FEDERAL RESERVE BANK of ST. LOUIS

RESEARCH DIVISION

Working Paper Series

An Empirical Investigation of Direct and Iterated Multistep Conditional Forecasts

Michael W. McCracken
and
Joseph McGillicuddy

Working Paper 2017-040A
https://doi.org/10.20955/wp.2017.040

November 2017

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

An Empirical Investigation of Direct and Iterated Multistep Conditional Forecasts *

Michael W. McCracken
Federal Reserve Bank of St. Louis

Joseph T. McGillicuddy
Federal Reserve Bank of St. Louis

November 2017

Abstract

When constructing unconditional point forecasts, both direct- and iterated-multistep (DMS and IMS) approaches are common. However, in the context of producing conditional forecasts, IMS approaches based on vector autoregressions (VAR) are far more common than simpler DMS models. This is despite the fact that there are theoretical reasons to believe that DMS models are more robust to misspecification than are IMS models. In the context of unconditional forecasts, Marcellino, Stock, and Watson (MSW, 2006) investigate the empirical relevance of these theories. In this paper, we extend that work to conditional forecasts. We do so based on linear bivariate and trivariate models estimated using a large dataset of macroeconomic time series. Over comparable samples, our results reinforce those in MSW: the IMS approach is typically a bit better than DMS with significant improvements only at longer horizons. In contrast, when we focus on the Great Moderation sample we find a marked improvement in the DMS approach relative to IMS. The distinction is particularly clear when we forecast nominal rather than real variables where the relative gains can be substantial.

JEL Nos.: C53, C52, C12, C32
Keywords: Prediction, forecasting, out-of-sample

McCracken (corresponding author): Research Division; Federal Reserve Bank of St. Louis; P.O. Box 442; St. Louis, MO 63166; michael.w.mccracken@stls.frb.org. *McGillicuddy*: Research Division; Federal Reserve Bank of St. Louis; P.O. Box 442; St. Louis, MO 63166; joseph.t.mcgillicuddy.

The authors gratefully acknowledge helpful comments from the Bundesbank Workshop on Forecasting 2017 and the St. Louis Fed Central Bank Forecasting conference.

The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of St. Louis, Federal Reserve System, or any of its staff.

1 Introduction

Conditional point forecasts are a useful means of evaluating the impact of hypothetical scenarios. In these exercises the goal is to predict variables such as GDP growth or inflation conditional on, for example, an assumed path of monetary or fiscal policy variables. As an example of the former, Dokko et al. (2009) use both the FRB/US model as well as a VAR to produce forecasts of the housing market conditional on various paths of the federal funds rate. As an example of the latter, Christoffel, Coenen, and Warne (2008) use their New Area-Wide (DSGE) Model to evaluate conditional forecasts of Euro Area GDP growth conditional on paths for a variety of series including government spending. In addition, as discussed in Sarychev (2014) and Hirtle et al. (2016), conditional forecasts have become an important component of bank stress testing. Conditional forecasts are also used in academic research including Giannone et al. (2014) who use VARs to construct forecasts of inflation conditional on paths for oil and other price indicators; Caruso, Reichlin, and Ricco (2015) who construct forecasts of fiscal variables conditional on IMF projections for GDP growth and inflation; and Baumeister and Kilian (2014) who consider forecasts of oil prices conditioned on a range of scenarios.

Regardless of whether these conditional forecasts were constructed using frequentist or Bayesian methods, the recursive nature of the VAR was ultimately used to produce the forecast.[1] Put differently, the VAR was first estimated as a model with one-step-ahead forecast errors and then its recursive structure was used to produce conditional forecasts at the desired horizon given the assumed scenario. This VAR-based IMS approach to producing conditional forecasts is by far the most common.[2]

Within the literature, a few alternative approaches exist. Instead of using a VAR, Guerrieri and Welch (2012) estimate a scalar autoregressive distributed lag (ARDL) model and use it to produce forecasts of bank net charge-offs conditional on those macroeconomic series that the Federal Reserve releases in its annual bank stress testing exercise. As above, their model is also estimated to have one-step-ahead forecast errors and is iterated forward to produce the conditional forecast − but without accounting for the joint evolution of all variables in the system. At first glance this approach seems unlikely to perform well,

[1]Or near-VARs in the case of DSGE models. See Giacomini (2013) for a discussion on the relationship between DSGE models and VARs.

[2]Throughout we will use the phrase "conditional forecast" in the same context used by Waggoner and Zha (1999). One can, of course, also interpret an impulse response function as a type of conditional forecast. As such, the local projections method of Jorda (2005) is a special case of DMS-based conditional forecasting.

especially at longer horizons, because it does not provide a complete characterization of the joint dynamics among the variables as would a VAR. On the other hand, far fewer parameters are estimated, and hence, this ARDL-based IMS approach to conditional forecasting may be more accurate in a mean-squared-error (MSE) sense by taking advantage of a bias-variance trade-off. A handful of others, including Arseneau (2017) and Kapinos and Mitnik (2016) have also used this approach to conditional forecasting. In the context of bank net charge-offs, Bolotnyy et al. (2013) perform a direct comparison of the accuracy of conditional forecasts made using fully specified VARs and those made by simpler ARDL models and find little evidence to recommend one over the other.

This "simpler might be better" approach to forecasting is reminiscent of an issue common in the literature on producing unconditional forecasts. Specifically, rather than estimate a fully specified VAR, it is quite common to use DMS models to construct point forecasts. When taking this approach, the predictors are lagged such that a distinct model is estimated for each horizon. Since the model-implied forecast error is horizon-specific, the model is used directly – no iteration is required. While this DMS approach is less fully specified than the VAR-based IMS approach, Bhansali (1997), Findley (1983), and Schorfheide (2005) each argue that, under certain assumptions, DMS models can be more robust to model misspecification than are IMS models. More recently, Chevillon (2017) shows that DMS models have an advantage when balancing a bias-variance trade-off at longer horizons.

It is therefore surprising that there appear to be almost no empirical examples of DMS approaches to conditional forecasting.[3] For a sophisticated forecasting agent at a central bank the intuition is obvious – the model would clearly be misspecified and would not account for all the general equilibrium feedback among the variables in the system. While this is true, it is also true that any empirical model is likely misspecified in some way. This point is emphasized by Bidder, Giacomini, and McKenna (2016) in the context of the New York Fed's CLASS model – a model used to produce conditional forecasts of bank stress under various severely adverse scenarios. At a minimum, nearly all models are only known up to a collection of unknown parameters that are estimated, which in turn introduces estimation error into the forecast.

In this paper we provide empirical evidence on the accuracy of VAR-based IMS conditional forecasts relative to ARDL-based DMS conditional forecasts. In no small part our investigation parallels Marcellino, Stock, and Watson (MSW, 2006) who compare both a

[3]There is a discussion of DMS-based conditional forecasting in Jorda and Marcellino (2010). Their approach is distinct from ours and is derived assuming Gaussian forecast errors.

large number of univariate and bivariate IMS models to comparable DMS models. Like them, we begin with a large dataset of monthly frequency macroeconomic time series that includes real, nominal, and financial time series dating back to 1959. From this database, 2,000 randomly selected bivariate VARs/ARDLs are estimated and used to construct a sequence of pseudo-out-of-sample conditional forecasts. With these forecasts in hand, MSEs are constructed and the accuracy of the forecasts are compared. In order to emphasize the methods used rather than the scenarios chosen, ex-post realized values are used when forming the conditioning paths. Like MSW, in our bivariate results we abstract from real-time data issues and conduct the exercise using a single vintage of data taken from FRED-MD (McCracken and Ng, 2016).

We then narrow our evidence to a smaller collection of 150 trivariate systems that always include one real, nominal, and financial variable. Our primary reason for choosing this collection of models is that they are closer in spirit to the types of monetary VARs used by central banks when producing conditional forecasts. In addition, this smaller collection of model makes it considerably easier to implement bootstrap-based inference when we investigate the role model misspecification plays for our results.

Regardless of whether bivariate or trivariate models are used, some of our results reinforce those found in MSW but other results do not. For example, when estimating the models and evaluating the forecasts over the time frame used in MSW we also find that VAR-based IMS conditional forecasts are generally more accurate though improvements are often quite modest. Empirically relevant improvements only arise at the longer horizons and are dependent on whether short or long lags are used. In addition, there is evidence that DMS methods provide specific benefits when the variable being forecasted is nominal (e.g. prices, wages, and money) rather than real or financial.

Our results begin to deviate from those in MSW when we either extend the out-of-sample period to include the more recent 2003-2016 period or when we restrict our sample to the Great Moderation. In both cases we observe a substantial improvement in the relative performance of the DMS approach. In fact, across both our bivariate and trivariate results we find that ARDL-based DMS methods are clearly the preferred choice when forecasting nominal variables. In many cases the improvements are quite large when using DMS methods to predict nominal variables. For both real and financial variables the results are much less clear: neither DMS nor IMS is particularly better than the other.

While the Great Moderation has a large impact on our results, the reasons for that impact are not obvious. In MSW the authors argue, in footnote 7, that DMS methods improve relative to IMS when a larger sample is used to estimate the model parameters. They base this on a comparison of relative MSEs constructed using the first and second halves of their 1979-2002 out-of-sample period in which they observe that DMS-based MSEs are relatively lower in the latter period. Instead, we find that DMS methods improve relative to IMS methods even when we shorten the estimation sample – so long as that sample consists of the period identified as the Great Moderation (e.g. the in-sample period starts in 1984 rather than 1959).

We consider two potential explanations for our results. In the first, based on the premise that DMS models are considered robust to model misspecification, we conduct a variety of tests of model misspecification for each of the trivariate VARs. Counterintuitively, we find less evidence of model misspecification in the Great Moderation sample than the sample used by MSW. In the second, based on the premise that the DMS models have less parameter estimation error and thus balance a bias-variance trade-off better, we compare the relative changes in information criteria across the two samples for each of the DMS models and the associated VAR. Here we find some evidence that the fit of DMS models has improved (deteriorated) at a higher (lower) rate than the corresponding VARs following the Great Moderation, suggesting that perhaps the simplicity of the DMS models allows them to handle better the lower levels of predictive content present during the Great Moderation.

The remainder of the paper proceeds as follows. Section 2 provides a simple example of the models considered and motivates why DMS models may be useful for conditional forecasting. Section 3 describes the modeling approaches more generally and discusses the data. Sections 4 and 5 discuss our results. Section 6 concludes. An appendix contains additional detail on the data used.

2 A Simple Example of the Models

To better understand the comparison of interest, and how model misspecification can make conditional forecasts from ARDL models more accurate than those from VAR models, consider a very simple example adapted from Clark and McCracken (2017) in which we forecast inflation (y_t) one period ahead conditioned on a known value for the federal funds rate (x_t). The data-generating process (DGP) of inflation and the funds rate is a zero-mean stationary VAR(1) taking the form

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0 & b \\ 0 & c \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_t \\ v_t \end{pmatrix},$$

with i.i.d. N(0,1) errors with contemporaneous correlation ρ. Using this DGP we provide three comparisons. In the first, the VAR is correctly specified while in the second and third it is not. In the latter two cases the DMS model is trivial, and yet we will see that it can still provide more accurate conditional forecasts. In each case we abstract from finite sample estimation error and construct forecasts using the pseudo-true parameters of the respective model. As a practical matter, estimation error certainly plays a role but we abstract from that in order to emphasize that model misspecification affects conditional forecasts as well as unconditional forecasts.

2.1 Correct Specification

In our first example, the VAR-based conditional forecasts are constructed after using OLS to estimate a VAR(1) for $(y_t, x_t)'$. The residuals are then used to estimate the error covariance matrix. Note that these regression parameter estimates and residual variance estimates are all consistent for the parameters of the DGP. From Waggoner and Zha (1999) we know that the time-t minimum-mean-square-error one-step-ahead conditional forecast of y_{t+1} given x_{t+1} takes the form

$$\begin{aligned} \hat{y}^c_{t,1} &= \hat{y}^u_{t,1} + \rho(\hat{x}^c_{t,1} - \hat{x}^u_{t,1}) \\ &= bx_t + \rho(x_{t+1} - cx_t) \end{aligned}$$

where the superscripts c and u denote conditional and unconditional forecasts, respectively. The conditional forecast of y is comprised of the standard, unconditional MSE-optimal forecast $\hat{y}^u_{t,1}$, plus an additional term that captures the impact of conditioning on the future value of the federal funds rate, $\hat{x}^c_{t,1} = x_{t+1}$. With this forecast in hand, straightforward algebra implies that, for $e^{IMS}_{t,1} = y_{t+1} - \hat{y}^c_{t,1}$, $E(e^{IMS}_{t,1})^2 = 1 - \rho^2$.

Now consider a very simple DMS forecast based on a model in which y_t is regressed on x_t, and hence the model takes the form

$$y_t = \gamma x_t + \varepsilon_t.$$

This model yields a forecast of the form

$$\begin{aligned} \hat{y}^c_{t,1} &= \gamma \hat{x}^c_{t,1} \\ &= (bc + \rho(1 - c^2))x_{t+1}. \end{aligned}$$

Given this forecast, straightforward algebra implies that, for $e_{t,1}^{DMS} = y_{t+1} - \hat{y}_{t,1}^c$, $E(e_{t,1}^{DMS})^2 = 1 - \rho^2 + (b - c\rho)^2$.

If we take the difference between the two MSEs we find that the IMS forecasts are more accurate than the DMS forecasts if and only if $(b - c\rho)^2 > 0$. Since this is trivially true, we reach the expected conclusion that the minimum-MSE approach to conditional forecasting provides more accurate forecasts.

2.2 Incorrect Specification of Conditional Mean

In our second example, everything remains the same except that the equation for x_t in the VAR is misspecified as $x_t = \alpha x_{t-2} + \eta_t$. In this framework, the regression parameters for the y equation remain consistent for their population values – including the residual variance. For the x equation, it is clear that $\alpha = c^2$. In addition, the residual variance for the x equation is $1 + c^2$ while the residual covariance across equations remains ρ. Together these imply that the IMS forecast takes the form

$$\begin{aligned} \hat{y}_{t,1}^c &= \hat{y}_{t,1}^u + \hat{\rho}(\hat{x}_{t,1}^c - \hat{x}_{t,1}^u) \\ &= bx_t + \frac{\rho}{\sqrt{1+c^2}}(x_{t+1} - c^2 x_{t-1}). \end{aligned}$$

Straightforward algebra implies $E(e_{t,1}^{IMS})^2 = 1 + \rho^2 - \frac{2\rho^2}{\sqrt{1+c^2}}$ which is, not surprisingly, larger than under correct specification. This opens the door for the trivial DMS example to become more accurate than the "optimal" VAR-based conditional forecast. Specifically, we find that $E(e_{t,1}^{IMS})^2$ is now less than $E(e_{t,1}^{DMS})^2$ if and only if $2\rho^2(1 - \frac{1}{\sqrt{1+c^2}}) < (b - \rho c)^2$. This is easily violated. For example, if we set $b = \rho^2$ and $c = \rho$ the inequality takes the form $2\rho^2(1 - \frac{1}{\sqrt{1+\rho^2}}) < 0$, which is false, and hence the simplistic DMS approach provides more accurate conditional forecasts.

2.3 Incorrect Specification of Residual Variance

In our third example, we return to the first example in which the conditional mean of the VAR is correctly specified. The sole difference is that we allow the contemporaneous correlation between the model errors to change from $\rho = \rho_0$ to ρ_1 over the one-step-ahead horizon T to $T + 1$. Since this change is unknown at time T, the point forecasts remain the same and take the form

$$\begin{aligned} \hat{y}_{T,1}^c &= \hat{y}_{T,1}^u + \rho_0(\hat{x}_{T,1}^c - \hat{x}_{T,1}^u) \\ &= bx_T + \rho_0(x_{T+1} - cx_T) \end{aligned}$$

and

$$\begin{aligned}\hat{y}^c_{T,1} &= \gamma \hat{x}^c_{T,1} \\ &= (bc + \rho_0(1 - c^2))x_{T+1}.\end{aligned}$$

for the VAR and DMS methods respectively. Straightforward algebra reveals that $E(e^{IMS}_{T,1})^2 = 1 + \rho_0^2 - 2\rho_0\rho_1$ and $E(e^{DMS}_{T,1})^2 = 1 + b^2 + \rho_0^2(1 - c^2) - 2bc\rho_1 - 2\rho_0\rho_1(1 - c^2)$. Both mean squared errors differ from what we obtained above because of the change in the contemporaneous correlation in the model errors. This again opens the door for the DMS example to be more accurate than the "optimal" VAR-based conditional forecast. Specifically, for all parameterizations for which $b > \rho_0 c$, $E(e^{IMS}_{t,1})^2$ is now less than $E(e^{DMS}_{t,1})^2$ if and only if $2c\rho_1 < b + \rho_0 c$. This too is easily violated. For example, if we set $b = 0$ and $c \in (0, 1)$, $\rho_0 \in (-1, 0)$, and $\rho_1 \in (0, 1)$, the inequality is false, and we find that the simple DMS approach provides a more accurate forecast in expectation.

3 Models and Data

The three examples from the previous section are obviously stylized and simplistic. Even so, very minor misspecification in either the conditional mean of the VAR or its residual variance matrix led to conditional forecasts that were potentially less accurate than the DMS conditional forecasts. In practice, all models will be misspecified at some level, and hence, as MSW emphasize, whether one method provides more accurate conditional forecasts than another is purely an empirical matter. In this section we describe the collection of models, both IMS and DMS, that we consider as well as the data used throughout our experiments.

3.1 Modeling Approaches

We produce a variety of results used to evaluate conditional forecasts. Many more results could have been produced, but we restricted some choices in order to either (a) allow comparison of our results to those in MSW or (b) focus attention on the accuracy of the various forecasting methods and the robustness thereof. In the following set of bullets we delineate the choices in the context of a bivariate system $Z_t = (Y_t, X_t)'$ consisting of two series that may be in levels or log-levels. Extensions to higher order systems are straightforward. The goal is to forecast Y_{t+h}. Let y_t denote the stationary transform of Y_t which may be in levels or consists of taking first or second differences. Finally, let x_t denote the comparably transformed value of X_t, but note that y and x need not have the same stationary transform.

- We consider two distinct methods for constructing $\hat{Y}^c_{t,h}$, the h-step-ahead conditional forecast of Y_{t+h}.

Method 1: Under the VAR-based IMS approach, at each forecast origin $t = R, ..., T-h$ we use OLS to estimate a vector autoregressive model of the form

$$z_t = C + A(L)z_{t-1} + \varepsilon_t,$$

where $z = (y, x)'$, $\varepsilon = (\varepsilon_y, \varepsilon_x)'$, and $A(L) = \sum_{j=0}^{p-1} A_j L^j$. Following Waggoner and Zha (1999), and specifically the formulas provided in Jarocinski (2010), the h-step-ahead conditional forecast of y_{t+h} is obtained by the standard minimum-MSE approach and thus takes the form

$$\hat{y}^c_{t,h} = \hat{y}^u_{t,h} + \sum_{1\leq i\leq h} \hat{\gamma}_{i,t}(x_{t+i} - \hat{x}^u_{t,i})$$

for a collection of constants $\hat{\gamma}_{i,t}$ that are non-stochastic functions of $\hat{A}_{i,t}$ and $\hat{\Sigma}_t = (t - 2p-1)^{-1}\sum_{s=1}^{t-1} \hat{\varepsilon}_{s+1}\hat{\varepsilon}'_{s+1}$. The values for $\hat{y}^u_{t,h}$ and $\hat{x}^u_{t,i}$ are those one would obtain from the standard unconditional forecasts of y or x constructed using the recursive structure of the VAR. Forecasts of Y_{t+h} are then computed by accumulating the sequence of forecasts $\hat{y}^c_{t,i}$ for $i = 1, ..., h$ in accordance with the order of integration of Y:

$$\hat{Y}^c_{t,h} = \left\{ \begin{array}{cl} \hat{y}^c_{t,h} & \text{if } Y_t \text{ is } I(0) \\ Y_t + \sum_{i=1}^h \hat{y}^c_{t,i} & \text{if } Y_t \text{ is } I(1) \\ Y_t + h\Delta Y_t + \sum_{i=1}^h \sum_{j=1}^i \hat{y}^c_{t,j} & \text{if } Y_t \text{ is } I(2) \end{array} \right\}.$$

Method 2: Under the ARDL-based DMS approach, at each forecast origin we use OLS to estimate a horizon-specific linear regression of the form

$$y^{(h)}_t = \alpha + \sum_{j=0}^{p-1} \beta_j y_{t-h-j} + \sum_{j=0}^{p-1} \delta_j x_{t-h-j} + \sum_{1\leq i\leq h} \gamma_i x_{t-h+i} + \varepsilon_t$$

where

$$y^{(h)}_t = \left\{ \begin{array}{cl} Y_t & \text{if } Y_t \text{ is } I(0) \\ Y_t - Y_{t-h} & \text{if } Y_t \text{ is } I(1) \\ Y_t - Y_{t-h} - h\Delta Y_t & \text{if } Y_t \text{ is } I(2) \end{array} \right\}.$$

The h-step-ahead conditional forecast of $y^{(h)}_{t+h}$ is then constructed as

$$\hat{y}^{c(h)}_{t,h} = \hat{\alpha} + \sum_{j=0}^{p-1} \hat{\beta}_{j,t} y_{t-j} + \sum_{j=0}^{p-1} \hat{\delta}_{j,t} x_{t-j} + \sum_{1\leq i\leq h} \hat{\gamma}_{i,t} x_{t+i}.$$

Forecasts of Y_{t+h} are then computed in accordance with the order of integration of Y:

$$\hat{Y}^c_{t,h} = \left\{ \begin{array}{cl} \hat{y}^{c(h)}_{t,h} & \text{if } Y_t \text{ is } I(0) \\ Y_t + \hat{y}^{c(h)}_{t,h} & \text{if } Y_t \text{ is } I(1) \\ Y_t + h\Delta Y_t + \hat{y}^{c(h)}_{t,h} & \text{if } Y_t \text{ is } I(2) \end{array} \right\}.$$

- We consider four distinct approaches to selecting the lag order p. In the first two, all lags are fixed at 4 or 12 respectively. In the second two, at each forecast origin t either AIC or BIC is used to select the number of lags $p \in \{0, ..., 12\}$.[4] In order to facilitate comparison across methods, the lag order is the same for both the autoregressive terms (y) and the distributed lag terms (x).

- We consider four distinct forecast horizons: $h = 3$, 6, 12, and 24. For brevity, we do not report output for $h = 6$, but note that these results typically lie between those associated with horizons 3 and 12.

- In the reported results, for a given forecast horizon h, we construct forecasts conditional on the path $x_{t+1}, ..., x_{t+h}$. (For the reported trivariate results, the forecasts are made conditional on the full path of just one series.) In unreported results we have also (i) produced forecasts conditional on just the value x_{t+h} and (ii) for our trivariate systems, produced forecasts conditioning on paths of two series rather than just one. The pattern of results remains the same, so we do not report them to conserve space.

- In order to facilitate comparison to the results in MSW, we only consider a recursive approach to model estimation. That is, for each forecast origin $t = R, ..., T - h$, observations $s = 1, .., t$ are used to estimate the model parameters.

- We consider a variety of different samples for model estimation and forecast evaluation. In some samples we use observations dating back to 1959:01 to estimate parameters and in others we only use Great Moderation data dating back to 1984:01 to estimate parameters. Similarly, in some samples we align our out-of-sample forecasting exercise with that of MSW and form forecasts from $R = 1979{:}01 + h$ to $T = 2002{:}12$ while in others our out-of-sample period ranges from $R = 2002{:}12 + h$ to $T = 2016{:}12$.

- Like MSW we require that, for a given pair/triple of series, horizon, and forecast origin, at least 120 observations are used to estimate every regression (i.e. across every lag-selection method and across both forecasting approaches, IMS and DMS) before incorporating the forecast in the results. This is really only applicable for one series, the trade-weighted exchange rate (which starts in 1973:01 and is only used in the trivariate exercises), when the out-of-sample period begins in $1979{:}01 + h$.

[4] To facilitate comparison with the results in MSW, we use the standard formulas for AIC and BIC. See Hansen (2010) for a discussion of information criteria and DMS models when the horizon is greater than 1.

3.2 Data Used

All of our results are based on models estimated using data from vintages of FRED-MD (McCracken and Ng, 2016). This dataset consists of 128 monthly macroeconomic series and is designed to emulate that used in Stock and Watson (2005). We use the June 2017 vintage when performing the bivariate exercises. Observations for most series are first available starting in 1959:01. For the exceptions, we impose the restriction that observations must be available starting no later than 1967:01 to be consistent with the dataset in MSW. As a result, we remove four series that fail to meet this condition. We also drop two series for ending prior to the desired end date of 2016:12 and one series due to significant outliers in the series post transformation. This leaves us with a final total of 121 series that we consider for the bivariate exercises. In order to facilitate comparison of our results with those in MSW we have organized the series into the same five groups used by MSW: income, output, sales, and capacity utilization; employment and unemployment; construction, inventories and orders; interest rates and asset prices; and nominal prices, wages, and money. The variables that we use, along with the relevant transformations used to induce stationarity, are delineated in the appendix (Table A1).

It is worth emphasizing that our dataset is distinct from that of MSW for two reasons. First, our dataset is a much more recent vintage than theirs, and hence, some differences are due to data revisions. Perhaps more importantly, our dataset of 121 series is smaller than the 170 that they use. The bulk of the missing series are from group 1 (income, output, sales, and capacity utilization) and group 3 (construction, inventories and orders). To get a feel for how important these differences are, in Table 1 we use our dataset to replicate their Table 5. In this exercise 2,000 random pairs of y and x are selected from the database such that y and x come from distinct groups and an equal number of series pairs (y, x) come from each of the 10 possible group pairings. For each pair, horizon, and lag-selection method, the out-of-sample MSEs from the VAR-based IMS unconditional forecasts and ARDL-based DMS unconditional forecasts are constructed and their ratio is taken. These ratios are then placed in bins associated with various quantiles of their empirical distribution. Despite having fewer series and having a distinct vintage of data, the results are remarkably similar, with few differences greater than the second decimal.

For the trivariate exercises, we focus on just 16 series within FRED-MD, which we separate into three variable groups: real, nominal, and financial. These variables, along with the relevant transformations used to induce stationarity, are likewise delineated in the appendix (Table A2). The June 2017 vintage was utilized for these exercises as well. In unreported results we also considered using real-time vintages of this smaller dataset of 16 series. This made little difference, and so we omit it for brevity.

4 Empirical Results

In this section we consider the relative accuracy of VAR-based IMS and ARDL-based DMS conditional forecasts. Our approach is directly comparable to that from Table 1 with the exception that we are now assessing the relative accuracy of conditional rather than unconditional forecasts. In contrast to Table 1, in the following we only report the mean and median functionals of the distribution of these ratios. This allows us to report results across multiple subsamples in a single table and thus facilitates comparison. There are three subsamples: (1) that associated with MSW, (2) one that extends the MSW sample but forecasts over the 2003-2016 period, and (3) another that also forecasts over the 2003-2016 period but only uses Great Moderation data to estimate model parameters. In addition, to get a better feel for the magnitude of the differences in MSEs, for each pairwise comparison we construct a simple t-test of equal MSE à la Diebold and Mariano (1995) and West (1996) at the 5% level.[5] If the test rejects in the upper tail we characterize the VAR-based forecasts as "better" and if the test rejects in the lower tail we characterize the ARDL-based forecasts as "better." This is a crude approach to inference, and certainly ignores issues associated with multiple testing, but we believe it conveys a useful guide to the statistical significance of the differences.

4.1 Bivariate Comparisons

We begin with comparisons based on bivariate systems. For the same $2,000$ pairs of y and x used in Table 1, for each horizon, for each lag-selection method, and for each sample, the out-of-sample MSEs from the VAR-based IMS conditional forecasts and ARDL-based DMS conditional forecasts, of both y and x, are constructed and their ratio is taken (DMS over IMS).[6] In panel A of Table 2 we report the mean, median, and percent better of these $4,000$ ratios of MSEs separately for each permutation of horizon, lag-selection method, and sample. Recall however that MSW identified variables from the prices, wages, and money (PWM) group as being distinct from the others in their slightly stronger preference for DMS

[5] The standard errors are constructed using a Newey-West (1987) HAC with the lag-length set to $[h * 1.5]$ (rounded up to 5 when $h = 3$).

[6] That is, we forecast y conditional on x and then forecast x conditional on y.

models. We therefore decompose out results in a similar manner. In panel B we report the results associated with the $2,400$ ratios arising from models that exclude any pairs with a series from the PWM group. Panel C does the same for the 800 ratios from pairs with a single element from the PWM group but when the variable being forecasted is not from the PWM group. The remaining 800 ratios, for which the variable being forecasted is from the PWM group, are reported in the final panel.

First consider the MSW sample in the left-most sub-panels of panels A through D. These results largely reinforce those in MSW for unconditional forecasts. So long as the variable being forecasted is not known to be from the PWM group, the mean and median relative MSE is typically greater than or equal to one suggesting a preference for the VAR-based IMS approach to conditional forecasting. That said, the gains are typically meager with statistically significant differences largely regulated to the longest horizons. The biggest differences clearly arise when a PWM variable is being forecasted. Especially when the lag order is shorter (BIC or 4) both the mean and median ratios are less than one and hence the DMS approach to conditional forecasting dominates. In fact, the magnitudes are large enough that for some horizons roughly 50% of the comparisons are considered statistically significant by our crude metric. Larger lag lengths make the IMS approach more competitive, though any gains are rarely statistically significant.

To get a feel for whether these MSW results are robust to different subsamples, we extend the MSW in-sample period through 2002 and then forecast out-of-sample from 2002:12 $+ h$ through 2016. Across each of the middle sub-panels of Table 2, the vast majority of the reported mean and median values are lower than those in the left sub-panel. That said, the relative improvement of the DMS approach is modest for most cases other than those associated with forecasts of a variable from the PWM group. For this group, the DMS approach continues to dominate with nominal improvements over the IMS approach of 10% or more, especially at the longer horizons and when a short lag order is chosen. To be fair, fewer of these larger differences are considered statistically significant by our rough metric, but even so, there are very few cases in which the VAR-based IMS approach is providing statistically significant improvements.

In the right-most sub-panel we again use an out-of-sample period from 2002:12 $+ h$ through 2016 but change the in-sample period to only include Great Moderation data which we date as starting in 1984:01. Once again we observe another modest improvement in the DMS approach relative to the IMS approach in most instances. The smallest gains arise when the pairing does not include a series from the PWM groups while the largest gains again arise when the variable being forecasted is from the PWM group. For this latter group the nominal gains of DMS over IMS grow to 20% or even more, especially at the longest forecast horizons.

The transition across the three sub-panels clearly indicates a sequence of modest improvements in the DMS approach relative to the IMS approach across most variables, lag-selection methods, and especially the longer horizons. In MSW, the authors, in response to a referee suggestion, do a subset comparison akin to ours and find that, relative to the IMS approach, the DMS approach improves in the second half of their 1979-2002 out-of-sample period. They conclude that this arises because DMS forecasts become less variable as the in-sample size increases (see footnote 7 of MSW). Our results are less supportive of this conclusion. While it is the case that the results in the middle sub-panels come from models estimated using a longer in-sample period than used by MSW, that is not the case in our Great Moderation subsample. An alternative interpretation of the MSW subsample analysis, in conjuction with our results, is that the Great Moderation itself has made the DMS approach perform relatively better and not the length of the sample used to estimate the model parameters, particularly for the PWM series.

4.2 Trivariate Comparisons

We now extend our bivariate evidence to trivariate environments. For these results we focus on a smaller number of 150 ($6 \times 5 \times 5$) systems each of which consists of a single real (6), nominal (5), and financial (5) series. The individual series are delineated in Table A2 (in the appendix) and were chosen, in broad terms, to align with the types of series one might observe in a standard monetary VAR. In Table 3 we again report means and medians of the distributions of relative MSEs along with our crude metric for determining significance. We also decompose our results into separate panels in order to identify if the results vary by whether the variable being forecasted is real, nominal, or financial.

As we did for the bivariate results, in the left-most sub-panel we begin by reporting the relative MSEs of our trivariate models over the same in- and out-of-sample periods used by MSW. In panel A, which aggregates the results from all 900 model comparisons (150 systems $\times$ 3 variables to forecast $\times$ 2 variables to condition on), we see a similar pattern to that observed for the bivariate results. For the fixed lag orders and when AIC is used for lag selection we again find that the mean and median ratios are greater than or equal to one. In addition, the few models that are significantly more accurate tend to be from the VARs rather than the ARDLs. But when BIC is used there are a number of comparisons for which the DMS approach is more accurate and significantly so.

The reason for this dichotomy becomes clear in panels B through D where we decompose the results by the type of variable being forecasted. When the real variables are being forecasted nearly all the quantiles are greater than or equal to one and there are almost no instances in which the DMS provides significant improvements over the IMS approach. In contrast, the conditional forecasts of the nominal variables are completely dominated by the DMS approach, especially when BIC is used for lag selection but to a lesser extent also when a fixed lag of 4 is used. At the longer lags the two methods are rarely that different. For the financials the distinction is less stark but probably leans towards the IMS approach unless BIC is used for lag selection. Even then the gains to DMS are nowhere near as large as was the case for the nominal series.

In the second sub-panel we transition to the latter out-of-sample period but continue estimating the models using the full dataset extending back to 1959:01. Akin to the bivariate results in Table 2, the means and medians in the middle subpanel of panel A are nearly all less than those in the left-most subpanel, suggesting a relative improvement in the DMS approach broadly across the variables. For the real variables in panel B, the DMS gains are generally modest and, as a practical matter, only reduce the advantage the IMS had in the previous out-of-sample period. In very few cases are the differences between the DMS and IMS approach significant.

For the nominals in panel C, there too is a bit of improvement in the relative strength of DMS to IMS. Almost every mean and median ratio is less than or equal to one even for the longer lag lengths. By our crude metric of significance, there are almost no cases in which the IMS is significantly better than DMS. That said, there are also fewer instances in which the DMS is significantly distinct from the IMS unless the shorter lag lengths are chosen. The financials again fall somewhere in between the nominal and real panels in terms of the relative strength of IMS and DMS. In broad terms, the results still lean in favor of the IMS approach, though the gains are rarely large and less significantly so.

In the right-most subpanels we now restrict the in-sample period to that of the Great Moderation while continuing to use the latter out-of-sample period. The pattern continues: for almost every mean or median in panel A, the values continue to decline, suggesting improvements for the DMS approach relative to the IMS approach. For the real variables, the relative improvement of the DMS approach continues but is not large and serves only to reinforce the fact that there are few if any statistically significant differences between the two methods. When forecasting the nominal series, the relative gains from using the DMS approach sometimes reach incredible levels of 50% or more, especially at the longer horizons or when information criteria are used to select the lag lengths. Across all lag selection methods and all horizons there exists almost no statistically significant advantages to using the IMS approach to forecasting nominal series. The financials once again lie in between the nominal and real panels in terms of the relative strength of IMS and DMS. One might argue that the results start to lean in favor of the DMS approach but the gains are infrequently large and are significantly so only in isolated instances.

Table 4 provides another perspective on the relative accuracy of IMS and DMS methods across sub-samples but this time with an eye towards identifying the corresponding best lag orders. For each horizon, method of lag selection, trivariate system, and for both the DMS and IMS methods, we construct the MSEs and report them relative to the MSE from a VAR(4). As in the previous tables, we report the mean and median of the distribution of these ratios. We also report the fraction of all permutations for which the given lag length performed best. Note that when the forecasts are based on IMS methods, the mean and median of the distribution of ratios for a lag length of 4 are one by construction. Note as well that, due to ties in lag selection, the sum of the fractions best can be greater than one.[7] For brevity, we only report results for the sample used by MSW and that associated with the Great Moderation.

Consider the left panel, that associated with the MSW sample. As we've seen before, across all variables and especially for the real and financial series the fraction that perform best tends to be higher when using VARs. In particular, when VARs are used to forecast either the real or financial series the fraction that perform best is typically highest when the lag lengths are short (i.e. a fixed lag of 4 or BIC). This pattern softens a bit at the longest horizon where we start to see longer lags becoming useful as well. In contrast, when VARs are used to forecast nominal series, a fixed lag of 12 or AIC is preferred. The only instances in which DMS-based forecasts show signs of life is again when forecasting nominals at the longer horizons and when longer lags are used.[8]

The sharp improvement of DMS-based methods during the Great Moderation sample

[7]For example, if AIC always chooses the lag length of 12 then the AIC and AR(12) columns are double counting the same models.

[8]Though, as we saw in the previous tables, these gains may not be large or statistically significant.

becomes apparent as we move to the right-hand panel. The fraction of times that the DMS is best is now typically higher than that for the IMS approach, though the amount varies by series. For real series, the IMS advantage over DMS is significantly reduced and, at the longest horizon, is essentially gone. For the financial series, DMS typically is best, though not in any way near the level in which DMS is best for nominals. For nominals, except for the shortest horizon, there are almost no instances in which IMS is better than DMS. In addition, when longer lags are chosen, the mean and median improvements relative to using a VAR(4) are enormous, reaching levels of 50% or more.

4.3 Unconditional Forecasts

Throughout this section, we have focused on the relative accuracy of DMS- and IMS-based conditional forecasts. We have done so in large part due to our relative surprise at how infrequently DMS-based methods are used for conditional forecasting despite their common usage for unconditional forecasting. In the previous sections, we have shown that not only are DMS methods potentially useful but their applicability may have improved over time, in part due to the Great Moderation.

This of course begs the question of whether the same is true in the context of unconditional forecasts of the kind analyzed in MSW. In this section, we delineate a limited set of results that support that view: as we transition from the samples used in MSW to the Great Moderation period, the DMS method improves for unconditional forecasts as well. Table 5 provides results associated with the bivariate comparisons while Table 6 provides comparable trivariate results.

For both the bivariate and trivariate cases, as we transition across the three samples we find marked improvements in the accuracy of DMS-based methods relative to IMS-based methods except at the shortest horizons. As was the case for the conditional forecasts, for the bivariate results in Table 5 the non-PWM series transition from leaning in favor of the IMS approach when using the MSW sample towards being indifferent between IMS or DMS in the Great Moderation sample. Similarly, the benefits to using DMS methods for the PWM series grow as we move across the three subsamples. The same also holds for the trivariate results in Table 6. The real and financial series transition from leaning in favor of using the IMS approach towards being indifferent between DMS and IMS, as indicated by the lack of significant differences. In contrast, except at the shortest horizons, the DMS approach becomes even more useful when forecasting the nominals, with gains that are often large and statistically significant.

5 Understanding the Relative Improvement of DMS

The results from Section 4 all point towards an improvement of ARDL-based DMS forecasts relative to IMS forecasts from VARs. The cause for this improvement, while obviously related to the Great Moderation, is not clear. In this section, we investigate a few issues that may point us in the right direction.

5.1 Tests of Predictive Ability

As we saw in Section 2, the DMS approach to conditional forecasting can outperform the minimum-MSE VAR-based approach when either the conditional mean of the VAR or its residual variance is misspecified. To investigate whether misspecification of the VAR plays an important role in our results, we conduct three distinct tests of predictive ability for each of the 150 trivariate VARs in Section 4. Each test focuses on properties of the scalar h-step-ahead forecast errors $\varepsilon^i_{t,h}$ $i = c, u$ implied by the VAR.

For a fixed target variable y_{t+h}, the first two test statistics are the t-statistics associated with regression-based tests of bias (α_0) and efficiency (α_1) of the form

$$\hat{\varepsilon}^u_{t,h} = \hat{g}'_{t,h}\alpha + error$$

with $\hat{g}'_{t,h} = (1, \hat{y}^u_{t,h})$ and $\alpha' = (\alpha_0, \alpha_1)$. We apply the tests separately for each of the three target variables in the VAR and for each horizon h. Note that the test uses the unconditional, rather than conditional forecasts from the VAR. We do so based on simulation evidence provided in Clark and McCracken (2017). There they show that the test had much higher power to detect misspecification in the conditional mean when using the unconditional rather than conditional forecasts.

The third is a normalized test of equal MSE developed in Clark and McCracken (2017) and is designed to detect misspecification in both the conditional mean of the VAR and residual variance. Note that, under minimum-MSE conditioning, correct specification of the VAR implies the existence of a non-negative constant k satisfying $E(\varepsilon^u_{t,h})^2 - E(\varepsilon^c_{t,h})^2 = k$. This constant depends on the VAR regression parameters A_i and residual variance Σ in much the same way as they do for the weights γ_i used to produce the conditional forecasts. Specifically, following the notation in Jarocinski (2010), first define $\Psi_j\Sigma^{1/2}$ as the matrix of orthogonalized impulse responses after j periods and let

$$R = \begin{pmatrix} \Sigma^{1/2} & 0 & & 0 & 0 \\ \Psi_1\Sigma^{1/2} & \Sigma^{1/2} & & 0 & 0 \\ & & \cdots & \Sigma^{1/2} & 0 \\ \Psi_{h-1}\Sigma^{1/2} & \Psi_{h-2}\Sigma^{1/2} & & \Psi_1\Sigma^{1/2} & \Sigma^{1/2} \end{pmatrix}.$$

Now let $\tilde{R}$ denote the matrix formed by those rows in R associated with a conditioning restriction. Straightforward algebra then implies $k = \iota' R\tilde{R}'(\tilde{R}\tilde{R}')^{-1}\tilde{R}R'\iota$ where ι is a vector that selects the single row associated with the variable being forecasted at the relevant horizon.[9] The test statistic takes the form of a centered Diebold and Mariano (1995) and West (1996)-type test of predictive ability based on the regression

$$(\hat{\varepsilon}^u_{t,h})^2 - (\hat{\varepsilon}^c_{t,h})^2 - \hat{k}_T = \alpha + error$$

where $\hat{k}_T$ denotes the plug-in estimator of k using full sample estimates of A_i, and Σ.

In each case, the standard t-statistic associated with the elements of α are asymptotically normal with zero mean when the VAR is correctly specified. However, especially for the centered test of equal MSE, the estimated standard error is not asymptotically valid due to the presence of parameter estimation error coming from both the regression parameters $\hat{A}_{i,t}$ and the residual variance parameters $\hat{\Sigma}_t$. For that reason, following Clark and McCracken (2017), we conduct inference using a percentile bootstrap applied directly to the t-statistics. In particular we use a residual-based moving block bootstrap developed in Bruggemann et al. (2016). In brief, this procedure is the VAR-equivalent of the sieve bootstrap but where we draw blocks of residuals rather than drawing residuals one at a time. All results are based on 299 bootstrap replications of the t-statistics using a block length of 40 for the residuals. Once we obtain the bootstrapped t-statistics, we center each based on the average across all draws and use their associated empirical distribution to estimate the relevant critical values.

One weakness of this approach to inference is that it requires selecting a fixed lag length for the VAR. This is perfectly reasonable when evaluating our VARs based on fixed lag lengths of 4 and 12 but is less intuitive for those results based on recursive application of AIC or BIC. In unreported results, we find that BIC selects a lag order of 2 a large portion of the time regardless of which trivariate VAR is being considered, and hence, we also apply our bootstrap at a fixed lag length of 2. AIC was less consistent in its lag selection with mass spread between 4 and 12 lags. For brevity we only report results for lags 2, 4, and 12.

[9]Because the models are trivariate, this vector will be zeros everywhere except one of the three final elements depending on which variable is being forecasted.

In Table 7 we report the results of the tests of predictive ability associated with all 150 trivariate VARs. Much like the previous tables, for each horizon and lag order we report the mean and median of the empirical distribution of the t-statistics associated with each test and do so separately for the real, nominal, and financial series. We also report the number of rejections obtained at the 5% level, though once again one needs to keep in mind that these are not adjusted to account for multiple testing and are intended solely as a rough guide. In the left-hand panel we report results for VARs estimated over the sample used by MSW while in the right-hand panel we do the same but estimated over the Great Moderation sample.

In the top left-hand sub-panel we report results for all VARs estimated using the MSW sample. For the bias and efficiency tests there are $3 \times 150 = 450$ test statistics while for the normalized equal MSE test there are twice as many since, for a given target variable y, the test is constructed conditioning on future values of x and z separately. Across all three tests there is considerable evidence of model misspecification. Particularly at the longer lag lengths, the number of rejections associated with the slope coefficient in the efficiency regression is substantial, ranging from 20% to nearly 95% of all 450 tests considered. Tests associated with the intercept exhibit significantly fewer rejections, though still more than one might expect at the 5% level. For the MSE tests, the number of rejections are typically on the order of 25% of the 900 tests, though that rises to over 50% when the lag order is 12 and at the longer horizons. In the remaining left-hand subpanels we decompose the results based on whether the variable being forecasted is real, nominal, or financial. Across these sub-panels, evidence of model misspecification is wide-spread and not concentrated solely on any specific subset of variables, lag lengths, or horizons.

That said, one should certainly be concerned about the degree of data mining exhibited across the four left hand subpanels. With so many tests applied to so many series and VARs it is hard to take any specific test seriously. For that reason we emphasize not the number of rejections in the left hand panel so much as the overall reduction of rejections as we transition to the right hand panel. While not uniform, the number of rejections reported in the right hand panels are typically lower than those reported in the left hand panel. This is especially true for the MSE test, for which the number of rejections is uniformly lower when the VARs are estimated using the Great Moderation sample.

As a whole it therefore seems reasonable to conclude that evidence of model misspecification is lower in the Great Moderation sample than in the sample used by MSW. This is somewhat surprising given that DMS models have become more accurate relative to VAR-based IMS models during the Great Moderation. Given the theoretical results recommending the use of DMS models when the corresponding VAR is misspecified, we would have expected more, not less evidence of misspecified VARs over the Great Moderation sample. In short, it is not obvious that the relative improvement of DMS models is being driven by model misspecification, either in the conditional mean or residual variances.

5.2 The Evolution of Model Fit

Of course, when using any parametric model the accuracy of the associated point forecasts also depend on the degree to which the model manages finite sample estimation error. That is, one explanation for the relative improvement of DMS models is simply that their simplicity reduces the effect parameter estimation error has on their accuracy in a mean-squared-error sense. In this section we report evidence associated with the evolution of MSE-based model fit as we transition from the sample used by MSW to a Great Moderation sample. To be clear, we do not necessarily expect to find much evidence of absolute improvements in model fit whether it be for the DMS models or for the VARs. It is well established in Campbell (2007) and Stock and Watson (2007) that predictive content has declined during the Great Moderation. We simply conjecture that lower levels of predictive content favor simple DMS models relative to more complex VARs.

To provide evidence of this hypothesis, for each trivariate system, for each forecast horizon, and for fixed lag lengths of 2, 4, and 12 we calculate the value of BIC associated with unconditional DMS models and the associated VARs.[10] For a given trivariate system and lag length, the VAR has a single value for BIC. In contrast, for the same lag lengths and trivariate system, the DMS models have distinct values of BIC for each pairing of horizon and target variable. The BIC values are all calculated twice: once using a pre-Great Moderation sample ranging from 1959:01-1983:12 and once using a Great Moderation sample ranging from 1984:01-2008:12. We use these subsamples, rather than those in our previous results, in order to make clear comparisons between pre- and Great Moderation (GM) samples but also to keep the sample sizes the same. Since the sample sizes are the same, for a fixed model configuration we can measure the evolution of model fit by comparing the two values of BIC estimated over distinct samples. Specifically, for each model configuration, we measure the degree of improvement (or deterioration) in model

[10] Due to data limitations, the trade-weighted exchange rate was removed from the collection of financial series. This reduced the number of potential triplets from 150 to 120.

fit based on $100(BIC(pre\text{-}GM) - BIC(GM))/|BIC(pre\text{-}GM)|$. Positive values indicate improved model fit while negative values indicate poorer model fit.

While interesting, these measures of model fit are insufficient for comparing DMS models to VAR-based IMS models across subsamples. As we noted earlier, a priori we expect to see at least some negative values for this metric due to the decline in predictive ability during the Great Moderation. What we need is to show how these measures of model fit have evolved across samples for DMS models relative to those associated with the VARs. We do this in Figure 1. In each sub-figure, a given point on the real plane represents the percent change in BIC across subsamples for both DMS and VAR-based IMS for a fixed trivariate system. A point above the diagonal means that the fit of the DMS model has improved (deteriorated) at a higher (lower) rate than the associated VAR. The opposite holds for points below the diagonal. Since the pattern of results was insensitive to the lag length, we only report figures based on a fixed lag order of 4. In addition, for the real and financial series the results were insensitive to the horizon, and hence, we only report results for $h = 12$. In contrast, for the nominals the results do depend on the horizon, and so we report separate figures for both $h = 3$ and 24. Results for $h = 12$ were intermediate to those for $h = 3$ and 24. Note that, to make the figures more readable, large improvements or declines are truncated and hence lie on the edges of the figure.

By this graphical measure, we begin to see some evidence of what might be causing the relative improvement of DMS-based forecasts over IMS-based forecasts. The evidence is most stark for the real series. Every point lies above the x-axis indicating that the fit of the DMS models has improved as we transition from the pre- to the Great Moderation sample. In contrast, a third of the values associated with IMS models are to the left of the y-axis indicating a decline in their model fit. All together, 86% of the points lie above the diagonal indicating that, relative to the IMS models, DMS models have improved their fitness as we transition from the pre- to the Great Moderation sample.

For the nominal series, the evidence is less clear and is horizon dependent. At the shortest horizon, where the unconditional and conditional DMS forecasts have seen little-to-no improvements in MSE, the vast majority of DMS models have actually exhibited a decline in model fit while the corresponding VARs have either improved their fitness or deteriorated at a lower rate. In contrast, at the longest horizon, where the nominal series have exhibited large gains in forecast accuracy relative to IMS models, roughly 80% of the DMS models have exhibited improvements in model fit. And while a comparable percent of IMS models have improved as well, the magnitude of their improvement is dominated by that of the DMS models. In total, roughly 70% of the points lie above the diagonal suggesting a widespread improvement in the fit of DMS models relative to the fit of the IMS models.

Finally, as we've seen in earlier tables, the results for the financial series lie somewhere between those of the real and nominal series. Among the DMS models, half exhibit an improved fit while half have deteriorated. This is less than the two-thirds of IMS models that have improved their fitness. Nevertheless, the gains achieved by the DMS models outweigh the gains of many of the IMS models and hence roughly 50% or the points lie above the diagonal. As such, we would expect to see some improvements in DMS forecasts relative to IMS models but nowhere near as prevalent as those for the nominal series.

6 Conclusions

Motivated by the increasing attention given to conditional forecasts, we provide empirical evidence on the relative accuracy of VAR-based IMS and ARDL-based DMS conditional forecasting models. Our approach follows that taken in MSW: we generate forecasts from a large number of models based on a large macroeconomic dataset and then compare the MSEs from the IMS and DMS models. In some ways our results emulate theirs but in others they do not. For example, when estimating the models and evaluating the forecasts over the sample used in MSW, we also find that IMS methods are generally more accurate though improvements are often quite modest. There is some evidence that DMS methods may be useful when the variable being forecasted is nominal rather than real.

Our results begin to deviate from those in MSW when we either extend the out-of-sample period to include the more recent 2003-2016 period or when we restrict our sample to the Great Moderation. In both cases, we observe a substantial increase in the relative performance of the DMS approach. Our results are robust to whether we evaluate bivariate or trivariate systems, whether we use fixed vintage or real-time vintage data, and whether we consider conditional or unconditional forecasts. While the theory suggests that the benefits of using DMS methods is driven by robustness to model specification, our results suggest that the reason may be their robustness to lower levels of predictability prevalent during the Great Moderation.

References

Arseneau, David M. (2017), "How Would U.S. Banks Fare in a Negative Interest Rate Environment?," FEDS Working Paper 2017-030, Federal Reserve Board of Governors.

Baumeister, Christiane, and Lutz Kilian (2014), "Real-Time Analysis of Oil Price Risks Using Forecast Scenarios," *IMF Economic Review*, 62, 119-145.

Bhansali, R.J. (1997), "Direct Autoregressive Predictors for Multistep Prediction: Order Selection and Performance Relative to Plug-In Predictors," *Statistica Sinica*, 7, 425-449.

Bidder, Rhys, Raffaella Giacomini, and Andrew McKenna (2016), "Stress Testing with Misspecified Models," Federal Reserve Bank of San Francisco Working Paper 2016-26.

Bolotnyy Valentin, Rochelle M. Edge, and Luca Guerrieri (2013), "Stressing Bank Profitability for Interest Rate Risk," FEDS Working Paper, Federal Reserve Board of Governors.

Bruggemann, Ralf, Carsten Jentsch, and Carsten Trenkler (2016), "Inference in VARs with Conditional Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 191, 69-85.

Campbell, Sean D. (2007), "Macroeconomic Volatility, Predictability, and Uncertainty in the Great Moderation: Evidence from the Survey of Professional Forecasters," *Journal of Business and Economic Statistics*, 25, 191-200.

Caruso, Alberto, Lucrezia Reichlin, and Giovanni Ricco (2015), "The Legacy Debt and the Joint Path of Public Deficit and Debt in the Euro Area," Discussion Paper 010, European Commission's DG ECFIN Fellowship Initiative 2014-2015.

Chevillon, Guillaume (2017), "Robustness of Multistep Forecasts and Predictive Regressions at Intermediate and Long Horizons," ESSEC Working Paper 1710.

Christoffel, Kai, Gunter Coenen, and Anders Warne (2008), "The New Area-Wide Model of the Euro Area: A Micro-Founded Open-Economy Model for Forecasting and Policy Analysis," ECB Working Paper Series No. 944.

Clark, Todd E., and Michael W. McCracken (2017), "Tests of Predictive Ability for Vector Autoregressions Used for Conditional Forecasting," *Journal of Applied Econometrics*, 32, 533-553.

Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.

Dokko, Jane, Brian Doyle, Michael T. Kiley, Jinill Kim, Shane Sherlund, Jae Sim, and Skander Van den Heuvel (2009), "Monetary Policy and the Housing Bubble," FEDS Working Paper 2009-49, Federal Reserve Board of Governors.

Findley, David F. (1983), "On the Use of Multiple Models for Multi-Period Forecasting," *American Statistical Association: Proceedings of Business and Economic Statistics*, 528-531.

Giacomini, Raffaella (2013), "The Relationship Between DSGE and VAR Models," in *Advances in Econometrics: VAR models in Macroeconomics — New Developments and Applications*, Thomas Fomby, Lutz Kilian, and Anthony Murphy (eds.), Emerald, 1-25.

Giannone, Domenico, Michele Lenza, Daphne Momferatou, and Luca Onorante (2014), "Short-Term Inflation Projections: A Bayesian Vector Autoregressive Approach," *International Journal of Forecasting*, 30, 635-644.

Guerrieri, Luca, and Michelle Welch (2012), "Can Macro Variables Used in Stress Testing Forecast the Performance of Banks?," FEDS Working Paper 2012-49, Federal Reserve Board of Governors.

Hansen, Bruce E. (2010), "Multi-Step Forecast Model Selection," University of Wisconsin Working Paper.

Hirtle, Beverly, Anna Kovner, James Vickery, and Meru Bhanot (2016), "Assessing Financial Stability: The Capital and Loss Assessment under Stress Scenarios (CLASS) Model," *Journal of Banking and Finance*, 69, S35-S55.

Jarocinski, Marek (2010), "Conditional Forecasts and Uncertainty about Forecast Revisions in Vector Autoregressions," *Economics Letters*, 108, 257-259.

Jorda, Oscar (2005), "Estimation and Inference of Impulse Responses by Local Projections," *The American Economic Review*, 95, 161-182.

Jorda, Oscar, and Massimiliano Marcellino (2010), "Path Forecast Evaluation," *Journal of Applied Econometrics*, 25, 635-662.

Kapinos, Pavel, and Oscar A. Mitnik (2016), "A Top-Down Approach to Stress-Testing Banks," *Journal of Financial Services Research*, 49, 229-264.

Marcellino, Massimiliano, James H. Stock, and Mark W. Watson (2006), "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series," *Journal of Econometrics*, 135, 499-526.

McCracken, Michael W., and Serena Ng (2016), "FRED-MD: A Monthly Database for Macroeconomic Research," *Journal of Business and Economic Statistics*, 34, 574-589.

Newey, Whitney K., and Kenneth D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Sarychev, Andrei (2014), "Scenario Generation vs. Forecasting: Predictive Performance Criteria and the Role of Vague Priors," Bank of England Working Paper.

Schorfheide, Frank (2005), "VAR Forecasting under Misspecification," *Journal of Econometrics*, 128, 99-136.

Stock, James H. and Mark W. Watson (2005), "Implications of Dynamic Factor Models for VAR Analysis," NBER Working Paper 11467.

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

Waggoner, Daniel F., and Tao Zha (1999), "Conditional Forecasts in Dynamic Multivariate Models," *The Review of Economics and Statistics*, 81, 639-651.

West, Kenneth D. (1996), "Asymptotic Inference about Predictive Ability," *Econometrica*, 64, 1067-1084.

Tables and Figures

Table 1: Comparison between FRED-MD and MSW – Distribution of Relative MSEs of ARDL-Based DMS vs. VAR-Based IMS Bivariate Unconditional Forecasts

Model	Mean/percentile	FRED-MD				MSW			
		Forecast horizon				Forecast horizon			
		3	6	12	24	3	6	12	24
AR(4)	Mean	1.00	1.00	1.02	1.07	1.00	1.00	1.02	1.09
	0.10	0.96	0.91	0.88	0.84	0.96	0.90	0.85	0.82
	0.25	0.98	0.96	0.95	0.95	0.99	0.97	0.96	0.96
	0.50	1.00	1.01	1.02	1.06	1.00	1.01	1.02	1.06
	0.75	1.02	1.03	1.08	1.14	1.02	1.04	1.08	1.19
	0.90	1.03	1.08	1.16	1.32	1.03	1.07	1.16	1.37
AR(12)	Mean	1.01	1.04	1.07	1.15	1.02	1.04	1.07	1.16
	0.10	0.99	0.98	0.96	0.93	0.99	0.97	0.95	0.91
	0.25	1.00	1.00	1.01	1.02	1.00	1.00	1.01	1.03
	0.50	1.01	1.02	1.05	1.11	1.01	1.03	1.06	1.13
	0.75	1.02	1.06	1.11	1.22	1.02	1.06	1.12	1.28
	0.90	1.05	1.12	1.19	1.41	1.04	1.10	1.20	1.45
BIC	Mean	0.96	0.95	0.97	1.02	0.98	0.97	0.99	1.06
	0.10	0.80	0.71	0.69	0.68	0.88	0.79	0.78	0.79
	0.25	0.93	0.89	0.90	0.94	0.96	0.93	0.92	0.94
	0.50	1.00	1.00	1.01	1.04	1.00	1.00	1.00	1.04
	0.75	1.02	1.03	1.05	1.13	1.02	1.03	1.06	1.15
	0.90	1.05	1.08	1.13	1.27	1.05	1.08	1.15	1.31
AIC	Mean	1.01	1.03	1.06	1.13	1.01	1.02	1.05	1.15
	0.10	0.94	0.93	0.91	0.88	0.94	0.91	0.89	0.87
	0.25	0.98	0.98	0.99	0.99	0.98	0.98	0.98	1.00
	0.50	1.01	1.02	1.05	1.09	1.01	1.02	1.05	1.11
	0.75	1.03	1.06	1.12	1.22	1.04	1.07	1.13	1.26
	0.90	1.08	1.14	1.23	1.41	1.08	1.13	1.23	1.47

Notes: The left-hand panel is our attempt to replicate Table 5 from MSW, which is displayed in the right-hand panel, using our dataset. Entries correspond to the indicated summary measure (i.e. mean, 10th percentile, 25th percentile, etc.) of the distribution of the ratio of the MSE for the ARDL-based DMS forecast to the MSE for the VAR-based IMS forecast for the given lag-selection method, horizon, and grouping. Per the procedure followed in MSW, these measures are computed over 2,000 randomly selected pairs of series (4,000 sets of forecasts) as described in the text, with each set of forecasts constructed using an in-sample period starting in 1959:01 over an out-of-sample period ranging from 1979:01+h to 2002:12.

Table 2: Distribution of Relative MSEs of ARDL-Based DMS vs. VAR-Based IMS Bivariate Conditional Forecasts

	Data:	Full			Full			Great Moderation			Full			Full			Great Moderation		
	Sample:	MSW			2003-2016			2003-2016			MSW			2003-2016			2003-2016		
	Forecast Horizon:	3	12	24	3	12	24	3	12	24	3	12	24	3	12	24	3	12	24
		(A) All variables									*(B) Pairs not including PWM*								
AR(4)	Mean	1.00	1.04	1.11	1.00	1.00	1.06	1.00	0.98	1.01	1.00	1.05	1.12	1.01	1.00	1.04	1.00	1.00	1.04
	Median	1.00	1.02	1.07	1.00	0.99	1.00	1.00	0.97	0.95	1.01	1.04	1.09	1.00	1.00	1.01	1.00	0.99	0.99
	VAR Better	3.2%	7.5%	11.2%	6.3%	3.4%	5.7%	2.5%	2.3%	4.3%	3.5%	8.6%	12.8%	7.7%	3.7%	5.9%	2.9%	2.5%	5.5%
	ARDL Better	5.7%	8.2%	5.9%	4.8%	7.6%	9.6%	1.5%	3.8%	7.0%	5.2%	4.3%	2.8%	5.4%	5.5%	7.3%	1.8%	2.2%	4.0%
AR(12)	Mean	1.02	1.08	1.18	1.01	1.01	1.07	1.01	1.00	1.03	1.02	1.08	1.18	1.01	1.03	1.07	1.02	1.03	1.06
	Median	1.01	1.06	1.13	1.01	1.01	1.03	1.01	1.01	0.99	1.01	1.06	1.13	1.01	1.03	1.05	1.02	1.03	1.03
	VAR Better	5.3%	10.4%	15.3%	4.6%	5.2%	8.3%	4.0%	4.1%	5.5%	4.8%	11.5%	16.7%	5.3%	5.5%	6.4%	4.5%	4.8%	6.8%
	ARDL Better	0.9%	1.3%	2.0%	1.9%	3.0%	5.1%	0.6%	1.6%	3.6%	0.8%	0.9%	1.3%	1.9%	3.3%	5.3%	1.0%	1.4%	3.1%
BIC	Mean	0.95	0.96	1.04	0.96	0.94	0.99	0.97	0.92	0.94	0.98	0.99	1.09	0.97	0.95	1.01	0.98	0.98	1.01
	Median	0.99	0.98	1.04	0.99	0.97	0.98	0.99	0.95	0.93	0.99	0.99	1.06	0.99	0.97	0.98	0.99	0.97	0.97
	VAR Better	3.7%	6.3%	9.6%	4.2%	3.0%	5.3%	2.3%	2.7%	3.5%	4.5%	7.0%	11.6%	4.3%	2.2%	4.9%	2.6%	3.0%	4.1%
	ARDL Better	17.7%	16.1%	11.3%	10.2%	9.0%	10.4%	5.5%	8.4%	10.3%	12.6%	8.3%	4.4%	11.5%	5.6%	7.0%	5.6%	2.0%	3.4%
AIC	Mean	1.01	1.06	1.16	1.00	1.01	1.05	1.01	1.00	1.02	1.02	1.07	1.17	1.00	1.02	1.06	1.01	1.03	1.05
	Median	1.01	1.05	1.11	1.00	1.01	1.03	1.00	1.00	0.97	1.01	1.06	1.12	1.00	1.02	1.04	1.01	1.02	1.00
	VAR Better	4.2%	7.7%	12.1%	3.0%	3.8%	8.4%	2.9%	3.5%	4.8%	5.0%	8.8%	13.8%	3.2%	3.8%	7.2%	3.0%	3.5%	5.9%
	ARDL Better	2.8%	2.9%	2.8%	3.9%	3.5%	6.3%	2.6%	2.9%	5.2%	1.9%	2.1%	1.7%	4.5%	3.3%	6.2%	2.0%	1.3%	2.9%
		(C) Non-PWM variables in pairs with PWM variable									*(D) PWM variables*								
AR(4)	Mean	1.01	1.11	1.23	1.02	1.15	1.31	1.01	1.06	1.15	0.98	0.92	0.96	0.98	0.88	0.88	0.98	0.82	0.78
	Median	1.01	1.08	1.14	1.01	1.07	1.10	1.01	1.01	1.00	0.98	0.91	0.94	0.98	0.85	0.80	0.98	0.80	0.73
	VAR Better	4.8%	11.5%	16.8%	7.9%	5.6%	9.3%	3.6%	3.0%	3.8%	0.8%	0.0%	1.0%	0.4%	0.4%	1.6%	0.1%	1.0%	1.3%
	ARDL Better	2.1%	0.8%	1.5%	2.1%	2.3%	2.9%	0.4%	0.1%	1.5%	10.5%	27.3%	19.6%	5.5%	18.9%	23.0%	2.0%	12.3%	21.6%
AR(12)	Mean	1.02	1.12	1.26	1.02	1.07	1.20	1.02	1.08	1.18	1.01	1.03	1.09	0.99	0.90	0.90	0.99	0.83	0.82
	Median	1.01	1.09	1.18	1.01	1.03	1.10	1.02	1.04	1.02	1.01	1.03	1.06	0.99	0.90	0.84	0.99	0.85	0.77
	VAR Better	5.6%	11.1%	22.0%	4.5%	7.4%	17.6%	5.8%	5.6%	5.5%	6.4%	6.4%	4.5%	2.8%	2.0%	4.6%	0.8%	0.3%	1.5%
	ARDL Better	0.8%	0.8%	0.9%	1.5%	2.6%	2.6%	0.0%	0.5%	0.9%	1.6%	3.3%	5.3%	2.1%	2.4%	7.1%	0.0%	3.4%	8.0%
BIC	Mean	1.00	1.09	1.19	1.01	1.14	1.23	1.01	1.07	1.13	0.84	0.72	0.76	0.88	0.70	0.68	0.89	0.59	0.55
	Median	1.01	1.06	1.12	1.01	1.07	1.12	1.01	1.02	1.00	0.83	0.70	0.73	0.88	0.62	0.54	0.85	0.53	0.43
	VAR Better	4.1%	10.1%	12.4%	7.6%	8.1%	10.0%	3.1%	3.5%	3.0%	0.5%	0.1%	0.8%	0.6%	0.4%	1.6%	0.8%	0.9%	1.9%
	ARDL Better	4.6%	2.0%	2.0%	3.5%	2.5%	2.0%	1.0%	0.8%	1.4%	46.3%	53.5%	41.0%	13.1%	25.4%	28.9%	9.8%	34.9%	39.9%
AIC	Mean	1.01	1.13	1.26	1.02	1.08	1.20	1.01	1.10	1.18	0.99	0.98	1.04	0.99	0.90	0.90	0.97	0.79	0.76
	Median	1.01	1.09	1.18	1.01	1.05	1.12	1.01	1.05	1.03	0.99	0.98	1.02	1.00	0.89	0.83	0.97	0.77	0.67
	VAR Better	4.1%	10.0%	16.8%	3.1%	6.3%	16.9%	3.6%	6.5%	5.3%	1.6%	1.8%	2.1%	2.0%	1.4%	3.6%	1.5%	0.5%	1.0%
	ARDL Better	5.1%	2.0%	1.0%	2.0%	2.0%	2.4%	1.3%	0.6%	1.6%	3.4%	6.4%	8.0%	3.8%	5.4%	10.3%	5.9%	10.3%	15.8%

Notes: The first row of the column headers denotes the data used, with "Full" indicating an in-sample period starting in 1959:01 and "Great Moderation" indicating an in-sample period starting in 1984:01. The second row of the column headers denotes the out-of-sample period, with"MSW"indicating a range of 1979:01+h to 2002:12 and "2003-2016" indicating a range of 2002:12+h to 2016:12. The non-percent entries correspond to the mean or median of the distribution of the ratio of the MSE for the ARDL-based DMS forecast to the MSE for the VAR-based IMS forecast for the given lag-selection method, horizon, grouping, data, and sample. The entries reported as percentages correspond to the fraction of comparisons for the given lag-selection method, horizon, grouping, data, and sample rejected in favor of either the VAR model or the ARDL model based on a simple t-test of equal MSE at the 5 percent level. For these t-tests, the standard errors were constructed using a Newey-West (1987) HAC with the lag length set to $h*1.5$ (rounded up to 5 in the case of $h = 3$).

Table 3: Distribution of Relative MSEs of ARDL-Based DMS vs. VAR-Based IMS Trivariate Conditional Forecasts

	Data:	Full			Full			Great Moderation			Full			Full			Great Moderation		
	Sample:	MSW			2003-2016			2003-2016			MSW			2003-2016			2003-2016		
	Forecast Horizon:	3	12	24	3	12	24	3	12	24	3	12	24	3	12	24	3	12	24
		(A) All variables									*(B) Real variables*								
AR(4)	Mean	1.02	1.05	1.11	1.01	0.99	1.04	1.00	0.91	0.89	1.03	1.12	1.24	1.00	1.03	1.09	1.00	0.94	0.97
	Median	1.01	1.04	1.08	1.00	0.96	1.02	1.00	0.88	0.87	1.02	1.09	1.19	1.01	1.03	1.09	1.00	0.93	0.93
	VAR Better	6.2%	8.1%	13.2%	6.7%	6.4%	6.7%	2.8%	1.4%	3.6%	13.0%	11.7%	19.3%	5.7%	10.7%	10.3%	6.3%	1.7%	5.0%
	ARDL Better	0.6%	3.8%	3.6%	0.4%	6.4%	6.8%	0.2%	7.8%	12.4%	0.0%	0.0%	0.0%	1.3%	0.3%	1.7%	0.7%	0.0%	1.7%
AR(12)	Mean	1.02	1.12	1.25	1.02	1.03	1.07	1.02	0.96	0.91	1.04	1.22	1.48	1.03	1.10	1.10	1.03	1.06	1.01
	Median	1.02	1.11	1.15	1.01	1.02	1.03	1.01	0.97	0.93	1.03	1.22	1.42	1.02	1.09	1.10	1.03	1.04	0.98
	VAR Better	8.7%	15.2%	17.0%	2.0%	2.3%	6.1%	3.0%	2.8%	3.2%	9.7%	32.0%	29.0%	3.3%	6.0%	11.3%	3.3%	1.0%	5.3%
	ARDL Better	0.0%	0.6%	0.8%	1.2%	0.8%	2.9%	0.3%	1.1%	3.6%	0.0%	1.3%	0.0%	1.3%	0.0%	2.3%	1.0%	0.0%	0.0%
BIC	Mean	0.92	0.89	0.93	0.92	0.86	0.90	0.93	0.78	0.77	1.00	1.07	1.15	0.95	1.05	1.10	0.96	0.99	1.02
	Median	0.98	0.97	0.93	0.95	0.89	0.98	0.95	0.86	0.83	1.00	1.06	1.09	0.97	1.03	1.09	0.97	0.99	1.00
	VAR Better	3.2%	4.4%	9.8%	3.2%	5.0%	6.3%	1.3%	1.7%	1.1%	4.7%	6.7%	11.7%	2.0%	6.0%	10.7%	0.3%	4.7%	2.3%
	ARDL Better	30.3%	31.9%	22.6%	4.6%	22.9%	24.2%	7.7%	24.9%	29.3%	2.0%	1.7%	0.3%	1.7%	0.3%	2.3%	0.3%	0.3%	0.0%
AIC	Mean	1.02	1.10	1.22	1.01	1.01	1.06	0.96	0.87	0.84	1.06	1.23	1.48	1.03	1.11	1.12	1.03	1.01	1.02
	Median	1.01	1.08	1.13	1.00	0.99	1.04	0.96	0.90	0.88	1.03	1.19	1.43	1.02	1.08	1.12	1.02	1.00	1.02
	VAR Better	5.0%	9.1%	13.3%	1.1%	2.0%	5.9%	1.8%	2.8%	4.9%	9.7%	18.0%	22.3%	1.3%	3.7%	9.3%	4.3%	4.7%	8.0%
	ARDL Better	2.6%	1.4%	3.2%	3.6%	2.3%	3.8%	4.4%	4.2%	10.4%	0.7%	0.3%	0.0%	1.0%	0.0%	0.3%	0.7%	0.7%	1.3%
		(C) Nominal variables									*(D) Financial variables*								
AR(4)	Mean	1.00	0.97	1.01	1.00	0.85	0.86	1.00	0.76	0.69	1.02	1.06	1.08	1.02	1.08	1.18	1.01	1.03	1.03
	Median	1.00	0.95	0.98	0.99	0.85	0.85	1.00	0.78	0.70	1.01	1.06	1.04	1.01	1.02	1.05	1.00	0.96	1.00
	VAR Better	0.0%	0.0%	1.0%	1.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.7%	12.7%	19.3%	13.3%	8.7%	9.7%	2.0%	2.7%	5.7%
	ARDL Better	1.7%	6.7%	3.7%	0.0%	17.7%	17.7%	0.0%	23.0%	31.0%	0.0%	4.7%	7.0%	0.0%	1.3%	1.0%	0.0%	0.3%	4.7%
AR(12)	Mean	1.01	1.05	1.12	1.00	0.91	0.88	1.00	0.74	0.61	1.01	1.08	1.15	1.03	1.08	1.23	1.02	1.08	1.12
	Median	1.01	1.06	1.06	1.00	0.88	0.85	1.00	0.70	0.56	1.01	1.08	1.08	1.01	1.04	1.09	1.02	1.06	1.07
	VAR Better	4.7%	4.7%	2.0%	0.7%	0.3%	1.7%	0.0%	0.0%	0.0%	11.7%	9.0%	20.0%	2.0%	0.7%	5.3%	5.7%	7.3%	4.3%
	ARDL Better	0.0%	0.3%	1.3%	2.0%	2.3%	6.0%	0.0%	3.3%	8.7%	0.0%	0.0%	1.0%	0.3%	0.0%	0.3%	0.0%	0.0%	2.0%
BIC	Mean	0.76	0.63	0.64	0.82	0.55	0.52	0.84	0.43	0.33	0.98	0.96	0.99	0.99	0.99	1.09	0.99	0.91	0.95
	Median	0.77	0.62	0.64	0.83	0.54	0.51	0.85	0.45	0.33	1.01	1.02	1.00	1.00	0.96	1.05	1.00	0.92	0.94
	VAR Better	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.0%	6.7%	17.7%	7.7%	9.0%	8.3%	3.7%	0.3%	1.0%
	ARDL Better	77.7%	75.7%	55.0%	11.3%	64.3%	64.7%	21.0%	69.3%	74.0%	11.3%	18.3%	12.3%	0.7%	4.0%	5.7%	1.7%	5.0%	14.0%
AIC	Mean	1.01	1.01	1.03	0.98	0.89	0.87	0.90	0.57	0.43	1.00	1.05	1.13	1.01	1.04	1.20	0.96	1.03	1.07
	Median	1.01	1.01	1.04	0.99	0.89	0.85	0.90	0.55	0.39	1.00	1.05	1.08	0.99	0.99	1.08	0.96	0.97	0.95
	VAR Better	5.3%	2.3%	3.0%	0.3%	0.0%	1.0%	0.3%	0.0%	0.0%	0.0%	7.0%	14.7%	1.7%	2.3%	7.3%	0.7%	3.7%	6.7%
	ARDL Better	0.0%	1.3%	6.7%	4.7%	5.3%	9.3%	7.7%	11.0%	26.7%	7.0%	2.7%	3.0%	5.0%	1.7%	1.7%	5.0%	1.0%	3.3%

Notes: See notes to Table 2.

Table 4: Distribution of MSEs of ARDL-Based DMS and VAR-Based IMS Trivariate Conditional Forecasts Relative to Those from VAR(4)

		Full Data, MSW Sample										Great Moderation Data, 2003-2016 Sample									
		Iterated forecasts					Direct forecasts					Iterated forecasts					Direct forecasts				
Horizon	Summary variable	AR(4)	AR(12)	BIC	AIC	Sum	AR(4)	AR(12)	BIC	AIC	Sum	AR(4)	AR(12)	BIC	AIC	Sum	AR(4)	AR(12)	BIC	AIC	Sum
(A) All																					
3	Mean	1.00	1.07	1.10	1.02		1.02	1.10	0.99	1.05		1.00	1.02	1.10	1.01		1.00	1.04	1.01	0.97	
	Median	1.00	1.07	1.07	1.01		1.01	1.09	0.99	1.02		1.00	1.00	1.10	1.02		1.00	1.00	1.00	0.97	
	Fraction best	0.21	0.11	0.22	0.16	0.70	0.04	0.09	0.12	0.07	0.32	0.09	0.17	0.12	0.11	0.50	0.09	0.16	0.11	0.15	0.51
12	Mean	1.00	0.98	1.18	0.95		1.05	1.12	1.01	1.04		1.00	0.92	1.14	0.96		0.91	0.91	0.85	0.84	
	Median	1.00	1.02	1.12	0.97		1.04	1.14	1.03	1.05		1.00	0.88	1.10	0.96		0.88	0.90	0.89	0.87	
	Fraction best	0.22	0.16	0.14	0.26	0.78	0.02	0.16	0.04	0.12	0.33	0.06	0.09	0.10	0.10	0.36	0.09	0.35	0.16	0.20	0.80
24	Mean	1.00	0.94	1.18	0.94		1.11	1.19	1.04	1.15		1.00	0.89	1.15	0.96		0.89	0.85	0.82	0.82	
	Median	1.00	0.95	1.11	0.95		1.08	1.12	1.04	1.09		1.00	0.90	1.05	0.98		0.87	0.83	0.86	0.84	
	Fraction best	0.18	0.23	0.11	0.24	0.77	0.01	0.14	0.10	0.12	0.36	0.05	0.09	0.12	0.08	0.35	0.07	0.42	0.10	0.29	0.87
(B) Real																					
3	Mean	1.00	1.17	1.03	1.06		1.03	1.21	1.03	1.13		1.00	1.02	1.12	1.00		1.00	1.05	1.08	1.03	
	Median	1.00	1.12	1.02	1.05		1.02	1.15	1.03	1.09		1.00	1.02	1.13	1.01		1.00	1.04	1.07	1.03	
	Fraction best	0.43	0.05	0.19	0.06	0.73	0.07	0.00	0.18	0.02	0.27	0.14	0.21	0.13	0.21	0.69	0.12	0.07	0.09	0.03	0.31
12	Mean	1.00	1.10	1.09	1.00		1.12	1.36	1.15	1.23		1.00	0.98	1.06	0.97		0.94	1.03	1.04	0.98	
	Median	1.00	1.08	1.06	1.00		1.09	1.30	1.13	1.20		1.00	1.00	1.04	1.00		0.93	1.01	1.05	0.99	
	Fraction best	0.42	0.10	0.07	0.35	0.94	0.01	0.07	0.01	0.04	0.13	0.04	0.18	0.18	0.19	0.59	0.17	0.15	0.06	0.04	0.41
24	Mean	1.00	1.01	1.07	0.98		1.24	1.51	1.22	1.46		1.00	1.00	1.03	1.00		0.97	1.01	1.04	1.01	
	Median	1.00	1.02	1.06	0.99		1.19	1.46	1.18	1.36		1.00	1.00	1.03	1.01		0.93	0.96	1.04	0.99	
	Fraction best	0.30	0.22	0.11	0.29	0.92	0.00	0.06	0.03	0.03	0.13	0.04	0.15	0.20	0.11	0.50	0.15	0.28	0.02	0.07	0.52
(C) Nominal																					
3	Mean	1.00	0.92	1.24	0.92		1.00	0.93	0.94	0.92		1.00	0.89	1.15	1.00		1.00	0.89	0.96	0.90	
	Median	1.00	0.90	1.21	0.91		1.00	0.90	0.93	0.90		1.00	0.89	1.13	1.00		1.00	0.89	0.95	0.89	
	Fraction best	0.06	0.27	0.00	0.37	0.70	0.04	0.10	0.05	0.13	0.33	0.04	0.30	0.00	0.06	0.40	0.00	0.34	0.01	0.24	0.60
12	Mean	1.00	0.80	1.40	0.85		0.97	0.85	0.87	0.85		1.00	0.66	1.31	0.89		0.76	0.48	0.56	0.50	
	Median	1.00	0.78	1.37	0.82		0.95	0.85	0.88	0.85		1.00	0.66	1.31	0.90		0.78	0.46	0.57	0.48	
	Fraction best	0.05	0.33	0.00	0.29	0.68	0.02	0.23	0.04	0.21	0.50	0.00	0.07	0.00	0.00	0.07	0.00	0.85	0.06	0.36	1.27
24	Mean	1.00	0.78	1.42	0.86		1.01	0.90	0.89	0.89		1.00	0.62	1.38	0.89		0.69	0.37	0.45	0.38	
	Median	1.00	0.78	1.40	0.83		0.98	0.84	0.86	0.85		1.00	0.61	1.38	0.91		0.70	0.35	0.44	0.36	
	Fraction best	0.06	0.32	0.00	0.26	0.64	0.00	0.28	0.09	0.28	0.65	0.00	0.05	0.00	0.00	0.05	0.00	0.83	0.04	0.65	1.52
(D) Financial																					
3	Mean	1.00	1.13	1.03	1.09		1.02	1.14	1.00	1.09		1.00	1.14	1.01	1.04		1.01	1.16	1.00	1.00	
	Median	1.00	1.10	0.99	1.03		1.01	1.12	1.00	1.02		1.00	1.14	1.01	1.03		1.00	1.17	0.99	0.99	
	Fraction best	0.13	0.00	0.46	0.06	0.66	0.02	0.16	0.14	0.05	0.37	0.08	0.01	0.25	0.06	0.40	0.16	0.07	0.21	0.17	0.61
12	Mean	1.00	1.05	1.05	1.00		1.06	1.14	0.99	1.05		1.00	1.12	1.05	1.01		1.03	1.23	0.93	1.05	
	Median	1.00	1.09	1.01	1.01		1.06	1.19	1.02	1.07		1.00	1.10	1.04	0.98		0.96	1.11	0.90	0.97	
	Fraction best	0.19	0.04	0.35	0.15	0.74	0.02	0.16	0.06	0.12	0.37	0.15	0.04	0.13	0.11	0.43	0.10	0.06	0.36	0.20	0.71
24	Mean	1.00	1.02	1.05	0.98		1.08	1.17	1.02	1.12		1.00	1.05	1.06	0.99		1.03	1.17	0.97	1.07	
	Median	1.00	1.04	1.02	0.99		1.04	1.09	1.01	1.08		1.00	1.05	1.01	0.98		1.00	1.10	0.93	0.95	
	Fraction best	0.19	0.16	0.23	0.16	0.74	0.02	0.09	0.16	0.03	0.30	0.11	0.09	0.16	0.14	0.49	0.05	0.15	0.24	0.13	0.58

Notes: See notes to Table 2 for distinctions between data used ("Full" vs "Great Moderation") and between out-of-sample periods ("MSW" vs "2003-2016"). The "Mean" and "Median" entries correspond to the mean and median, respectively, of the distribution of the ratio of the MSE for the forecast from the given forecast method to the MSE of the forecast from the iterated VAR(4) for the given horizon, grouping, data, and sample. The "Fraction best" entries denote the fraction of all models considered for the given horizon, grouping, data, and sample in which the given forecast method has the smallest MSE among the eight possibilities; the sum of these fractions is reported in the "Sum" columns for all iterated and for all direct forecasts, respectively. The sum of the fraction best exceeds 1 in some cases because of ties.

Table 5: Distribution of Relative MSEs of ARDL-Based DMS vs. VAR-Based IMS Bivariate Unconditional Forecasts

	Data:	Full			Full			Great Moderation			Full			Full			Great Moderation		
	Sample:	MSW			2003-2016			2003-2016			MSW			2003-2016			2003-2016		
	Forecast Horizon:	3	12	24	3	12	24	3	12	24	3	12	24	3	12	24	3	12	24
		(A) All variables									*(B) Pairs not including PWM*								
AR(4)	Mean	1.00	1.02	1.07	1.00	0.99	1.01	1.00	0.93	0.89	1.01	1.05	1.12	1.01	1.01	1.05	1.00	0.96	0.93
	Median	1.00	1.02	1.06	1.00	0.99	1.01	1.00	0.95	0.92	1.01	1.04	1.09	1.00	1.01	1.03	1.01	0.96	0.94
	VAR Better	2.9%	5.7%	10.8%	6.5%	4.8%	6.3%	2.0%	1.5%	3.6%	3.0%	7.5%	13.0%	8.3%	5.7%	7.1%	2.6%	1.2%	4.3%
	ARDL Better	4.5%	13.3%	10.5%	4.6%	11.6%	13.4%	1.5%	8.3%	12.4%	3.2%	2.5%	0.9%	4.8%	6.3%	6.0%	1.4%	2.9%	6.2%
AR(12)	Mean	1.01	1.07	1.15	1.00	0.99	1.02	1.01	0.93	0.87	1.02	1.07	1.18	1.01	1.01	1.06	1.02	0.96	0.92
	Median	1.01	1.05	1.11	1.01	1.00	1.02	1.02	0.94	0.90	1.01	1.05	1.13	1.01	1.01	1.03	1.02	0.95	0.91
	VAR Better	4.3%	7.8%	10.5%	5.0%	4.1%	5.7%	4.1%	1.9%	2.6%	3.5%	8.2%	12.9%	5.9%	4.2%	5.3%	5.0%	1.8%	2.8%
	ARDL Better	0.7%	1.1%	1.7%	2.2%	3.7%	4.9%	0.5%	1.7%	4.3%	0.5%	0.4%	0.3%	2.1%	4.0%	5.0%	0.8%	1.4%	3.4%
BIC	Mean	0.96	0.97	1.02	0.96	0.94	0.98	0.97	0.89	0.86	0.99	1.02	1.10	0.98	0.99	1.05	0.98	0.96	0.94
	Median	1.00	1.01	1.04	0.99	0.99	1.01	0.99	0.95	0.93	1.00	1.02	1.08	1.00	1.00	1.02	1.00	0.97	0.96
	VAR Better	3.7%	7.3%	10.3%	4.9%	4.7%	7.5%	1.8%	1.2%	2.6%	4.7%	8.9%	12.8%	5.5%	5.5%	8.8%	2.0%	1.1%	2.8%
	ARDL Better	15.6%	15.6%	11.5%	9.2%	11.7%	10.7%	4.8%	11.8%	15.8%	8.8%	4.8%	1.8%	9.0%	6.8%	5.7%	4.9%	3.5%	7.4%
AIC	Mean	1.01	1.06	1.13	1.00	0.99	1.03	1.01	0.94	0.89	1.02	1.08	1.18	1.00	1.02	1.07	1.02	0.99	0.96
	Median	1.01	1.05	1.09	1.00	1.01	1.02	1.00	0.96	0.91	1.01	1.06	1.13	1.00	1.02	1.04	1.01	0.98	0.96
	VAR Better	3.9%	6.7%	10.5%	3.4%	3.5%	5.7%	3.2%	1.3%	2.0%	4.6%	8.5%	13.2%	3.8%	4.0%	6.0%	3.8%	1.6%	2.2%
	ARDL Better	2.8%	2.3%	2.8%	4.2%	4.7%	6.3%	2.0%	4.1%	7.2%	1.9%	0.9%	0.6%	4.7%	3.9%	5.3%	1.7%	1.9%	4.3%
		(C) Non-PWM variables in pairs with PWM variable									*(D) PWM variables*								
AR(4)	Mean	1.01	1.06	1.11	1.01	1.02	1.07	1.01	0.97	0.95	0.98	0.89	0.88	0.97	0.86	0.84	0.98	0.80	0.72
	Median	1.01	1.03	1.08	1.01	1.01	1.03	1.01	0.98	0.96	0.98	0.89	0.86	0.98	0.86	0.81	0.98	0.79	0.71
	VAR Better	4.4%	5.9%	14.5%	7.5%	6.3%	9.6%	2.4%	2.5%	4.0%	0.8%	0.0%	0.4%	0.0%	0.4%	0.5%	0.0%	1.1%	1.1%
	ARDL Better	2.5%	2.4%	0.0%	2.8%	4.0%	5.9%	0.6%	1.3%	6.9%	10.5%	56.5%	49.9%	5.6%	34.9%	43.1%	2.4%	31.4%	36.6%
AR(12)	Mean	1.02	1.10	1.17	1.01	1.04	1.07	1.02	0.97	0.92	1.00	1.03	1.02	0.98	0.89	0.86	0.99	0.79	0.69
	Median	1.01	1.08	1.14	1.01	1.03	1.05	1.02	1.00	0.95	1.01	1.02	1.00	0.99	0.89	0.84	0.99	0.83	0.70
	VAR Better	5.1%	9.1%	9.3%	4.6%	6.1%	10.6%	4.6%	3.9%	3.9%	5.9%	5.1%	4.3%	2.9%	1.5%	2.1%	0.6%	0.3%	0.6%
	ARDL Better	0.8%	0.9%	0.0%	2.0%	2.5%	3.3%	0.3%	0.8%	2.6%	1.3%	3.3%	7.9%	2.8%	3.9%	6.3%	0.0%	3.4%	8.6%
BIC	Mean	1.00	1.04	1.08	1.00	1.03	1.09	1.00	0.98	0.96	0.85	0.72	0.72	0.88	0.69	0.65	0.90	0.59	0.51
	Median	1.00	1.03	1.06	1.01	1.02	1.03	1.00	0.99	0.98	0.84	0.71	0.70	0.88	0.62	0.57	0.88	0.55	0.42
	VAR Better	4.1%	9.8%	13.0%	7.1%	6.9%	9.1%	1.9%	1.0%	1.9%	0.1%	0.0%	0.4%	0.6%	0.3%	1.8%	0.9%	1.6%	2.4%
	ARDL Better	5.5%	2.5%	0.0%	6.3%	6.0%	2.4%	1.1%	1.1%	4.4%	46.1%	61.0%	52.0%	12.6%	31.9%	34.1%	8.1%	47.4%	52.6%
AIC	Mean	1.01	1.08	1.15	1.00	1.03	1.07	1.00	0.99	0.93	0.99	0.97	0.96	0.99	0.88	0.85	0.97	0.75	0.64
	Median	1.00	1.05	1.10	1.00	1.02	1.05	1.01	0.99	0.94	0.99	0.97	0.96	1.00	0.88	0.82	0.98	0.75	0.61
	VAR Better	4.0%	5.8%	10.9%	3.1%	4.3%	8.9%	3.5%	1.1%	2.0%	1.6%	2.3%	1.9%	2.1%	1.0%	1.8%	1.1%	0.4%	1.4%
	ARDL Better	4.9%	0.9%	0.0%	2.6%	3.5%	3.9%	1.3%	2.6%	4.0%	3.6%	7.9%	11.9%	4.0%	8.0%	11.8%	3.4%	12.3%	18.9%

Notes: See notes to Table 2.

Table 6: Distribution of Relative MSEs of ARDL-Based DMS vs. VAR-Based IMS Trivariate Unconditional Forecasts

	Data:	Full			Full			Great Moderation			Full			Full			Great Moderation		
	Sample:	MSW			2003-2016			2003-2016			MSW			2003-2016			2003-2016		
	Forecast Horizon:	3	12	24	3	12	24	3	12	24	3	12	24	3	12	24	3	12	24
					(A) All variables									*(B) Real variables*					
AR(4)	Mean	1.02	1.00	1.03	1.01	0.96	0.95	1.00	0.90	0.87	1.03	1.03	1.11	1.00	1.00	1.04	1.00	0.95	0.99
	Median	1.01	1.01	1.05	1.00	0.97	0.97	1.01	0.91	0.89	1.02	1.02	1.10	1.01	1.03	1.05	1.01	0.93	0.97
	VAR Better	6.7%	9.6%	8.4%	6.9%	4.9%	7.6%	0.9%	0.0%	1.1%	13.3%	10.0%	3.3%	4.7%	8.0%	10.7%	2.0%	0.0%	0.0%
	ARDL Better	0.7%	10.4%	10.9%	1.3%	16.2%	20.7%	0.0%	15.6%	21.1%	0.0%	0.0%	0.0%	4.0%	3.3%	0.7%	0.0%	0.0%	2.0%
AR(12)	Mean	1.02	1.11	1.16	1.01	0.99	1.00	1.02	0.89	0.77	1.04	1.21	1.34	1.03	1.12	1.09	1.03	1.05	0.94
	Median	1.02	1.11	1.10	1.01	0.96	0.97	1.01	0.90	0.80	1.03	1.19	1.28	1.02	1.11	1.07	1.03	1.03	0.95
	VAR Better	6.9%	10.7%	8.9%	2.9%	1.3%	2.2%	3.3%	0.7%	1.3%	8.0%	24.7%	10.7%	4.7%	3.3%	6.0%	3.3%	0.7%	4.0%
	ARDL Better	0.0%	0.2%	2.0%	0.9%	1.3%	4.0%	0.2%	0.4%	6.7%	0.0%	0.0%	0.0%	1.3%	0.0%	1.3%	0.7%	0.0%	1.3%
BIC	Mean	0.92	0.87	0.87	0.93	0.84	0.84	0.94	0.79	0.76	1.01	1.03	1.07	0.96	1.02	1.07	0.97	1.02	1.05
	Median	0.98	0.99	0.99	0.95	0.96	0.98	0.96	0.92	0.86	1.01	1.03	1.05	0.98	1.01	1.05	0.97	1.04	1.03
	VAR Better	3.1%	8.7%	4.4%	3.3%	4.2%	6.0%	0.7%	4.4%	1.8%	6.0%	23.3%	0.7%	2.0%	6.0%	15.3%	0.7%	13.3%	4.7%
	ARDL Better	29.6%	36.7%	28.9%	3.6%	29.1%	28.9%	6.0%	37.3%	38.2%	2.0%	2.7%	0.0%	1.3%	2.0%	0.0%	0.7%	1.3%	0.0%
AIC	Mean	1.02	1.06	1.10	1.00	0.95	0.98	0.96	0.83	0.78	1.05	1.18	1.26	1.02	1.09	1.09	1.03	1.01	1.01
	Median	1.01	1.05	1.06	0.99	0.95	0.97	0.96	0.85	0.83	1.04	1.15	1.13	1.02	1.06	1.07	1.03	0.97	1.00
	VAR Better	4.9%	4.7%	8.4%	1.8%	1.1%	1.3%	1.3%	0.4%	0.4%	8.7%	9.3%	8.0%	4.0%	3.3%	2.7%	2.7%	1.3%	1.3%
	ARDL Better	2.9%	3.3%	5.3%	4.0%	1.3%	9.1%	7.8%	8.2%	14.9%	0.7%	1.3%	0.0%	0.7%	0.0%	4.7%	0.7%	1.3%	4.7%
					(C) Nominal variables									*(D) Financial variables*					
AR(4)	Mean	1.00	0.92	0.90	0.99	0.86	0.84	1.00	0.77	0.68	1.02	1.05	1.08	1.02	1.02	0.98	1.01	0.98	0.94
	Median	1.00	0.92	0.90	0.99	0.86	0.84	1.00	0.76	0.69	1.01	1.04	1.06	1.01	1.02	0.98	1.01	0.99	0.93
	VAR Better	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.7%	18.7%	22.0%	16.0%	6.7%	12.0%	0.7%	0.0%	3.3%
	ARDL Better	2.0%	21.3%	30.0%	0.0%	38.7%	44.0%	0.0%	40.0%	44.7%	0.0%	10.0%	2.7%	0.0%	6.7%	17.3%	0.0%	6.7%	16.7%
AR(12)	Mean	1.01	1.04	0.98	1.00	0.88	0.83	1.00	0.69	0.52	1.01	1.08	1.15	1.02	0.99	1.08	1.02	0.92	0.85
	Median	1.01	1.05	1.00	0.99	0.87	0.79	1.00	0.65	0.45	1.01	1.10	1.10	1.00	0.96	0.97	1.02	0.90	0.82
	VAR Better	4.0%	0.0%	1.3%	0.7%	0.0%	0.7%	0.7%	0.0%	0.0%	8.7%	7.3%	14.7%	3.3%	0.7%	0.0%	6.0%	1.3%	0.0%
	ARDL Better	0.0%	0.7%	6.0%	1.3%	3.3%	8.0%	0.0%	0.7%	13.3%	0.0%	0.0%	0.0%	0.0%	0.7%	2.7%	0.0%	0.7%	5.3%
BIC	Mean	0.77	0.61	0.57	0.83	0.56	0.54	0.87	0.46	0.37	0.98	0.96	0.97	0.99	0.94	0.93	0.99	0.89	0.86
	Median	0.77	0.60	0.61	0.84	0.57	0.55	0.88	0.46	0.37	1.00	1.02	1.02	1.00	0.99	0.98	0.99	0.94	0.86
	VAR Better	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.3%	2.7%	12.7%	8.0%	6.7%	2.7%	1.3%	0.0%	0.7%
	ARDL Better	76.7%	87.3%	78.7%	8.7%	77.3%	70.0%	17.3%	87.3%	90.7%	10.0%	20.0%	8.0%	0.7%	8.0%	16.7%	0.0%	23.3%	24.0%
AIC	Mean	1.01	0.97	0.89	0.98	0.86	0.83	0.90	0.56	0.41	1.00	1.03	1.16	0.99	0.91	1.02	0.94	0.91	0.91
	Median	1.02	0.99	0.92	0.99	0.87	0.83	0.91	0.54	0.38	1.00	1.01	1.09	0.98	0.93	0.96	0.95	0.89	0.85
	VAR Better	5.3%	0.7%	1.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%	4.0%	16.0%	1.3%	0.0%	1.3%	1.3%	0.0%	0.0%
	ARDL Better	0.0%	3.3%	16.0%	4.7%	3.3%	14.7%	8.7%	15.3%	26.0%	8.0%	5.3%	0.0%	6.7%	0.7%	8.0%	14.0%	8.0%	14.0%

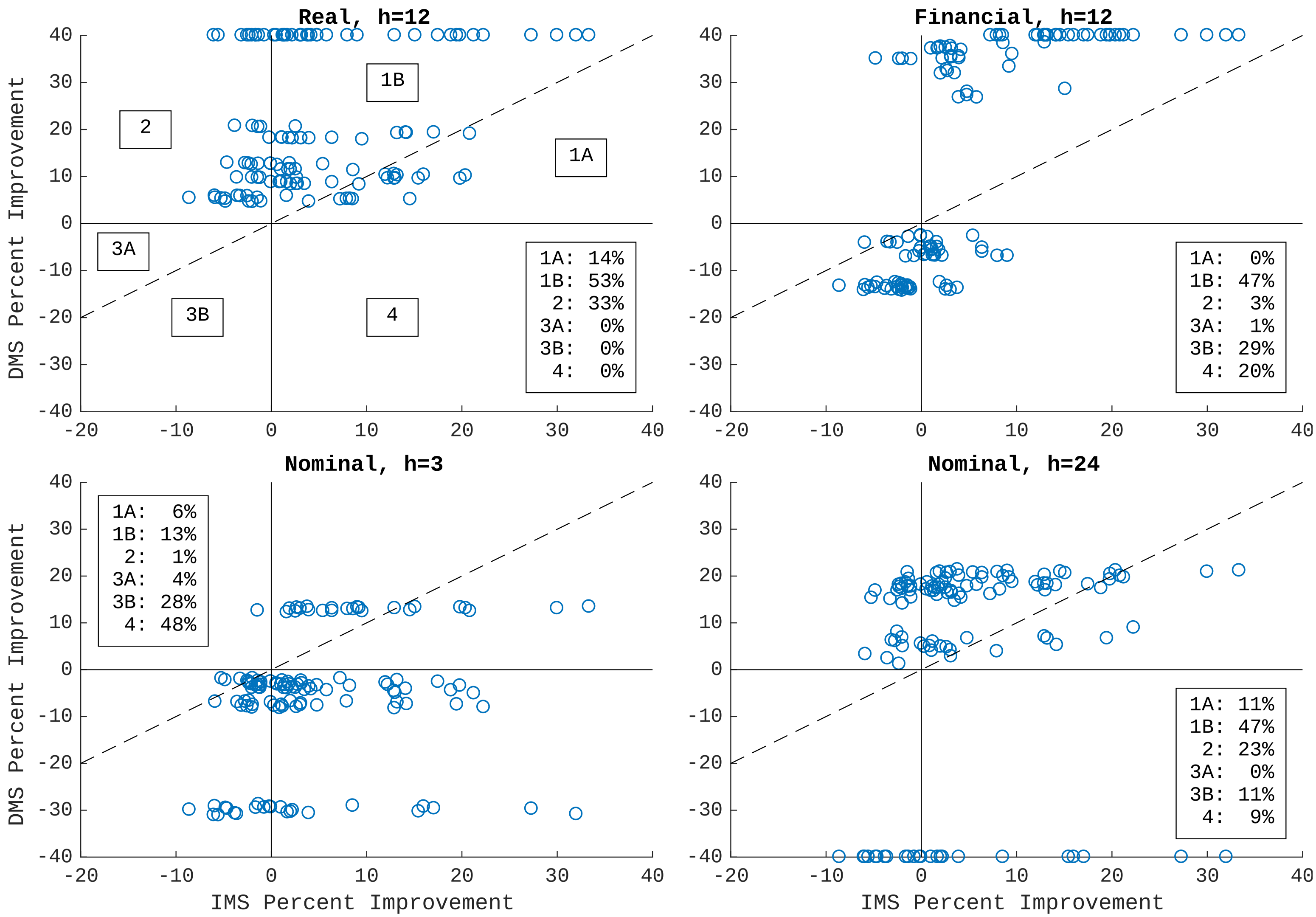
Notes: See notes to Table 2.

Table 7: Distribution of t-statistics from Bias, Efficiency, and Equal MSE Tests

Data/ Sample:		Full/ MSW									Great Moderation/ 2003-2016								
Test:		Bias			Efficiency			Equal MSE			Bias			Efficiency			Equal MSE		
Forecast Horizon:		3	12	24	3	12	24	3	12	24	3	12	24	3	12	24	3	12	24
(A) All variables																			
AR(2)	Mean	0.08	0.89	0.74	-1.37	-1.80	-1.34	-1.10	-1.19	-1.36	-0.06	0.14	0.06	-1.51	-1.66	-1.61	-0.91	-0.85	-0.73
	Median	-0.12	0.32	0.20	-1.55	-1.68	-0.88	-1.19	-1.18	-1.35	-0.03	0.33	0.34	-1.53	-1.26	-0.81	-0.93	-0.77	-0.69
	Rejected	7.8%	19.8%	12.2%	26.4%	41.8%	18.7%	26.1%	23.6%	23.7%	4.9%	12.4%	17.1%	46.7%	30.4%	31.8%	14.1%	9.1%	6.1%
AR(4)	Mean	0.52	0.60	0.77	-3.47	-1.46	-1.40	-1.06	-1.58	-1.66	0.17	0.57	0.12	-2.38	-1.04	-1.69	-0.85	-1.07	-0.96
	Median	0.01	0.02	0.14	-3.48	-1.26	-1.33	-1.11	-1.57	-1.65	-0.04	0.41	0.14	-2.42	-1.24	-0.89	-0.89	-1.02	-0.99
	Rejected	14.2%	16.7%	12.9%	67.8%	27.1%	20.2%	28.6%	34.9%	31.9%	3.8%	6.7%	14.4%	41.6%	24.4%	22.0%	14.9%	10.8%	7.3%
AR(12)	Mean	1.03	0.62	0.50	-6.81	-3.50	-2.42	-1.00	-2.22	-2.59	0.52	0.77	0.49	-4.31	-3.09	-1.87	-0.70	-1.45	-1.69
	Median	0.01	-0.05	0.03	-6.79	-3.24	-1.94	-1.09	-2.26	-2.68	0.00	0.18	0.06	-4.22	-2.66	-1.69	-0.67	-1.44	-1.72
	Rejected	24.4%	16.2%	11.8%	92.7%	50.9%	25.6%	27.7%	53.8%	52.3%	5.6%	6.4%	6.2%	57.8%	38.9%	21.3%	14.3%	18.6%	17.9%
(B) Real variables																			
AR(2)	Mean	0.44	2.42	2.21	-0.48	-2.73	-2.56	-1.13	-1.08	-1.34	-0.73	1.28	1.30	0.66	-2.54	-2.76	-1.10	-0.71	-0.56
	Median	0.63	2.23	1.80	-0.65	-2.59	-2.37	-1.21	-1.03	-1.12	-0.14	1.64	1.20	0.64	-2.57	-2.50	-1.04	-0.67	-0.56
	Rejected	8.7%	39.3%	33.3%	12.7%	57.3%	48.7%	25.0%	20.0%	20.7%	9.3%	13.3%	15.3%	14.7%	34.0%	35.3%	19.7%	8.7%	4.7%
AR(4)	Mean	1.28	2.01	2.35	-2.21	-2.31	-2.77	-1.09	-1.49	-1.52	0.06	1.46	1.44	-0.84	-2.40	-2.75	-0.99	-0.71	-0.60
	Median	1.09	1.72	1.84	-2.15	-1.91	-2.38	-1.13	-1.57	-1.50	-0.28	1.55	1.20	-0.76	-2.11	-2.61	-0.94	-0.77	-0.64
	Rejected	22.7%	27.3%	34.7%	42.0%	36.7%	50.7%	27.7%	38.0%	35.3%	0.7%	15.3%	20.0%	6.7%	36.7%	40.0%	19.0%	5.0%	3.7%
AR(12)	Mean	2.95	1.92	1.99	-6.05	-3.25	-2.85	-1.01	-2.45	-2.86	1.09	1.96	1.55	-3.38	-4.52	-3.35	-0.77	-0.97	-1.18
	Median	3.30	1.72	1.68	-6.51	-3.19	-2.46	-1.08	-2.42	-2.85	0.21	1.91	1.45	-3.20	-4.58	-3.65	-0.64	-0.91	-1.13
	Rejected	53.3%	23.3%	22.7%	85.3%	54.0%	36.0%	26.7%	56.7%	57.7%	16.7%	18.7%	12.0%	37.3%	58.7%	32.7%	21.0%	14.0%	11.7%
(C) Nominal variables																			
AR(2)	Mean	-0.10	0.15	0.04	-1.04	-0.65	-0.33	-1.41	-1.50	-1.37	-0.02	-0.02	0.15	-4.01	-0.69	0.24	-0.73	-0.82	-0.66
	Median	-0.04	0.07	0.04	-1.09	-0.70	-0.36	-1.35	-1.50	-1.53	-0.03	0.01	0.08	-4.02	-0.31	0.40	-0.78	-0.76	-0.62
	Rejected	0.7%	0.0%	0.0%	9.3%	24.7%	4.7%	36.0%	34.0%	27.3%	0.0%	1.3%	19.3%	96.7%	19.3%	28.0%	13.3%	11.0%	8.3%
AR(4)	Mean	-0.08	-0.24	0.05	-4.43	-0.18	-0.29	-1.39	-1.94	-1.79	-0.04	0.21	-0.05	-3.39	0.88	-0.11	-0.72	-1.10	-0.98
	Median	-0.08	-0.17	0.05	-4.27	-0.35	-0.26	-1.49	-1.94	-1.91	-0.04	0.09	0.00	-3.53	0.49	-0.08	-0.71	-0.89	-0.93
	Rejected	0.0%	3.3%	0.0%	91.3%	4.0%	5.3%	41.3%	43.3%	38.7%	0.0%	1.3%	8.0%	68.7%	14.7%	5.3%	11.3%	13.3%	7.7%
AR(12)	Mean	-0.08	-0.15	-0.07	-7.32	-2.62	-1.39	-1.40	-2.53	-2.71	-0.02	0.10	0.03	-4.05	-1.82	-0.05	-0.69	-1.65	-2.01
	Median	-0.08	-0.15	-0.06	-7.29	-2.63	-1.30	-1.42	-2.45	-2.79	-0.03	0.11	0.02	-3.79	-1.85	-0.09	-0.56	-1.58	-1.92
	Rejected	0.7%	7.3%	1.3%	98.0%	35.3%	10.0%	40.3%	66.7%	57.7%	0.0%	0.7%	2.7%	57.3%	15.3%	1.3%	11.7%	24.0%	25.3%
(D) Financial variables																			
AR(2)	Mean	-0.11	0.11	-0.05	-2.58	-2.01	-1.12	-0.77	-1.01	-1.36	0.57	-0.85	-1.25	-1.17	-1.76	-2.32	-0.90	-1.02	-0.96
	Median	-0.70	0.08	-0.56	-2.58	-1.74	-1.24	-0.74	-1.04	-1.50	0.07	0.07	-0.45	-1.23	-1.30	-0.90	-0.94	-0.99	-0.80
	Rejected	14.0%	20.0%	3.3%	57.3%	43.3%	2.7%	17.3%	16.7%	23.0%	5.3%	22.7%	16.7%	28.7%	38.0%	32.0%	9.3%	7.7%	5.3%
AR(4)	Mean	0.36	0.03	-0.10	-3.77	-1.89	-1.13	-0.71	-1.30	-1.67	0.49	0.05	-1.03	-2.89	-1.61	-2.21	-0.85	-1.39	-1.30
	Median	-0.04	0.01	-0.66	-3.75	-1.80	-1.31	-0.77	-1.30	-1.58	0.25	0.00	-0.37	-2.88	-1.70	-1.15	-1.07	-1.64	-1.52
	Rejected	20.0%	19.3%	4.0%	70.0%	40.7%	4.7%	16.7%	23.3%	21.7%	10.7%	3.3%	15.3%	49.3%	22.0%	20.7%	14.3%	14.0%	10.7%
AR(12)	Mean	0.22	0.09	-0.43	-7.07	-4.64	-3.03	-0.58	-1.68	-2.20	0.48	0.26	-0.12	-5.51	-2.94	-2.21	-0.66	-1.71	-1.89
	Median	-0.70	-0.49	-0.87	-6.40	-4.71	-2.37	-0.77	-1.82	-2.38	0.05	0.00	-0.07	-5.61	-2.41	-2.12	-0.80	-1.61	-1.92
	Rejected	19.3%	18.0%	11.3%	94.7%	63.3%	30.7%	16.0%	38.0%	41.7%	0.0%	0.0%	4.0%	78.7%	42.7%	30.0%	10.3%	17.7%	16.7%

Notes: See notes to Table 2 for definitions of the data and samples used. See Section 5.1 for definitions of the three tests performed. The non-percent entries correspond to the mean or median of the distribution of the t-statistics associated with the given test, lag-selection method, horizon, grouping, data, and sample. The entries reported as percentages correspond to the fraction of models rejected at the 5 percent level for the given test, lag-selection method, horizon, grouping, data, and sample, with critical values determined using 299 bootstrap replications from a residual-based moving block bootstrap with a block length of 40.

Figure 1: Improvements in BIC between Samples 1959:01-1983:12 and 1984:01-2008:12 when $p = 4$



Notes: Each panel above plots the percent improvements in the BIC from sample period 1959:01-1983:12 to sample period 1984:01-2008:12 for the DMS models against those for the IMS models for the given series grouping and horizon and when the number of lags of all variables included in each model is 4. Percent improvement in the BIC is calculated as $(x_1 - x_2)/|x_1|$ where x_1 is the BIC from the first sample and x_2 is the BIC from the second. We divide the points into six regions as defined in the top-left panel, and we report the percent of points that fall into each region. The DMS values are capped at ± 40 percent to improve the clarity of the figures, which is why many points are observed at those limits. The two sample periods were chosen so that they both have the same number of observations and are separated by the start of the Great Moderation at the beginning of 1984. We require all series to have the same number of observations in both samples, so the series TWEXMMTH is excluded from these exercises as it is missing observations prior to 1973:01. This leaves us with 120 trivariate systems instead of the 150 used in the forecasting exercises.

Appendix

Below is the set of 121 series, organized by group, that we use in the bivariate exercises. The column "Series" contains the series identifier in FRED-MD. The column "Trans." denotes one of the following data transformations for a series x: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$.

Table A1: Series Used in the Bivariate Exercises

Group 1: Income, Output, Sales, and Capacity Utilization

	Series	Trans.	Sample Period	Description
1	RPI	5	1959:01-2016:12	Real personal income
2	W875RX1	5	1959:01-2016:12	Real personal incme ex transfer receipts
3	INDPRO	5	1959:01-2016:12	IP: total
4	IPFPNSS	5	1959:01-2016:12	IP: final products and nonindustrial supplies
5	IPFINAL	5	1959:01-2016:12	IP: final products (market group)
6	IPCONGD	5	1959:01-2016:12	IP: consumer goods
7	IPDCONGD	5	1959:01-2016:12	IP: durable consumer goods
8	IPNCONGD	5	1959:01-2016:12	IP: nondurable consumer goods
9	IPBUSEQ	5	1959:01-2016:12	IP: business equipment
10	IPMAT	5	1959:01-2016:12	IP: materials
11	IPDMAT	5	1959:01-2016:12	IP: durable materials
12	IPNMAT	5	1959:01-2016:12	IP: nondurable materials
13	IPMANSICS	5	1959:01-2016:12	IP: manufacturing (SIC)
14	IPB51222S	5	1959:01-2016:12	IP: residential utilities
15	IPFUELS	5	1959:01-2016:12	IP: fuels
16	CUMFNS	2	1959:01-2016:12	Capacity utilization: manufacturing
17	DPCERA3M086SBEA	5	1959:01-2016:12	Real personal consumption expenditures
18	CMRMTSPLx	5	1959:01-2016:12	Real mfg. and trade industries sales
19	RETAILx	5	1959:01-2016:12	Retail and food services sales

Group 2: Employment and Unemployment

	Series	Trans.	Sample Period	Description
1	CLF16OV	5	1959:01-2016:12	Civilian labor force
2	CE16OV	5	1959:01-2016:12	Civilian employment
3	UNRATE	2	1959:01-2016:12	Civilian unemployment rate
4	UEMPMEAN	2	1959:01-2016:12	Avg. duration of unemployment (weeks)
5	UEMPLT5	5	1959:01-2016:12	Civilians unemployed - less than 5 weeks
6	UEMP5TO14	5	1959:01-2016:12	Civilians unemployed - 5-14 weeks
7	UEMP15OV	5	1959:01-2016:12	Civilians unemployed - 15 weeks and over
8	UEMP15T26	5	1959:01-2016:12	Civilians unemployed - 15-26 weeks
9	UEMP27OV	5	1959:01-2016:12	Civilians unemployed - 17 weeks and over
10	CLAIMSx	5	1959:01-2016:12	Initial claims
11	PAYEMS	5	1959:01-2016:12	All employees: total nonfarm
12	USGOOD	5	1959:01-2016:12	All employees: goods-producing industries
13	CES1021000001	5	1959:01-2016:12	All employees: mining
14	USCONS	5	1959:01-2016:12	All employees: construction
15	MANEMP	5	1959:01-2016:12	All employees: manufacturing
16	DMANEMP	5	1959:01-2016:12	All employees: durable goods
17	NDMANEMP	5	1959:01-2016:12	All employees: nondurable goods
18	SRVPRD	5	1959:01-2016:12	All employees: service-producing industries
19	USTPU	5	1959:01-2016:12	All employees: trade, transp., and utilities
20	USWTRADE	5	1959:01-2016:12	All employees: wholesale trade
21	USTRADE	5	1959:01-2016:12	All employees: retail trade
22	USFIRE	5	1959:01-2016:12	All employees: financial activities
23	USGOVT	5	1959:01-2016:12	All employees: government

24	CES0600000007	1	1959:01-2016:12	Avg. weekly hours: goods-producing
25	AWOTMAN	2	1959:01-2016:12	Avg. weekly overtime hours: manufacturing
26	AWHMAN	1	1959:01-2016:12	Avg. weekly hours: manufacturing

Group 3: Construction, Inventories and Orders

	Series	Trans.	Sample Period	Description
1	HOUST	4	1959:01-2016:12	Housing starts: total new privately owned
2	HOUSTNE	4	1959:01-2016:12	Housing starts: Northeast
3	HOUSTMW	4	1959:01-2016:12	Housing starts: Midwest
4	HOUSTS	4	1959:01-2016:12	Housing starts: South
5	HOUSTW	4	1959:01-2016:12	Housing starts: West
6	PERMIT	4	1960:01-2016:12	New private housing permits (SAAR)
7	PERMITNE	4	1960:01-2016:12	New private housing permits: Northeast (SAAR)
8	PERMITMW	4	1960:01-2016:12	New private housing permits: Midwest (SAAR)
9	PERMITS	4	1960:01-2016:12	New private housing permits: South (SAAR)
10	PERMITW	4	1960:01-2016:12	New private housing permits: West (SAAR)
11	AMDMNOx	5	1959:01-2016:12	New orders for durable goods
12	AMDMUOx	5	1959:01-2016:12	Unfilled orders for durable goods
13	BUSINVx	5	1959:01-2016:12	Total business inventories
14	ISRATIOx	2	1959:01-2016:12	Total business inventories to sales ratio

Group 4: Interest Rates and Asset Prices

	Series	Trans.	Sample Period	Description
1	FEDFUNDS	2	1959:01-2016:12	Effective federal funds rate
2	CP3Mx	2	1959:01-2016:12	3-month AA financial commercial paper rate
3	TB3MS	2	1959:01-2016:12	3-month treasury bill
4	TB6MS	2	1959:01-2016:12	6-month treasury bill
5	GS1	2	1959:01-2016:12	1-year treasury yield
6	GS5	2	1959:01-2016:12	5-year treasury yield
7	GS10	2	1959:01-2016:12	10-year treasury yield
8	AAA	2	1959:01-2016:12	Moody's seasoned AAA corporate bond yield
9	BAA	2	1959:01-2016:12	Moody's seasoned BAA corporate bond yield
10	COMPAPFFx	1	1959:01-2016:12	3-month commercial paper minus fed funds
11	TB3SMFFM	1	1959:01-2016:12	3-month treasury minus fed funds
12	TB6SMFFM	1	1959:01-2016:12	6-month treasury minus fed funds
13	T1YFFM	1	1959:01-2016:12	1-year treasury minus fed funds
14	T5YFFM	1	1959:01-2016:12	5-year treasury minus fed funds
15	T10YFFM	1	1959:01-2016:12	10-year treasury minus fed funds
16	AAAFFM	1	1959:01-2016:12	Moody's AAA corporate minus fed funds
17	BAAFFM	1	1959:01-2016:12	Moody's BAA corporate minus fed funds
18	EXSZUSx	5	1959:01-2016:12	Switzerland/U.S. foreign exchange rate
19	EXJPUSx	5	1959:01-2016:12	Japan/U.S. foreign exchange rate
20	EXUSUKx	5	1959:01-2016:12	U.S./U.K. foreign exchange rate
21	EXCAUSx	5	1959:01-2016:12	Canada/U.S. foreign exchange rate
22	S&P 500	5	1959:01-2016:12	S&P's common stock price index: composite
23	S&P: indust	5	1959:01-2016:12	S&P's common stock price index: industrials
24	S&P div yield	2	1959:01-2016:12	S&P's composite common stock: dividend yield
25	S&P PE ratio	5	1959:01-2016:12	S&P's composite common stock: price-earnings ratio
26	VXOCLSx	1	1962:07-2016:12	CBOE S&P 100 volatility index: VXO

Group 5: Nominal Prices, Wages, and Money

	Series	Trans.	Sample Period	Description
1	M1SL	6	1959:01-2016:12	M1 money stock
2	M2SL	6	1959:01-2016:12	M2 money stock
3	M2REAL	5	1959:01-2016:12	Real M2 money stock
4	AMBSL	6	1959:01-2016:12	St. Louis adjusted monetary base
5	TOTRESNS	6	1959:01-2016:12	Total reserves of depository institutions
6	BUSLOANS	6	1959:01-2016:12	Commercial and industrial loans

7	REALLN	6	1959:01-2016:12	Real estate loans at all commercial banks
8	NONREVSL	6	1959:01-2016:12	Total nonrevolving credit
9	CONSPI	2	1959:01-2016:12	Ratio of nonrevolving credit to personal income
10	MZMSL	6	1959:01-2016:12	MZM money stock
11	DTCOLNVHFNM	6	1959:01-2016:12	Consumer motor vehicle loans outstanding
12	DTCTHFNM	6	1959:01-2016:12	Total consumer loans and leases outstanding
13	INVEST	6	1959:01-2016:12	Securities in bank credit at all commercial banks
14	WPSFD49207	6	1959:01-2016:12	PPI: finished goods
15	WPSFD49502	6	1959:01-2016:12	PPI: finished consumer goods
16	WPSID61	6	1959:01-2016:12	PPI: intermediate materials
17	WPSID62	6	1959:01-2016:12	PPI: crude materials
18	OILPRICEx	6	1959:01-2016:12	Crude oil, spliced WTI and Cushing
19	PPICMM	6	1959:01-2016:12	PPI: metals and metal products
20	CPIAUCSL	6	1959:01-2016:12	CPI: all items
21	CPIAPPSL	6	1959:01-2016:12	CPI: apparel
22	CPITRNSL	6	1959:01-2016:12	CPI: transportation
23	CPIMEDSL	6	1959:01-2016:12	CPI: medical care
24	CUSR0000SAC	6	1959:01-2016:12	CPI: commodities
25	CUSR0000SAD	6	1959:01-2016:12	CPI: durables
26	CUSR0000SAS	6	1959:01-2016:12	CPI: services
27	CPIULFSL	6	1959:01-2016:12	CPI: all items less food
28	CUSR0000SA0L2	6	1959:01-2016:12	CPI: all items less shelter
29	CUSR0000SA0L5	6	1959:01-2016:12	CPI: all items less medical care
30	PCEPI	6	1959:01-2016:12	PCE: chain-type price index
31	DDURRG3M086SBEA	6	1959:01-2016:12	PCE: durable goods
32	DNDGRG3M086SBEA	6	1959:01-2016:12	PCE: nondurable goods
33	DSERRG3M086SBEA	6	1959:01-2016:12	PCE: services
34	CES0600000008	6	1959:01-2016:12	Avg. hourly hearnings: goods-producing
35	CES2000000008	6	1959:01-2016:12	Avg. hourly earnings: construction
36	CES3000000008	6	1959:01-2016:12	Avg. hourly earnings: manufacturing

Below is the set of 16 series, organized by group, that we use in the trivariate exercises. The column "Series" contains the series identifier in FRED-MD. The column "Trans." denotes one of the following data transformations for a series x: (1) no transformation; (2) Δx_t; (3) $\Delta^2 x_t$; (4) $log(x_t)$; (5) $\Delta \log(x_t)$; (6) $\Delta^2 \log(x_t)$. Note that all but one variable (TWEXMMTH) are in the set of series that we use for the bivariate exercises shown above.

Table A2: Series Used in the Trivariate Exercises

	Series	Trans.	Sample Period	Description
	Group 1: Real Variables			
1	RPI	5	1959:01-2016:12	Real personal income
2	INDPRO	5	1959:01-2016:12	IP: total
3	CE16OV	5	1959:01-2016:12	Civilian employment
4	UNRATE	2	1959:01-2016:12	Civilian unemployment rate
5	AWHMAN	1	1959:01-2016:12	Avg. weekly hours: manufacturing
6	DPCERA3M086SBEA	5	1959:01-2016:12	Real personal consumption expenditures
	Group 2: Nominal Variables			
1	CES3000000008	6	1959:01-2016:12	Avg. hourly earnings: manufacturing
2	WPSFD49207	6	1959:01-2016:12	PPI: finished goods
3	OILPRICEx	6	1959:01-2016:12	Crude oil, spliced WTI and Cushing
4	CPIAUCSL	6	1959:01-2016:12	CPI: all items
5	PCEPI	6	1959:01-2016:12	PCE: chain-type price index
	Group 3: Financial Variables			
1	M1SL	6	1959:01-2016:12	M1 money stock
2	FEDFUNDS	2	1959:01-2016:12	Effective federal funds rate
3	GS10	2	1959:01-2016:12	10-year treasury yield
4	TWEXMMTH	5	1973:01-2016:12	Trade-weighted U.S. dollar index: major currencies
5	S&P 500	5	1959:01-2016:12	S&P's common stock price index: composite