks.

The paper is structured as follows. The next section introduces the forecasting model and statistics. Section 3 discusses the model specifications and presents the time-series used. Section 4 and 5 present the empirical forecasting results for both aggregate M2 and households sector M2, respectively. Section 6 concludes.

2 Forecasting models and statistics

2.1 The ARDL model and pseudo out-of-sample forecasting

We follow the original work by Diebold and Mariano (1995), Clark and McCracken (2001) and others by comparing models according to their forecasting performance in a pseudo out-of-sample forecasting environment. Our forecasting models are built on an autoregressive distributed lag (ARDL) model for out-of-sample forecasting money growth. This framework is well-known and is expected to be more robust to potential misspecification errors compared to a VAR model as noted by Stock and Watson (2003, p. 791).

In general the h-step ahead linear money demand forecasting regression can be written

$$\Delta m^h_{t+h} = \alpha_0 + \sum_{j=1}^{p} \beta_j \Delta m_{t+1-j} + \sum_{j=1}^{q} \theta_j \Delta y_{t+1-j} + \sum_{j=1}^{k} \phi_j i_{t+1-j} + \sum_{j=1}^{l} \rho_j U_{t+1-j} + e^h_{t+h} \quad (1)$$

where Δm, Δy, i, U and e denote the logarithmic (log-)change in real money, the log-change in real income, an opportunity cost measure, some uncertainty measure and the h-step ahead forecast error. The h-step ahead forecast of growth in real money is a linear combination of its own lagged values as well as lagged values of the remaining regressors.

The h-multi-step ahead forecast will be computed by means of the iterated (or plug-in) forecast method. Suppose we want to forecast Δm for period t using a dynamic model, say ARDL(1) for example. If we have data on Δm available only up to period $t-2$, we can apply the chain rule of forecasting

$$\begin{aligned} \widehat{\Delta m}_{t-1} &= \widehat{\alpha}_0 + \widehat{\beta}_1 \Delta m_{t-2} + \widehat{\beta}_2 X_{t-2} \\ \widehat{\Delta m}_{t} &= \widehat{\alpha}_0 + \widehat{\beta}_1 \widehat{\Delta m}_{t-1} + \widehat{\beta}_2 X_{t-1} \end{aligned}$$

where we always have access to the historical realisations of the exogenous regressor X. Hence, as in Rapach and Strauss (2008), we do not model a separate data-generating process for y_t, i_t and U_t in eq. 1 but use their historical realisations for computing the iterated forecast for m_t, instead. Marcellino et al. (2006) have shown that iterated forecasts typically outperform the direct forecasts. Furthermore, the relative performance of the iterated forecasts improves with the forecast horizon.

Possible cointegrating relationships among the variables are ignored, and all variables are transformed to eliminate stochastic and deterministic trends. This assumption is not problematic, as Größl and Tarassow (2015) have recently shown that there exists no stable long-run money demand relationship between the level variables m_t (using aggregate M2), y_t and i_t in the U.S. economy for the period between 1978 and 2013. We treat the logarithm of M2 and the logarithm of the income measure as $I(1)$, and assume that the opportunity cost measure as well as additional uncertainty measures are stationary for each rolling window.

2.2 Forecast model evaluation method

The empirical evidence of parameter instability is widespread in macroeconomic forecasting (Stock and Watson, 2003). In this work, we apply the rolling-window forecasting scheme using a fixed number of the most recent data at each point of time to generate sequences of out-of-sample forecasts which are used for evaluating competing models (Tashman, 2000). Pesaran and Timmermann (2007) have shown that neglecting existing structural breaks may result in biased forecasts. The rolling-window approach is more robust against structural breaks compared to a recursive forecasting scheme with an extending forecast window.

As it remains ambiguous how to select the relevant variables of the ARDL model for each training set, we combine the rolling-window forecasting scheme by two different indicator selection procedures. In a first approach, we conduct a *general-to-specific* (G2S) analysis, and sequentially eliminate irrelevant variables. The second *specific-to-general* (S2G) approach starts from a parsimonious model, and sequentially adds relevant variables. To be more concrete, the two algorithms work as follows:

(i) General-to-Specific (*G2S*): Regress initially Δm_t on a candidate ARDL model with 12 lags for each regressor using OLS. Omit the variable with the highest p-value to estimate the new candidate model. Repeat the steps based on the latest candidate model, until all remaining variables have a p-value no greater than 5 percent (two-sided).

(ii) Specific-to-General (*S2G*): Regress initially Δm_t on an intercept using OLS. Scan the list of candidate variables from a candidate ARDL model with 12 lags for each regressor that, if added, improves the selected AIC criterion to the greatest extent. If improvement is possible, add the new regressor to the initial model and remove it from the candidate model, and go to step 1; if not, stop.

Additionally, we compute the equally-weighted mean of all forecasts as proposed by the model combination literature (see e.g. Rapach and Strauss, 2008). The mean forecast, $\widehat{m}_{t+h}^{mean} = K^{-1}\sum_{j=1}^{K}\widehat{m}_{j,t+h}^{k}$, is the simple average using all K separate out-of-sample forecasts. Our set of individual forecast models comprises $K = 7$ different ARDL model specifications (which will be described below) for which we (i) set the maximum lag length to twelve, (ii) conduct the G2S and (iii) S2G indicator selection, respectively. This amounts in total to 21 model specifications for which we compute $\widehat{m}_{t+h}^{mean}$ for each forecast horizon.

As parameters might be changing over time, and because we want to give more weight to recent observations, we use the rolling-window forecasting scheme. The algorithm works as follows:

(i) The total sample of $T+1$ observations is split into the first 1 to T_s ($T_s < T+1$) observations.

(ii) Determine the relevant set of regressors of the ARDL model using either the G2S or S2G indicator selection procedure based on T_s observations. Estimate the model parameters.

(iii) Compute the h-multi-step iterated forecast, Δm_{T_s+h}.

(iv) Move the initial and final sample observations to 2 and $T_s + 1$, respectively.

(v) Repeat the steps (ii) to (iv) until you arrive at time T, at which the parameters of interest are re-estimated based on the sample of observations $T - T_s + 1$ to T to compute the forecast for $T + 1$.

In the following applications, the training set consists of $T_s = 96$ monthly observations (8 years) to do any meaningful forecasting. However, as recently shown by Inoue et al. (2017), it remains under debate how many recent observations should be used in estimation. Pesaran and Timmermann (2007) and Inoue et al. (2017) have suggested different criteria to determine the optimal window size under structural changes. Instead of applying these rather complex algorithms, we will experiment with two alternative window sizes, T_s, in the robustness section to show the validity of our results.

2.2.1 Evaluating forecasting accuracy

As our measure for forecast comparison, we use the ratio of the h-step-ahead root mean squared error (RMSE) for a given h-step-ahead forecast of model k made at time t (denoting the last observation of the training set), $\Delta\widehat{m}^k_{t+h}$, relative to the corresponding forecast of the benchmark model, $\Delta\widehat{m}^b_{t+h,t}$. The equation is given by

$$R^h_k = \frac{\sqrt{F^{-1}\sum_{j=1}^{F}(\Delta m_{t+h} - \Delta\widehat{m}^k_{t+h})^2}}{\sqrt{F^{-1}\sum_{j=1}^{F}(\Delta m_{t+h} - \Delta\widehat{m}^b_{t+h})^2}} \tag{2}$$

and where F denotes the number of out-of-sample forecasts. If model k outperforms the competitor on average, R^h_k is smaller one, and vice versa. For the forecast analysis, we compute the relative RMSE as defined in eq. (2), relative to the ARDL(12,12,12) benchmark (and hence without any automatic indicator selection) of the standard money demand model specification.[2] This standard money demand specification is described in eq. (1) and consists on the RHS only lags of Δm_t, Δy_t and i_t. All forecasts are based on the rolling-window forecasting scheme. The competitor model is also based on the rolling-window forecasting scheme combined, in contrast to the benchmark case, with either G2S or S2G indicator selection.

However, the statistics R^h_k does not measure whether the loss differentials between two competing forecasts are significantly different in a statistical sense. Diebold and Mariano (1995) proposed an intuitive test of equal predictive accuracy of two models. The test by Diebold and Mariano (DM, henceforth) relies on assumptions made directly on the forecast error loss differential. Denote the h-periods ahead forecast error at time t by e_t^{t-h}, and the, for instance, quadratic loss by $L(e_t^{t-h}) = e_t^{2,t-h}$. The loss differential of two forecasts for observation t is given by $d_{12t} = e_{1t}^{2,t-h} - e_{2t}^{2,t-h}$. The null of equal predictive accuracy corresponds to $E(d_{12}) = 0$, in

[2] Usually forecasts are compared to a simple AR(p) or Random-walk model. As our focus is on comparing the standard money demand specification against some uncertainty-augmented model, it is not our focus to compare these theoretically-guided money demand models against some pure statistical time-series specification.

which case the test statistics is

$$DM_{12} = \frac{\bar{d}_{12}}{\widehat{\sigma}_{\bar{d_{12}}}} \rightarrow N(0,1) \tag{3}$$

and where $\bar{d}_{12} = F^{-1}\sum_{j=1}^{F} d_{12j}$ refers to the sample mean loss differential and $\widehat{\sigma}_{\bar{d}_{12}}$ is a consistent estimate of the standard deviation of $\bar{d}_{12}$ based on F number of forecasts available.

However, as the forecasts errors may be serially correlated, one can alternatively regress the loss differential on an intercept by OLS using HAC standard errors

$$d_{12t} = \beta_0 + u_t \tag{4}$$

where u_t is an *i.i.d.* zero-mean error term. The null hypothesis is that there is no difference in the point estimates between the two forecasts, i.e $\beta = 0$. The estimate of β follows asymptotically a standard normal distribution (Diebold and Mariano, 1995). Additionally, we apply the small sample correction as proposed by Harvey et al. (1997).

It should be noted that the Diebold-Mariano approach is only valid if the estimated forecast models are non-nested. With nested models, however, under the null, the forecast errors are asymptotically equal and therefore perfectly correlated (Mc Cracken, 2000; Clark and McCracken, 2001). Typically, those tests on either equal MSE or encompassing explicitly account for uncertainty of parameter estimation. However, as noted by Stock and Watson (2003), the null distribution of the Clark and McCracken (2001) test can only be computed if the number of lags in the models does not change over time. As we will apply the rolling forecasting scheme combined with either G2S and S2G indicator selection potentially yielding a time-varying lag variation, there is no possibility to apply the Clark-McCracken test approach. Hence, we will rely on the Diebold-Mariano framework following Stock and Watson, instead.

2.2.2 Directional forecast evaluation

Another issue concerns the analysis of the correct prediction of the direction-of-change in growth of the real money stock (e.g. Ups, Downs). The Kuipers score (KS) is an evaluation criterion defined as the difference between the hit rate (H, correctly predicted Ups) and the false alarm rate (F, wrongly predicted Ups) of model k for each h-step-ahead forecast, $KS_h^k = H_{t+h}^k - F_{t+h}^k$ (Pesaran, 2015, pp. 396).

Pesaran and Timmermann (1992) (PT) proposed a statistical test of market timing which is based on

$$PT = \frac{\hat{P} - \hat{P}_*}{\sqrt{\hat{V}(\hat{P}) - \hat{V}(\hat{P}_*)}} \tag{5}$$

where $\hat{P}$ is the proportion of correctly predicted Ups and $\hat{P}_*$ is the estimate of the probability of correctly predicting the events assuming predictions and realizations are independently distributed. $\hat{V}(\hat{P})$ and $\hat{V}(\hat{P}_*)$ are consistent estimates of the variances of $\hat{P}$ and $\hat{P}_*$, respectively.

In the following, we will apply the regression based version of the PT test, testing whether the predicted Ups, $x_t = I(X_t)$ are related to the actual Ups, $y_t = I(Y_t)$ using a sample of observations and where $I(A)$ is an indicator function that takes the value of unity if $A > 0$ and zero otherwise. The PT statistics can be approximated by the t-ratio of the coefficient of $x_t = I(X_t)$ in the following OLS regression

$$y_t = \alpha + \beta x_t + u_t \tag{6}$$

where $y_t = I(Y_t)$ and α is the intercept. Under the null hypothesis, $H_0 : \beta = 0$, the two stochastic variables are independently distributed, and vice versa. Serial correlation in the errors, u_t, are likely to occur but can be dealt with by using Bartlett weights to compute HAC standard errors (see Pesaran, 2015, pp. 398).

We will compute both the KS statistics and the test results of the PT test for each model as well as different multi-step forecast horizons h.

3 Estimation results

In this section we present and discuss the model specifications as well as underlying variables. Our dataset comprises monthly observations for the U.S. economy, and the selected sample starts in 1971m1 and ends in 2014m12. We employ the monthly FRED-MD dataset published by McCracken and Ng (2015) for most variables. For details we refer to the Data Appendix.[3]

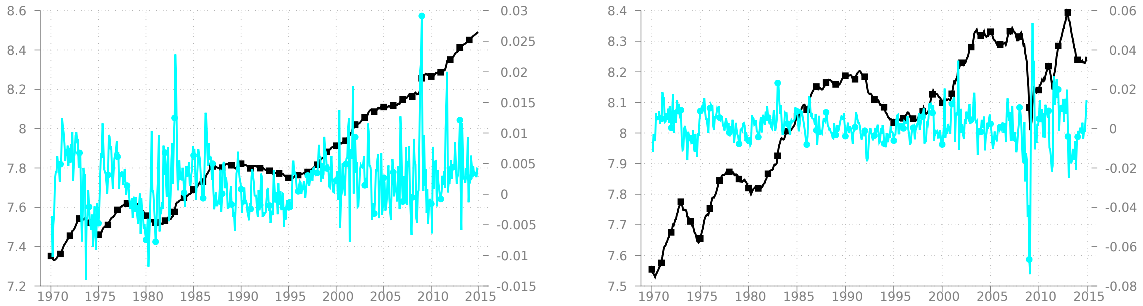
3.1 Model specifications

Our benchmark model can be seen as a variant of the standard money demand specification used in the literature. As the depend variable we take the official measure of aggregate M2 deflated by the CPI price index before taking the first difference of the logarithm, $\Delta m_t = \Delta \ln(M_t/P_t)$. Apart from real aggregate M2, we also study U.S. real household-sector M2 which is also deflated by the personal consumption expenditure price index, PCE. In the robustness section we also employ the (PCE deflated) M4 Divisia money stock measure as published by the *Center for Financial Stability* (Barnett et al., 2013). For more details on the latter series we refer to the Data Appendix.

The aggregate as well as the household-sector M2 time-series are both depicted in Figure 1. Both period growth rates exhibit a few substantial amplitudes over time. For growth of real household-sector M2 the massive negative value in 2009 is striking. Overall, we see evidence for time-variation in the unconditional variance in both growth series.

The benchmark model consists on the RHS of an income measure for which we use the first difference of the logarithm of real personal income, Δy_t. Our opportunity cost measure, i_t, is defined as the difference between the 3-Month Treasury Bill rate and the own rate of M2. We treat Δm_t, Δy_t and i_t as stationary $I(0)$ variables.

[3] All computations are done by the open-source econometric software `Gretl` (Cottrell and Lucchetti, 2016).



(a) aggregate M2

(b) households-sector M2

Figure 1: Log of real M2 and its first difference. Sample: 1971m1–2014m12.

We proceed with the specification of the following seven dynamic ARDL models:

$$\text{(Model I)}: \ \Delta m_t = \alpha_0 + \sum_{j=1}^{q1} \beta_j \Delta m_{t-j} + \sum_{j=1}^{q2} \theta_j \Delta y_{t-j} + \sum_{j=1}^{q3} \phi_j i_{t-j} + e_t$$

$$\text{(Model II)}: \ \Delta m_t = \alpha_0 + \sum_{j=1}^{q1} \beta_j \Delta m_{t-j} + \sum_{j=1}^{q2} \theta_j \Delta y_{t-j} + \sum_{j=1}^{q3} \phi_j i_{t-j} + \sum_{j=1}^{q4} \rho_j U_{\pi t-j} + e_t$$

$$\text{(Model III)}: \ \Delta m_t = \alpha_0 + \sum_{j=1}^{q1} \beta_j \Delta m_{t-j} + \sum_{j=1}^{q2} \theta_j \Delta y_{t-j} + \sum_{j=1}^{q3} \phi_j i_{t-j} + \sum_{j=1}^{q4} \rho_j U_{st-j} + e_t$$

$$\text{(Model IV)}: \ \Delta m_t = \alpha_0 + \sum_{j=1}^{q1} \beta_j \Delta m_{t-j} + \sum_{j=1}^{q2} \theta_j \Delta y_{t-j} + \sum_{j=1}^{q3} \phi_j i_{t-j} + \sum_{j=1}^{q4} \rho_j U_{vt-j} + e_t$$

$$\text{(Model V)}: \ \Delta m_t = \alpha_0 + \sum_{j=1}^{q1} \beta_j \Delta m_{t-j} + \sum_{j=1}^{q2} \theta_j \Delta y_{t-j} + \sum_{j=1}^{q3} \phi_j i_{t-j} + \sum_{j=1}^{q4} \rho_j U_{ft-j} + e_t$$

$$\text{(Model VI)}: \ \Delta m_t = \alpha_0 + \sum_{j=1}^{q1} \beta_j \Delta m_{t-j} + \sum_{j=1}^{q2} \theta_j \Delta y_{t-j} + \sum_{j=1}^{q3} \phi_j i_{t-j} + \sum_{j=1}^{q4} \rho_j U_{et-j} + e_t$$

$$\text{(Model VII)}: \ \Delta m_t = \alpha_0 + \sum_{j=1}^{q1} \beta_j \Delta m_{t-j} + \sum_{j=1}^{q2} \theta_j \Delta y_{t-j} + \sum_{j=1}^{q3} \phi_j i_{t-j} + \sum_{j=1}^{q4} \rho_j U_{pt-j} + e_t$$

Model I marks the standard money demand model specification. Model II is augmented by the inflation uncertainty measure, U_π, as recently applied by Größl and Tarassow (2015) in their money demand model. This inflation uncertainty index is the standard deviation of the permanent component estimated by an unobserved component stochastic volatility model (UC-SV), as proposed by Stock and Watson (2005), to the CPI inflation rate. The UC-SV models heteroskedasticty in inflation explicitly and might be preferred to standard homoskedastic (S)VAR models (Chua et al., 2011). Following the perspective of Grimme et al. (2011), the time-varying standard deviation of the time-varying permanent component of inflation reflects uncertainty in the inflation rate.

In Model III, we consider the stock market risk premium, U_s, as previously applied by Greiber and Lemke (2005) and Größl and Tarassow (2015) in their money demand studies. U_s is defined as the ratio of the dividend yield on the S&P 500 stock price index over ten-year U.S. Treasury notes (Fama and French, 1988). In Model IV we introduce the implied volatility VXO measure, U_v, as popularized by Bloom (2009). This series is assumed to capture expected stock market volatility. Variable U_v was applied in the money demand literature among others by Carpenter and Lange (2003) and Cook and Choi (2007) before. In Model V we replace the VXO series by an alternative financial market uncertainty index, U_f, as recently estimated by Ludvigson et al. (2015). To be more concrete, we use the series to which Ludvigson et al. refer as the $h = 1$–series. To our knowledge this series has not been applied in the money demand context, yet. In Model VI we include the macroeconomic uncertainty index, U_e, as proposed by Jurado et al. (2015). Also in this case we use their so called $h = 1$–measure.[4] Again, we are not aware of any money demand study which considered this series before. Lastly, in Model VII we consider the popular economic policy uncertainty index, U_p, as compiled by Baker et al. (2013). Also this series has not been applied in the money demand literature before. For more details we refer to the Data Appendix.

3.2 Uncertainty series

Figure 2 depicts the various uncertainty measures in standardized units. The horizontal bar corresponds to 1.65 standard deviations above the unconditional mean of each series (standardized to zero). From the literature on macroeconomic uncertainty measures, it is well known that macro uncertainty is strongly countercyclical, and exhibits large spikes in the deepest recession (Bloom, 2009; Jurado et al., 2015). The correlation with growth of aggregate M2 is less clear, however.

Panel (a) depicts the inflation uncertainty series, U_π. It exceeds 1.65 standard deviations mainly during the oil price crisis between mid 1973 and 1975m9, almost in 1982m3 and again during the period between 2008m9 and 2009m12. Inflation uncertainty is weakly negatively correlated with growth in M2 ($\rho = -0.06$) using the full sample. However, the rolling-window correlation coefficients (using a window-size of 120 months) reveals some time-varying unconditional contemporaneous correlation between the series. While the correlation is negative between 1975 and 1986 (on average $\rho \approx -0.25$), it turns positive in the missing money period during the mid 1990s ($\rho \approx 0.20$) reaching a peak in 1998 before turning almost zero again. Another hike in correlation (up to $\rho \approx 0.45$) can be seen between 2008 and 2009. In their recent study, Größl and Tarassow (2015) have found that U.S. households increase their demand for M2 in response to positive changes in this inflation uncertainty measure.

Economic policy uncertainty (see Panel (b)) exceeds the 1.65 standard deviations in October 1987, the Iraq War, in the beginning of 2000 and frequently between 2009 and 2013. The unconditional correlation between economic policy uncertainty and growth of M2 is almost negligible though.

The macroeconomic uncertainty index was constructed by Jurado et al. (2015). The series

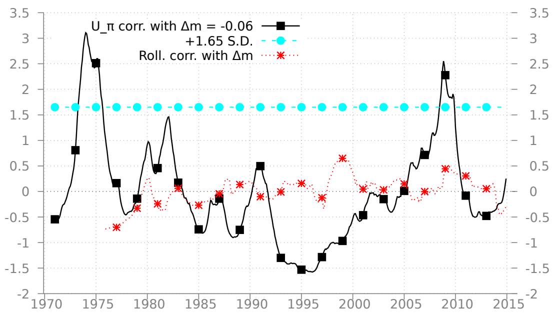
[4]For both the financial market uncertainty index (U_f) as well as the macroeconomic uncertainty index (U_e) our results are robust against the use of different horizons of both U_f and U_e, respectively.

(see Panel (c)) indicates that uncertainty peaks and exceeds the 1.65 standard deviation in 1974, between 1980 and 1982 as well as between 2007 and 2009. While this measure is known to be highly countercyclical, the full sample correlation indicates a negligible link between macroeconomic uncertainty and growth in M2 as $\rho = -0.02$. However, the rolling-window correlation directs attention to some time-variation of this statistics. While the correlation is negative during the two oil price crises episodes in the 1970s and 1980s—for which the correlation ranges between -0.5 to -0.7—the correlation turns slightly positive during the Volcker-era ($\rho \approx 0.15$) and almost zero during the Clinton administration period in the 1990s. However, one can see that the relationship has turned positive for most periods since 2000 reaching its highest level between 2008 and 2013 as the correlation is about 0.3. This latter finding is in line with the argument by Jurado et al. (2015, p. 1992) that between 2007m12 and 2009m6, uncertainty is highest for the monetary base apart from non-borrowed reserves and total reserves.

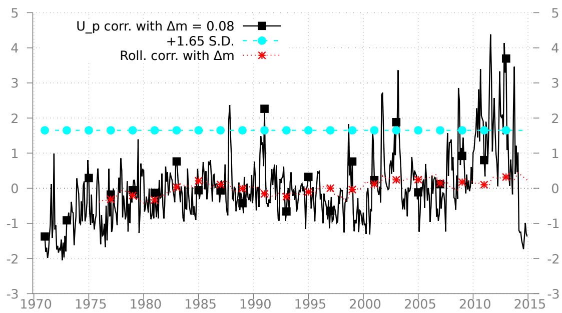
The VXO options-based volatility index was popularized by Bloom (2009). Panel (d) depicts the standardized VXO series. The series exceeds the 1.65 standard deviation in 1974/75, the Black Monday in Oct. 1987, the late 1990s and early 2000s during the New Economy boom and bust episode, and again in early 2009 and 2012. This measure is positively correlated with growth in M2 as the correlation coefficient is 0.22. The rolling-window correlation indicates some switch in the unconditional dynamics between the two series since the mid 1980s: while the correlation is negative before mid 1985 (average $\rho \approx -0.25$) it has turned positive with an average correlation of about $\rho \approx 0.25$ since then. Carpenter and Lange (2003) and Cook and Choi (2007) found that the VXO series is positively correlated with M2 money demand.

Given that some authors have classified the VXO index as a problematic measure of true uncertainty as it contains a large component attributable to changes in the variance risk premium that are unrelated to common notions of uncertainty (Bollerslev et al., 2009), Ludvigson et al. (2015) proposed the construction of an alternative financial market uncertainty measure. We depict their time-series in Panel (e). This series (U_f) identifies almost the same episodes of extremely uncertain financial market episodes as the implied volatility measure (extreme episodes are those exceeding the historical 1.65 S.D.). Surprisingly, the full sample correlation with growth in M2 is almost zero ($\rho = 0.04$). However, the time-varying correlation indicates a strong negative unconditional correlation between U_f and Δm before the mid 1980s (average $\rho \approx -0.5$), and a near-zero correlation between 1986 and 1996. The correlation has turned positive for most years after 1998 with an average correlation of $\rho \approx 0.25$. Thus, the unconditional dynamic relationship is similar to the one between growth of M2 and the VXO measure, U_v.

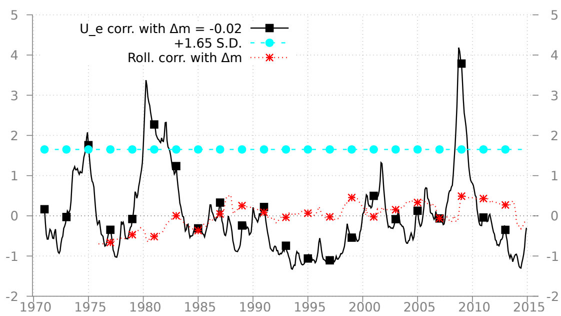
Finally, Panel (f) displays the stock market risk premium (U_s) as previously considered for instance by Greiber and Lemke (2005) and Größl and Tarassow (2015) in the money demand literature. While this series seems to follow a stationary process between 1971 and the mid 1990s, its DGP has changed since then towards a more persistent (probably non-stationary) process. The U_s time-series fluctuates widely and is successively increasing since 2000. The full sample correlation with growth in M2 is weakly positive ($\rho = 0.11$). However, there is again evidence for time-varying unconditional correlation between U_s and growth in M2: while the correlation was negative during the first oil price crises, it turned positive during the Volcker-era in the 1980s, negative again during the missing money period in the mid 1990s and positive since 2000. In their extensive study, Größl and Tarassow (2015) document a positive response of U.S. households' money demand to an increase in U_s.



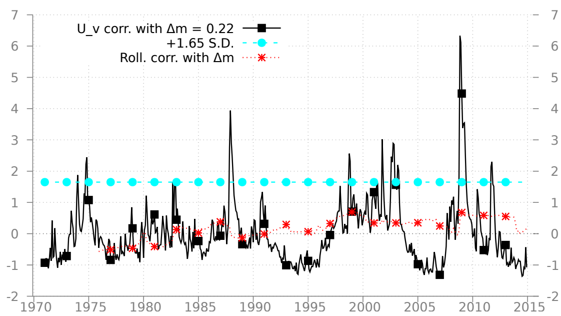
(a) Inflation uncert.: U_π



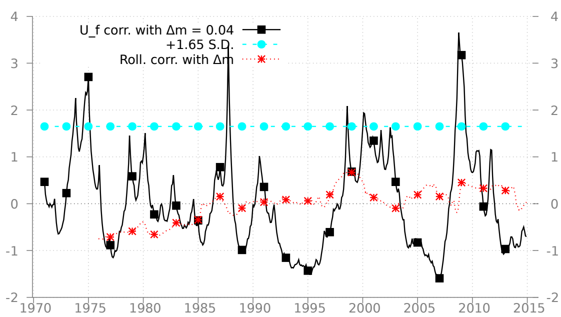
(b) Economic policy uncert.: U_p



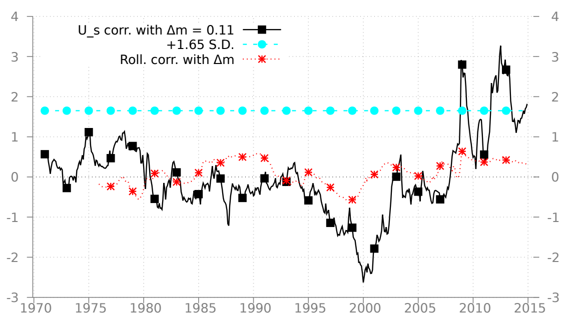
(c) Macroeconomic uncert.: U_e



(d) VXO: U_v



(e) Financ.-market uncert.: U_f

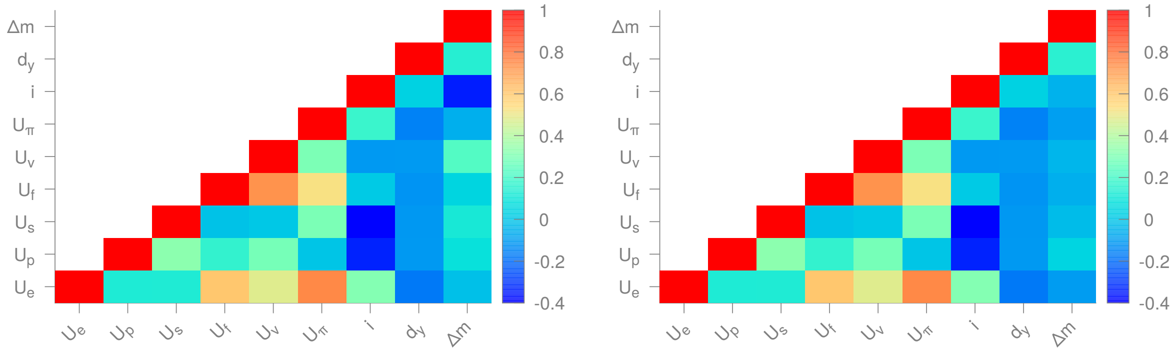


(f) Stock market risk premium: U_s

NOTE: Risk series are expressed in standardized units. The horizontal line corresponds to 1.65 standard deviations above the unconditional mean of each standardized series (normalized to zero). Full sample correlations with monthly growth of aggregate M2 money demand are reported in the legend. Rolling-window correlations with the monthly growth of M2 money demand are based on 120 months.

Figure 2: Economic and financial uncertainty over time. Sample: 1971m1 to 2014m12.

For the sake of completeness, we provide the unconditional correlation between all variables used in our study in Figure 3 for both growth of aggregate M2 and growth of household sector M2, respectively. The correlation structures between the uncertainty series and monetary growth are almost equal for both money definitions.



(a) growth of aggregate M2 (b) growth of household sector M2

NOTE: The first difference operator is denoted by Δ. The abbreviations m, y, i denote the log of real M2 (aggregate or households sector), log of real income and the opportunity cost measure. The abbreviations of the uncertainty series are defined in Figure 2.

Figure 3: Contemporaneous, unconditional correlations between all variables used. Sample: 1971m1 to 2014m12.

4 Forecasting results for aggregate M2

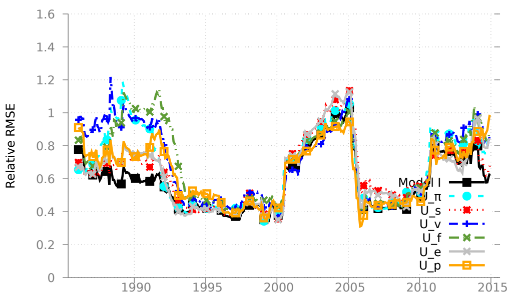
In this section we report the results of the out-of-sample rolling-window forecasting exercise, as described in Section 2.2 for growth of aggregate M2. The out-of-sample forecasts and forecast errors are estimated by means of a rolling-window approach with window width $T_s = 96$ monthly observations. In the robustness section we show that the main results are robust against the variation of this window width. Additionally, we show that the outcomes are robust against the use of the alternative Divisia M4 index series.

4.1 ARDL forecast accuracy

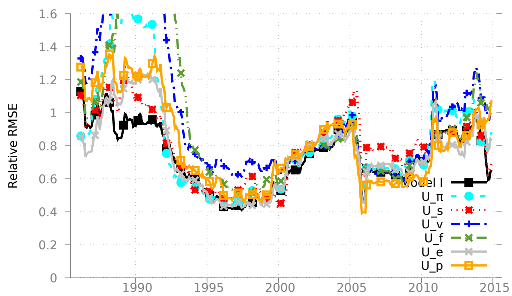
Forecast accuracy is measured by the root mean squared error (RMSE) criteria. To be more concrete, we compute the RMSE of a model relative to the RMSE of the benchmark ARDL(12,12,12) Model I specification (comprising lags of Δm_t, Δy_t and i_t). The relative RMSE is defined as $\left(RMSE_h^k / RMSE_h^b\right)$ where $RMSE_h^b$ denotes the RMSE at forecast horizon h of the benchmark Model I and $RMSE_h^k$ refers to forecast of some competitor model k for the same horizon. Furthermore, we apply the Diebold-Mariano test on equal predictive accuracy, as described before. In the subsequent analysis we are mainly interested in the forecast accuracy performance of different uncertainty augmented money demand models since the year 2000.

As said, all out-of-sample forecasts are based on rolling-window ARDL models with a width of $T_s = 96$ monthly observations. However, in order to visualise the time-variation in relative RMSEs, we run a second rolling-window with width $T_e = 60$ monthly observations over the out-of-sample forecasts estimated in the first step. To get a first impression about the time-varying forecast performances, Figure 4 depicts the rolling-window relative RMSE for each model (based on S2G indicator selection which performs best in most cases) relative to the benchmark ARDL(12,12,12) model for the 1-, 2-, 6- and 9-month ahead forecast, respectively.

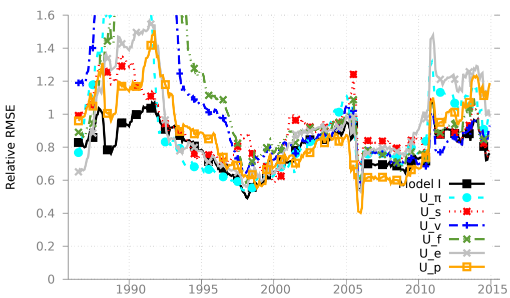
The results indicate substantial time-variation in the relative forecasting performance for all models. For the 1-month ahead forecast horizon, one can observe that the majority of competitor models, including Model I combined with S2G indicator selection, tend to outperform the benchmark for most periods considered. We observe three episodes during which all or at least some models performed weaker compared to the remaining periods. The first episode is between 1988 and 1993 for which we observe large differences across models even though most models outperform the benchmark. The common drop in relative RMSE (and hence improvement in forecast accuracy relative to the benchmark) among the competitor models between 1994 and 1999 is associated with a period of low unconditional variance of growth of aggregate M2, as depicted in Figure 1. However, in 2000 we observe an immediate deterioration in relative forecast accuracy across all models followed by a continuous rise in relative RMSE which lasts until 2004—the period comprising the New Economy Bust and its repercussions. However, most models have re-gained their relative forecast accuracy in 2004. Lastly, the recent GFC period has triggered another deterioration in relative RMSE since 2009. Very similar dynamics are observed for the 2-, 6- and 9-month forecast horizons.



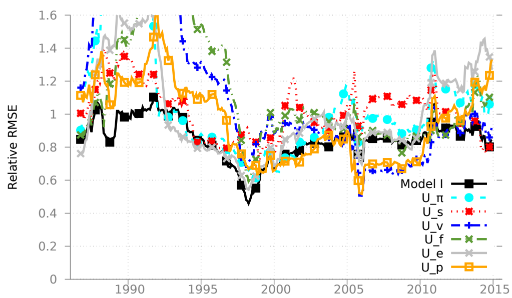
(a) 1-month ahead



(b) 2-month ahead



(c) 6-month ahead



(d) 9-month ahead

Note: These are the RMSEs relative to the RMSE of the benchmark ARDL(12,12,12) model. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 96$ observations. Based on 420 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1980m1, we run another rolling-window with width $T_e = 60$ observations to construct the time-varying relative RMSE with 359 observations. The first observation of the relative RMSE is given for 1985m2. Values on the y-axis exceeding 1.6 are truncated.

Figure 4: Out-of-sample rolling-window forecasting accuracy results after *S2G* indicator selection for growth in aggregate M2.

In the following Table 1 we provide a more compact summary of forecast accuracy for each model with a special focus on the period since 2000. To be more concrete, we report information on the relative RMSE compared to the benchmark ARDL(12,12,12) model for the 1-, 2-, 6- and 9-month horizon, respectively. Furthermore, statistics are reported for (i) the full model specifications (based on an ARDL(12,12,12,12), except Model I which is an ARDL(12,12,12)), (ii) the selected models after both G2S and S2G indicator selection, respectively, and (iii) for the forecast combination *mean model* which is an equally-weighted average of all forecasts available. Additionally, we provide information on the Diebold-Mariano test results under quadratic loss. Lastly, these statistics are computed (a) using all forecasts available between 2000m1 and 2014m12 as well (b) between 2007m1 and 2014m12. This sub-sample evaluation allows us to analyse the effects of the GFC period for the recent forecasting performance of our models.

Based on all forecasts available between 2000m1 and 2014m12, we find that Model I combined with S2G indicator selection yields a reduction in RMSE by about 38% relative to its ARDL(12,12,12) counterpart. The second most accurate model includes the policy uncertainty measure (U_p) and yields a relative RMSE of 0.666 while the mean model still yields a relative RMSE of 0.674. We observe that S2G indicator selection clearly outperforms G2S indicator selection at all horizons. The full model specifications of the competitor models yield less accurate forecasts compared to the benchmark. Most importantly, we can reject the null of the Diebold-Mariano test on equal predictive accuracy relative to the benchmark model at the 5% for all top-3 models at the 1-month horizon.

At the 2-month horizon, Model I combined with S2G indicator selection provides the most accurate forecasts (0.748) again closely followed by the policy uncertainty model (U_p; 0.763) and the macroeconomic uncertainty (U_e; 0.769) model. However, the null of the Diebold-Mariano test can only be rejected at least at the 10% level for the winning model and the mean model (0.776). Very similar results are obtained at the 6-month horizon where again Model I combined with S2G indicator selection yields the most precise forecasts (0.785). Interestingly, at the 9-month horizon we find that the VXO model yields a relative RMSE of 0.796 and hence outperforms Model I combined with S2G indicator selection as well as the mean model by almost 10%. At the 9-month horizon, we can reject the null of the Diebold-Mariano test for all three models at least at the 10% level and for the mean model even at the 1% level.

Next, we analyse all forecasts for the period from 2007m1 to 2014m12. Again Model I combined with S2G indicator selection yields the lowest relative RMSE at both the 1- (0.665) and 9-month (0.795) ahead forecast horizons and the second best forecast accuracy results at the 2- (0.76) and 6-month (0.779) horizon, respectively. While the macroeconomic uncertainty model wins at the 2-month horizon (0.75), the model including the VXO measure provides the most accurate forecast at the 6-month horizon (0.778).

Among the top-3 models we also find the one including the stock market risk premium, U_s, (at the 1-month horizon) and the mean model (at the 9-month horizon). At the 1-, 2- and 9-month horizons, all top-3 models provide systematically more accurate forecasts compared to the benchmark while at the 6-month horizon the null of the Diebold-Mariano test can at least be rejected at the 10% level only for the VXO model and Model I combined with S2G indicator selection.

Overall, in terms of forecast accuracy the standard money demand model tends to perform very well for short- and medium-term out-of-sample forecasts of growth of aggregate M2 since 2000.

This holds for both the period since 200m1 as well as for the sub-sample period between 2007m1 and 2014m12. Nevertheless, we also find evidence that the models including the VXO measure and the macroeconomic uncertainty measure provide some additional forecasting content.

In the robustness section, we show that these findings are robust against the use of a shorter or longer rolling-window width T_s as well as against the use of the real M4 divisia series as an alternative measure of the aggregate money stock.

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
	2000m1–2014m12				**2007m1–2014m12**			
			A: Full Model					
U_π	1.147**	1.104	1.168**	1.232**	1.106	1.103	1.108	1.159
U_s	1.489**	1.434**	1.434**	1.255**	1.608**	1.673*	1.759**	1.466**
U_v	1.305**	1.345**	1.448**	1.389***	1.369*	1.390*	1.537	1.431*
U_f	1.325***	1.363**	1.382***	1.401***	1.414**	1.578**	1.514**	1.570***
U_e	1.182**	1.150*	1.163*	1.182*	1.167**	1.210**	1.231*	1.221
U_p	1.318***	1.405***	1.438***	1.434***	1.501***	1.739***	1.675***	1.859***
			B: G2S					
Mod. I	0.714**	0.810	0.826	0.924	0.758	0.839	0.886	0.885
U_π	0.801	0.828	0.987	0.959	0.888	0.877	1.124	1.010
U_s	0.926	1.049	1.030	0.953	1.092	1.295	1.253	1.041
U_v	0.836	0.975	0.964	0.980	1.040	1.100	1.107	0.948
U_f	0.908	1.111	1.210	1.216*	0.996	1.218	1.193	1.238
U_e	0.830	0.990	0.961	0.985	0.889	1.101	1.129	0.987
U_p	0.979	1.089	1.064	1.295*	1.219	1.402**	1.401**	1.625**
			C: S2G					
Mod. I	$\mathbf{0.624}^{**}$	$\mathbf{0.748}^{*}$	$\mathbf{0.785}^{*}$	$\underline{0.856^{*}}$	$\mathbf{0.665}^{**}$	$\underline{0.760}$	$\underline{0.779^{*}}$	$\mathbf{0.795}^{**}$
U_π	0.693**	0.800	0.886	0.997	0.792	0.837	0.921	0.946
U_s	0.689**	0.834	0.881	1.027	$\underline{0.671^{**}}$	*0.762*	*0.794*	0.864
U_v	0.696**	0.846	*0.816*	$\mathbf{0.796}^{*}$	0.795	0.948	$\mathbf{0.778}^{*}$	$\underline{0.835^{*}}$
U_f	0.706*	0.834	0.841	0.954	0.797	0.916	0.821	0.926
U_e	0.704*	*0.769*	0.941	1.006	$\mathit{0.726}^{*}$	$\mathbf{0.750}^{*}$	0.985	1.018
U_p	$\underline{0.666^{**}}$	$\underline{0.763}$	0.828	0.879	0.745	0.833	0.951	0.977
			D: Mean Model					
Mean	$\mathit{0.674}^{**}$	0.776*	$\underline{0.804^{*}}$	$\mathit{0.862}^{***}$	0.754**	0.844	0.874	$\mathit{0.842}^{**}$

Note: The table reports the RMSE relative to an ARDL(12,12,12) comprising lagged values of the endogenous money growth variable, a real income measure and an opportunity cost measure. We also report the test results on equal predictive accuracy using the test proposed by Diebold and Mariano (1995) under a quad-quad loss function. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Diebold-Mariano test. The 1st, 2nd and 3rd lowest relative RMSEs are highlighted by bold, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 96$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

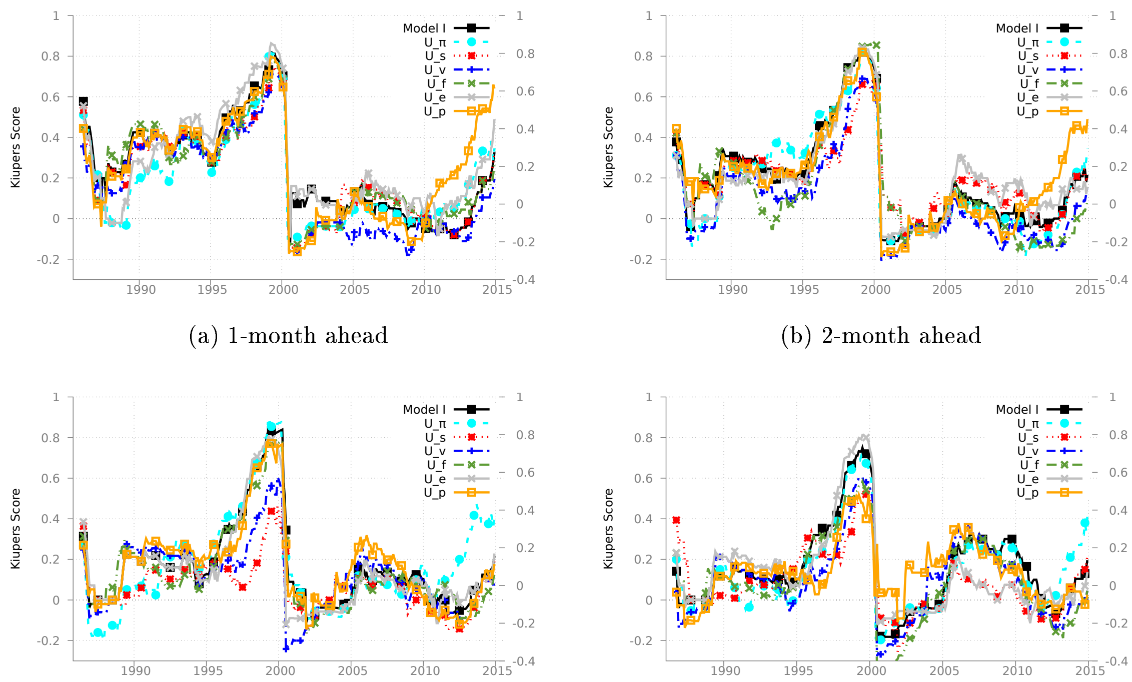
Table 1: Relative RMSE and Diebold-Mariano test results for growth of aggregate M2 using a rolling-window with width $T_s = 96$ observations.

4.2 Direction-of-change forecasts based on ARDL forecasts

Next we present the direction-of-change forecast results, as described on page 9. We will report both the Kuipers Score (KS) and the test results of the Pesaran-Timmermann (PT) test on predictive failure.

To get an idea about eventual time-variation in the direction-of-change forecasting performance of our models, again we compute the KS statistics based on all available forecasts using a second rolling-window of width $T_e = 60$ monthly observations. Figure 5 depicts the results for all models based on S2G indicator selection for the 1-, 2-, 6- and 9-month horizon, respectively.

We observe the following basic patterns: After an initial drop in the KS values for all models and horizons between 1986 and 1987, we observe a succeeding increase in the KS values reaching its maximum in 1999 reflecting the fact that it is easier to predict direction-of-changes for periods marked by rather low unconditional variance (see again Figure 1). The New Economy Bust phase around 2000 was accompanied by a drastic fall in the average KS value among all models and at all horizons. The period since 2000 was associated with a substantial increase in the unconditional variance in growth of aggregate M2. For instance, at the 1-month horizon the average KS across models is 0.8 in 1999 but close to zero in 2000. Qualitatively, similar dynamics can be observed for the remaining horizons. This shows that it has become much harder to forecast the direction-of-change of growth of real aggregate M2 since 2000. Nevertheless, we still observe that some models yield higher KS values compared to others.



(a) 1-month ahead

(b) 2-month ahead

(c) 6-month ahead

(d) 9-month ahead

NOTE: For details, see note below Figure 4.

Figure 5: Out-of-sample rolling-window Kuipers Score after *S2G* model selection for growth in aggregate M2.

Table 2 replicates Table 1 but reports the Kuipers Score and the Pesaran-Timmermann test results instead. For the period between 2000m1 and 2014m12 we see that the full ARDL(12,12,12,12) model including inflation uncertainty (U_{π}) yields the highest KS values at both the 1- (0.18) and 2-month (0.203) forecast horizon, respectively, and the second highest KS values at the 6- (0.187) and 9-month (0.18) horizons. The winning model at the 6-month horizon is the VXO model (U_v; 0.189) and at the 9-month horizon the financial uncertainty model (U_f; 0.203) performs best. The policy uncertainty model (U_p) yields two times the third rank. However, Model I, either with or without automatic indicator selection, performs rather poorly in direction-of-change forecasting. For instance, the benchmark ARDL(12,12,12) models yields only a KS value of 0.134, 0.118, 0.172 and 0.104 at the 1-, 2-, 6- and 9-month horizon, respectively.

The null of the PT test of predictive failure can be rejected at least at the 10% level for each top-3 model at any forecast horizon. The PT test also indicates some weaknesses of the standard money demand specification in terms of direction-of-change forecasting.

The weak performance of the standard money demand specification is confirmed when analysing the recent GFC period between 2007m1 and 2014m12. At the 1-month horizon, the model including the policy uncertainty series combined with S2G indicator selection yields the highest KS value of 0.282 which is almost three times larger than the KS value of the benchmark model (0.103). The inflation uncertainty model (S2G; 0.155) and the macroeconomic policy model (S2G; 0.142) rank second and third, respectively. At the 6-month horizon the inflation uncertainty model (S2G) clearly outperforms all remaining models with a KS of 0.205 while the policy uncertainty model yields the second highest KS of 0.126. The financial uncertainty model yields a KS of 0.203 at the 9-month horizon while the inflation uncertainty model (S2G) yields a KS of 0.142. For the remaining models we find KS values close to zero or even negative.

The PT test confirms that direction-of-change forecasting growth of real aggregate M2 has become harder during the GFC period, as most models suffer from predictive failure at all horizons. For instance, at the 1-month horizon we can only reject the null of the PT test for the two best performing models including the inflation uncertainty (at the 10% level) and macroeconomic policy model (at the 5% level). Surprisingly, we cannot reject the null for any model at the 2-month horizon. At the 6-month horizon, the null can be rejected for the inflation uncertainty model, and at the 9-month horizon for the *mean* model (both at the 10% level).

Overall, the results strongly show that the standard money demand model performs rather poor in terms of direction-of-change forecasting. However, the inclusion of inflation uncertainty, policy uncertainty, the VXO measure and financial uncertainty yield a substantial gain in forecast performance. This finding is robust against the use of a shorter or longer rolling-window width, as will be shown below.

4.3 Robustness exercise

Recently, Pesaran and Timmermann (2007) have shown that the mean squared forecast error may depend on the window-width of the rolling-window forecasting scheme under the assumption

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
		2000m1–2014m12				**2007m1–2014m12**		
			A: Full Model					
Mod. I	0.134	0.118	0.172^{*}	0.104	0.103	-0.011	0.089	0.076
U_π	$\mathbf{0.180}^{**}$	$\mathbf{0.203}^{**}$	$\underline{0.187^{**}}$	$\underline{0.180^{**}}$	$\mathit{0.142}$	$\underline{0.179}$	$\mathit{0.116}$	$\mathit{0.126}$
U_s	0.059	0.097	0.068	0.067	0.053	0.039	-0.013	-0.039
U_v	0.120	0.084	$\mathbf{0.189}^{*}$	0.150	0.024	-0.066	0.034	0.024
U_f	0.082	0.083	0.118	$\mathbf{0.203}^{**}$	0.076	-0.026	0.016	$\mathbf{0.203}$
U_e	0.106	0.091	0.037	0.039	0.039	-0.024	-0.074	-0.024
U_p	$\mathit{0.158}^{*}$	$\underline{0.189^{**}}$	$\mathit{0.172}^{**}$	0.135	0.126	0.113	$\underline{0.126}$	0.037
			B: G2S					
Mod. I	0.072	0.088	0.043	0.013	0.105	0.103	-0.024	-0.061
U_π	0.058	0.005	0.051	0.127	0.129	-0.021	0.016	0.053
U_s	0.126^{*}	0.120	0.141^{*}	0.134	0.092	$\mathit{0.139}$	0.029	0.026
U_v	0.052	0.021	0.128	0.127	0.050		0.050	0.026
U_f	0.080	0.090	0.080	0.081	0.053	-0.013	0.016	-0.021
U_e	-0.002	0.082	-0.009	0.104	0.005	0.026	-0.111	0.018
U_p	$\underline{0.179^{**}}$	$\mathit{0.172}^{**}$	0.088	0.104	0.103	0.103	0.016	-0.050
			C: S2G					
Mod. I	0.086	0.086	0.093	0.161^{**}	0.082	0.095	0.071	0.058
U_π	0.094^{*}	0.072	0.132^{*}	$\mathit{0.178}^{**}$	$\underline{0.155^{*}}$	0.092	$\mathbf{0.205}^{*}$	$\underline{0.142}$
U_s	0.102	0.132^{*}	0.056	0.064	0.082	0.082	0.018	0.045
U_v	0.004	0.005	0.102	0.118	0.032	-0.008	0.032	-0.058
U_f	0.087	0.004	0.057	0.102	0.092	-0.071	0.053	-0.071
U_e	0.156^{**}	0.133	0.080	0.049	$\mathit{0.142}$	0.055	0.055	0.018
U_p	0.132^{*}	0.103	0.110^{*}	0.125^{*}	$\mathbf{0.282}^{**}$	$\mathbf{0.192}$	-0.008	-0.045
			D: Mean Model					
Mean	0.102	0.140	0.148^{*}	0.034	0.105	0.016	0.029	-0.121^{*}

Note: The table reports the Kiupers score as the difference between the hit-rate and the false-alarm rate. We also report the test results on predictive failure using the test proposed by Pesaran (2015) with HAC standard errors. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Pesaran-Timmermann test. The 1st, 2nd and 3rd hightest Kiupers Scores are highlighted by bold, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 96$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table 2: Kiupers score and Pesaran-Timmermann test results for growth of aggregate M2 using a rolling-window with width $T_s = 96$ observations.

that a structural break has in fact occurred. In order to account for this potential issue, we will repeat our analysis for two different window-widths.

In Tables A1 and A2 in the Appendix, we report the relative RMSE and Diebold-Mariano test results based on a rolling-window forecast with width $T_s = 84$ and $T_s = 120$ monthly observations, respectively.

Using the smaller rolling-window with width of 84 monthly observations, we can confirm that in terms of forecast accuracy Model I after S2G indicator selection still performs well. For the period between 2000 and 2014, this model ranks first both at the 1- and 2-month forecast horizon, respectively, and second at both the 6- and 9-month forecast horizon, respectively. The winning model at the 6- and 9-month horizons is the VXO model which, however, outperforms Model I only by a small margin. Among the top-3 models we also find the macroeconomic uncertainty model, the mean model, the financial uncertainty model and the policy uncertainty model. Considering only the period since 2007m12 underlines the forecast accuracy performance of Model I combined with S2G indicator selection: this model ranks first at the 1-, 2- and 9-month horizon and second at the 6-month horizon at which the VXO model ranks first. Lastly, we can always reject the null of the Diebold-Mariano test for all top-3 models at every horizon considered.

Table A2 reports the results using a longer rolling-window width of 120 monthly observations. Most importantly, this time the Diebold-Mariano test cannot be rejected for any of the models with a relative RMSE below unity indicating that none of the competitor models systematically outperforms the benchmark. Second, one can see that the gain in automatic indicator selection decreases substantially such that the gain in forecast accuracy does not exceed 10% compared to the benchmark ARDL(12,12,12) model. Nevertheless, for the period between 2000m1 and 2014m12, Model I combined with automatic indicator selection ranks best one time, second one time and third two times. The mean model yields the most accurate results at the 1- and 9-month horizon, respectively, while Model I combined with G2S indicator selection wins at the 2-month horizon and the VXO measure ranks first at the 6-month horizon. However, for the period since 2007m1, Model I ranks again two times best followed by the stock market risk premium (U_s) model.

In terms of direction-of-change forecasts we can confirm the rather weak performance of the standard money demand model specification. Tables A3 and A4 summarize the results of the Kuipers Score and Pesaran-Timmermann test, respectively, for both the rolling-window based forecasts with width $T_s = 84$ and $T_s = 120$ monthly observations. As depicted in Table A3, Model I ranks only once on the third place (and never better) when considering the period between 2000 and 2014 at the 6-month horizon. The winning models are the policy uncertainty model, the stock market risk premium model and the inflation uncertainty model at the 1-, 2-, 6- and 9-month forecast horizon, respectively. Among the second best models we find the financial uncertainty model, the economic policy model and the mean model. Also for the period since 2007 we find that the policy uncertainty model and the inflation uncertainty model rank best while the standard money demand model performs rather poorly. These results are robust against the use of a larger window of width $T_s = 120$ monthly observations as reported in Table A4. In this latter case we find even stronger support for the relevance of the stock market premium and the macroeconomic uncertainty series for direction-of-change forecasts.

Lastly, we replace growth of aggregate M2 by growth of M4 Divisia as computed by Barnett et al. (2013). Table A5 shows the relative RMSE and Diebold-Mariano test results based on a rolling-window with width $T_s = 96$. The economic uncertainty model dominates at all horizons for the period between 2000m1 and 2014m12, and outperforms Model I after S2G indicator selection by about 3% to 5% at each horizon. This latter model ranks second at the 1-, 2- and 6-month horizon, respectively. Among the top-3 models we also find the mean model and the financial uncertainty model. Very similar results are obtained for the recent GFC period between 2007m1 and 2014m12 even though the inflation uncertainty model can also be found among the top-3 models during this period.

With regard to direction-of-change forecasts we can confirm the poor performance of standard money demand formulations as shown in Table A6. Irrespective of the chosen sample, one can see that the macroeconomic uncertainty model, the financial uncertainty model and the stock market model dominate the race. Among the top-3 models we also find the policy uncertainty model and, at least for the period between 2000m1 and 2014m12, the mean model. Furthermore, the PT test can always be rejected at least at the 10% level for the top-3 models at any horizon and for both sub-samples.

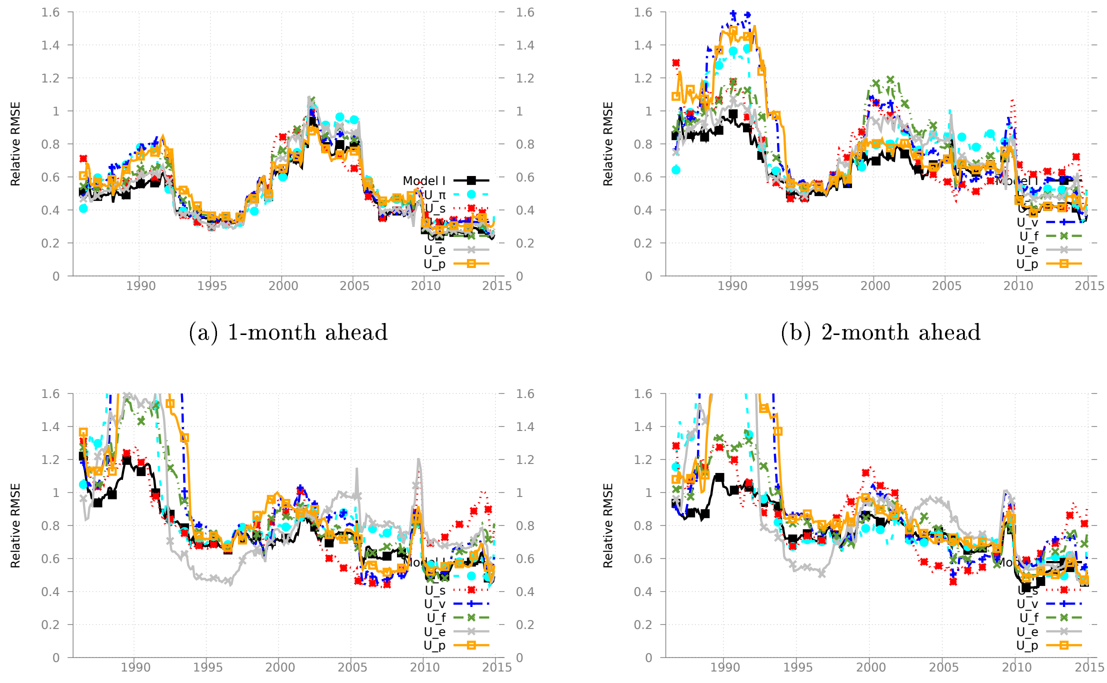
In contrast, while the PT test can be rejected for the standard money demand Model I with G2S indicator selection at each horizon for the sample since 2000, all Model I variants suffer from predictive failure for the period between 2007m1 and 2014m12. This underlines the relevance of specific uncertainty series for forecasting direction-of-changes of growth of the aggregate money stock in the U.S.—especially during the recent GFC period.

5 Forecasting results for households' demand for M2

Recently, Cook and Choi (2007) reported evidence that the money demand behaviour for M2 is not homogeneous across the different sectors in the U.S. The authors show that U.S. households react differently in their demand for M2 to certain types of financial market risk compared to the U.S. non-financial firm sector. In this section, we follow the arguments put forward by Cook and Choi (2007), and proceed with our forecasting exercise for growth of real U.S. household sector M2. Regarding the definition of this sub-aggregate of M2, we closely follow Cook and Choi (2007) and Größl and Tarassow (2015). See the Data Appendix for further details.

5.1 ARDL forecasts

Figure 6 replicates Figure 4 for growth of real household-sector M2. While all competitor models clearly outperform the benchmark ARDL(12,12,12) in terms of RMSE at the 1-month forecast horizon over the whole time span considered, this is not necessarily the case for longer forecast horizons. As can be seen, there is large variance across models in relative RMSE between 1985 and 1994, around the high-volatility periods in 2000/1, and lastly between 2009 and 2010 for forecast horizons longer than 1 month. However, most competitor models tend to yield more accurate forecasts, independent of the horizon considered, relative to the benchmark for most periods since 1995.



(a) 1-month ahead

(b) 2-month ahead

(c) 6-month ahead

(d) 9-month ahead

Note: These are the RMSEs relative to the RMSE of the benchmark ARDL(12,12,12) model. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 96$ observations. Based on 420 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1980m1, we run another rolling-window with width $T_e = 60$ observations to construct the time-varying relative RMSE with 359 observations. The first observation of the relative RMSE is given for 1985m2.

Figure 6: Out-of-sample rolling-window forecasting accuracy results after *S2G* model selection for growth in household-sector M2.

In Table 3 we replicate Table 1 and report the results on forecast accuracy and the Diebold-Mariano test on equal predictive accuracy relative to the benchmark ARDL(12,12,12) Model I. Again, we conduct this analysis for all forecasts between 2000m1 and 2014m12 as well as 2007m1 and 2014m12 to allow for eventual breaks in the forecast performance of models due to the GFC event.

For the period between 2000m1 and 2014m12 we see that indicator selection results in a substantial improvement in forecast accuracy for all models and horizons. For instance, at the 1-month forecast horizon, Model I combined with S2G indicator selection yields a reduction in RMSE by about 65% relative to the ARDL(12,12,12) counterpart. The winning model, however, is the mean model with a relative RMSE of 0.344. The mean model also provides the most accurate forecasts (0.478) at the 2-month horizon closely followed by Model I combined with G2S indicator selection (0.49). Model I combined with S2G indicator selection also dominates at both the 6- (0.634) and 9-month (0.611) forecast horizon, respectively, closely followed by the inflation uncertainty model, the mean model and the policy uncertainty model. We can reject the null of the Diebold-Mariano test for almost all models at the 1-month horizon. At the 2-month horizon, the null can only be rejected for the winning mean model, Model I combined with G2S indicator selection and the inflation uncertainty model (G2S). Surprisingly, both for the 6- and 9-month forecast horizon, the DM test cannot be rejected for any model with a relative RMSE below unity.

We obtain very similar results when focusing on the period between 2007 and 2014. Again the mean model ranks first at the 1- and 2-month horizons, the inflation uncertainty model dominates at the 6-month horizon and Model I combined with S2G indicator selection ranks first at the 9-month horizon. This latter model ranks second at the 1- and 6-month horizon and third at the 2-month horizon. While we can reject the null of the DM test for the top-3 models at the 1-month horizon at least at the 5% level, the null of this test can only be rejected for the inflation uncertainty model (G2S) and the mean model at the 2-month forecast horizon (both at least at the 10% level). However, one of the models provides statistically different forecast accuracy compared to the benchmark Model I at longer horizons.

In total, at the first two forecast horizons, the mean model, Model I combined with automatic indicator selection and the inflation uncertainty model dominate. At longer horizons, Model I combined with S2G indicator selection and the inflation uncertainty model yield substantial improvements in RMSE even though these latter results must be interpreted carefully given that the null of the DM test cannot be rejected for any forecasting model at both the 6- or 9-month horizon, respectively.

In our robustness Section 5.3 we show that in terms of forecast accuracy of growth of household-sector M2, the top-3 models are the mean model, the inflation uncertainty model (with some automatic indicator selection) and Model I (with some automatic indicator selection). This holds for shorter as well as longer rolling-window widths.

5.2 Direction-of-change forecasts based on ARDL

Figure 7 replicates Figure 5 for growth of household-sector M2. We depict the rolling-window based Kuipers Score for the period between 1985 and 2014. Again we find some differences across models in the KS values over time.

At the 1-month forecast horizon the KS values vary around a cross-sectional mean value of about 0.4 for the period between 1988 and 2000. After 2000 one can observe a decline in the average KS values reaching its historical low close to zero in 2007. This deterioration in direction-of-change forecasting remains for most models until the end of the sample. Interestingly, at least at the 1-month horizon some models are able to improve their KS values after 2010 reaching again a level close to 0.4 again.

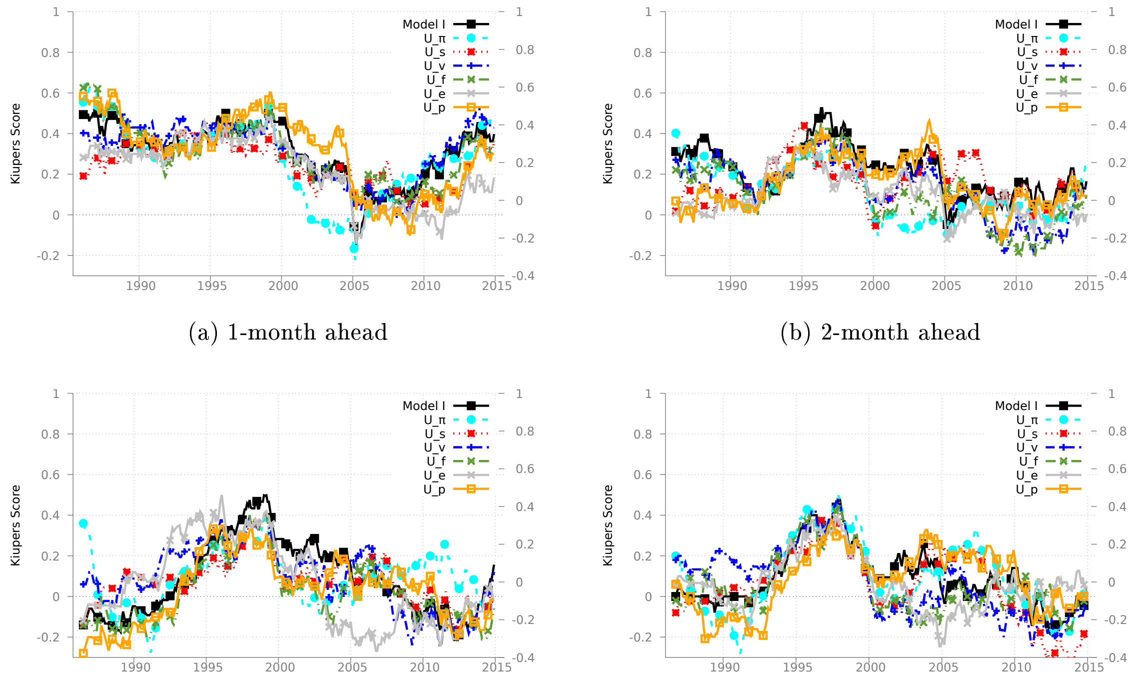
For longer forecast horizons, we observe that most models yield very low KS values close to zero or even negative for the period between 1986 and 1993. Most models are able to improve their KS values to about 0.3 during the rather tranquil (in terms of low unconditional variance)

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
	2000m1–2014m12				**2007m1–2014m12**			
			A: Full Model					
U_π	1.167	1.193*	0.914	0.873	1.206	1.208	0.837	0.798
U_s	1.319***	1.321*	1.545**	1.429**	1.380**	1.373*	1.663**	1.565***
U_v	1.465**	1.686***	1.789**	1.836*	1.651**	1.844***	1.897**	1.990*
U_f	1.360***	1.404**	0.998	0.944	1.435**	1.475*	0.895	0.863
U_e	1.303***	1.319**	1.123	1.064	1.302**	1.362**	1.102	1.018
U_p	1.123	1.131	1.223*	1.238	1.111	1.159	1.247*	1.279
			B: G2S					
Mod. I	*0.375***	0.490*	0.803	0.861	0.329*	0.458	0.861	0.869
U_π	0.439*	0.566*	0.786	0.769	0.371*	0.499*	0.779	0.744
U_s	0.468*	0.652	1.099	1.158	0.415*	0.591	1.191	1.242
U_v	0.498	0.775	1.076	1.393	0.460	0.757	1.102	1.507
U_f	0.410*	0.663	0.760	0.870	0.365*	0.631	0.729	0.818
U_e	0.471*	0.670	0.848	0.811	0.380*	0.593	0.800	0.744
U_p	0.386**	0.580	1.036	0.991	*0.315***	0.524	1.077	1.023
			C: S2G					
Mod. I	0.355**	*0.528*	**0.634**	**0.611**	0.311**	*0.481*	0.609	**0.578**
U_π	0.400**	0.597	0.640	0.643	0.344*	0.529	**0.586**	0.606
U_s	0.414*	0.679	0.856	0.830	0.387*	0.683	0.914	0.870
U_v	0.394**	0.608	0.680	0.714	0.353*	0.586	0.691	0.717
U_f	0.385**	0.608	0.749	0.730	0.328**	0.560	0.742	0.717
U_e	0.380**	0.602	0.795	0.698	0.331*	0.558	0.764	0.663
U_p	0.416*	0.570	0.685	*0.672*	0.372*	0.527	0.691	*0.644*
			D: Mean Model					
Mean	**0.344****	**0.478***	*0.666*	0.684	**0.298****	**0.433***	*0.661*	0.664

Note: The table reports the RMSE relative to an ARDL(12,12,12) comprising lagged values of the endogenous money growth variable, a real income measure and an opportunity cost measure. We also report the test results on equal predictive accuracy using the test proposed by Diebold and Mariano (1995) under a quad-quad loss function. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Diebold-Mariano test. The 1st, 2nd and 3rd lowest relative RMSEs are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 96$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table 3: Relative RMSE and Diebold-Mariano test results for growth of household-sector M2 using a rolling-window with width $T_s = 96$ observations.

period of real money growth during the mid 1990s. Since 2000, however, the average KS values are close to zero for forecast horizons larger than 1 month. This indicates that direction-of-change forecasting growth of real household-sector M2 is hard.



(a) 1-month ahead

(b) 2-month ahead

(c) 6-month ahead

(d) 9-month ahead

NOTE: These are the forecasting accuracy results using a rolling-window estimation method. For the sample 1971m1 to 2014m12 the initial training set uses 120 observations from 1971m1 to 1980m12 to determine the optimal lag length before computing the h-multi-step iterated forecasts. Next, the beginning and end of the training set is extended by one additional observation such that it ranges from 1971m2 to 1981m1. Again, the optimal lag length based on the extended training set is determined before computing the h-multi-step iterated forecasts. The last training set is reached for the sample end 2014m11 to compute at maximum $h = 1$ multi-step iterated forecasts. The number of the out-of-sample observations is 396. Based on the sequence of 396 out-of-sample forecasts, we compute the average RMSE for a sequence of h-multi-step forecasts for a specific period, e.g. 1971m1 to 1999m12.

Figure 7: Out-of-sample rolling-window Kuipers Score after *S2G* model selection for growth in household-sector M2.

Table 4 summarises the KS statistics and Pesaran-Timmermann test results and replicates Table 2 for household-sector M2. For the period between 2000m1 and 2014m12 and at the 1-month forecast horizon, Model I combined with G2S indicator selection ranks first with a KS value of 0.349, closely followed by the mean model (0.333) and the benchmark model (0.323). Interestingly, the benchmark model yields the highest KS value (0.223) at the 2-month horizon followed by the policy uncertainty model combined with G2S indicator selection (0.221). For all top-3 models and at both the 1- and 2-month forecast horizon, respectively, we can reject the null of the PT test at the 1% level.

The winning model at the 6-month horizon is the inflation uncertainty model (0.128) followed by the VXO model combined with G2S indicator selection (0.117). At the 9-month horizon the policy uncertainty model (S2G) ranks first with a rather low KS value of 0.099 closely followed by the financial uncertainty model (G2S; 0.089). However, at both the 6- and 9-month forecast horizons we cannot reject the null of the PT test for any model indicating serious predictive failures in terms direction-of-change forecasts for all model specifications.

These results are robust against the use of the smaller sample ranging from 2007 to 2014. Again, at the 1-month horizon Model I combined with S2G indicator selection yields a rather high KS value of 0.402 closely followed by the mean model (0.379) and the policy uncertainty model (0.36). At the 2-month horizon, though, the policy uncertainty model (G2S) outperforms the benchmark with a KS value of 0.234 against 0.191. Again the PT test can be rejected for all top-3 models at both the 1- and 2-month forecast horizons, respectively. At the 6-month horizon, the stock market risk premium model ranks first with a low KS value of 0.04 while the financial uncertainty model wins at the 9-month horizon with a KS value of 0.127. However, the PT test indicates serious predictive failures for all models at the 6- and 9-month horizon, respectively. Nevertheless, the standard money demand specification model never ranks among the top-3 at both the 6- and 9-month horizons.

In total, these results indicate that in terms of direction-of-change forecasts for growth of real household sector M2 the inflation uncertainty index, the financial uncertainty index, the policy uncertainty index and the mean model provide relevant forecasting information.

The robustness exercise in Section 5.3 will show that especially the policy uncertainty measure, the inflation uncertainty measure, the stock market risk premium and partly the VXO series provide additional information for forecasting growth in household-sector M2 since 2000. This conclusion can be drawn for different widths on which our rolling-window forecasting exercise is based on.

5.3 Robustness exercise

Tables A7 and A8 in the Appendix report the relative RMSE and Diebold-Mariano test results for both a rolling window of width $T_s = 84$ and $T_s = 120$, respectively.

For the shorter window width we find that in terms of forecast accuracy Model I combined with S2G indicator selection ranks first at the 1-, 6- and 9-month horizon and second at the 2-month horizon using all forecasts between 2000m1 and 2014m12. Among the top-3 models we also find the mean model (ranks first at the 2-month horizon), the VXO model, the inflation uncertainty model and the economic uncertainty model. The results for the period between 2007m1 and 2014m12 are similar with the only exception being the VXO model which does not rank among the top-3 models at any horizon now. Again the stock market risk premium model does not rank among the top-3 models in any case.

According to the DM test, we can reject the null for all models at the 1-month horizon for which the relative RMSE is below one. Interestingly, at the 2-month horizon, the DM test can only be rejected at least at the 10% level for the mean model and Model I after S2G indicator selection. Furthermore, none of the competitor models yields systematically more accurate forecasts at the 6- and 9-month horizons compared to the benchmark Model I according to the DM test. These findings hold for both samples considered.

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
	2000m1–2014m12				**2007m1–2014m12**			
	A: Full Model							
Mod. I	*0.323****	**0.223*****	−0.039	−0.042	0.359***	*0.191**	−0.208*	−0.145
U_π	0.185**	0.136*	**0.128**	0.015	0.252**	0.024	*0.023*	−0.061
U_s	0.224***	0.149*	*0.080*		0.188*	0.061	**0.040**	−0.106
U_v	0.295***	0.111	0.013	0.037	0.290***	0.021	−0.170	−0.042
U_f	0.241***	0.127*	0.045	0.002	0.290***	0.145	−0.002	−0.023
U_e	0.108	0.085	−0.123	−0.036	0.126	0.086	−0.225*	−0.015
U_p	0.256***	*0.189****	−0.045	−0.025	0.316***	0.212**	−0.119	−0.054
	B: G2S							
Mod. I	**0.349*****	0.152**	−0.111	−0.152*	**0.402*****	0.067	−0.247**	−0.267**
U_π	0.264***	0.183**	0.021	−0.027	0.338***	0.065	−0.060	−0.123
U_s	0.146**	0.128*	−0.027	−0.113	0.106	0.062	−0.147	−0.269**
U_v	0.194**	0.144	0.117	0.033	0.232**	0.127	−0.041	−0.083
U_f	0.217***	0.017	−0.043	0.089	0.313***	0.020	−0.079	**0.127**
U_e	0.079	0.108	−0.040	0.006	0.127	0.149	0.028	−0.031
U_p	0.269***	0.221***	−0.033	*0.081*	*0.360****	**0.234*****	−0.119	0.008
	C: S2G							
Mod. I	0.199***	0.102	0.040	0.018	0.315***	0.046	−0.062	−0.082
U_π	0.211***	0.084	0.065	0.046	0.357***	0.065	0.019	−0.102
U_s	0.162**	0.139*	0.019	−0.013	0.171*	0.004	−0.124	−0.250**
U_v	0.235***	0.030	0.024	−0.051	0.336***	−0.038	−0.082	−0.123
U_f	0.219***	0.031	−0.004	−0.007	0.251**	−0.040	−0.085	−0.001
U_e	0.057	−0.009	−0.111	−0.053	0.090	−0.074	−0.120	−0.055
U_p	0.141*	0.099	0.027	**0.099**	0.191**	0.067	−0.058	−0.017
	D: Mean Model							
Mean	0.333***	0.144*	0.002	−0.071	0.379***	0.086	−0.163	−0.226*

Note: The table reports the Kiupers score as the difference between the hit-rate and the false-alarm rate. We also report the test results on predictive failure using the test proposed by Pesaran (2015) with HAC standard errors. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Pesaran-Timmermann test. The 1st, 2nd and 3rd hightest Kiupers Scores are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 96$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table 4: Kiupers score and Pesaran-Timmermann test results for growth of household-sector M2 using a rolling-window with width $T_s = 96$ observations.

Using a longer window width of $T_s = 120$ observations, we find that in terms of forecast accuracy the mean model dominates at the 1- and 2-month horizons, respectively (see Table A8), and outperforms Model I (with our without indicator selection) by at least 10%. At both the 6- and 9-month horizon, respectively, the inflation uncertainty model dominates closely followed Model I with S2G indicator selection. Among the top-3 models we also find the policy uncertainty model and the VXO model. Again these findings hold for both samples considered.

While the null of the DM test can be rejected for all models combining one of the two indicator selection algorithms at the 1-month horizon, the picture is much more heterogeneous at longer horizons. For instance, at the two month horizon the null of the DM test can only be rejected by Model I (G2S and S2G), the inflation uncertainty model, the VXO model, the economic policy uncertainty model and the mean model. Interestingly, the null of the DM test cannot be rejected for any model at the 6-month horizon even though we can reject this test for the inflation uncertainty model at the 9-month horizon. This latter results shows that inflation uncertainty provides statistically relevant information for medium-term forecasts of growth of real household-sector M2 since 2000. These results also hold for the smaller sample period since 2007.

The direction-of-change forecast evaluation results are reported in Tables A9 and A10. Based on a rolling-window with width $T_s = 84$ monthly observations and evaluating the sample between 2000m1 and 2014m12, we see that Model I after G2S indicator selection ranks first closely followed by the VXO model. At the 2-month horizon the stock market risk premium model (S2G) wins before the macroeconomic policy uncertainty model. At both the 6- and 9- month horizons the economic policy uncertainty model (S2G) yields the highest positive KS values. Considering only forecasts for the period since 2007m1, the inflation uncertainty model ranks first at the 1-month horizon, the policy uncertainty model wins at the 2-month horizon while the policy uncertainty model provides the highest KS values again at both the 6- and 9-month forecast horizons, respectively.

Irrespective of the chosen sub-sample, the standard money demand specification suffers from predictive failures for forecast horizons longer than 1 month. Again, we find that considering one of the before-mentioned uncertainty variables provides relevant information for direction-of-change forecasts. These results are confirmed when using a longer rolling-window width, as reported in Table A10.

6 Conclusion

This paper investigates the predictive power of different uncertainty measures to out-of-sample forecast growth of U.S. aggregate real M2, U.S. aggregate real Divisia M4 as well as household-sector monetary real M2 using monthly data between 1971m1 and 2014m12. Our findings support recent arguments that private actors' liquidity preferences are correlated with economic uncertainty dynamics. However, instead of evaluating the in-sample fit of risk-augmented money demand models, we focus on the out-of-sample forecast content of various economic uncertainty measures.

In particular, we compare both forecast accuracy and direction-of-change forecast ability of an otherwise standard benchmark ARDL money demand model comprising lagged values of a real monetary measure, a real income measure and an opportunity cost measure with an augmented money demand model additionally including lagged values of some uncertainty measure. The list of potential uncertainty measures comprises recent time series heavily applied in the business cycle literature and others which have been used in the money demand literature before. Lastly, we combine two different indicator selection procedures with a rolling-window forecasting scheme to detect the optimal ARDL model specification in a recursive manner.

With regard to forecasting growth of real aggregate M2, we find that the benchmark money demand specification provides reasonable forecast accuracy at the short- and medium-term horizon for the period between 2000 and 2014. Nevertheless, the implied volatility VXO capital market risk measure and the macroeconomic uncertainty series help to improve forecast accuracy. These findings are confirmed for forecasting growth of real aggregate M4 Divisia. This standard money demand specification, however, performs poorly in terms of direction-of-change forecasts since 2000. Especially the inflation uncertainty series, the financial market uncertainty series and economic policy uncertainty yield an improvement regarding this type of forecast.

The standard money demand specification also provides reasonable short- and medium-term forecast accuracy for growth of real household-sector M2 since the year 2000. However, the inflation uncertainty series, the macroeconomic uncertainty index and simple model combination help to improve point forecast accuracy. The standard model again performs poorly for this monetary aggregate in terms of direction-of-change forecast. However, the consideration of inflation uncertainty, economic policy uncertainty, stock market risk premium or financial market uncertainty provides relevant forecasting information content—especially for the period since 2007 covering the recent GFC event.

In total our results stress the relevance of economic uncertainty variables for forecasting growth of monetary aggregates in the U.S. We find that specific economic uncertainty measures are helpful to improve forecast accuracy as well as direction-of-change forecasts of money growth. Thus, improved forecasts of monetary growth may strengthen the purpose of (expected) monetary growth dynamics to provide timely information about central macroeconomic variables such as the inflation gap or the output gap which are measured imperfectly Coenen et al. (2005). Furthermore, and as recently stressed by Ball (2012), understanding money demand in a zero interest rate and quantitative easing environment will tell the central bank how much it must reduce the monetary base to raise interest rates.

References

Bachmann, R., S. Elstner, and E. R. Sims (2013, April). Uncertainty and Economic Activity: Evidence from Business Survey Data. *American Economic Journal: Macroeconomics 5*(2), 217–249.

Baker, S. R., N. Bloom, and S. J. Davis (2013). Measuring Economic Policy Uncertainty. Technical report.

Ball, L. (2012). Short-run money demand. *Journal of Monetary Economics 59*(7), 622–633.

Barnett, W., J. Liu, R. Mattson, and J. Noort (2013, February). The New CFS Divisia Monetary Aggregates: Design, Construction, and Data Sources. *Open Economies Review 24*(1), 101–124.

Beck, G. W. and V. Wieland (2007, 04-05). Money in Monetary Policy Design: A Formal Characterization of ECB-Style Cross-Checking. *Journal of the European Economic Association 5*(2-3), 524–533.

Benhabib, J., S. Schmitt-Grohe, and M. Uribe (2001, January). The Perils of Taylor Rules. *Journal of Economic Theory 96*(1-2), 40–69.

Bernanke, B. S. (2006). Monetary aggregates and monetary policy at the federal reserve: A historical perspective. In *At the Fourth ECB Central Banking Conference, Frankfurt.*

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica 3*, 623–685.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. Terry (2013). Really uncertain business cycles. LSE Research Online Documents on Economics 51526, London School of Economics and Political Science, LSE Library.

Bollerslev, T., G. Tauchen, and H. Zhou (2009, November). Expected Stock Returns and Variance Risk Premia. *Review of Financial Studies 22*(11), 4463–4492.

Bossone, B. (2014). Secular stagnation. Economics Discussion Papers 2014-47, Kiel Institute for the World Economy.

Caballero, R. J. and E. Farhi (2013, January). A Model of the Safe Asset Mechanism (SAM): Safety Traps and Economic Policy. NBER Working Papers 18737, National Bureau of Economic Research, Inc.

Carpenter, S. B. and J. Lange (2003). Money Demand and Equity Markets. FEDS Working Paper 2003-03, Board of Governors of the Federal Reserve System.

Carstensen, K. (2006). Stock Market Downswing and the Stability of European Monetary Union Money Demand. *Journal of Business and Economic Statistics 24*(4), 395–402.

Christiano, L. J. and M. Rostagno (2001, October). Money Growth Monitoring and the Taylor Rule. NBER Working Papers 8539, National Bureau of Economic Research, Inc.

Chua, C. L., D. Kim, and S. Suardi (2011). Are empirical measures of macroeconomic uncertainty alike? *Journal of Economic Surveys 4*, 801–827.

Clark, T. E. and M. W. McCracken (2001, November). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics 105*(1), 85–110.

Coenen, G., A. Levin, and V. Wieland (2005). Data Uncertainty and the Role of Money as an Information Variable for Monetary Policy. *European Economic Review 49*(4), 975–1006.

Cook, D. and W. G. Choi (2007, April). Financial Market Risk and U.S. Money Demand. IMF Working Papers 07/89, International Monetary Fund.

Cottrell, A. and R. Lucchetti (2016). *Gretl User's Guide.*

Cronin, D., R. Kelly, and B. Kennedy (2011, May). Money growth, uncertainty and macroeconomic activity: a multivariate GARCH analysis. *Empirica 38*(2), 155–167.

D'Amico, S. and A. Orphanides (2008). Uncertainty and disagreement in economic forecasting. Finance and Economics Discussion Series 2008-56, Board of Governors of the Federal Reserve System (U.S.).

de Bondt, G. J. (2009). Euro area money demand - empirical evidence on the role of equity and labour markets. Working Paper Series 1086, European Central Bank.

Diebold, F. X. and R. S. Mariano (1995, January). Comparing predictive accuracy. *Journal of Business and Economic Statistics 20*(1), 134–144.

Fama, E. F. and K. R. French (1988, October). Dividend yields and expected stock returns. *Journal of Financial Economics 22*(1), 3–25.

Greiber, C. and W. Lemke (2005). Money Demand and Macroeconomic Uncertainty. Discussion Paper: Series 1 26/2005, Deutsche Bundesbank.

Grimme, C., S. Henzel, and E. Wieland (2011). Inflation uncertainty revisited: Do different measures disagree? Ifo Working Paper Series Ifo Working Paper No. 111, Ifo Institute for Economic Research at the University of Munich.

Größl, I. and U. Fritsche (2006). The store-of-value-function of money as a component of household risk management. Macroeconomics and Finance Series 200606, Hamburg University, Department Wirtschaft und Politik.

Größl, I. and A. Tarassow (2015, May). A Microfounded Model of Money Demand Under Uncertainty, and its Empirical Validation Using Cointegration and Rolling-Window Dynamic Multiplier Analysis. Macroeconomics and Finance Series 201504, Hamburg University, Department Wirtschaft und Politik.

Harvey, D., S. Leybourne, and P. Newbold (1997, June). Testing the equality of prediction mean squared errors. *International Journal of Forecasting 13*(2), 281–291.

Higgins, M. and S. Majin (2009). Inflation uncertainty and money demand. *Applied Economics Letters 16*(13), 1323–1328.

Inoue, A., L. Jin, and B. Rossi (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics 196*(1), 55–67.

Jurado, K., S. C. Ludvigson, and S. Ng (2015, March). Measuring Uncertainty. *American Economic Review 105*(3), 1177–1216.

Ludvigson, S. C., S. Ma, and S. Ng (2015, December). Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response? NBER Working Papers 21803, National Bureau of Economic Research, Inc.

Marcellino, M., J. H. Stock, and M. Watson (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of Econometrics 135*(1-2), 499–526.

Mc Cracken, M. W. (2000, December). Robust out-of-sample inference. *Journal of Econometrics 99*(2), 195–223.

McCracken, M. W. and S. Ng (2015, June). FRED-MD: A Monthly Database for Macroeconomic Research. Working Papers 2015-12, Federal Reserve Bank of St. Louis.

Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Number 9780198759980 in OUP Catalogue. Oxford University Press.

Pesaran, M. H. and A. Timmermann (1992, October). A Simple Nonparametric Test of Predictive Performance. *Journal of Business & Economic Statistics 10*(4), 561–65.

Pesaran, M. H. and A. Timmermann (2007, March). Selection of estimation window in the presence of breaks. *Journal of Econometrics 137*(1), 134–161.

Rapach, D. E. and J. K. Strauss (2008). Forecasting US employment growth using forecast combining methods. *Journal of Forecasting 27*(1), 75–93.

Rossi, B., T. Sekhposyan, and M. Soupre (2016, July). Understanding the Sources of Macroeconomic Uncertainty. Working Papers 920, Barcelona Graduate School of Economics.

Seitz, F. and J. von Landesberger (2014, November). Household Money Holdings in the Euro Area: An Explorative Investigation. *Journal of Banking and Financial Economics 2*(2), 83–115.

Stock, J. and M. Watson (2005). Has inflation become harder to forecast? mimeo, Harvard University.

Stock, J. H. and M. W. Watson (2003). Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature 41*(3), 788–829.

Stock, J. H. and M. W. Watson (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit, and Banking 39*, 3–33.

Stock, J. H. and M. W. Watson (2012). Disentangling the Channels of the 2007-09 Recession. *Brookings Papers on Economic Activity 44*(1 (Spring)), 81–156.

Söderström, U. (2005, November). Targeting Inflation with a Role for Money. *Economica 72*(288), 577–596.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting 16*(4), 437–450.

Wright, J. H. (2011, June). Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset. *American Economic Review 101*(4), 1514–34.

Data Appendix

Most series were collected from the FRED-MD monthly Database for Macroeconomic Research provided by McCracken and Ng (2015) or from the Financial Accounts of the United States. The variables are defined as follows:

Real income, y_t, is the logarithm of real personal income (FRED: RPI, monthly, SA).

The opportunity cost measure, i_t, is the difference between the 3-Month Treasury Bill (FRED: TB3MS, NSA monthly) minus the own rate of M2 (FRED: M2OWN, NSA, monthly).

The real monetary aggregate M2, m_t, is the logarithm of the real M2 money stock deflated by the CPI price index (FRED: M2REAL, SA, monthly).

The real monetary aggregate M2 of the household sector, m_t, is the logarithm of the difference between the nominal aggregate money stock M2 (FRED: M2NS, NSA, monthly) minus the sum of the following M2 components of the four non-household sectors:

1. Non-financial business sector
   - Checkable deposits and currency (FoF: FL143020005, quarterly, NSA)
   - Time and savings deposits (FoF: FL143030005, quarterly, NSA)
   - Money market fund shares (FoF: FL143034005, quarterly, NSA)
2. General government sector
   - Checkable deposits and currency (FoF: FL363020005, quarterly, NSA)
   - Time and savings deposits (FoF: FL363030005, quarterly, NSA)
   - Money market fund shares (FoF: FL213034003, quarterly, NSA)
3. Domestic financial sector
   - Foreign deposits (FoF: FL633091003, quarterly, NSA)
   - Checkable deposits and currency (FoF: FL793020005, quarterly, NSA)
   - Time and savings deposits (FoF: FL793030005, quarterly, NSA)
   - Money market fund shares (FoF: FL793034005, quarterly, NSA)
4. Rest of the World
   - U.S. checkable deposits and currency (FoF: FL263020005, quarterly, NSA)
   - U.S. time deposits (FoF: FL263030005, quarterly, NSA)
   - Money market fund shares (FoF: FL793034005, quarterly, NSA)

The quarterly series are expanded to monthly frequency using a cubic spline method. The resulting nominal series is deflated by the PCE (FRED: PCEPI, SA, monthly) price deflator before taking the logarithm. The logarithmic series is de-seasonalised by means of the X-12-Arima procedure.

The stock market risk premium, U_s, is the logarithm of the ratio of the dividend yield over the 10-year Treasury Rate (FRED: GS10, NSA, monthly). The dividend yield is computed by the dividends paid (from Shiller `http://www.econ.yale.edu/~shiller/data/ie_data.xls`, NSA, monthly) divided by the lagged S&P 500 stock price index (FRED: SP500, NSA, monthly).

The economic policy uncertainty measure, U_e, is constructed by Baker et al., and can be downloaded from `http://www.policyuncertainty.com/media/US_Policy_Uncertainty_Data.xlsx` (monthly, NSA).

The VXO implied volatility index, U_v, is the CBOE S&P 100 volatility index (FRED: VXOCLSx, NSA, monthly).

The macroeconomic uncertainty index, U_e, proposed by Jurado et al. (2015) can be obtained from `https://www.sydneyludvigson.com/s/MacroFinanceUncertainty_2016Aug_update.zip`

The financial market uncertainty index, U_f, proposed by Ludvigson et al. (2015) can be obtained from `https://www.sydneyludvigson.com/s/MacroFinanceUncertainty_2016Aug_update.zip`

The inflation uncertainty measure, U_π, is estimated by means of the Stock and Watson (2007) unobserved-component-stochastic-volatility (UCSV) model. Inflation, x_t, is measured by the annualized growth rate of the CPI price index (FRED: CPIAUCSL, SA, monthly). The setup of the UCSV model can be described as follows: It is assumed that the series of interest, x_t, can be decomposed into a permanent and transitory component with time-varying volatility. Allowing for time-variations is based on the empirical fact that parameter shifts in the estimated variances of the components have occurred over time for the U.S. economy (Stock and Watson, 2007). The dynamics of inflation closely follow an integrated moving-average process which can be re-written as an unobserved component model. It is assumed that x_t is driven by a stochastic trend, τ_t, with serially uncorrelated innovations η_t. The stochastic trend is driven by another white noise innovation ϵ_t:

$$x_t = \tau_t + \eta_t \tag{7}$$

$$\tau_t = \tau_{t-1} + \epsilon_t \; . \tag{8}$$

Both innovations η_t and ϵ_t are *i.i.d* normally distributed. Furthermore, the logarithms of the variances of both the transitory part, $\sigma^2_{\eta,t}$ ($\eta_t \sim N(0, \sigma^2_{\eta,t})$), as well as permanent part, $\sigma^2_{\epsilon,t}$ ($\epsilon_t \sim N(0, \sigma^2_{\epsilon,t})$), evolve as separate random-walks according to:

$$log\ \sigma^2_{\eta,t} = log\ \sigma^2_{\eta,t-1} + \nu_{\eta,t} \tag{9}$$

$$log\ \sigma^2_{\epsilon,t} = log\ \sigma^2_{\epsilon,t-1} + \nu_{\epsilon,t} \; . \tag{10}$$

The innovations to the variances, $\nu_t = (\nu_{\eta,t}, \nu_{\epsilon,t})'$, are *i.i.d.* $N(0, \gamma I_2)$ and orthogonal to each other. The parameter γ controls the smoothness of the stochastic volatilities $\sigma^2_{*,t}$. The model is estimated using the Gibbs sampling approach. We fit the UCSV(0.2) model to our CPI price inflation time series, π_t using a prior for the initial condition of $\gamma = 0.2$. This prior was also used by Stock and Watson (2007) for GDP inflation. We found that the results were robust against different prior values.

A Robustness section for growth of real aggregate M2

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
		2000m1–2014m12				**2007m1–2014m12**		
				A: Full Model				
U_π	1.208*	1.141	1.219**	1.468***	1.082	1.111	1.102	1.446
U_s	1.503***	1.470***	1.736***	1.479***	1.535**	1.730**	2.114**	1.592**
U_v	1.429***	1.445***	1.500**	1.493***	1.493**	1.432**	1.623	1.520**
U_f	1.400***	1.387**	1.427***	1.549***	1.488**	1.707**	1.636***	1.786***
U_e	1.294*	1.223*	1.226**	1.251**	1.388	1.357	1.387***	1.218*
U_p	1.331***	1.331**	1.486***	1.664**	1.403***	1.504***	1.773***	2.135**
				B: G2S				
Mod. I	0.633**	0.742**	0.881	0.859**	0.656***	0.739***	0.892	0.880*
U_π	0.705*	0.934	0.911	1.011	0.657***	0.898	0.940	1.099
U_s	0.973	1.110	1.456**	1.282*	1.065	1.178	1.745**	1.461*
U_v	0.909	1.028	0.959	1.180	1.131	1.101	1.130	1.460**
U_f	1.023	1.168	1.277**	1.290***	1.210	1.386	1.366*	1.439**
U_e	0.901	1.051	1.056	1.107	1.045	1.110	1.185*	1.215**
U_p	0.890	1.098	1.163	1.267	1.031	1.128	1.526*	1.602**
				C: S2G				
Mod. I	**0.526***	**0.670**	0.702**	0.705**	**0.536***	**0.640***	0.673***	**0.668***
U_π	0.615**	0.770	0.842	0.880	0.657**	0.782	0.877	0.941
U_s	0.639***	0.865	0.843	0.898	0.573***	0.804	*0.810*	0.850
U_v	*0.609***	0.753	**0.701**	**0.704**	0.649**	0.797	**0.652***	0.683***
U_f	0.700**	0.816	*0.773*	0.833	0.684*	0.824	0.814	0.873
U_e	0.598**	0.773	0.821	0.832	*0.634****	0.735*	0.843	*0.839*
U_p	0.679**	0.795	0.788	*0.725**	0.733*	0.840	0.930	0.851
				D: Mean Model				
Mean	0.629***	*0.744***	0.773**	0.811**	0.659***	*0.738***	0.843	0.865

Note: The table reports the RMSE relative to an ARDL(12,12,12) comprising lagged values of the endogenous money growth variable, a real income measure and an opportunity cost measure. We also report the test results on equal predictive accuracy using the test proposed by Diebold and Mariano (1995) under a quad-quad loss function. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Diebold-Mariano test. The 1st, 2nd and 3rd lowest relative RMSEs are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 84$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table A1: Relative RMSE and Diebold-Mariano test results for growth of aggregate M2 using a rolling-window with width $T_s = 84$ observations.

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
	2000m1–2014m12				**2007m1–2014m12**			
			A: Full Model					
U_π	1.083	1.074	1.160*	1.195**	1.062	1.054	1.159	1.193
U_s	1.313**	1.360*	1.313**	1.302***	1.431*	1.502	1.345*	1.265
U_v	1.219***	1.216***	1.182***	1.170***	1.280***	1.275***	1.189**	1.188**
U_f	1.212***	1.259**	1.262***	1.230***	1.275**	1.373**	1.304**	1.288***
U_e	1.157*	1.129	1.222*	1.185	1.176*	1.187	1.261	1.164
U_p	1.209**	1.246**	1.259**	1.276**	1.383***	1.434***	1.478***	1.505***
			B: G2S					
Mod. I	0.937	**0.932**	0.999	0.998	**0.898**	1.036	1.147	1.109
U_π	0.926	*0.949*	1.085	1.051	0.948	1.046	1.286	1.138
U_s	1.125	1.176	1.267*	1.199**	1.228	1.307	1.466**	1.148
U_v	1.079	1.138	1.206*	1.089	1.136	1.215	1.307**	1.140
U_f	1.085	1.215	1.185*	1.166*	1.074	1.289	1.241	1.255*
U_e	1.033	1.095	1.158	1.086	1.008	1.105	1.336**	1.158
U_p	1.123	1.310**	1.190	1.155	1.271*	1.489***	1.470***	1.371***
			C: S2G					
Mod. I	*0.929*	0.963	*0.989*	0.977	*0.916*	*1.009*	*1.034*	**0.974**
U_π	0.969	0.987	1.145	1.078	0.945	1.022	1.314	1.236
U_s	0.965	0.996	1.024	1.115	0.940	**0.979**	**1.009**	1.090
U_v	0.954	1.013	**0.957**	*0.982*	0.982	1.106	1.031	0.980
U_f	1.009	1.088	0.993	1.039	1.006	1.203	1.036	1.058
U_e	0.970	1.015	1.115	1.106	0.905	0.984	1.264	1.125
U_p	1.010	1.116	1.070	1.052	1.070	1.193	1.159	1.077
			D: Mean Model					
Mean	**0.898**	0.943	0.969	**0.946**	0.929	1.022	1.082	*1.001*

Note: The table reports the RMSE relative to an ARDL(12,12,12) comprising lagged values of the endogenous money growth variable, a real income measure and an opportunity cost measure. We also report the test results on equal predictive accuracy using the test proposed by Diebold and Mariano (1995) under a quad-quad loss function. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Diebold-Mariano test. The 1st, 2nd and 3rd lowest relative RMSEs are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 120$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table A2: Relative RMSE and Diebold-Mariano test results for growth of aggregate M2 using a rolling-window with width $T_s = 120$ observations.

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
	2000m1–2014m12				**2007m1–2014m12**			
	A: Full Model							
Mod. I	0.112	0.120	*0.150*	0.128	0.229*	0.076	0.089	0.037
U_π	0.098	0.174*	**0.212****	**0.265*****	0.116	*0.166*	**0.339*****	**0.289****
U_s	0.120	*0.197***	0.129	0.143*	0.155	0.129	−0.011	0.003
U_v	0.046	0.100	0.114	0.151	0.047	0.047	−0.016	0.021
U_f	0.097	0.143*	0.113	0.128	0.063	0.061	0.124	*0.147*
U_e	*0.156***	<u>0.210**</u>	0.053	0.130*	0.168	0.155	−0.024	0.013
U_p	**0.173****	0.159*	0.136*	0.122	0.187	<u>0.200</u>	0.074	−0.016
	B: G2S							
Mod. I	0.117*	0.133*	0.089	0.178*	0.155	0.103	0.013	0.103
U_π	0.074	0.037	<u>0.158**</u>	<u>0.205**</u>	0.179*	0.089	<u>0.189*</u>	<u>0.239*</u>
U_s	0.104	**0.233*****	0.023	0.164*	0.155	0.153	−0.174**	−0.034
U_v	0.059	0.060	0.045	0.120	0.163	−0.013	−0.013	−0.053
U_f	0.082	0.075	0.098	0.090	0.013	−0.089	0.037	0.024
U_e	0.005	0.082	0.007	0.143*	−0.021	0.016	−0.087	0.029
U_p	0.007	0.142*	0.038	0.076	0.113	**0.216**	0.024	0.011
	C: S2G							
Mod. I	0.056	0.064	0.041	0.101*	0.068	0.055	0.018	0.008
U_π	0.057	−0.017	0.051	0.164**	0.105	−0.061	*0.129*	0.116
U_s	0.088	0.051	0.013	0.035	0.029	0.039	−0.047	−0.084
U_v	0.066	0.006	0.013	0.072	0.076	0.013	−0.047	−0.045
U_f	<u>0.171**</u>	0.172*	−0.009	0.013	0.229**	0.153	−0.037	−0.097
U_e	0.042	0.066	0.066	0.088	−0.008	−0.034	−0.071	0.029
U_p	0.132*	0.095	0.035	0.034	**0.258****	0.082	−0.097	−0.008
	D: Mean Model							
Mean	0.133*	0.126	0.119	*0.179***	0.218	0.105	0.053	0.076

Note: The table reports the Kiupers score as the difference between the hit-rate and the false-alarm rate. We also report the test results on predictive failure using the test proposed by Pesaran (2015) with HAC standard errors. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Pesaran-Timmermann test. The 1st, 2nd and 3rd hightest Kiupers Scores are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 84$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table A3: Kiupers score and Pesaran-Timmermann test results for growth of aggregate M2 using a rolling-window with width $T_s = 84$ observations.

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
	2000m1–2014m12				**2007m1–2014m12**			
			A: Full Model					
Mod. I	0.141*	*0.134**	0.088	0.066	0.092	*0.079*	−0.071	−0.058
U_π	0.059	0.052	0.091	0.106	0.039	0.076	*0.050*	−0.013
U_s	0.141**	0.112	0.074	0.090	0.103	0.076	0.039	
U_v	<u>0.150</u>	0.105	<u>0.136</u>	0.151*	0.003	0.039	−0.026	−0.026
U_f	0.082	0.044	0.090	0.120	−0.061	−0.124	−0.084	−0.021
U_e	0.097	0.090	0.053	0.039	0.042	0.055	−0.047	−0.061
U_p	0.076	0.015	0.074	0.128	−0.003	−0.037	−0.047	−0.113
			B: G2S					
Mod. I	0.042	0.035	0.058	0.043	−0.008	−0.061	−0.097	−0.111
U_π	0.013	0.059	0.007	0.151*	0.053	0.076	−0.074	0.050
U_s	0.088	0.043	0.058	*0.164***	0.066	−0.047	−0.084	*0.089*
U_v	0.037	0.059	0.014	0.021	0.003	−0.011	−0.087	−0.161**
U_f	0.058	0.081	*0.112*	**0.211****	−0.071	−0.021	−0.111	−0.147**
U_e	−0.025	0.103	0.046	0.061	−0.005	0.008	−0.011	−0.097
U_p	0.096	−0.047	0.043	−0.031	0.066	−0.124	−0.071	−0.213**
			C: S2G					
Mod. I	0.072	0.109	0.088	0.088	0.082	−0.005	0.045	0.008
U_π	0.109**	0.049	**0.201****	0.149*	<u>0.182*</u>	0.016	**0.153**	**0.139**
U_s	**0.193****	<u>0.186**</u>	0.088	<u>0.185**</u>	**0.255***	<u>0.205</u>	0.042	<u>0.132</u>
U_v	−0.003	0.126	0.096	0.104	−0.045	0.032	0.018	−0.018
U_f	0.057	0.042	0.110	0.074	−0.032	−0.071	0.005	−0.097
U_e	0.080	**0.208*****	0.096	0.119*	*0.145*	**0.282****	<u>0.082</u>	0.055
U_p	0.065	0.042	0.013	0.066	0.068	0.005	−0.058	−0.032
			D: Mean Model					
Mean	0.064	0.124*	0.087	0.050	−0.032	0.045	−0.045	−0.071

Note: The table reports the Kiupers score as the difference between the hit-rate and the false-alarm rate. We also report the test results on predictive failure using the test proposed by Pesaran (2015) with HAC standard errors. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Pesaran-Timmermann test. The 1st, 2nd and 3rd hightest Kiupers Scores are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 120$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table A4: Kiupers score and Pesaran-Timmermann test results for growth of aggregate M2 using a rolling-window with width $T_s = 120$ observations.

B Growth of real aggregate M4 Divisia

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
	2000m1–2014m12				**2007m1–2014m12**			
			A: Full Model					
U_π	1.214**	1.269**	1.372**	1.415***	1.123	1.131	1.177	1.226
U_s	1.322**	1.382***	1.687***	1.683***	1.337**	1.411**	1.578**	1.409***
U_v	1.306**	1.442***	1.379***	1.273**	1.349**	1.437**	1.269	1.188
U_f	1.332***	1.455***	1.468***	1.440***	1.341**	1.547***	1.487***	1.482***
U_e	1.052	1.042	1.079	1.215*	1.017	0.965	1.014	1.241
U_p	1.177**	1.186**	1.265***	1.342***	1.194*	1.171	1.205**	1.318***
			B: G2S					
Mod. I	0.956	1.042	1.059	0.945	0.958	1.031	0.957	0.895
U_π	1.176*	1.211	1.359**	1.125	1.098	1.040	1.168	1.044
U_s	1.240**	1.377***	1.647***	1.171	1.256*	1.342**	1.492**	1.096
U_v	1.148	1.371**	1.342***	1.190*	1.182	1.355*	1.087	1.035
U_f	1.215*	1.472**	1.533***	1.279**	1.182	1.429*	1.338**	1.228
U_e	1.007	1.107	1.096	1.151	0.979	1.032	0.971	1.106
U_p	1.074	1.072	1.155	1.100	1.107	1.058	1.061	1.041
			C: S2G					
Mod. I	$\underline{0.816^{**}}$	$\underline{0.872^{*}}$	$\underline{0.845^{**}}$	0.863**	$\underline{0.764^{**}}$	$\underline{0.788^{**}}$	$\mathit{0.840}^{**}$	0.856*
U_π	0.875	0.933	*0.854*	0.983	0.842	*0.865*	$\underline{0.832}$	0.988
U_s	0.864	0.893	0.913	0.962	*0.820*	0.925	0.895	0.943
U_v	0.880	0.914	0.904	0.914	0.909	0.972	0.932	0.950
U_f	*0.860*	0.955	0.986	$\mathit{0.855}^{*}$	0.858	1.009	0.972	$\mathit{0.845}^{*}$
U_e	$\mathbf{0.780}^{**}$	**0.824**	$\mathbf{0.801}^{**}$	$\mathbf{0.845}^{**}$	$\mathbf{0.694}^{**}$	$\mathbf{0.722}^{*}$	$\mathbf{0.793}^{**}$	$\underline{0.839}$
U_p	0.956	1.039	0.968	1.012	0.916	1.002	0.917	1.044
			D: Mean Model					
Mean	0.860**	$\mathit{0.887}^{**}$	0.927	$\underline{0.846^{***}}$	0.853**	0.873*	0.886	$\mathbf{0.836}^{***}$

Note: The table reports the RMSE relative to an ARDL(12,12,12) comprising lagged values of the endogenous money growth variable, a real income measure and an opportunity cost measure. We also report the test results on equal predictive accuracy using the test proposed by Diebold and Mariano (1995) under a quad-quad loss function. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Diebold-Mariano test. The 1st, 2nd and 3rd lowest relative RMSEs are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 96$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table A5: Relative RMSE and Diebold-Mariano test results for growth of $M4$ divisia using a rolling-window with width $T_s = 96$ observations.

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
	2000m1–2014m12				**2007m1–2014m12**			
			A: Full Model					
Mod. I	0.144	0.121	0.080	0.198***	0.117	0.083	0.002	0.144
U_π	0.141**	0.064	0.206***	0.138**	0.127	0.012	**0.215****	0.151
U_s	0.129	0.124	0.132*	0.158**	0.131	0.080	0.080	0.036
U_v	0.099	0.114	0.072	*0.206***	0.117	0.080	0.053	<u>0.229*</u>
U_f	0.078	0.127*	0.037	0.140*	0.043	0.070	0.057	0.199*
U_e	0.195**	0.144	*0.228****	0.130*	0.205*	0.134	0.154	0.114
U_p	0.166*	0.158*	0.176**	0.184**	0.110	0.070	0.056	0.144
			B: G2S					
Mod. I	0.179**	0.152*	0.125*	0.185**	0.175	0.114	0.120	0.080
U_π	0.114	0.133	0.075	0.154**	0.100	0.077	0.066	0.107
U_s	0.122	0.132*	0.194***	**0.220*****	0.164	0.097	0.127	0.168
U_v	0.108	0.111	0.091	0.100	0.148	0.097	0.090	0.114
U_f	0.135	0.063	0.019	0.170**	0.161	0.053	0.047	**0.236****
U_e	0.195**	0.125	**0.246*****	0.204**	0.127	0.117	*0.191**	*0.222**
U_p	0.182**	0.217**	0.195**	0.133	0.097	0.158	0.137	0.087
			C: S2G					
Mod. I	0.133	0.110	0.155**	0.037	0.090	0.107	0.100	−0.079
U_π	0.107	0.091	0.096	0.058	0.124	0.097	0.026	0.013
U_s	0.121	<u>0.232***</u>	<u>0.231***</u>	<u>0.217***</u>	0.151	**0.222****	<u>0.208*</u>	0.154
U_v	0.099	0.075	0.044	0.143*	0.131	0.114	0.060	0.097
U_f	<u>0.225***</u>	0.099	0.107	0.133**	<u>0.273***</u>	0.053	0.114	0.050
U_e	**0.256*****	0.185**	0.155**	0.152**	**0.313*****	<u>0.195*</u>	0.168	0.121
U_p	0.184**	*0.226****	0.151**	0.124**	*0.225***	*0.181*	0.144	0.029
			D: Mean Model					
Mean	0.116	**0.234*****	0.124*	0.204***	0.073	0.161	0.097	0.168*

Note: The table reports the Kuipers score as the difference between the hit-rate and the false-alarm rate. We also report the test results on predictive failure using the test proposed by Pesaran (2015) with HAC standard errors. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Pesaran-Timmermann test. The 1st, 2nd and 3rd hightest Kuipers Scores are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 96$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table A6: Kuipers score and Pesaran-Timmermann test results for growth of $M4$ divisia using a rolling-window with width $T_s = 96$ observations.

C Robustness section for growth of real household-sector M2

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
	2000m1–2014m12				**2007m1–2014m12**			
			A: Full Model					
U_π	1.290***	1.415***	1.177	1.158	1.246**	1.410***	1.077	1.047
U_s	1.213*	1.331**	1.882**	2.011***	1.160	1.330*	2.015**	2.244***
U_v	1.510**	1.653**	1.840***	2.129**	1.667**	1.809***	1.946***	2.291**
U_f	1.358***	1.420**	1.526*	1.699**	1.426***	1.507**	1.532	1.786**
U_e	1.351*	1.461*	1.177	1.293*	1.343	1.513	1.165	1.286
U_p	1.194*	1.155	1.295**	1.271*	1.179	1.174	1.350**	1.298*
			B: G2S					
Mod. I	0.512**	*0.640*	0.952	0.928	0.470*	**0.539***	0.980	0.948
U_π	0.604*	0.827	1.023	0.982	0.524*	0.747	0.989	0.922
U_s	0.672	0.942	1.646*	1.551**	0.639	0.895	1.774*	1.714**
U_v	0.690	0.876	1.351	1.175	0.698	0.896	1.461	1.239
U_f	0.722	0.970	1.338	1.325*	0.709	0.944	1.398	1.349*
U_e	0.692	0.929	1.147	1.154	0.629	0.902	1.164	1.155
U_p	0.625	0.794	1.620	1.336	0.591	0.774	1.791	1.458
			C: S2G					
Mod. I	**0.425****	<u>0.607*</u>	**0.756**	**0.700**	**0.378****	*0.567**	<u>0.752</u>	**0.712**
U_π	0.468**	0.680	*0.809*	0.764	<u>0.390**</u>	0.610	**0.729**	*0.748*
U_s	0.526*	0.877	1.049	0.998	0.495*	0.889	1.131	1.085
U_v	*0.456***	0.698	0.823	*0.758*	0.423**	0.684	0.872	0.796
U_f	0.510**	0.797	1.009	0.996	0.477**	0.758	1.036	1.013
U_e	0.508**	0.755	0.916	<u>0.735</u>	0.443**	0.716	0.943	<u>0.735</u>
U_p	0.476**	0.682	0.826	0.789	0.437**	0.674	0.824	0.775
			D: Mean Model					
Mean	<u>0.454**</u>	**0.582***	<u>0.806</u>	0.794	*0.419***	<u>0.540*</u>	*0.818*	0.812

Note: The table reports the RMSE relative to an ARDL(12,12,12) comprising lagged values of the endogenous money growth variable, a real income measure and an opportunity cost measure. We also report the test results on equal predictive accuracy using the test proposed by Diebold and Mariano (1995) under a quad-quad loss function. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Diebold-Mariano test. The 1st, 2nd and 3rd lowest relative RMSEs are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 84$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table A7: Relative RMSE and Diebold-Mariano test results for growth of household-sector M2 using a rolling-window with width $T_s = 84$ observations.

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
	2000m1–2014m12				**2007m1–2014m12**			
			A: Full Model					
U_π	1.115	1.051	0.836	0.789	1.127	1.058	0.779	0.717
U_s	1.170	1.289	1.541	1.584*	1.233	1.360	1.647	1.715**
U_v	1.245*	1.328**	1.528*	1.602	1.314*	1.397**	1.625**	1.733*
U_f	1.128	1.074	0.901	0.849	1.142	1.062	0.864	0.810
U_e	1.137	1.111	1.033	0.984	1.052	1.049	0.980	0.938
U_p	1.108	1.161*	1.331	1.298	1.127	1.191	1.386	1.370
			B: G2S					
Mod. I	0.436**	0.590**	0.958	0.973	0.378**	0.565*	1.001	1.008
U_π	0.447**	<u>0.537**</u>	0.760	**0.644***	0.381**	<u>0.508**</u>	0.749	**0.621***
U_s	0.533**	0.724	1.244	1.100	0.498*	0.730	1.338	1.184
U_v	0.545*	0.754	1.126	1.276	0.506*	0.765	1.199	1.374
U_f	0.462**	0.612*	0.920	0.773	0.416**	0.594*	0.960	0.781
U_e	0.534**	0.725	0.941	0.873	0.438**	0.678	0.949	0.847
U_p	0.473**	0.674	1.198	1.174	0.414**	0.659	1.277	1.234
			C: S2G					
Mod. I	<u>0.400**</u>	0.586*	<u>0.685</u>	*0.675*	<u>0.343**</u>	0.548*	<u>0.676</u>	<u>0.641</u>
U_π	0.424**	0.595*	**0.670**	0.719	*0.359***	0.543**	**0.643**	0.694
U_s	0.447**	0.689	0.803	0.801	0.402**	0.688	0.819	0.816
U_v	*0.419***	0.615*	0.779	0.776	0.372**	0.588*	0.791	0.782
U_f	0.454**	0.648	0.815	0.726	0.412**	0.626	0.808	0.722
U_e	0.450**	0.672	0.780	0.746	0.388**	0.651	0.782	0.724
U_p	0.460**	*0.562***	0.755	<u>0.667</u>	0.410**	*0.541**	0.767	*0.648*
			D: Mean Model					
Mean	**0.387****	**0.506****	*0.714*	0.701	**0.335****	**0.477****	*0.720*	0.696

Note: The table reports the RMSE relative to an ARDL(12,12,12) comprising lagged values of the endogenous money growth variable, a real income measure and an opportunity cost measure. We also report the test results on equal predictive accuracy using the test proposed by Diebold and Mariano (1995) under a quad-quad loss function. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Diebold-Mariano test. The 1st, 2nd and 3rd lowest relative RMSEs are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 120$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table A8: Relative RMSE and Diebold-Mariano test results for growth of household-sector M2 using a rolling-window with width $T_s = 120$ observations.

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
	2000m1–2014m12				**2007m1–2014m12**			
			A: Full Model					
Mod. I	*0.229****	0.094	−0.071	−0.066	0.234***	0.109	−0.228*	−0.165
U_π	0.081	0.136*	*0.066*	0.022	0.151	0.172*	−0.037	−0.145
U_s	0.208***	0.053	0.006	−0.003	0.231**	0.043	−0.083	−0.102
U_v	0.244***	0.062	−0.025	−0.034	0.272***	−0.040	−0.184*	−0.226**
U_f	0.152**	0.084	−0.027	−0.031	0.207**	0.082	−0.084	−0.124
U_e	0.066	0.048	−0.146*	−0.065	0.088	0.086	−0.203*	−0.118
U_p	0.217***	0.165***	0.070	0.070	0.276***	**0.215****	0.070	0.070
			B: G2S					
Mod. I	**0.247*****	0.084	−0.109	−0.131*	0.340***	0.111	−0.268**	−0.185
U_π	0.099	0.084	−0.068	−0.025	0.193*	0.091	−0.182	−0.040
U_s	0.199***	*0.138**	0.003	−0.130*	0.232**	0.063	−0.146	−0.224**
U_v	0.148**	0.090	−0.009	*0.026*	0.146	0.025	−0.162	−0.120
U_f	0.070	−0.016	−0.143**	−0.190**	0.168	−0.041	−0.245**	−0.327***
U_e	0.114	0.017	−0.143*	0.015	0.233**	0.090	−0.139	−0.056
U_p	0.202***	0.108	−0.104	0.015	0.277***	0.195**	−0.117	*0.028*
			C: S2G					
Mod. I	0.148*	0.103	−0.053	−0.047	*0.337****	0.107	−0.146	−0.125
U_π	0.198**	0.122*	−0.040	0.010	**0.398*****	0.167	−0.040	−0.020
U_s	0.154*	**0.166****	0.045	−0.094	0.274***	*0.190*	−0.084	−0.274**
U_v	0.117	−0.058	−0.071	0.002	0.168	−0.063	−0.104	−0.082
U_f	0.053	−0.049	−0.272***	−0.051	0.144	0.021	−0.292**	−0.101
U_e	0.090	0.003	−0.047	−0.037	0.174**	−0.033	−0.073	−0.012
U_p	0.159**	0.107	**0.116**	**0.156****	0.235**	0.067	**0.127**	**0.108**
			D: Mean Model					
Mean	0.194***	0.107	−0.075	−0.083	0.276***	0.109	−0.205*	−0.142

Note: The table reports the Kiupers score as the difference between the hit-rate and the false-alarm rate. We also report the test results on predictive failure using the test proposed by Pesaran (2015) with HAC standard errors. $***$, $**$ and $*$ indicates significance at the 1, 5 and 10 percent level for the Pesaran-Timmermann test. The 1st, 2nd and 3rd hightest Kiupers Scores are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 84$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table A9: Kiupers score and Pesaran-Timmermann test results for growth of household-sector M2 using a rolling-window with width $T_s = 84$ observations.

Model	h=1	h=2	h=6	h=9	h=1	h=2	h=6	h=9
		2000m1–2014m12				**2007m1–2014m12**		
			A: Full Model					
Mod. I	0.218***	*0.181***	−0.037	−0.045	0.293***	*0.212***	−0.102	−0.123
U_π	0.250***	0.164**	<u>0.058</u>	0.019	0.317***	0.068	0.025	0.001
U_s	<u>0.277***</u>	0.147	0.025	−0.047	0.250**	0.082	−0.064	−0.251**
U_v	0.209**	0.052	0.011	0.024	0.334***	0.000	−0.108	−0.129
U_f	0.196***	0.117	0.043	0.017	0.251***	0.146	**0.063**	0.000
U_e	0.181**	0.056	−0.093	−0.026	0.274***	0.087	−0.140	−0.056
U_p	0.233***	0.137*	0.049	<u>0.097</u>	0.317***	<u>0.215**</u>	<u>0.049</u>	<u>0.090</u>
			B: G2S					
Mod. I	0.175**	0.171**	−0.029	−0.073	0.337***	0.171*	−0.099	−0.146
U_π	0.193***	**0.210*****	−0.009	*0.072*	**0.359*****	0.150	−0.078	0.000
U_s	0.178**	0.128	−0.039	−0.012	0.169	0.065	−0.144	−0.209*
U_v	0.218**	0.098	−0.022	−0.024	0.333***	0.124	−0.109	−0.169
U_f	**0.288*****	0.103	−0.077	0.017	0.334***	0.146	−0.143	−0.083
U_e	0.074	−0.049	−0.039	−0.075	0.171*	0.003	−0.100	−0.077
U_p	0.238***	<u>0.185**</u>	*0.054*	**0.131***	0.274***	**0.232****	*0.047*	**0.131**
			C: S2G					
Mod. I	0.231***	0.135*	0.021	−0.067	0.274***	0.087	−0.080	−0.145
U_π	0.230***	0.111	**0.065**	−0.047	*0.357****	0.149	−0.040	−0.143
U_s	0.203**	0.141**	0.013	−0.004	0.232**	0.006	−0.125	−0.272**
U_v	0.174**	0.024	−0.099	−0.086	0.232***	−0.020	−0.185	−0.228*
U_f	0.264***	0.082	0.003	0.013	0.313***	0.063	−0.062	−0.018
U_e	0.187***	0.086	−0.090	0.012	0.257***	0.052	−0.095	0.010
U_p	0.196***	0.144**	0.010	0.057	0.274***	0.150	−0.082	*0.043*
			D: Mean Model					
Mean	*0.276****	0.156**	−0.037	−0.008	**0.359*****	0.192*	−0.144	−0.104

Note: The table reports the Kiupers score as the difference between the hit-rate and the false-alarm rate. We also report the test results on predictive failure using the test proposed by Pesaran (2015) with HAC standard errors. ∗∗∗, ∗∗ and ∗ indicates significance at the 1, 5 and 10 percent level for the Pesaran-Timmermann test. The 1st, 2nd and 3rd hightest Kiupers Scores are highlighted by fat, underlined and italic font, respectively. All model forecasts are estimated by means of a rolling-window regression with width $T_s = 120$ observations. Based on 384 out-of-sample forecast errors for which the first observation for the forecast horizon $h = 1$ is given for 1983m1, we compute the statistics of interest.

Table A10: Kiupers score and Pesaran-Timmermann test results for growth of household-sector M2 using a rolling-window with width $T_s = 120$ observations.