



Conference on

# Forecasting and Monetary Policy

Berlin, 23-24 March 2009

**Valentina Corradi**

University of Warwick

**Andres Fernandez**

Rutgers University and Universidad de Los Andes

**Norman R. Swanson**

Rutgers University and the Federal Reserve Bank of Philadelphia

**„Real-Time Datasets Really Do Make a Difference: Definitional Change, Data Release and Forecasting“**

# Real-Time Datasets Really Do Make a Difference: Definitional Change, Data Release, and Forecasting\*

Valentina Corradi<sup>1</sup>, Andres Fernandez<sup>2</sup>, and Norman R. Swanson<sup>3</sup>

<sup>1</sup>University of Warwick, <sup>2,3</sup>Rutgers University, <sup>2</sup>Universidad de Los Andes, and the <sup>3</sup>Federal Reserve Bank of Philadelphia

first draft - January 2009

## Abstract

In this paper we ask the following three questions. First, does the usefulness of a real-time dataset depend upon which release of data we are interested in forecasting; and are real-time datasets useful, in general? Second, what are the trade-offs associated with: early release inefficiency; early release measurement error; the impact that definitional change has on time series consisting of only early releases; and the calibration of prediction models using mixed releases of data, as occurs when one uses the latest vintage of data in recursive real-time forecasting experiments? Third, what is the informational content of the revision process for the variables that we examine (i.e. for output, money, and prices), and when does the revision process increase the predictability of a given release of our variables? In order to answer these questions, we construct a variety of different real-time prediction models, and evaluate their performance in a series of ex-ante prediction experiments that are designed to mimic forecasting approaches that are used when constructing forecasts in real-time for the purpose of policy setting and other sorts of real-time decision making. We also propose a new early release data efficiency test based on our prediction experiments. The test, when used in conjunction with extant rationality tests in the literature, sheds important new light on the issues raised in the three above questions. Finally, we carry out an empirical illustration of our methodology in which we examine the real-time predictive content of money for income, building on the work of Garratt, Koop, Mise and Vahey (2009). Our findings point to the importance of making real-time datasets available to forecasters. Moreover, we uncover a new evidence of the efficiency of early release money data, the inefficiency of price and output data, and the importance of the choice between using "first available" or "latest available" releases of data in prediction construction. Interestingly, for prices, first available data are mean square forecast error (MSFE) "best" for predicting *any* release of prices, while models estimated using "latest release" data are MSFE-best when predicting *any* release of money. This finding is explained in some part by definitional change problems associated with money, and by the usefulness of the revision process when predicting output. In our empirical illustration, we find little marginal predictive content in money; but we note that vector autoregressions with money do not perform significantly worse than autoregressions, when predicting output in the last 20 years. This in turn suggests that money may actually be a useful control variable in policy applications, at least in recent years.

*Keywords:* bias; efficiency; generically comprehensive tests; rationality; preliminary, final, and real-time data.

*JEL classification:* C32, C53, E01, E37, E47.

---

\* Valentina Corradi, Department of Economics, University of Warwick, Coventry, CV4 7AL, UK, v.corradi@warwick.ac.uk. Andres Fernandez, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, afernandez@fas-econ.rutgers.edu; and Department of Economics, Universidad de Los Andes, Cra. 1 No. 18A-10 Edificio C. Bogota, Colombia. Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. Corradi gratefully acknowledges ESRC grant RES-062-23-0311, and Swanson acknowledges financial support from a Rutgers University Research Council grant.

# 1 Introduction

In an important paper, Aruoba (2006) finds that for most US macroeconomic series, revision errors have a positive bias, and are highly predictable using information available at the time of the first release. This is an interesting finding, as it suggests that using multiple vintages of data in prediction models may yield improved predictions, in real-time; leading in turn to improved real-time policy setting behavior in governmental organizations as well as improved real-time private decision making behavior for individuals. One way to account for the predictive content of revision errors noted by Aruoba (2006) is to estimate prediction models that employ all vintages (releases) of available variables in a model estimated using the Kalman filter or some other filtering procedure (see e.g. Mariano and Tanazaki (1995)). However, such approaches do not explicitly account for definitional changes that characterize many U.S. series, and that can “corrupt” early releases of real-time datasets. Moreover, the literature has little to say concerning which release of data to predict; and whether or not it is preferable to use mixed releases of data when estimating models, such as when “latest available” data are used, or to use only early releases, or releases pertaining to the release of data being predicted. In this paper, we attempt to shed new light on these and other related issues. Precedents to the research carried out in this paper include: Diebold and Rudebusch (1991), Hamilton and Perez-Quiros (1991), Robertson and Tallman (1998), Gallo and Marcellino (1999), Swanson, Ghysels, and Callan (1999), McConnell and Perez-Quiros (2000), Amato and Swanson (2001), Croushore and Stark (2001,2003), Ghysels, Swanson and Callan (2002), Bernanke and Boivin (2003), Croushore (2006), and Swanson and van Dijk (2006), and the papers cited therein.

To be more specific, in this paper we ask the following three questions. First, does the usefulness of a real-time dataset depend upon which release of data we are interested in forecasting; and are real-time datasets useful, in general? Second, what are the trade-offs associated with: early release inefficiency; early release measurement error; the impact that definitional change has on time series consisting of only early releases; and the calibration of prediction models using mixed releases of data, as occurs when one uses the latest vintage of data in recursive real-time forecasting experiments? Third, what is the informational content of the revision process for the variables that we examine (i.e. for output, money, and prices), and when does the revision process increase the predictability of a given release of our variables?

In order to answer these questions, we construct a variety of different real-time prediction models, and evaluate their performance in a series of *ex – ante* prediction experiments that are

designed to mimic forecasting approaches that are used when constructing forecasts in real-time for the purpose of policy setting and other sorts of real-time decision making. The prediction models include one that uses only first release data and one that uses the latest available data (i.e. uses a mixture of different releases), among others. We also propose a new early release data efficiency test based on our prediction experiments. The test, when used in conjunction with extant linear and nonlinear rationality tests in the literature, sheds important new light on the issues raised in the two above questions. In particular, our test is a Diebold and Mariano (1995) type test that is constructed by comparing mean square forecast errors (*MSFEs*) associated with sequences of *ex-ante* predictions, where each sequence of predictions is for a different release of the same calendar dated observations. The availability of real-time datasets for the variables that we examine, namely output, prices and money make the construction of our test feasible. We view this test as an important additional to the current arsenal of tests that are available for testing early release rationality (see e.g. Corradi, Fernandez and Swanson (2009) for a discussion of these tests).

Our findings point clearly to the importance of making real-time datasets available to forecasters. Moreover, we uncover new evidence concerning the efficiency (and lack thereof) of early release data, and concerning whether or not one should use "first available" or "latest available" releases of data in prediction construction. Interestingly, for prices, first available data are "*MSFE*-best" for predicting *any* release (supporting a conclusion that prices are efficient), while models estimated using "latest release" data are *MSFE*-best when predicting *any* release of money (supporting a conclusion of inefficiency). Interestingly, one of the main reasons for the latter finding turns out not to be a lack of efficiency, but rather a problem associated with the presence of definitional changes to the data that result in the "contamination" of time series constructed using only early releases. In effect, if definitional changes are too substantial, then we argue that use of early release data is tantamount to forming time series by concatenating data from various "definitionally" different versions of a variable. These definitional change problems do not appear to be an issue for prices and output, however.

Our finding concerning the usefulness of early release data for predicting prices (and hence our finding that price data are efficient), is mitigated by the fact that our price variable is the only one for which we find clear and systematic evidence pointing to the importance of the revision process for *ex-ante* prediction. This points to a trade-off between using inefficient early release data versus using mixtures of different releases of data when forming prediction models, and is taken

as evidence that prices are actually “mildly inefficient”, in agreement with earlier findings in the literature, such as those discussed in Corradi, Fernandez and Swanson (2009).

One of the more thought provoking findings of our prediction experiments is that prediction models estimated and implemented using *only* preliminary data are *MSFE*-best for prices, while use of “latest available” data leads to *MSFE*-best prediction in the case of money, and these findings are robust, *regardless of the release of data that we are forecasting*. Thus, any debate concerning which release of data to predict and which model might be most useful for predicting that release appears to be vacuous. One can simply predict the release of choice, given one’s knowledge concerning the rationality of the variable - the model to use in our experimental setup remains fixed.

In order to illustrate the implementation of our methodology, we carry out an empirical illustration in which we examine the real-time predictive content of money for income, building on the work of Amato and Swanson (2001), Garratt, Koop, Mise and Vahey (2009), and others. While we find little marginal predictive content in money, we note that vector autoregressions with money do not perform significantly worse than autoregressions, when predicting output in the last 20 years. This in turn suggests that money may actually be a useful control variable in policy applications, at least in recent years.

The rest of the paper is organized as follows. In Section 2, we outline some notation, and in Section 3 we discuss the empirical methodology used in the remainder of the paper. In particular, we outline the variety of real-time prediction models that we analyze, and we discuss our new test of early release variable efficiency. Section 4 contains a discussion of the data used, and our empirical findings. In Section 5, we summarize the results from our empirical illustration, and concluding remarks are given in Section 6. Tables, figures, and appendices are collected at the end of the paper.

## 2 Setup

Let  ${}_{t+k}X_t$  denote a variable (reported as an annualized growth rate) for which real-time data are available, where the subscript  $t$  denotes the time period to which the datum pertains, and the subscript  $t+k$  denotes the time period during which the datum becomes available. In this setup, if we assume a one month reporting lag, then first release or “preliminary” data are denoted by  ${}_{t+1}X_t$ . In addition, we denote fully revised or “final” data, which is obtained as  $k \rightarrow \infty$ , by  ${}_fX_t$ . Finally, data are grouped into so-called *vintages*, where the first vintage is preliminary data, the

second release is  $2^{nd}$  available data, and so on.

To further set notation, let  ${}_{t+2}u_t^{t+1} = {}_{t+2}X_t - {}_{t+1}X_t$ , and  ${}_{t+1}W_t = ({}_{t+1}X_t, {}_{t+1}u_{t-1}^t)$ , where  ${}_{t+1}u_{t-1}^t = {}_{t+1}X_{t-1} - {}_tX_{t-1}$ . Thus,  ${}_{t+2}u_t^{t+1}$  and  ${}_{t+1}u_{t-1}^t$  denote the errors between the  $2^{nd}$  and the  $1^{st}$  releases at time  $t+2$  and at time  $t+1$ , respectively. Furthermore, let  $\mathcal{F}_t^{t+1} = \sigma({}_{s+1}W_s; 1 \leq s \leq t)$ , so that  $\mathcal{F}_t^{t+1}$  contains information available at the time of the first release, assuming a one month lag before the first datum becomes available.

### 3 Empirical Methodology

#### 3.1 Prediction

In this sub-section, we discuss the prediction models that will be used for addressing the questions outlined above. In particular, we consider the issue of prediction using various variable/vintage combinations as defined in the following set of models.

*Model A (First Available Data):*  ${}_{t+k}X_{t+1} = \alpha_{t+1,t}^A + \sum_{i=1}^{p^A} \beta_{i,t+1,t}^A {}_{t+2-i}X_{t+1-i} + {}_{t+k}\epsilon_{t+1}^A$ ;

*Model B ( $k^{th}$  Available Data)* :  ${}_{t+k}X_{t+1} = \alpha_{t+1,t}^B + \sum_{i=1}^{p^B} \beta_{i,t+1,t}^B {}_{t+2-i}X_{t+3-k-i} + {}_{t+k}\epsilon_{t+1}^B$ ;

*Model C (Real-Time Data)* :  ${}_{t+k}X_{t+1} = \alpha_{t+1,t}^C + \sum_{i=1}^{p^C} \beta_{i,t+1,t}^C {}_{t+1}X_{t+1-i} + {}_{t+k}\epsilon_{t+1}^C$ .

Our analysis is carried out by recursively estimating the above models, and subsequently constructing sequences of *ex-ante* 1-step ahead predictions, for various values of  $k$ . Notice that when  $k = 2$ , we are assuming that the “target variable” to be forecast is the first release. On the other hand, when  $k = 24$ , it can be assumed that the target variable is effectively the final release (see Faust and Wright (2009) and Swanson and van Dijk (2006) for discussions of what constitutes final release data).

Model A has explanatory variables that are formed using only  $1^{st}$  available data. Thus, the first model corresponds to the approach of simply using first available data and ignoring all later releases, regardless of which release of data is being forecasted. As discussed above, this model should be expected to perform well if data revisions are “news”, and if definitional changes do not contaminate the data. If definitional changes affect our data, then one might expect that Model A would perform well for very small values of  $k$ , but would perform increasingly worse as  $k$  increases.

Model B is specified using explanatory variables that are available  $k - 1$  months ahead, corresponding to  $(k - 1)^{st}$  available data. Thus, this model uses data that have been revised  $k - 1$  times in order to predict data that likewise have been revised  $k - 1$  times. In this sense, Model B is included only as a “reality check”, as the model uses stale information in all instances other than the case where  $k = 2$  (in which case Models A and B are equivalent). Put differently, one might imagine

that Model B, where only  $k^{th}$  release data are used as explanatory variables in models where the target variable is the  $k^{th}$  release, is a reasonable model. However, when  $k > 2$ , the calendar date of information used to predict one-step ahead is at least two periods prior to the prediction period. Hence, the “cost” of using Model B, is the inclusion of “stale” data, and hence the model should be expected to perform poorly.

In Model C, the latest release of each observation is used in prediction, so that the dataset is fully updated prior to each new prediction being made. We refer to this model as our “real-time” model, as policy makers and others who construct new predictions each period, after updating their datasets and re-estimating their models, generally use this type of model. If useful information accrues via the revision process, then one might expect that using real-time data (Model C) would yield a better predictor of  ${}_{t+k}X_t$  than when only “stale”  $1^{st}$  release data are used (Model A), for example. Of course, the last statement has a caveat. Namely, it is possible that  $1^{st}$  release data are best predicted using only  $1^{st}$  release regressors,  $2^{nd}$  release using  $2^{nd}$  release regressors, etc. This might arise if the use of real-time data as in Model C results in an “informational mix-up”, due to the fact that every observation used to estimate the model is a different release, and only one of these releases corresponds to the release being predicted, at any point in time (see discussion in the introduction for further details). However, Model C is not subject to the sort of definitional or structural change problems that may plague Models A and B. Because of this fact, it is of interest to assess whether the “cost” of definitional change problems is greater than the “cost” of using mixed release datasets, when predicting a single release,  $k$ . We are thus interested in finding out whether Model C “wins” or not because it will tell us whether using latest release data that avoids definitional change issues offsets the “mixture of releases” issues associated with the use of “real-time” models. For example, if  $1^{st}$  release data are “best” predicted using our “real-time” model, then one should conclude that definitional changes are important, and that real-time datasets that include many releases of data, such as that used by Mariano and Tanazaki (1995) have definitional change problems that have not been addressed in the literature, thus far.

It is crucial to note that the models given above cannot be directly implemented in practice in order to construct real-time predictions. The reason for this can be illustrated easily by examining Model C. Say that the intercept is zero,  $p = 1$  and  $k = 3$ . Then, the natural least squares estimator is:

$$\widehat{\beta}_{1,t+k-1,t} = \frac{\sum_{j=2}^t {}_{t+k-1}X_j {}_tX_{j-1}}{\sum_{j=2}^t {}_tX_{j-1}^2}.$$

Here, predictions would then be calculated using  ${}_{t+k}\widehat{X}_{t+1}^f = \widehat{\beta}_{t+k-1,t} {}_{t+1}X_t$ . The problem with

this formulation is that while  $\widehat{\beta}_{t+k-1,t}$  uses data that is calendar dated  $t$  or earlier, the *vintage* of data used in the estimator when  $j = t$  is  $t + 2$ , which is available only in calendar period  $t + 2$ . On one hand, this approach is sensible, given that we are interested in predicting the second release of data at calendar date  $t + 1$ . On the other hand, this approach is not sensible, because we would be using future information to predict the past (i.e. the future information appears in  $\widehat{\beta}_{t+k-1,t}$ ). In order to carry out true real-time prediction, we must assume that we have observations for calendar dates only up until period  $t$ , using only vintages  $t + 1$  and earlier. For this reason, implementation of the models in practice requires that the models be estimated only for  $k = 2$ , and that these estimators be used for predictions for all data releases. Alternatively, predictors could be calculated using calendar dated observations ending  $k - 2$  periods prior to the most recently available calendar observation. This latter approach is clearly sub-optimal (except for the case where  $k = 2$ ), given the importance of the first order autoregressive coefficients in our models, and hence is not implemented. The key insight necessary in order to understand the above argument is that real-time prediction involves using *all* information available up until the period just prior to the prediction period, when forming 1-step ahead forecasts. This in turn implies that one *must* use first release data, at least for the very latest calendar dated observation. If one constructs estimators in the spirit of  $\widehat{\beta}_{1,t+k-1,t}$ , then in order to ensure that the prediction is real-time, one must use data only up until  $k - 2$  periods prior to the most recently available calendar observation. As should be expected, the experiments that we have carried out suggests that this approach leads to poorer predictions.

Given the above considerations, one natural approach is to compare the following prediction equations (least squares estimators for the intercept and first order slope parameter in each model are also given):

*Model A Type Predictions:*  ${}_{t+k}\widehat{X}_{t+1}^f = \widehat{\alpha}_{t+1,t}^A + \sum_{i=1}^{p^A} \widehat{\beta}_{i,t+1,t}^A {}_{t+2-i}X_{t+1-i}$ , for  $t = R, \dots, T - k$ ,  
where

$$\begin{bmatrix} \widehat{\alpha}_{t+1,t}^A \\ \widehat{\beta}_{1,t+1,t}^A \end{bmatrix} = \begin{bmatrix} t-1 & \sum_{j=2}^t {}_jX_{j-1} \\ \sum_{j=2}^t {}_jX_{j-1} & \sum_{j=2}^t {}_jX_{j-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=2}^t {}_{j+1}X_j \\ \sum_{j=2}^t {}_{j+1}X_j {}_jX_{j-1} \end{bmatrix}$$

*Model B Type Predictions* :  ${}_{t+k}\widehat{X}_{t+1}^f = \widehat{\alpha}_{t+1,t}^B + \sum_{i=1}^{p^B} \widehat{\beta}_{i,t}^B {}_{t+k-i}X_{t+1-i}$ , for  $t = R, \dots, T - k$ ,  
where

$$\begin{bmatrix} \widehat{\alpha}_{t+1,t}^B \\ \widehat{\beta}_{1,t+1,t}^B \end{bmatrix} = \begin{bmatrix} t-1 & \sum_{j=2}^t {}_jX_{j+1-k} \\ \sum_{j=2}^t {}_jX_{j+1-k} & \sum_{j=2}^t {}_jX_{j+1-k}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=2}^t {}_{j+1}X_{j+2-k} \\ \sum_{j=2}^t {}_{j+1}X_{j+2-k} {}_jX_{j+1-k} \end{bmatrix}$$

*Model C Type Predictions* :  ${}_{t+k}\widehat{X}_{t+1}^f = \widehat{\alpha}_{t+1,t}^C + \sum_{i=1}^{p^C} \widehat{\beta}_{i,t}^C {}_{t+1-i}X_{t+1-i}$ , for  $t = R, \dots, T - k$ ,

where

$$\begin{bmatrix} \widehat{\alpha}_{t+1,t}^C \\ \widehat{\beta}_{1,t+1,t}^C \end{bmatrix} = \begin{bmatrix} t-1 & \sum_{j=2}^t {}_{t+1}X_{j-1} \\ \sum_{j=2}^t {}_{t+1}X_{j-1} & \sum_{j=2}^t {}_{t+1}X_{j-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=2}^t {}_{t+1}X_j \\ \sum_{j=2}^t {}_{t+1}X_j {}_{t+1}X_{j-1} \end{bmatrix}$$

Notice that the estimators used in the three prediction models are indeed quite different. Moreover, analogous least squares estimators for all other parameters in the prediction equations follow immediately (by simply setting  $k = 2$  in the original models). Given the above formulation, it is also clear that the straw-man random walk with drift prediction model that we also consider in our experiments has intercept parameter that differs across the three models.

In our predictive experiments, we set: (i)  $p = 1$ ; (ii)  $p = SIC$ ; (iii)  $p = AIC$ ; (iv)  $p = 0$  (random walk with drift model).<sup>1</sup> Additionally, we set  $k = \{2, 3, 4, 6, 12, 24\}$ . Furthermore, in addition to the basic regression model, we consider models where we include additional revision error regressors of the form:  ${}_{t+1}W'_t = {}_{t+1}u_{t-1}^t$ ,  ${}_{t+1}W'_t = ({}_{t+1}u_{t-1}^t, {}_{t+1}u_{t-2}^{t-1})$ ,  ${}_{t+1}W'_t = {}_{t+1}u_{t-2}^{t-1}$ , and  ${}_{t+1}W'_t = ({}_{t+1}u_{t-1}^t, {}_{t+1}u_{t-2}^t)$ , where the notation used in these regressors is defined at the end of the previous section.

All experiments are based on the examination of the *MSFEs* associated with 1-step ahead predictions constructed using recursively estimated models, where  $R$  observations are used in our first estimation,  $R + 1$  observations are used in our second estimation, etc. We thus construct sequences of  $P - k$  *ex-ante* predictions and prediction errors, where  $T = R + P$  is the sample size. We set  $R$  to be 1969:4, so that our first prediction is calendar date 1970:1. The start calendar date of our dataset is 1959:4 and we have vintages of data from 1965:4.

In an empirical example, we also estimate multivariate versions of all of the models described above, where we include (i) money, income, prices, and interest rates; and (ii) income, prices, and interest rates. In these models it is assumed that the target variable of interest is output growth. Thus, we are examining, in real-time, the marginal predictive content of money for output, using regressions including various data vintages, various revision errors, and for a target variable that corresponds to various releases of output growth.

*MSFEs* are in turn examined via the use of Diebold and Mariano (DM: 1995) and Clark and McCracken (2001) predictive accuracy test (see also Clark and McCracken (2005), Clark and McCracken (2008)).<sup>2</sup> The test has a null hypothesis of equal predictive accuracy, and is defined as

<sup>1</sup>Models with lags selected using the SIC yielded more accurate predictions, on average, and hence we do not report findings for cases (i) and (iii). Complete results have been tabulated, though, and are available upon request from the authors.

<sup>2</sup>Clark and McCracken (2008) reconsider tests for comparing non-nested as well as nested forecasting models, when forecasts are produced using real time data. They show that, under the news hypothesis, data revisions do not affect the limiting distributions of tests for predictive evaluation. On the other hand, the use of real time data plays a crucial role whenever revisions are noisy, and effects are different depending whether we are comparing non-nested or nested models.

follows:

$$DM = \sqrt{P-k} \frac{\frac{1}{P} \sum_{t=R}^{T-k} \widehat{d}_{t,k}}{\frac{1}{P-k} \sum_{j=-\bar{j}}^{\bar{j}} \sum_{t=R+j}^{T-k} K\left(\frac{j}{M}\right) \left(\widehat{d}_t - \bar{d}\right) \left(\widehat{d}_{t-j} - \bar{d}\right)},$$

where  $\widehat{d}_{t,k} = l(\widehat{\varepsilon}_{1,t,k}) - l(\widehat{\varepsilon}_{2,t,k})$  is a random variable defined to be the difference between the prediction errors of two models that are being compared, when transformed according to a given loss function,  $l$ ,  $\bar{d} = \frac{1}{P-k} \sum_{t=R}^{T-k} \widehat{d}_{t,k}$ , and the denominator is a heteroskedasticity and autocorrelation consistent covariance estimator, such as the Newey-West estimator. The limiting distribution of the DM statistic is given in Theorems 3.1 and 3.2 in Clark and McCracken (2005) under quadratic loss, so that  $\widehat{d}_{t,k} = \widehat{\varepsilon}_{1,t,k}^2 - \widehat{\varepsilon}_{2,t,k}^2$ , and is  $N(0, 1)$  in cases where the prediction models are nonnested and parameter estimation error vanishes (or the in-sample and out-of-sample loss functions are the same - see also Corradi and Swanson (2006)). In the sequel, we consider only quadratic loss, and hence report mean square forecast errors (*MSFEs*) as well as DM test statistics based on quadratic loss.

### 3.2 Efficiency Tests

Perhaps the most frequently implemented rationality tests in the literature are based on the following variety of regression model:

$${}_f X_t = \alpha + {}_{t+1} X_t \beta + {}_{t+1} W_t' \gamma + \epsilon_{t+1}, \quad (1)$$

where  ${}_{t+1} W_t$  is an  $m \times 1$  vector of variables representing the conditioning information set available at time period  $t+1$  and  $\epsilon_{t+1}$  is an error term assumed to be uncorrelated with  ${}_{t+1} X_t$  and  ${}_{t+1} W_t$ . The null hypothesis is  $H_0 : \alpha = 0, \beta = 1$ , and  $\gamma = 0$ , and corresponds to the idea of testing for the rationality of  ${}_{t+1} X_t$  for  ${}_f X_t$ , by finding out whether the conditioning information in  ${}_{t+1} W_t$ , available to the data issuing agency at the time of first release, has been efficiently used (see Mankiw, Runkle, and Shapiro (1984), Mankiw and Shapiro (1986), Kavajecz and Collins (1995), Mork (1987), Keane and Runkle (1990), and Rathjens and Robins (1995) for further details). A summary of this sort of test is given in Swanson and van Dijk (SD: 2006). Additionally, SD consider the entire revision history for each variable, and hence discuss the “timing” of data rationality by generalizing (1) as follows:

$${}_{t+k} X_t - {}_{t+1} X_t = \alpha + {}_{t+1} X_t \beta + {}_{t+1} W_t' \gamma + \epsilon_{t+k}, \quad (2)$$

where  $k = 1, 2, \dots$  defines the release (or vintage) of data. Rather than carrying out regression based tests using the above models, Corradi, Fernandez and Swanson (CFS: 2009) consider moment type

tests that assess whether revision errors can be “explained” using contemporaneous information. The CFS test that we use in this paper is defined as follows:

$$M_T = \sup_{\gamma \in \Gamma} |m_{1,T}(\gamma)|, \quad (3)$$

where, for example:  $m_T(\gamma) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T-2} {}_{t+2}u_t^{t+1} w \left( \sum_{j=0}^{t-1} \gamma'_j \Phi({}_{t+1-j}W_{t-j}) \right)$ ,

$$\Gamma = \left\{ \gamma_j : a_j \leq \gamma_j \leq b_j, j = 1, 2; |a_j|, |b_j| \leq B j^{-\kappa}, \kappa \geq 2 \right\}, \quad (4)$$

${}_{t+1-j}W_{t-j}$  is information available in real-time, and  $\Phi$  is a so-called generically comprehensive function, allowing for the construction of a test that is consistent against generic nonlinear alternatives. In this test, the null hypothesis is  $H_0 : E({}_{t+2}u_t^{t+1} | \mathcal{F}_t^{t+1}) = 0$  *a.s.*, which states that the first release is rational and unbiased, as the revision error in this case is a martingale difference sequence, adapted to the filtration generated by the entire history of past revision errors and past values of the variable to be predicted. This is consistent with the “news” version of rationality, according to which subsequent data revisions only take into account news that was not available at the time of the first release. Thus, if we fail to reject  $H_0$ , it means that the first data release already incorporates all available information at the current time, and hence “improved” predictions cannot be constructed. Should  $H_0$  be rejected, CFS outline another test for assessing whether the rejection is due to non-zero bias or inefficiency. Moreover, and as shown in Lemma 1 of de Jong,  $(\Gamma, \|\gamma - \gamma'\|)$  is a compact metric space, with  $\|\gamma - \gamma'\| = \left( \sum_{j=1}^{\infty} j^\kappa |\gamma_j - \gamma'_j|^2 \right)^{1/2}$ , where  $|\cdot|$  denotes the Euclidean norm. In practice, one can allow for  $\kappa = 2$  and choose  $a_j = a(j+1)^{-2}$  and  $b_j = b(j+1)^{-2}$ , where  $a$  and  $b$  belong to some compact set in  $\mathcal{R}^q$ . It is immediate to see that the weight attached to past observations decreases over time. Indeed, as stated in the assumptions in CFS, the revision error is assumed to be a short memory process, and therefore it is “independent” of its distant past.

In this paper, we consider a new test for the news and noise versions of rationality that take advantage of the particular form of the regressions that are discussed in the previous subsection. In order to motivate the test, consider conditions under which  $\widehat{\beta}_{1,t+1,t}^A = \frac{\sum_{j=2}^t {}_{j+1}X_j {}_jX_{j-1}}{\sum_{j=2}^t {}_jX_{j-1}^2} \equiv \widehat{\beta}_{1,t+1,t}^C = \frac{\sum_{j=2}^t {}_{t+1}X_j {}_{t+1}X_{j-1}}{\sum_{j=2}^t {}_{t+1}X_{j-1}^2}$ .

If there are no issues due to definitional change problems associated with the data and there is no measurement error of the variety where:

$${}_jX_{j-1} = {}_fX_{j-1} + \epsilon_{1,t}$$

and

$${}_{j+1}X_{j-1} = {}_fX_{j-1} + \epsilon_{2,t},$$

say, where  $\sigma_{\epsilon_1}^2 > \sigma_{\epsilon_2}^2$ , then if there is additionally no inefficiency, we must have that  $\widehat{\beta}_{1,t+1,t}^A \equiv \widehat{\beta}_{1,t+1,t}^C$ . In this case, all prediction are identical, regardless of whether Model A or C is used, and then we must have that:

$$MSFE_A = MSFE_C, \quad \forall k = 2, 3, \dots$$

and

$$MSFE_A(i) = MSFE_A(j) \quad \forall i, j = 2, 3, \dots, k$$

and

$$MSFE_C(i) = MSFE_C(j) \quad \forall i, j = 2, 3, \dots, k$$

However, now assume that in fact there is measurement error, which we know *must* be the case if early release data are efficient, otherwise we would observe revision errors that are identically zero in the data, everywhere. This means that we have no need to test the null hypothesis:

$H_0'''$ : *no measurement error and efficiency.*

Even casual observation of the results in Tables 3-5 suggests that there is no need to test this null, as the *MSFEs* in a particular row of entries in these tables vary by up to 45% in value for different values of  $k$ . This means that we can construct a test with the null:

$H_0''$ : *measurement error and efficiency.*

Now, under this null hypothesis, even as  $P, T \rightarrow \infty$ , notice that  $\widehat{\beta}_{1,t+1,t}^A \rightarrow \beta^{A,*}$  and  $\widehat{\beta}_{1,t+1,t}^C \rightarrow \beta^{C,*}$ , where  $\beta^{A,*}$  and  $\beta^{C,*}$  are the probability limits of  $\widehat{\beta}_{1,t+1,t}^A$  and  $\widehat{\beta}_{1,t+1,t}^C$ ; but clearly  $\beta^{A,*} \neq \beta^{C,*}$ . This follows because  $\widehat{\beta}_{1,t+1,t}^A$  is constructed using only first release data and the variance of the measurement error in that first release data does not diminish as the sample size (either in- or out-of-sample) increases, whereas the measurement error of many of the observations used in the construction of  $\widehat{\beta}_{1,t+1,t}^C$  diminishes more and more as the sample increases, as a larger and larger portion of the data in this estimator become data that have been revised many times and hence have very little measurement error left in them (we might assume that after 24 releases the measurement error vanishes, for example).

For this reason, we have a testing problem somewhat analogous to that considered in Clark and McCracken (2009), except that parameter estimation error in the denominator of Diebold-Mariano (1995) type tests does not vanish, regardless of loss function. More specifically, it follows immediately that we can write  $H_0$  and its alternative as:

$H_0'$ :  $E(l(\epsilon_{1,t,k}) - l(\epsilon_{2,t,k})) = 0, \quad \forall k = 2, 3, \dots$

$H_A$ : *the negation of  $H_0'$ .*

Assuming quadratic loss, we shall test the following version of this hypothesis:

$$H_0 : MSFE_A = MSFE_C , \quad \forall k = 2, 3, \dots$$

This hypothesis can be tested using the DM test discussed in the previous subsection. Now, say we reject  $H_0$ . Then, either: (i) There is no need to test further, as we now know we have inefficiency; or (ii) we want to know if we have measurement error in addition to inefficiency; or (iii) problems with the data in the form of definitional changes may in some way be causing rejection of the null. We shall discuss these alternative causes of test rejection further in the next section.

## 4 Empirical Results

### 4.1 Data

Our real-time dataset includes real GDP (seasonally adjusted), the GDP chain-weighted price index (seasonally adjusted), the money stock (measured as M1, seasonally adjusted) and the interest rate (measured as the rate on the 3-month Treasury bill). All series have a quarterly frequency and our real-time dataset for each of the four variables was obtained from the Federal Reserve Bank of Philadelphia’s real-time dataset for Macroeconomists (RTDSM). The RTDSM can be accessed on-line at <http://www.phil.frb.org/econ/forecast/readow.html>. The series were obtained from the “*by-variable*” files of the “*core variables/quarterly observations/quarterly vintages*” dataset, and are discussed in detail in Croushore and Stark (2001). Note also that interest rates are not revised, and hence our interest rate dataset is a vector rather than a matrix (see Ghysels, Swanson, and Callan (2002) for a detailed discussion of the calendar date/vintage structure of real-time datasets).

The first vintage in our sample is 1965:4, for which the first calendar observation is 1959:3. This means that the first observation in our dataset is the observation that was available to researchers in the fourth quarter of 1965, corresponding to calendar dated data for the third quarter of 1953. The datasets range up to the 2006:4 vintage and the corresponding 2006:3 calendar date, allowing us to keep track of the exact data that were available at each vintage for every possible calendar dated observation up to one quarter before the vintage date. This makes it possible to trace the entire series of revisions for each observation across vintages.

Various summary information about the datasets is depicted in the first 6 plots in each of Figures 1-3. For example, note that we use log-differences throughout our analysis (except for interest rates); and various releases of the log-differences of all variables, except the interest rate, are depicted in the plots. Also included are plots of the first and second revision errors measured as the difference between the first vintage (e.g. first available) of a calendar observation and the second

and third vintages, respectively; and cumulative revision errors for various releases. As can readily be seen upon inspection of the distributions of the revision errors, as well as via examination of the summary statistics reported in Table 1, the first revision (i.e. the difference between the first and second vintages) is fairly close to normally distributed. On the other hand, the distribution of the second revision errors is mostly concentrated near zero, implying that much of the revision process has already taken place in the first revision. Indeed, the distributional shape of revision errors beyond the first revision is very much the same as that reported for the second revision in these plots, with the exception of revision errors associated with definitional and other structural changes to the series. This is one of the reasons why much of our analysis focuses only on the impact of first and second revision errors - later revision errors appear to offer little useful information, other than signalling the presence of definitional and related methodological changes. Indeed, an important property of real-time datasets like the RTDSM is the possibility that calendar observations may vary across vintages for reasons other than because of “pure” revisions. This feature of the data is illustrated in plots with title “Calendar data across vintages”, where we have plotted early calendar dates (e.g. 1959:1; 1962:1; and 1965:3) across all available vintages in our sample. Of note is that the data for a particular calendar date sometimes varies significantly across vintages. For instance, looking at the 1959:Q4 calendar observation for output across all vintages, one can observe several discrete movements driving the value of that particular observation from a monthly growth of 1% for the earlier vintages to 0.5% for the later vintages. It seems reasonable to argue that most (if not all) of the discrete variations in that particular calendar observation are not due to “pure revisions”, but are solely a consequence of “definitional breaks” in the measurement of output. To verify this claim, we plotted and compared several other calendar observations across all vintages and we could identify nine clear breaks in the following dates: 1976:Q1; 1981:Q1; 1986:Q1; 1992:Q1, 1996:Q1, 1997:Q2; 1999:Q4; 2000:Q2; and, 2004:Q1. These are the breaks that define the sample periods for which summary statistics are reported in Table 2; and inspection of this table suggests that the basic properties of the series often change after such definitional breaks. This can be graphically illustrated by noticing that for output, the three calendar dates plotted often exhibit abrupt changes in the *same* vintages, corresponding to these dates. Not surprisingly, the same nine breaks were identified in our measure of prices, since our measure is a composite measure of GDP prices. However, it should be noted that the same procedure for the money series does not yield such well defined “definitional breaks”, as some of the breaks do not apply to all vintages.

In the introduction of this paper, we argued that the variable that one cares about predicting

may in some cases be a near-term vintage, and if one cares about *final* data, one may in some cases have a difficult time, as one would need then to predict unknowable future definitional and other methodological changes implicit to the data construction process. Of course, this argument can be checked by examining the performance of our prediction models. However, note that the above discussion suggests that this argument is worthy of investigation. Namely, “pure revision” appears to occur in the near-term, “definitional change” occurs in the long term, and little occurs between. Moreover, and as argued in the introduction, application of simple level shifts to datasets in order to address the “definitional change” issue may be inadequate in some cases, as the entire dynamic structure of a series might have changed after a “definitional change”, so that the current definition of GDP may define an entirely different time series that based on an earlier definition, say. It should be noted that the summary statistics reported in Table 2 suggest that there are indeed significant differences between the means and other measures associated with series in different “definitional change” sub-samples, at least for output. This evidence, thus, points to at least one benefit of using only the latest vintage of data in prediction, such as is done when using Model C. However, we shall see in the next section that Model C does not always yield the best predictions, suggesting that the definitional change problem is not always relevant to the analysis of real-time data.

## 4.2 Prediction Experiments

As discussed in Section 3, we carried out three types of simple autoregressive prediction experiments, where the objective was to forecast output, prices, and money. The methods involved fitting regression Models A, B, and C. Recall that Model B is our “strawman” model, and should be expected to perform increasingly poorly as  $k$  increases. Moreover, Model A involves constructing predictions using only first release data, and hence might be expected to perform poorly for predicting  $k^{th}$  releases, when  $k$  is large, assuming that either (i) data are inefficient, or (ii) definitional changes result in “contamination” of the earliest first release calendar dated observations used in the construction of prediction model parameter estimates. However, Model C uses a mixture of releases both in parameter estimation and in prediction construction, given parameter estimates. In this sense, even if there are efficiency and/or definitional “problems” associated with using first release data, these may be outweighed by the cost of using mixed releases of data, and hence Model A might still be *MSFE*-best.

Turning now to our results, note that Tables 3-5 report on the predictive accuracy of Models A-C using simple autoregressive models, and Table A1.1-A1.3 (see Appendix 1), contain the same summary statistics, except that additional regressors in the form of revision errors are added to the

models (as discussed above). Entries in the tables are *MSFEs* and DM tests statistics that are calculated assuming that Model A is the benchmark against which each other model is compared. In Tables 1-3, entries in bold denote lowest *MSFEs* for a particular value of  $k$ , across all models. A number of clear-cut conclusions emerge when the results reported in these tables are examined.

Upon inspection of Table 3, note that the *MSFE*-best output predictions result when using Model A for low values of  $k$ , and using Model C for high values of  $k$ . However, for price predictions (see Table 4), Model A is always *MSFE*-best, while for money, Model C is always *MSFE*-best, regardless of prediction sub-period (see Panels A, B, and C Table 5). These results are consistent with a number of conclusions.

First, note that if there are no problems with the data due to definitional change, and early releases are efficient, then one might expect that Model A *could* dominate Model C, assuming that the issue of using mixed releases of data in Model C sufficiently reduces predictive accuracy. This is exactly the case that occurs with prices. Of course, early price releases might still be "mildly inefficient", it is just that if definitional changes are not a problem, then the "cost" of using mixed releases of data for prediction model parameter estimation as well as for prediction construction, when the objective is to predict a particular release, may outweigh the "costs" associated with "mild" inefficiency and/or definitional change problems.

Second, the fact that Model C "wins" when predicting money suggests that inefficiency and/or definitional change problems associated with early release data outweigh any potential mixed release problems associated with using Model C. This suggests that an important role for our new efficiency test is to distinguish between efficiency problems and definitional change problems. In particular, if our efficiency test rejects, we know that one reason for rejection may actually be a failure of the maintained assumption of no definitional change. Thus, if the extant efficiency tests from the literature discussed in the previous section all accept the efficiency null, while our test rejects, then we have evidence that definitional change may be driving our test rejections. Regardless, though, we have clear evidence that whether or not one should use the latest or earlier releases of data in prediction is not just dependent upon the release of data to be predicted, but is also upon what the target variable of interest is. When the target variable is money, Model C is preferred (i.e. use the latest data available). However, when the target variable is prices, Model A is preferred. Most importantly, these findings concerning which model is preferred are *independent of the release,  $k$ , being predicted*. This suggests that the "target release" to be predicted is actually not very important, which is quite surprising. The reason why this finding is surprising is that one might

expect the cumulative effect that the combination of inefficiency, measurement error and definitional change has on model choice will result in different models being chosen when the target release to be predicted increases from preliminary to final data. Such does not appear to be the case for prices and money.

Taken together, these first two results give potent evidence that real-time datasets are indeed useful, and should be collected

Third, for some variables, the evidence can be mixed. In particular, for output, Model A “wins” when predicting early releases, while Model C “wins” when predicting later releases. This suggests that there is either sufficiently substantive measurement error or there is sufficient inefficiency that for predicting later releases, prediction models estimated using only first release data are not useful. However, the use of mixed releases of data in estimation of Model C is sufficiently costly that predicting early releases is best done with Model A.

Fourth, as expected, Model B performs poorly, and is particularly bad for larger values of  $k$ . Additionally, the real-time random walk with drift models that we estimate never yield predictions as accurate as those based upon our autoregressive type models.

Fifth, note that  $MSFEs$  associated with the “best” models for money (i.e. Model C) largely decrease as  $k$  increases. This is consistent with the view that using real-time data that has been revised as much as possible, when forming model coefficient estimates, leads to estimates that are more accurate (in the sense that releases used in coefficient construction are more efficient and/or may have less measurement error, and are hence “closer” to their “true” values), when the objective is the prediction of later release or even “final” data. This in turn suggests that later releases of data should be predicted more accurately using Model C (as is indeed the case for money, where Model C is actually the  $MSFE$ -best model), but not necessarily when using other models. Indeed, notice that for prices, where Model A wins, the  $MSFEs$  actually increase as one increases  $k$  from 2 to 3 to 4, before beginning to decrease. The same sort of mixed pattern of increasing and decreasing  $MSFEs$  characterizes output.

Tables A1.1-A1.3 repeat the above experiments, but add revision errors into the mix. If revision errors are useful in *ex – ante* predictions of the sort that we report on in our tables, then we have direct evidence of inefficiency. Consider the case of prices (see Table A1.2). Point  $MSFEs$  reported for Model A in this case are very frequently lower than the comparable  $MSFEs$  reported in Table 4. This is a very interesting result, suggesting that early price releases are actually inefficient, and that the reason Model A “wins” in Table 4 is that the use of mixed release data associated with

the estimation and implementation of Model C is simply "too costly" relative to the predictive accuracy losses associated with using inefficient data. Evidence in support of this finding would be forthcoming were we to find evidence of inefficiency using the extant efficiency tests from the literature discussed above. Even more importantly, we have evidence that there is information in the revision process for prices. This constitutes further evidence of the importance of collecting real-time datasets. However, although point *MSFEs* are lower in virtually every case (when considering Model A) when  $u_{C_1}$  and  $u_{C_2}$  are included, the absolute magnitude of the difference in *MSFEs* is rather small, suggesting that the difference is likely insignificant. Moreover, examination of Tables A1 and A3 suggests that there is little information in the revision processes for the other two variables.

### 4.3 Data Rationality

The approach for early release rationality testing in this paper involves testing for inefficiency via examination of the null hypothesis:

$$H_0 : MSFE_A = MSFE_C , \quad \forall k = 2, 3, \dots$$

using the DM test. Test results are reported in Panel B in Tables 3-5, where DM statistics constructed assuming that Model A is the benchmark against which each other individual model is to be compared, and assuming quadratic loss, are reported. The most important row of entries in these tables is the third row in each sub-panel, where Model A is compared with Model C. Since Model A is the benchmark, negative statistic values indicate that Model A is *MSFE*-best. Of note is the fact that the only statistics comparing Models A and C that are large in absolute value are those associated with money. This suggests that the only series for which our test rejects the null (based on the use of 10% level critical values drawn from the standard normal distribution - see above for further discussion) is money. Now, consider the rationality test results reported for extant tests from the literature in Appendix 2, Table A2.1. Since these tests fail to reject the null of rationality for money, our test is likely rejecting because of definitional change problems associated with using only first release data in Model A. Recall that we can get a feeling for the importance of definitional change by referring to the plot entitled "Calendar data across vintages" in Figure 3.

Interestingly, DM test statistics for prices and output fail to reject the null of efficiency.

Consider first the case of prices, and recall that we noted that there may still be "mild inefficiency" in the data if the costs associated with using Model C (i.e. the costs of using mixed releases of data) are sufficiently great, and if there is little problem with definitional change. The reason

for this is that in such cases, the costs associated with using mixed releases (which is implicit when implementing Model C) can be approximately offset by the costs of using relatively inefficient early releases of data when predicting later releases (as is done with Model A). In such cases, neither model can significantly “beat” the other, based on inference using our test. However, this does not mean that one model may not yield smaller point *MSFEs*, as this indeed happens in the case of prices, for *all* values of  $k$ . Namely, Model A is always *MSFE*-best, as discussed above. Our findings in Table A2.1 confirm that prices are probably “mildly inefficient”, as the null of rationality fails to reject based on the linear rationality test implemented using an F-statistic, but is rejected based upon implementation of the Chao Corradi and Swanson (2001) and the Corradi, Fernandez and Swanson (2009) tests. Moreover, we have additional evidence that prices are indeed “mildly inefficient”. Namely, our results from the experiments where we predict prices with revision errors as well as lagged prices (see Table A1.2) suggest that *MSFEs* are almost always slightly better when revision errors (i.e.  $u_{C_1}$  and  $u_{C_2}$ ) are included in the prediction models.

Now, consider output and note that the evidence for output reported in Table A2.1 is the same as that reported for prices in that table. This evidence of inefficiency is consistent with the fact that Model C outperforms Model A (at least based upon comparison of point *MSFEs*) for all values of  $k$  greater than 3. However, the level of inefficiency is again insufficient to cause a rejection of  $H_0$ , based on implementation of the test using standard significance levels, as in the case of prices.

Taken together, then, we have evidence that definitional changes are an issue in our money series, while efficiency problems are prevalent for our price and output series. Moreover, the way in which these different problems combine in the data determines whether Model A or Model C dominates, when forming predictions for any given release of a variable. Thus, again, we have strong evidence as to the importance and informational content of real-time datasets.

#### **4.4 Real-Time Marginal Predictive Content of Money for Output**

In an empirical illustration, we implemented vector versions of Models A,B, and C to examine whether money and revision errors from money and other variables have marginal predictive content for output. Results are gathered in Table 4, and correspond to those reported in Table 3, except that vector autoregressions are estimated rather than autoregressions, and the target variable to be predicted is output. Note that models with and without money (and revision errors of money) are estimated. The number of lags output, money, prices, and interest rates are selected using the Schwarz information criterion. This illustration is an extension of earlier work of Amato and Swanson (2001), Garratt, Koop, Mise and Vahey (2009), and others.

As Model B is a “strawman” model with little usefulness, we report results only for Models A and C. Thus, the models that we examine are:

**Model A (output equation of the vector autoregressive version):**

$${}_{t+k}Y_{t+1} = \alpha + \sum_{i=1}^p \beta_{i,t+2-i}^{A,Y} Y_{t+1-i} + \theta [{}_{t+1}Y_{t-1} - {}_tY_{t-1}] + \sum_{i=1}^p \beta_{i,t+2-i}^{A,P} P_{t+1-i} + \sum_{i=1}^p \beta_{i,t+2-i}^{A,M} M_{t+1-i} + \sum_{i=1}^p \beta_{i,t+2-i}^{A,R} R_{t+1-i} + \theta'_{t+1} W_t + \varepsilon_{t+k}$$

**Model C: (output equation of the vector autoregressive version):**

$${}_{t+k}Y_{t+1} = \alpha + \sum_{i=1}^p \beta_{i,t+1}^{C,Y} Y_{t+1-i} + \theta [{}_{t+1}Y_{t-1} - {}_tY_{t-1}] + \sum_{i=1}^p \beta_{i,t+1}^{C,P} P_{t+1-i} + \sum_{i=1}^p \beta_{i,t+1}^{C,M} M_{t+1-i} + \sum_{i=1}^p \beta_{i,t+1}^{C,R} R_{t+1-i} + \theta'_{t+1} W_t + \varepsilon_{t+k},$$

where  $W_t = {}_{t+1}u_{t-1}^t$  or  $W_t = ({}_{t+1}u_{t-1,t}^{t'} u_{t-2}^{t-1})'$ , corresponding to the two cases considered (the two cases are denoted  $u_{C_1}$  and  $u_{C_2}$  in Table 6, respectively),  ${}_{t+1}u_{t-1}^t = ({}_{t+1}u_{t-1}^{t,Y}, {}_{t+1}u_{t-1}^{t,P}, {}_{t+1}u_{t-1}^{t,M})'$ , and  ${}_{t+1}u_{t-k}^{t,X} = {}_{t+1}X_{t-k} - {}_tX_{t-k}$ , for  $X = Y, P$ , or  $M$  and  $k = 1, 2$ .

Upon inspection of Table 6, it is clear that it is always the case that (regardless of sample period, model, and vintage) the models with money yield higher *MSFEs* than the models without money (entries that are in bold denote *MSFE*-best models). This result holds for Models A and C, regardless of whether or not revision errors are included. Thus, at least based on the comparison of point *MSFEs*, there is very clear evidence that money does not contain any marginal predictive content for output. This result should be viewed with caution, however, unless the sole purpose of the modeler is to predict output as accurately as possible. In particular, note that autoregressive models yield lower *MSFEs* than their vector autoregressive counterparts for all prediction periods (compare the results reported in Table 3 with those reported in Table 6). Thus, it is likely the case that no variable (other than output itself) has marginal predictive content for output. However, when carrying out policy analysis, for example, one needs to include control variables that the government can manipulate. Simply specifying an AR model has little use. For this reason, a better measure of the usefulness of money and other variables might be whether these variables can be added *without worsening predictive performance*. If such is the case, then one has evidence that increased parameter and model uncertainty associated with including extra explanatory variables in the model is offset by increased predictive performance, hence suggesting that the “bigger” model is adequately specified. If one uses this measure of performance, then, given the lack of starred

entries in Table 6 for values of  $k$  greater than 3, we have evidence as to the “adequacy” of vector autoregression models for later release data. However, note that DM tests carried out on vector autoregressions without money (corresponding to all cases in the table where entries are in bold, since none of the models associated with bold entries include money) still tell us nothing about the “adequacy” of models with money. For this reason, we examine the “adequacy” of our models with money by italicizing the *MSFE*-best models with money for each release, and then placing a star on the entries for which the AR-best model is more accurate, based on implementation of the DM test. Interestingly, for our longer prediction periods beginning in 1971 and 1983, money does not appear adequate, as the AR model yields significantly more accurate predictions. However, for the shortest forecast period from 1990, the *MSFE*-best models with money are adequate for all releases except first release data.

Also interesting is the fact that when considering VAR models, output is always best predicted using varieties of Model A, *regardless of release being predicted*, placing this variable together with prices as being variables where use of preliminary data yields the most accurate predictions. Needless to say, the findings of this illustration suggest that there is much to be learned via analysis of real-time datasets, again underscoring the importance of building and maintaining such datasets.

## 5 Concluding Remarks

Real time data are used in this paper to shed new light on the importance of data rationality and definitional change when the objective is the construction of prediction models for a given release of a variable. Many of the conclusions from our analysis would not have been possible without the availability of real-time datasets, underscoring the importance of collecting and maintaining such datasets. For example, the data rationality test that we propose can only be implemented with real-time data.

Some of our findings include the following. For prices, first available data are “*MSFE*-best” for predicting *any* release (supporting a conclusion that prices are efficient), while models estimated using “latest release” data are *MSFE*-best when predicting *any* release of money (supporting a conclusion of inefficiency). Interestingly, one of the main reasons for the latter finding turns out not to be a lack of efficiency, but rather a problem associated with the presence of definitional changes to the data that result in the “contamination” of time series constructed using only early releases. In effect, if definitional changes are too substantial, then we argue that use of early release data is tantamount to forming time series by concatenating data from various “definitionally” different

versions of a variable. These definitional change problems do not appear to be a issue for prices and output, however.

Our finding concerning the usefulness of early release data for predicting prices (and hence our finding that price data are efficient), is mitigated by the fact that our price variable is the only one for which we find clear and systematic evidence pointing to the importance of the revision process for *ex – ante* prediction. This points to a trade-off between using inefficient early release data versus using mixtures of different releases of data when forming prediction models, and is taken as evidence that prices are actually “mildly inefficient”, in agreement with earlier findings in the literature, such as those discussed in Corradi, Fernandez and Swanson (2009).

One of the more thought provoking findings of our prediction experiments is that prediction models estimated and implemented using *only* preliminary data are *MSFE*-best for prices, while use of “latest available” data leads to *MSFE*-best prediction in the case of money, and these findings are robust, *regardless of the release of data that we are forecasting*. Thus, any debate concerning which release of data to predict and which model might be most useful for predicting that release appears to be vacuous. One can simply predict the release of choice, given one’s knowledge concerning the rationality of the variable - the model to use in our experimental setup remains fixed.

In an empirical illustration, we find little marginal predictive content in money. However, we also note that vector autoregressions with money do not perform significantly worse than autoregressions, when predicting output in the last 20 years. This in turn suggests that money may actually be a useful control variable in policy applications, at least in recent years.

Many problems in this literature remain unsolved. For example, from an empirical perspective it remains to extend the analysis that we carry out to more releases of data, to further examine the important problem of definitional change that is addressed in the paper, and to examine the properties of other real-time datasets. From a theoretical perspective, it remains to examine the properties of various predictive accuracy tests in the recursive and real-time framework employed in the paper.

## 6 References

Amato, J. and N.R. Swanson (2001), The Real-Time Predictive Content of Money for Output, *Journal of Monetary Economics* **48**, 3–24.

Aruoba, S.B. (2006), Data Revisions are not Well-Behaved, *Journal of Money, Credit and Banking*, forthcoming.

Bernanke, B.S. and J. Boivin (2003), Monetary Policy in a Data-Rich Environment, *Journal of Monetary Economics* **50**, 525–546.

Brodsky, N. and P. Newbold (1994), Late Forecasts and Early Revisions of United States GNP, *International Journal of Forecasting* **10**, 455–460.

Burns, A.F. and W.C. Mitchell (1946), *Measuring Business Cycles*, New York: NBER.

Chao, J., V. Corradi and N.R. Swanson (2001), An out-of-sample Test for Granger Causality, *Macroeconomic Dynamics* **5**, 598–620.

Clark, T.E., and M.W. McCracken (2001), Tests of Equal Forecast Accuracy and Encompassing for Nested Models, *Journal of Econometrics* **105**, 85–110.

Clark, T.E., and M.W. McCracken (2005), Evaluating Direct Multistep Forecasts, *Econometric Reviews* **24**, 369–404.

Clark, T. E. and M.W. McCracken (2009), Tests of Equal Predictive Ability with Real-Time Data, *Journal of Business & Economic Statistics*, forthcoming.

Corradi, V. A. Fernandez, and N.R. Swanson (2009), Information in the Revision Process of Real-Time Datasets, *Journal of Business & Economic Statistics*, forthcoming.

Corradi, V. and N.R. Swanson (2006a), Predictive Density Evaluation, in: **Handbook of Economic Forecasting**, eds. Clive W.J. Granger, Graham Elliot and Allan Timmermann, Elsevier, Amsterdam, pp. 197–284.

Croushore, D. (2006), Forecasting with Real-Time Macroeconomic Data, in: **Handbook of Economic Forecasting**, eds. Clive W.J. Granger, Graham Elliot and Allan Timmermann, Elsevier, Amsterdam, pp. 961–982.

Croushore, D. and T. Stark (2001), A Real-Time Dataset for Macroeconomists, *Journal of Econometrics* **105**, 111–130.

Croushore, D. and T. Stark (2003), A Real-Time Dataset for Macroeconomists: Does Data Vintage Matter?, *Review of Economics and Statistics* **85**, 605–617.

Diebold, F. X. and R.S. Mariano (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics* **13**, 253–263.

Diebold, F.X. and G.D. Rudebusch (1991), Forecasting Output with the Composite Leading Index: A Real-Time Analysis, *Journal of the American Statistical Association* **86**, 603–610.

Diebold, F.X. and G.D. Rudebusch (1996), Measuring Business Cycles: A Modern Perspective, *Review of Economics and Statistics* **78**, 67–77.

Faust, J. and J. Wright (2009), Comparing Greenbook and Reduced Form Forecasts using a Large

Realtime Dataset, *Journal of Business & Economic Statistics*, forthcoming.

Gallo, G.M. and M. Marcellino (1999), Ex Post and Ex Ante Analysis of Provisional Data, *Journal of Forecasting* **18**, 421–433.

Garratt, A., G. Koop, E. Mise and S.P. Vahey (2009), Real-Time Prediction with UK Monetary Aggregates in the Presence of Model Uncertainty, *Journal of Business and Economic Statistics*, forthcoming.

Ghysels, E., N.R. Swanson and M. Callan (2002), Monetary Policy Rules with Model and Data Uncertainty, *Southern Economic Journal* **69**, 239–265.

Hamilton, J.D. and G. Perez-Quiros (1996), What Do the Leading Indicators Lead?, *Journal of Business* **69**, 27–49.

Howrey, E.P. (1978), The Use of Preliminary Data in Econometric Forecasting, *Review of Economics and Statistics* **66**, 386–393.

Kavajecz, K.A. and S. Collins (1995), Rationality of Preliminary Money Stock Estimates, *Review of Economics and Statistics* **77**, 32–41.

Keane, M.P. and D.E. Runkle (1989), Are Economic Forecasts Rational?, Federal Reserve Bank of Minneapolis *Quarterly Review* **13** (Spring), 26–33.

Keane, M.P. and D.E. Runkle (1990), Testing the Rationality of Price Forecasts: New Evidence from Panel Data, *American Economic Review* **80**, 714–735.

Kennedy, J. (1993), An Analysis of Revisions to the Industrial Production Index, *Applied Economics* **25**, 213–219.

Mankiw, N.G., D.E. Runkle and M.D. Shapiro (1984), Are Preliminary Announcements of the Money Stock Rational Forecasts?, *Journal of Monetary Economics* **14**, 15–27.

Mankiw, N.G. and M.D. Shapiro (1986), News or Noise: an Analysis of GNP Revisions, *Survey of Current Business* **66**, 20–25.

Mariano, R.S. and H. Tanizaki (1995), Prediction of Final Data with Use of Preliminary and/or Revised Data, *Journal of Forecasting* **14**, 351–380.

McConnell, M.M. and G. Perez Quiros (2000), Output Fluctuations in the United States: What Has Changed Since the Early 1980s?, *American Economic Review* **90**, 1464–1476.

McCracken, M. W. (2007), Asymptotics for Out-of-Sample Tests of Causality, *Journal of Econometrics*, **140**, 719–752.

Milbourne, R.D. and G.W. Smith (1989), How Informative Are Preliminary Announcements of the Money Stock in Canada?, *Canadian Journal of Economics*, **22**, 595–606.

Morgenstern, O. (1963), *On The Accuracy of Economic Observations*, 2nd. ed., Princeton: Princeton University Press.

Mork, K.A. (1987), Ain't Behavin': Forecast Errors and Measurement Errors in Early GNP Estimates, *Journal of Business & Economic Statistics* **5**, 165–175.

Muth J.F. (1961), Rational Expectations and the Theory of Price Movements, *Econometrica* **29**, 315–335.

Neftçi, S.N. and Theodossiou (1991), Properties and Stochastic Nature of BEA's Early Estimates of GNP, *Journal of Economics and Business* **43**, 231–239.

Pierce, D.A. (1981), Sources of Error in Economic Time Series, *Journal of Econometrics* **17**, 305–321.

Ratjens, P. and R.P. Robins (1995), Do Government Agencies Use Public Data: the Case of GNP, *Review of Economics and Statistics*, **77**, 170–172.

Robertson, J.C. and E.W. Tallman (1998), Data Vintages and Measuring Forecast Performance, Federal Reserve Bank of Atlanta *Economic Review* **83** (Fourth Quarter), 4–20.

Runkle, D.E. (1998), Revisionist History: How Data Revisions Distort Economic Policy Research, Federal Reserve Bank of Minneapolis *Quarterly Review* **22** (Fall), 3–12.

Shapiro, M.D. and M.W. Watson (1988), Sources of Business-Cycle Fluctuations, *NBER Macroeconomics Annual* **3**, 111–148.

Stekler, H.O. (1967), Data Revisions and Economic Forecasting, *Journal of the American Statistical Association* **62**, 470–483.

Stinchcombe, M.B. and H. White (1998), Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative, *Econometric Theory*, **14**, 3, 295–325.

Swanson, N.R., E. Ghysels and M. Callan (1999), A Multivariate Time Series Analysis of the Data Revision Process for Industrial Production and the Composite Leading Indicator, in R.F. Engle and H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Oxford: Oxford University Press, pp. 45–75.

Swanson, N.R. and D. van Dijk (2006), Are Statistical Reporting Agencies Getting It Right? Data Rationality and Business Cycle Asymmetry, *Journal of Business & Economic Statistics* **24**, 24–42.

West, K. D. (1996), Asymptotic Inference About Predictive Ability, *Econometrica* **64**, 1067-1084.

Table 1: Growth Rate and Revision Error Summary Statistics – Output, Prices and Money<sup>(\*)</sup>

<i>Vrbl</i>	<i>Vint</i>	<i>R-Err</i>	<i>smp1</i>	$\bar{y}$	$\hat{\sigma}_y$	$\hat{\sigma}_{\bar{y}}$	<i>skew</i>	<i>kurt</i>	<i>LB</i>	<i>JB</i>	<i>ADF</i>
Output	1	–	65:4	0.00657	0.00790	0.00061	-1.259	6.734	111.5	134.5	-6.035
	2	–	65:4	0.00705	0.00851	0.00066	-1.148	6.678	107.1	123.8	-6.350
	–	1	65:4	0.00046	0.00189	0.00015	0.118	2.950	23.58	0.424	-6.050
	–	2	65:4	-0.00002	0.00106	0.00008	0.324	8.982	32.01	237.0	-5.471
	1	–	70:1	0.00626	0.00817	0.00068	-1.190	6.395	96.90	101.0	-5.729
	2	–	70:1	0.00675	0.00877	0.00073	-1.111	6.446	92.42	98.05	-4.911
	–	1	70:1	0.00048	0.00193	0.00016	0.041	2.873	28.17	0.208	-5.620
	–	2	70:1	-0.00001	0.00108	0.00009	0.300	9.087	27.70	216.9	-5.128
	1	–	83:1	0.00734	0.00486	0.00050	0.230	3.940	44.55	3.728	-6.299
	2	–	83:1	0.00766	0.00538	0.00056	0.379	3.958	53.06	5.179	-6.178
	–	1	83:1	0.00029	0.00171	0.00018	-0.301	2.677	25.53	1.944	-9.994
	–	2	83:1	0.00012	0.00098	0.00010	1.753	9.658	14.73	207.4	-9.753
1	–	90:1	0.00682	0.00463	0.00057	-0.376	3.470	30.12	1.892	-5.513	
2	–	90:1	0.00724	0.00510	0.00063	-0.195	3.277	38.83	0.490	-5.508	
–	1	90:1	0.00037	0.00161	0.00020	-0.091	2.077	17.02	2.755	-3.667	
–	2	90:1	0.00008	0.00093	0.00012	2.090	12.60	13.24	275.2	-8.226	
Prices	1	–	65:4	0.00958	0.00608	0.00047	1.163	4.093	1040	44.05	-1.613
	2	–	65:4	0.00988	0.00636	0.00049	1.245	4.272	925.4	51.79	-1.463
	–	1	65:4	0.00026	0.00114	0.00009	1.235	7.277	35.13	160.8	-5.384
	–	2	65:4	-0.00001	0.00054	0.00004	-1.335	13.30	10.61	745.4	-12.93
	1	–	70:1	0.00968	0.00638	0.00052	1.093	3.716	955.5	31.43	-1.411
	2	–	70:1	0.01001	0.00666	0.00055	1.173	3.884	852.6	37.00	-1.274
	–	1	70:1	0.00029	0.00118	0.00010	1.175	6.844	33.78	118.5	-4.891
	–	2	70:1	-0.00003	0.00055	0.00005	-1.484	13.28	12.24	669.0	-9.487
	1	–	83:1	0.00624	0.00297	0.00030	0.471	2.992	191.9	3.429	-1.998
	2	–	83:1	0.00646	0.00296	0.00030	0.424	2.693	192.4	3.256	-1.699
	–	1	83:1	0.00020	0.00096	0.00010	1.423	10.95	16.40	264.2	-8.831
	–	2	83:1	-0.00001	0.00053	0.00006	-1.506	17.27	14.99	783.4	-7.982
1	–	90:1	0.00528	0.00261	0.00032	0.872	4.549	50.94	13.71	-1.914	
2	–	90:1	0.00553	0.00252	0.00031	0.661	3.499	60.61	5.018	-1.285	
–	1	90:1	0.00023	0.00083	0.00010	2.748	17.87	13.34	644.4	-7.535	
–	2	90:1	-0.00004	0.00056	0.00007	-2.321	18.07	13.61	627.0	-8.304	
Money	1	–	65:4	0.01207	0.01200	0.00093	0.077	3.119	169.2	0.209	-3.364
	2	–	65:4	0.01240	0.01191	0.00093	0.097	3.169	173.2	0.374	-4.922
	–	1	65:4	0.00018	0.00135	0.00011	1.800	12.30	15.50	660.0	-13.33
	–	2	65:4	0.00009	0.00113	0.00009	1.316	14.54	11.80	923.2	-12.13
	1	–	70:1	0.01221	0.01244	0.00103	0.068	2.983	166.7	0.131	-4.488
	2	–	70:1	0.01256	0.01235	0.00103	0.088	3.029	170.3	0.187	-4.464
	–	1	70:1	0.00018	0.00140	0.00012	1.785	11.72	16.31	521.8	-12.67
	–	2	70:1	0.00005	0.00117	0.00010	1.344	14.26	10.67	782.4	-11.62
	1	–	83:1	0.01085	0.01403	0.00145	0.283	2.682	159.2	1.783	-3.108
	2	–	83:1	0.01120	0.01391	0.00144	0.301	2.728	163.1	1.805	-3.131
	–	1	83:1	0.00010	0.00138	0.00014	2.155	12.96	11.76	438.3	-10.04
	–	2	83:1	0.00009	0.00090	0.00009	-0.056	7.032	19.05	58.43	-9.292
1	–	90:1	0.00788	0.01302	0.00160	0.360	2.955	162.5	1.437	-2.218	
2	–	90:1	0.00827	0.01302	0.00161	0.454	3.170	159.7	2.188	-2.273	
–	1	90:1	0.00009	0.00139	0.00017	1.383	9.642	16.40	131.0	-8.506	
–	2	90:1	0.00001	0.00095	0.00012	-0.488	6.006	18.34	24.04	-7.682	

(\*) Summary statistics are reported for a generic variable denoted by  $y$ , where  $y$  = output, price, and money growth rates (see table rows where  $vint = 1, 2$ , corresponding to first and second available data), as well where  $y$  = the revision error associated with these variables (see table rows where  $R-Err = 1, 2$ , corresponding to revision errors associated with second and third available data - i.e.  ${}_{t+2}u_t^{t+1}$  and  ${}_{t+3}u_t^{t+2}$ ). Statistics are reported for samples beginning in 1970:1, 1983:1, and 1990:1. All samples end in 2006:4. Additionally,  $\bar{y}$  is the mean of the series,  $\hat{\sigma}_y$  is the standard error of the series,  $\hat{\sigma}_{\bar{y}}$  is the standard error of  $\bar{y}$ ,  $skew$  is skewness,  $kurt$  is kurtosis,  $LB$  is the Ljung-Box statistic,  $JB$  is the Jarques-Bera statistic, and  $ADF$  is the augmented Dickey-Fuller statistic, where lag augmentations are selected via use of the Schwarz Information Criterion. See Sections 4 and 5 for further details.

Table 2: Growth Rate and Revision Error Summary Statistics Based on Various Subsamples –  
Output<sup>(\*)</sup>

<i>Vrbl</i>	<i>Vint</i>	<i>R-Err</i>	<i>simpl</i>	$\bar{y}$	$\hat{\sigma}_y$	$\hat{\sigma}_{\bar{y}}$	<i>skew</i>	<i>kurt</i>	<i>LB</i>	<i>JB</i>	<i>ADF</i>
	1	–	65:4-06:4	0.00657	0.00790	0.00061	-1.259	6.734	111.5	134.53	-6.035
	2	–	65:4-06:4	0.00705	0.00851	0.00066	-1.148	6.678	107.1	123.88	-6.350
	–	1	65:4-06:4	0.00046	0.00189	0.00015	0.118	2.950	23.58	0.424	-6.050
	–	2	65:4-06:4	-0.00002	0.00106	0.00008	0.324	8.982	32.01	237.04	-5.471
	1	–	65:4-75:4	0.00628	0.01114	0.00027	-1.139	4.566	55.42	11.33	-3.424
	2	–	65:4-75:4	0.00613	0.01169	0.00029	-1.273	4.595	64.44	13.14	-3.082
	–	1	65:4-75:4	0.00035	0.00191	0.00005	0.498	3.306	8.750	1.572	-3.198
	–	2	65:4-75:4	-0.00008	0.00094	0.00002	-0.776	7.838	2.908	35.68	-6.000
	1	–	76:1-80:4	0.00655	0.00976	0.00049	-1.559	5.774	30.36	11.02	-
	2	–	76:1-80:4	0.00876	0.01139	0.00057	-1.125	5.292	53.24	6.246	-
	–	1	76:1-80:4	0.00096	0.00199	0.00010	0.478	2.565	9.755	1.002	-
	–	2	76:1-80:4	-0.00064	0.00134	0.00007	-1.722	4.246	20.36	8.506	-
	1	–	81:1-85:4	0.00632	0.00998	0.00050	-0.301	2.234	29.88	1.065	-
	2	–	81:1-85:4	0.00680	0.01044	0.00052	-0.074	2.425	25.95	0.564	-
	–	1	81:1-85:4	0.00062	0.00259	0.00013	0.150	2.024	7.350	1.169	-
	–	2	81:1-85:4	-0.00003	0.00072	0.00004	0.470	8.664	8.102	18.65	-
	1	–	86:1-91:4	0.00496	0.00450	0.00019	-0.938	4.512	25.33	4.410	-0.172
	2	–	86:1-91:4	0.00493	0.00466	0.00019	-0.831	3.318	39.97	2.434	0.480
	–	1	86:1-91:4	-0.00009	0.00184	0.00008	-0.764	2.941	4.842	2.051	-5.982
	–	2	86:1-91:4	0.00027	0.00104	0.00004	1.732	6.962	5.392	2.646	3.532
	1	–	92:1-95:4	0.00643	0.00375	0.00022	0.212	2.412	18.59	4.410	-
	2	–	92:1-95:4	0.00738	0.00406	0.00024	0.482	2.974	7.951	2.434	-
	–	1	92:1-95:4	0.00101	0.00134	0.00008	-1.289	4.629	8.588	4.045	-
	–	2	92:1-95:4	-0.00008	0.00045	0.00003	-3.262	12.266	2.044	58.55	-
	1	–	96:1-97:1	0.00854	-	-	-	-	-	-	-
	2	–	96:1-97:1	0.00569	-	-	-	-	-	-	-
	–	1	96:1-97:1	-0.00027	-	-	-	-	-	-	-
	–	2	96:1-97:1	0.00005	-	-	-	-	-	-	-
	1	–	97:2-99:3	0.00904	-	-	-	-	-	-	-
	2	–	97:2-99:3	0.00984	-	-	-	-	-	-	-
	–	1	97:2-99:3	0.00048	-	-	-	-	-	-	-
	–	2	97:2-99:3	-0.00008	-	-	-	-	-	-	-
	1	–	99:4-00:1	0.01293	-	-	-	-	-	-	-
	2	–	99:4-00:1	0.00922	-	-	-	-	-	-	-
	–	1	99:4-00:1	0.00202	-	-	-	-	-	-	-
	–	2	99:4-00:1	0.00000	-	-	-	-	-	-	-
	1	–	00:2-03:4	0.00639	0.00551	0.00037	0.638	2.193	17.55	1.570	-
	2	–	00:2-03:4	0.00641	0.00562	0.00037	0.369	2.397	12.48	0.796	-
	–	1	00:2-03:4	0.00001	0.00184	0.00012	0.477	1.995	6.526	1.380	-
	–	2	00:2-03:4	0.00050	0.00101	0.00007	1.438	3.144	10.22	3.911	-
	1	–	04:1-06:4	0.00786	0.00258	0.00022	-0.614	2.671	10.62	0.866	-
	2	–	04:1-06:4	0.00998	0.00386	0.00032	1.152	4.678	22.72	0.866	-
	–	1	04:1-06:4	0.00089	0.00068	0.00006	0.003	1.920	6.137	0.914	-
	–	2	04:1-06:4	-0.00008	0.00042	0.00003	-2.613	8.421	3.170	16.59	-

(\*) See notes to Table 1. Subsamples were chosen by examining plots of various calendar dates (including 1959:4, 1960:4, 1961:4, and 1962:4) across all vintages from 1964:4-2006:4, and by defining “breaks” to be points where the historical data changed. See Section 4 and 5 for further details.

Table 3: MSFEs Calculated Based on Simple Real-Time Autoregressions Without Revision Errors  
for Output<sup>(\*)</sup>

<i>Model</i>	<i>RevErr</i>	<i>k = 2</i>	<i>k = 3</i>	<i>k = 4</i>	<i>k = 6</i>	<i>k = 12</i>	<i>k = 24</i>
<i>Panel A: Mean Square Forecast Errors</i>							
<i>Begin Date of Forecast Period = 1970:1</i>							
A	None	<b>0.642</b>	<b>0.783</b>	<b>0.825</b>	0.841	0.839	0.836
B	None	0.642	0.983	1.028	1.232	1.090	1.134
C	None	0.661	0.792	<b>0.825</b>	<b>0.828</b>	<b>0.817</b>	<b>0.815</b>
RWD-A	None	0.768	0.879	0.890	0.861	0.833	0.843
RWD-B	None	0.768	0.896	0.916	0.905	0.846	0.925
RWD-C	None	0.766	0.874	0.884	0.856	0.829	0.838
<i>Begin Date of Forecast Period = 1983:1</i>							
A	None	0.212	<b>0.259</b>	0.270	0.303	0.322	0.354
B	None	0.212	0.303	0.322	0.490	0.466	0.573
C	None	<b>0.206</b>	<b>0.259</b>	<b>0.267</b>	<b>0.299</b>	<b>0.316</b>	<b>0.346</b>
RWD-A	None	0.275	0.337	0.345	0.374	0.382	0.419
RWD-B	None	0.275	0.326	0.337	0.359	0.345	0.363
RWD-C	None	0.248	0.309	0.314	0.343	0.356	0.387
<i>Begin Date of Forecast Period = 1990:1</i>							
A	None	<b>0.175</b>	<b>0.204</b>	<b>0.216</b>	0.278	<b>0.286</b>	<b>0.332</b>
B	None	0.175	0.247	0.236	0.343	0.432	0.491
C	None	0.176	0.214	<b>0.216</b>	<b>0.275</b>	0.289	0.329
RWD-A	None	0.224	0.275	0.270	0.322	0.327	0.371
RWD-B	None	0.224	0.271	0.266	0.324	0.344	0.379
RWD-C	None	0.208	0.258	0.250	0.304	0.319	0.356
<i>Panel B: Diebold-Mariano Test Statistics Corresponding to Entries in Panel A</i>							
<i>Begin Date of Forecast Period = 1970:1</i>							
A	None	—	—	—	—	—	—
B	None	—	-1.415	-1.423	-2.216	-1.148	-1.586
C	None	-0.349	-0.159	-0.008	0.213	0.348	0.368
RWD-A	None	-1.077	-0.718	-0.446	-0.125	0.035	-0.047
RWD-B	None	-1.077	-0.811	-0.589	-0.359	-0.041	-0.470
RWD-C	None	-1.070	-0.689	-0.411	-0.097	0.057	-0.012
<i>Begin Date of Forecast Period = 1983:1</i>							
A	None	—	—	—	—	—	—
B	None	—	-1.096	-1.056	-2.558	-2.467	-2.542
C	None	0.480	-0.053	0.255	0.344	0.378	0.562
RWD-A	None	-1.525	-1.826	-1.718	-1.688	-1.568	-1.582
RWD-B	None	-1.525	-1.638	-1.576	-1.421	-0.568	-0.200
RWD-C	None	-0.989	-1.366	-1.242	-1.181	-1.001	-0.936
<i>Begin Date of Forecast Period = 1990:1</i>							
A	None	—	—	—	—	—	—
B	None	—	-1.419	-0.683	-1.943	-1.743	-2.365
C	None	-0.060	-0.570	-0.039	0.166	-0.149	0.157
RWD-A	None	-1.316	-1.771	-1.499	-1.182	-1.188	-1.157
RWD-B	None	-1.316	-1.557	-1.336	-1.075	-1.027	-0.764
RWD-C	None	-0.852	-1.281	-0.976	-0.711	-0.845	-0.647

(\*) In Panel A, forecast mean square errors (MSFEs) are reported based on predictions constructed using recursively estimated models with estimation period beginning in 1965:4 and *ex ante* prediction periods beginning in 1970:1, 1983:1, or 1990:1. Corresponding Diebold-Mariano predictive accuracy test statistics are reported in Panel B. In all cases, Model A is set as the “benchmark” model, so that a negative statistic means that Model A is “MSFE-better” than the particular model against which it is being compared. All estimated models are either pure autoregressions or autoregressions with revision error(s) included as additional explanatory variables. Lags are selected using the Schwarz Information Criterion. The pure autoregression models are: Model A (*First Available Data*):  $t+kX_{t+1} = \alpha + \sum_{i=1}^p \beta_{it+2-i}X_{t+1-i} + \varepsilon_{t+k}$  Model B (*k<sup>th</sup> Available Data*) :  $t+kX_{t+1} = \alpha + \sum_{i=1}^p \beta_{it+2-i}X_{t+3-k-i} + \varepsilon_{t+k}$  Model C (*Real-Time Data*) :  $t+kX_{t+1} = \alpha + \sum_{i=1}^p \beta_{it+1}X_{t+1-i} + \varepsilon_{t+k}$  In the above models,  $X$  denotes the growth rate of either output, prices, or money. Also, RWD is the random walk with drift model in log levels,  $u_{C1} = t+1u_{t-k}^t$ ,  $k = 1$ ;  $u_{C2} = t+1u_{t-k}^t$ ,  $k = 1, 2$ ; and  $u_{C3} = t+1u_{t+1-k}^{t+2-k}$ ,  $k = 3$ . Further details are contained in Sections 3 and 4.

Table 4: MSFEs Calculated Based on Simple Real-Time Autoregressions Without Revision Errors

		for Prices <sup>(*)</sup>						
<i>Model</i>	<i>RevErr</i>	<i>k = 2</i>	<i>k = 3</i>	<i>k = 4</i>	<i>k = 6</i>	<i>k = 12</i>	<i>k = 24</i>	
<i>Panel A: Mean Square Forecast Errors</i>								
<i>Begin Date of Forecast Period =1970:1</i>								
A	None	<b>0.149</b>	<b>0.166</b>	<b>0.167</b>	<b>0.163</b>	<b>0.139</b>	<b>0.125</b>	
B	None	0.149	0.214	0.211	0.339	0.516	0.608	
C	None	0.160	0.179	0.182	0.178	0.151	0.137	
RWD-A	None	0.430	0.463	0.457	0.442	0.444	0.437	
RWD-B	None	0.430	0.486	0.489	0.495	0.557	0.739	
RWD-C	None	0.472	0.500	0.496	0.480	0.478	0.469	
<i>Begin Date of Forecast Period =1983:1</i>								
A	None	<b>0.072</b>	<b>0.076</b>	<b>0.075</b>	<b>0.069</b>	<b>0.063</b>	<b>0.063</b>	
B	None	0.072	0.077	0.076	0.088	0.257	0.556	
C	None	0.075	0.079	0.079	0.073	0.066	0.065	
RWD-A	None	0.395	0.376	0.374	0.362	0.349	0.360	
RWD-B	None	0.395	0.414	0.416	0.394	0.380	0.303	
RWD-C	None	0.512	0.488	0.486	0.474	0.458	0.469	
<i>Begin Date of Forecast Period =1990:1</i>								
A	None	<b>0.057</b>	<b>0.047</b>	<b>0.050</b>	<b>0.049</b>	<b>0.040</b>	<b>0.043</b>	
B	None	0.057	0.059	0.054	0.073	0.108	0.191	
C	None	0.061	0.051	0.053	0.053	0.042	<b>0.043</b>	
RWD-A	None	0.440	0.406	0.409	0.387	0.368	0.378	
RWD-B	None	0.440	0.447	0.455	0.426	0.419	0.388	
RWD-C	None	0.561	0.522	0.526	0.500	0.478	0.489	
<i>Panel B: Diebold-Mariano Test Statistics Corresponding to Entries in Panel A</i>								
<i>Begin Date of Forecast Period =1970:1</i>								
A	None	—	—	—	—	—	—	
B	None	—	-1.419	-1.222	-3.003	-4.030	-5.085	
C	None	-0.704	-0.758	-0.788	-0.826	-0.807	-0.741	
RWD-A	None	-5.584	-5.526	-5.408	-5.565	-5.575	-5.926	
RWD-B	None	-5.584	-5.802	-5.647	-5.648	-5.614	-5.647	
RWD-C	None	-6.462	-6.544	-6.417	-6.587	-6.677	-6.936	
<i>Begin Date of Forecast Period =1983:1</i>								
A	None	—	—	—	—	—	—	
B	None	—	-0.061	-0.074	-1.589	-2.627	-3.793	
C	None	-0.747	-0.782	-0.715	-0.971	-0.735	-0.452	
RWD-A	None	-8.740	-8.298	-8.206	-8.827	-8.798	-8.741	
RWD-B	None	-8.740	-8.787	-8.724	-9.020	-8.444	-6.042	
RWD-C	None	-9.928	-9.616	-9.594	-10.408	-10.37	-10.06	
<i>Begin Date of Forecast Period =1990:1</i>								
A	None	—	—	—	—	—	—	
B	None	—	-1.669	-0.280	-1.763	-3.278	-3.967	
C	None	-0.703	-0.983	-0.429	-0.628	-0.341	-0.171	
RWD-A	None	-7.913	-7.974	-7.956	-7.497	-7.640	-7.676	
RWD-B	None	-7.913	-8.341	-8.378	-7.720	-7.768	-7.149	
RWD-C	None	-8.813	-9.031	-9.185	-8.671	-8.850	-8.730	

(\*) See notes to Table 3.

Table 5: MSFEs Calculated Based on Simple Real-Time Autoregressions Without Revision Errors  
for Money<sup>(\*)</sup>

<i>Model</i>	<i>RevErr</i>	<i>k = 2</i>	<i>k = 3</i>	<i>k = 4</i>	<i>k = 6</i>	<i>k = 12</i>	<i>k = 24</i>
<i>Panel A: Mean Square Forecast Errors</i>							
<i>Begin Date of Forecast Period = 1970:1</i>							
A	None	1.030	1.024	1.008	0.937	0.984	0.998
B	None	1.030	1.253	1.284	1.253	2.000	1.692
C	None	<b>1.025</b>	<b>1.016</b>	<b>0.999</b>	<b>0.918</b>	<b>0.948</b>	<b>0.971</b>
RWD-A	None	1.633	1.630	1.612	1.556	1.565	1.567
RWD-B	None	1.633	1.650	1.646	1.611	1.666	1.752
RWD-C	None	1.650	1.643	1.623	1.560	1.562	1.561
<i>Begin Date of Forecast Period = 1983:1</i>							
A	None	1.168	1.140	1.099	1.007	1.022	1.040
B	None	1.168	1.366	1.475	1.682	2.824	2.308
C	None	<b>1.117</b>	<b>1.088</b>	<b>1.042</b>	<b>0.949</b>	<b>0.955</b>	<b>0.979</b>
RWD-A	None	2.257	2.235	2.199	2.140	2.142	2.148
RWD-B	None	2.257	2.277	2.281	2.285	2.344	2.314
RWD-C	None	2.323	2.300	2.260	2.201	2.198	2.208
<i>Begin Date of Forecast Period = 1990:1</i>							
A	None	1.058	1.058	1.035	0.916	0.907	0.955
B	None	1.058	1.155	1.219	1.591	3.001	2.480
C	None	<b>1.023</b>	<b>1.023</b>	<b>0.995</b>	<b>0.877</b>	<b>0.861</b>	<b>0.915</b>
RWD-A	None	2.383	2.395	2.387	2.319	2.279	2.312
RWD-B	None	2.383	2.452	2.499	2.541	2.589	2.487
RWD-C	None	2.539	2.551	2.542	2.475	2.431	2.464
<i>Panel B: Diebold-Mariano Test Statistics Corresponding to Entries in Panel A</i>							
<i>Begin Date of Forecast Period = 1970:1</i>							
A	None	—	—	—	—	—	—
B	None	—	-2.499	-2.621	-2.391	-3.486	-3.087
C	None	0.155	0.252	0.301	0.666	1.251	0.914
RWD-A	None	-3.635	-3.643	-3.624	-3.671	-3.379	-3.268
RWD-B	None	-3.635	-3.630	-3.583	-3.520	-3.395	-4.090
RWD-C	None	-3.419	-3.390	-3.370	-3.364	-3.078	-2.958
<i>Begin Date of Forecast Period = 1983:1</i>							
A	None	—	—	—	—	—	—
B	None	—	-2.099	-2.444	-3.821	-4.059	-3.758
C	None	1.423	1.474	1.622	1.805	1.917	1.802
RWD-A	None	-4.404	-4.439	-4.466	-4.577	-4.482	-4.380
RWD-B	None	-4.404	-4.455	-4.504	-4.562	-4.507	-4.695
RWD-C	None	-4.266	-4.279	-4.294	-4.392	-4.317	-4.232
<i>Begin Date of Forecast Period = 1990:1</i>							
A	None	—	—	—	—	—	—
B	None	—	-0.710	-1.014	-2.736	-3.411	-3.222
C	None	1.264	1.264	1.366	1.777	1.788	1.553
RWD-A	None	-3.770	-3.803	-3.840	-3.969	-3.850	-3.717
RWD-B	None	-3.770	-3.825	-3.892	-4.042	-4.001	-4.080
RWD-C	None	-3.796	-3.825	-3.858	-3.974	-3.874	-3.753

(\*) See notes to Table 3.

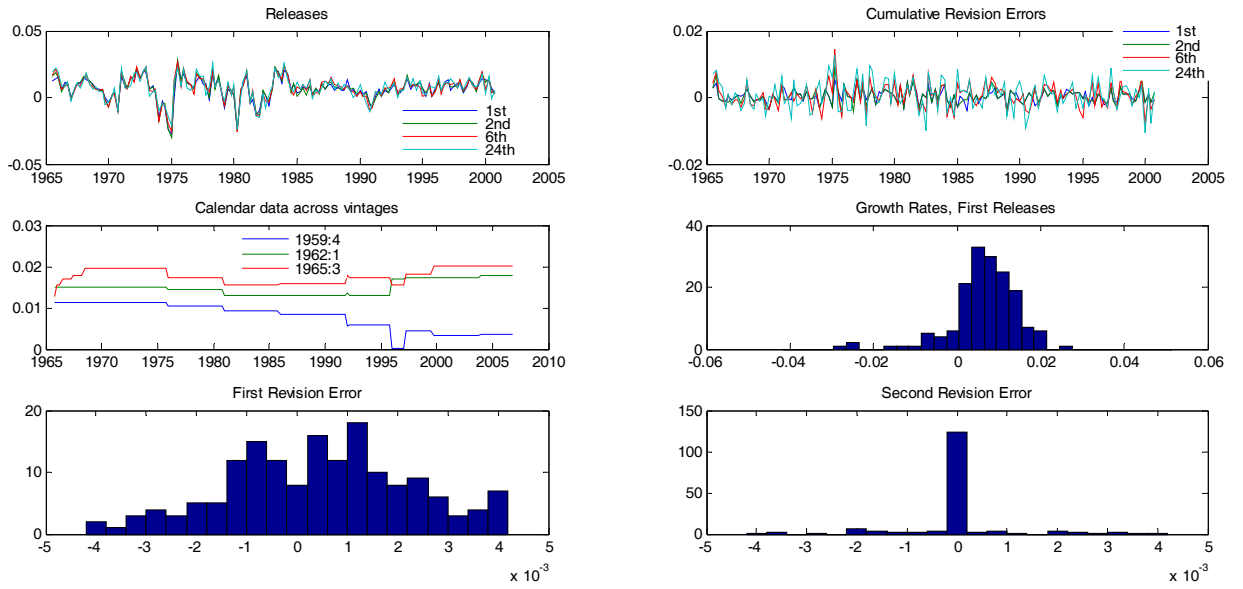
Table 6: MSFEs Calculated Based on Real-Time Vector Autoregressions With and Without

Money and Revision Errors <sup>(*)</sup>							
<i>Model</i>	<i>RevErr</i>	<i>k = 2</i>	<i>k = 3</i>	<i>k = 4</i>	<i>k = 6</i>	<i>k = 12</i>	<i>k = 24</i>
<i>Panel A: Mean Square Forecast Errors</i>							
<i>Begin Date of Forecast Period =1971:2</i>							
VAR - NoM Mod A	None	0.866	1.004	1.031	1.049	1.088	1.083
VAR - M Mod A	None	<i>1.069*</i>	<i>1.196*</i>	<i>1.190*</i>	<i>1.200*</i>	<i>1.226*</i>	<i>1.235*</i>
VAR - NoM Mod C	None	0.889	1.027	1.054	1.049	1.074	1.042
VAR - M Mod C	None	1.499	1.629	1.639	1.628	1.662	1.602
VAR - NoM Mod A	<i>u<sub>C1</sub></i>	<b>0.719</b>	<b>0.856</b>	<b>0.876</b>	<b>0.891</b>	<b>0.929</b>	<b>0.916</b>
VAR - M Mod A	<i>u<sub>C1</sub></i>	1.172	1.296	1.285	1.296	1.330	1.355
VAR - NoM Mod A	<i>u<sub>C2</sub></i>	1.036	1.168	1.219	1.235	1.281	1.271
VAR - M Mod A	<i>u<sub>C2</sub></i>	1.562	1.691	1.706	1.708	1.731	1.716
VAR - NoM Mod C	<i>u<sub>C1</sub></i>	0.840	0.982	1.006	1.009	1.037	1.014
VAR - M Mod C	<i>u<sub>C1</sub></i>	2.213	2.322	2.362	2.356	2.406	2.303
VAR - NoM Mod C	<i>u<sub>C2</sub></i>	1.107	1.267	1.317	1.327	1.372	1.341
VAR - M Mod C	<i>u<sub>C2</sub></i>	1.334	1.499	1.489	1.463	1.482	1.466
<i>Begin Date of Forecast Period =1983:1</i>							
VAR - NoM Mod A	None	0.270	<b>0.315*</b>	0.322	0.342	<b>0.338</b>	0.366
VAR - M Mod A	None	<i>0.324*</i>	<i>0.378*</i>	<i>0.377*</i>	<i>0.396</i>	<i>0.389</i>	<i>0.409</i>
VAR - NoM Mod C	None	0.274	0.323	0.326	0.334	0.345	<b>0.364*</b>
VAR - M Mod C	None	0.466	0.524	0.509	0.511	0.500	0.519
VAR - NoM Mod A	<i>u<sub>C1</sub></i>	<b>0.269*</b>	<b>0.315*</b>	0.322	0.340	0.339	0.366
VAR - M Mod A	<i>u<sub>C1</sub></i>	<i>0.324*</i>	<i>0.378*</i>	0.378	<i>0.396</i>	0.391	0.410
VAR - NoM Mod A	<i>u<sub>C2</sub></i>	0.293	0.341	0.348	0.361	0.358	0.382
VAR - M Mod A	<i>u<sub>C2</sub></i>	0.390	0.451	0.454	0.447	0.438	0.447
VAR - NoM Mod C	<i>u<sub>C1</sub></i>	<b>0.269*</b>	0.319	<b>0.320</b>	<b>0.328</b>	0.346	<b>0.364</b>
VAR - M Mod C	<i>u<sub>C1</sub></i>	0.454	0.512	0.496	0.505	0.504	0.518
VAR - NoM Mod C	<i>u<sub>C2</sub></i>	0.278	0.325	0.328	0.336	0.354	0.371
VAR - M Mod C	<i>u<sub>C2</sub></i>	0.516	0.585	0.567	0.562	0.546	0.561
<i>Begin Date of Forecast Period =1990:1</i>							
VAR - NoM Mod A	None	<b>0.227*</b>	<b>0.243</b>	<b>0.253</b>	0.313	<b>0.316</b>	<b>0.343</b>
VAR - M Mod A	None	<i>0.241*</i>	<i>0.252</i>	<i>0.260</i>	<i>0.316</i>	<i>0.319</i>	<i>0.344</i>
VAR - NoM Mod C	None	0.242	0.266	0.265	0.313	0.333	0.344
VAR - M Mod C	None	0.319	0.334	0.329	0.366	0.386	0.392
VAR - NoM Mod A	<i>u<sub>C1</sub></i>	0.230	0.245	0.255	0.315	0.318	0.347
VAR - M Mod A	<i>u<sub>C1</sub></i>	<i>0.241*</i>	0.253	0.261	0.317	<i>0.319</i>	0.346
VAR - NoM Mod A	<i>u<sub>C2</sub></i>	0.239	0.254	0.265	0.323	0.326	0.351
VAR - M Mod A	<i>u<sub>C2</sub></i>	0.253	0.264	0.273	0.327	0.329	0.351
VAR - NoM Mod C	<i>u<sub>C1</sub></i>	0.235	0.258	0.258	<b>0.308</b>	0.329	<b>0.343</b>
VAR - M Mod C	<i>u<sub>C1</sub></i>	0.326	0.338	0.333	0.378	0.404	0.413
VAR - NoM Mod C	<i>u<sub>C2</sub></i>	0.246	0.270	0.270	0.316	0.336	0.343
VAR - M Mod C	<i>u<sub>C2</sub></i>	0.332	0.349	0.344	0.387	0.412	0.417

(\*) See notes to Table 3. Vector autoregressions with and without money are used to predict real-time output. Entries are *MSFEs*, and starred entries denote rejection of the Diebold-Mariano null hypothesis of equal predictive accuracy at a 10% level, using standard normal critical values, assuming that the benchmark model is Model A from Table 3. Entries in bold are *MSFE*-best across all models, for a given value of *k*, and *MSFEs* in italics are *MSFE*-best across all models that include money, for a given value of *k*. See notes to Table 3 and Section 3 of the paper for complete details.

Figure 1: Output - Historical Data and Prediction Results

version/KE1FZ306.wmf



version/KE1FZ307.wmf

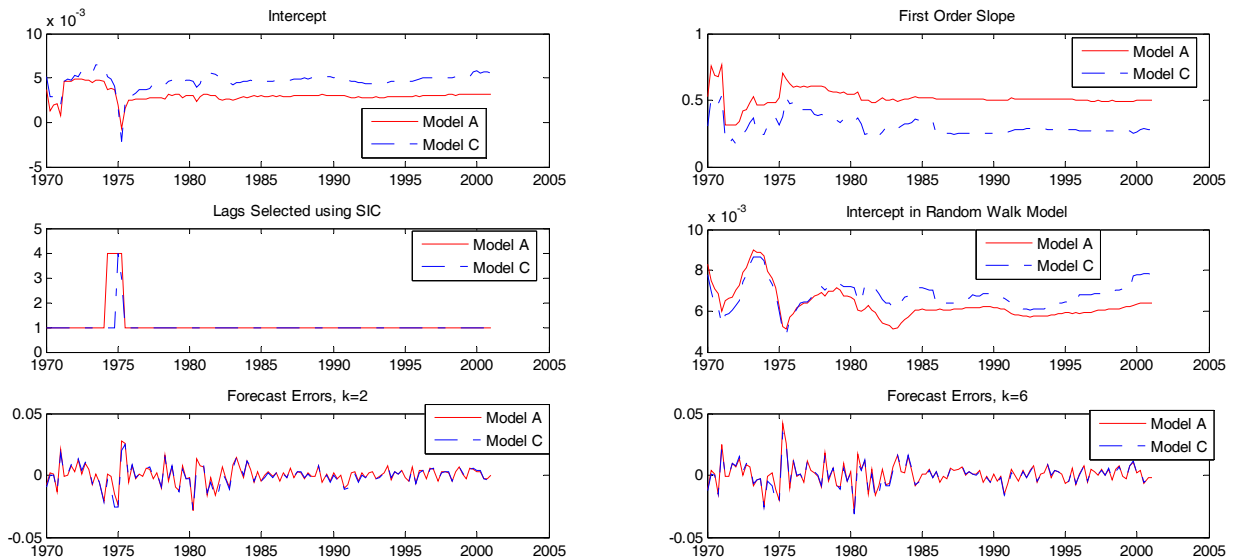
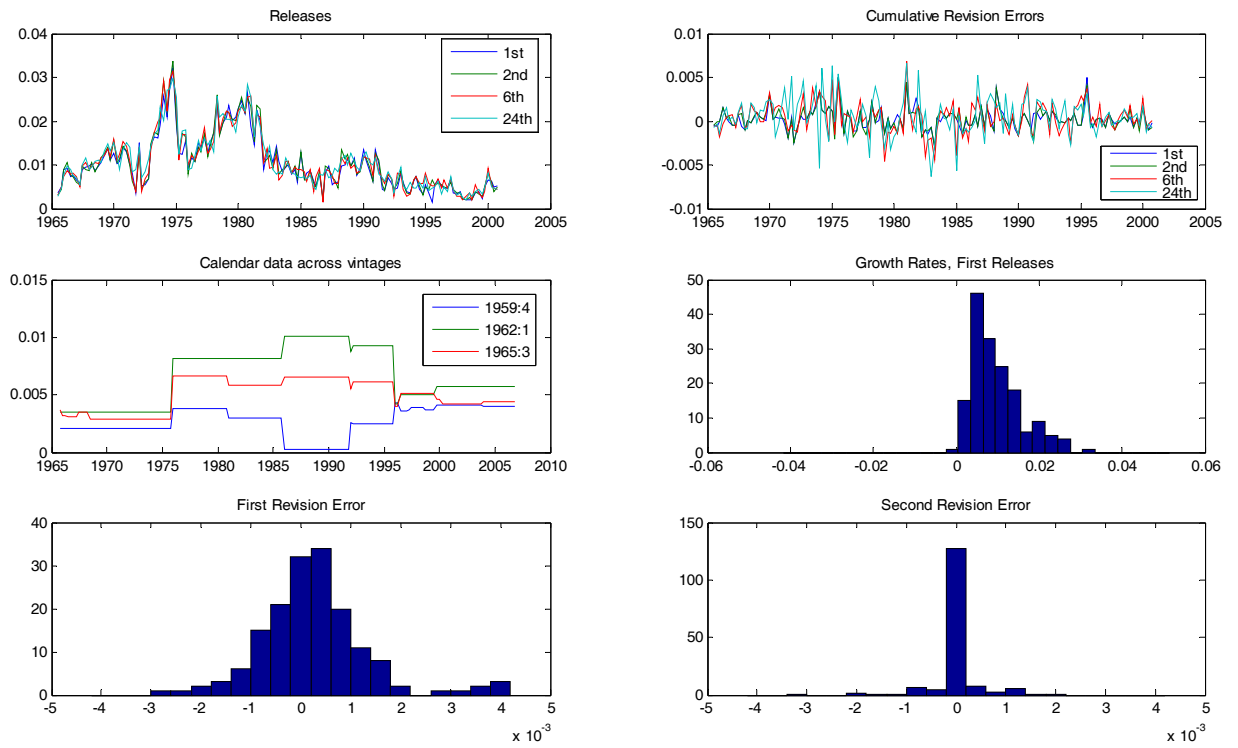


Figure 2: Prices - Historical Data and Prediction Results

version/KE1FZ308.wmf



version/KE1FZ309.wmf

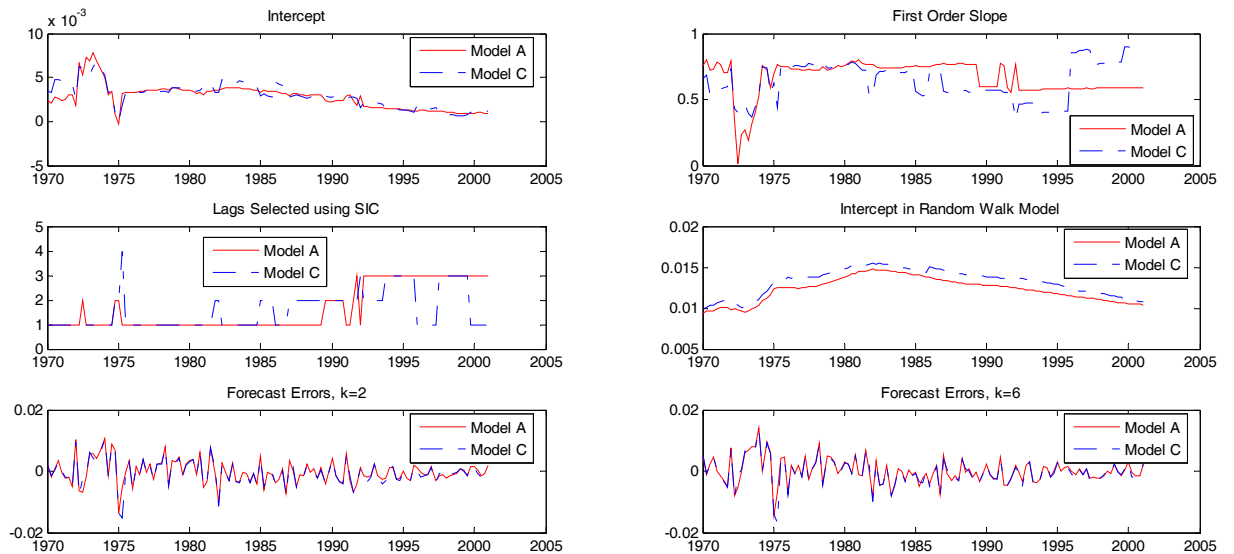
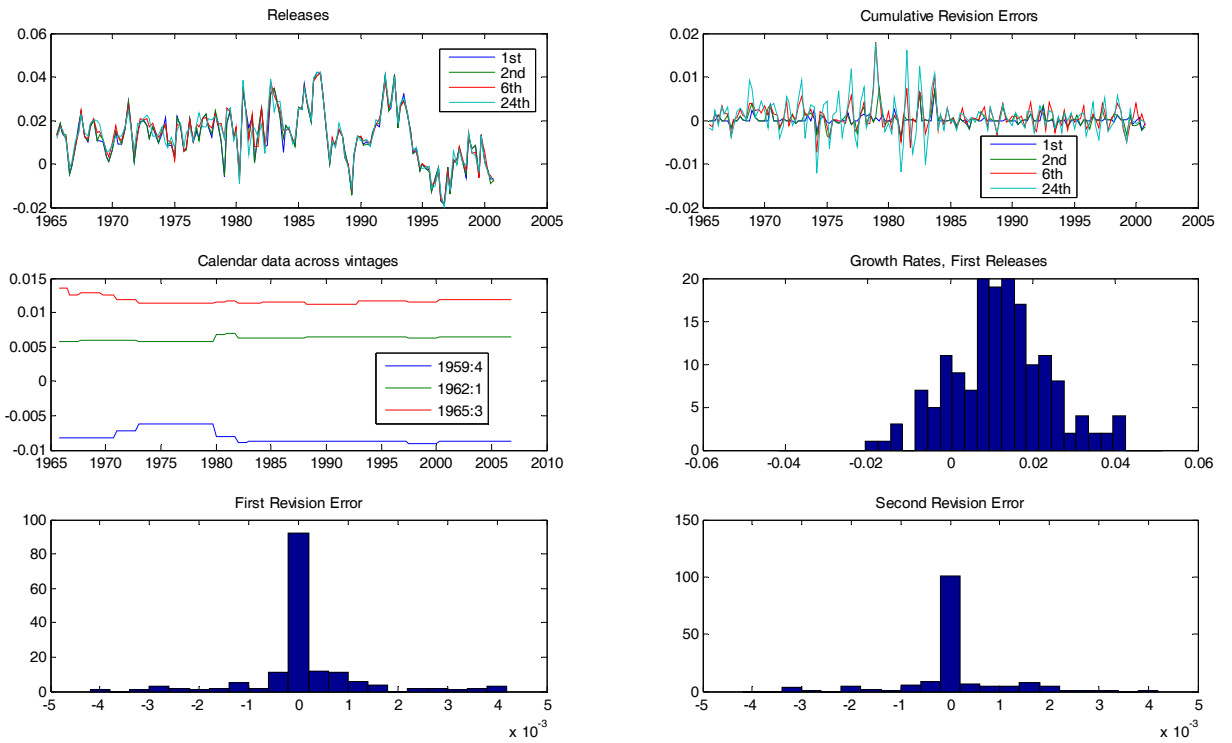
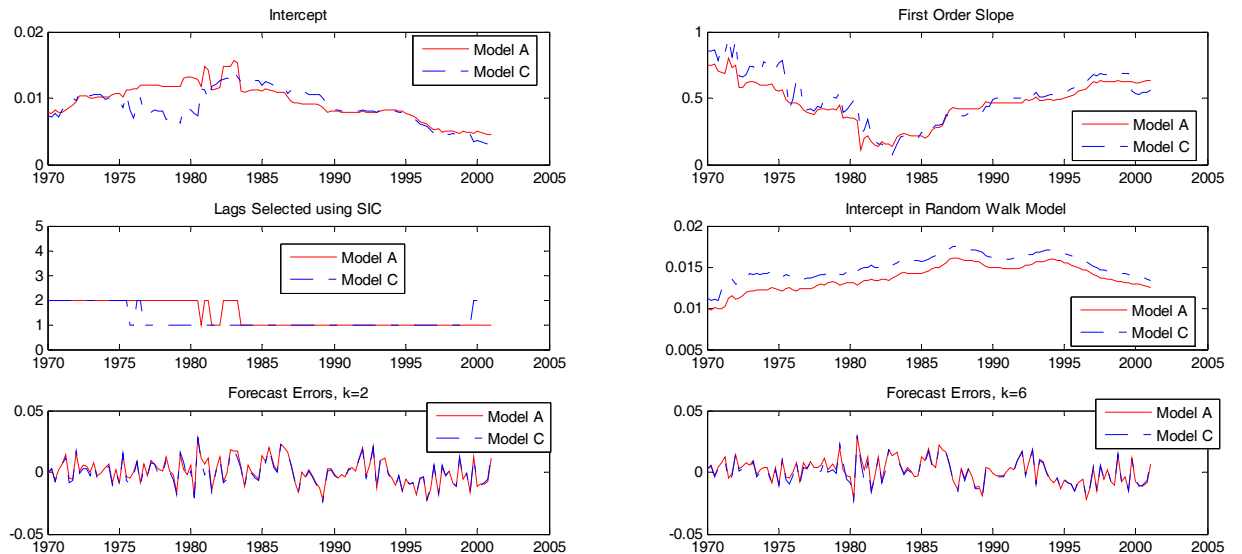


Figure 3: Money - Historical Data and Prediction Results

version/KE1FZ30A.wmf



version/KE1FZ30B.wmf



## Appendix 1 - Forecast Experiment Results When Revision Errors Included

Table A1.1: MSFEs Calculated Based on Simple Real-Time Autoregressions With Revision

		Errors for Output <sup>(*)</sup>					
<i>Model</i>	<i>RevErr</i>	<i>k = 2</i>	<i>k = 3</i>	<i>k = 4</i>	<i>k = 6</i>	<i>k = 12</i>	<i>k = 24</i>
<i>Panel A: Mean Square Forecast Errors</i>							
<i>Begin Date of Forecast Period = 1970:1</i>							
A	$u_{C1}$	0.636	0.778	0.822	0.848	0.848	0.848
A	$u_{C2}$	0.729	0.883	0.932	0.962	0.967	0.961
A	$u_{C3}$	0.681	0.830	0.873	0.889	0.887	0.883
A	$u_{C4}$	0.645	0.783	0.827	0.862	0.868	0.873
C	$u_{C1}$	0.656	0.798	0.828	0.845	0.841	0.852
C	$u_{C2}$	0.781	0.945	0.986	1.023	1.031	1.028
C	$u_{C3}$	0.618	0.749	0.781	0.827	0.804	0.818
C	$u_{C4}$	0.659	0.811	0.846	0.886	0.892	0.902
<i>Begin Date of Forecast Period = 1983:1</i>							
A	$u_{C1}$	0.220	0.268	0.278	0.310	0.332	0.364
A	$u_{C2}$	0.226	0.274	0.284	0.312	0.335	0.365
A	$u_{C3}$	0.216	0.263	0.274	0.304	0.322	0.354
A	$u_{C4}$	0.224	0.272	0.282	0.313	0.336	0.371
C	$u_{C1}$	0.247	0.301	0.307	0.335	0.366	0.393
C	$u_{C2}$	0.245	0.300	0.306	0.333	0.365	0.392
C	$u_{C3}$	0.230	0.285	0.290	0.321	0.337	0.370
C	$u_{C4}$	0.260	0.316	0.324	0.350	0.389	0.415
<i>Begin Date of Forecast Period = 1990:1</i>							
A	$u_{C1}$	0.184	0.211	0.222	0.286	0.295	0.342
A	$u_{C2}$	0.186	0.212	0.224	0.288	0.297	0.343
A	$u_{C3}$	0.176	0.205	0.217	0.279	0.288	0.332
A	$u_{C4}$	0.185	0.213	0.224	0.288	0.300	0.351
C	$u_{C1}$	0.206	0.238	0.241	0.306	0.324	0.366
C	$u_{C2}$	0.201	0.233	0.237	0.303	0.321	0.365
C	$u_{C3}$	0.193	0.227	0.231	0.293	0.302	0.348
C	$u_{C4}$	0.215	0.246	0.250	0.321	0.340	0.383
<i>Panel B: Diebold-Mariano Test Statistics Corresponding to Entries in Panel A</i>							
<i>Begin Date of Forecast Period = 1970:1</i>							
A	$u_{C1}$	0.277	0.209	0.111	-0.260	-0.335	-0.445
A	$u_{C2}$	-0.889	-0.881	-0.872	-0.863	-0.889	-0.915
A	$u_{C3}$	-1.748	-1.744	-1.646	-1.485	-1.418	-1.448
A	$u_{C4}$	-0.099	-0.001	-0.070	-0.445	-0.651	-0.814
C	$u_{C1}$	-0.217	-0.207	-0.044	-0.058	-0.034	-0.232
C	$u_{C2}$	-1.410	-1.438	-1.315	-1.294	-1.320	-1.381
C	$u_{C3}$	0.353	0.453	0.567	0.208	0.488	0.276
C	$u_{C4}$	-0.373	-0.543	-0.404	-0.815	-0.983	-1.224
<i>Begin Date of Forecast Period = 1983:1</i>							
A	$u_{C1}$	-1.084	-0.958	-0.884	-0.742	-1.118	-0.978
A	$u_{C2}$	-1.539	-1.404	-1.308	-0.853	-1.154	-1.022
A	$u_{C3}$	-1.309	-1.304	-1.277	-0.302	-0.104	-0.111
A	$u_{C4}$	-1.180	-1.045	-0.974	-0.851	-1.151	-1.253
C	$u_{C1}$	-1.284	-1.369	-1.200	-1.058	-1.362	-1.154
C	$u_{C2}$	-1.248	-1.342	-1.167	-1.018	-1.369	-1.174
C	$u_{C3}$	-0.954	-1.267	-0.996	-0.891	-0.704	-0.688
C	$u_{C4}$	-1.448	-1.544	-1.469	-1.214	-1.594	-1.430
<i>Begin Date of Forecast Period = 1990:1</i>							
A	$u_{C1}$	-1.720	-1.358	-1.277	-1.276	-1.299	-1.295
A	$u_{C2}$	-1.583	-1.263	-1.210	-1.149	-1.213	-1.162
A	$u_{C3}$	-0.466	-0.455	-0.382	-0.168	-0.426	-0.167
A	$u_{C4}$	-1.412	-1.042	-0.904	-1.031	-1.141	-1.360
C	$u_{C1}$	-1.084	-1.252	-0.965	-0.867	-1.041	-0.935
C	$u_{C2}$	-1.022	-1.195	-0.890	-0.833	-1.024	-0.978
C	$u_{C3}$	-0.791	-1.025	-0.685	-0.598	-0.569	-0.551
C	$u_{C4}$	-1.051	-1.178	-0.976	-0.918	-1.072	-1.003

(\*) See notes to Table 3. Revision errors included as additional regressors in the prediction equations reported on in Table 3 are:  $u_{C1} = {}_{t+1}u_{t-1}^t$ ,  $u_{C2} = ({}_{t+1}u_{t-1}^t, {}_{t+1}u_{t-2}^{t-1})$ ,  $u_{C3} = {}_{t+1}u_{t-2}^{t-1}$ , and  $u_{C4} = ({}_{t+1}u_{t-1}^t, {}_{t+1}u_{t-2}^t)$ . Further details are contained in Sections 3 and 4.

Table A1.2: MSFEs Calculated Based on Simple Real-Time Autoregressions With Revision

Errors for Prices <sup>(*)</sup>							
<i>Model</i>	<i>RevErr</i>	<i>k = 2</i>	<i>k = 3</i>	<i>k = 4</i>	<i>k = 6</i>	<i>k = 12</i>	<i>k = 24</i>
<i>Panel A: Mean Square Forecast Errors</i>							
<i>Begin Date of Forecast Period =1970:1</i>							
A	<i>u<sub>C1</sub></i>	0.145	0.163	0.164	0.161	0.136	0.124
A	<i>u<sub>C2</sub></i>	0.143	0.161	0.162	0.160	0.136	0.121
A	<i>u<sub>C3</sub></i>	0.153	0.170	0.170	0.167	0.141	0.124
A	<i>u<sub>C4</sub></i>	0.150	0.166	0.167	0.166	0.139	0.122
C	<i>u<sub>C1</sub></i>	0.155	0.174	0.177	0.173	0.146	0.132
C	<i>u<sub>C2</sub></i>	0.164	0.183	0.187	0.185	0.155	0.137
C	<i>u<sub>C3</sub></i>	0.159	0.178	0.181	0.178	0.150	0.136
C	<i>u<sub>C4</sub></i>	0.156	0.174	0.177	0.175	0.148	0.133
<i>Begin Date of Forecast Period =1983:1</i>							
A	<i>u<sub>C1</sub></i>	0.072	0.076	0.075	0.069	0.063	0.063
A	<i>u<sub>C2</sub></i>	0.070	0.073	0.073	0.069	0.064	0.061
A	<i>u<sub>C3</sub></i>	0.072	0.076	0.076	0.070	0.064	0.064
A	<i>u<sub>C4</sub></i>	0.073	0.076	0.076	0.071	0.062	0.060
C	<i>u<sub>C1</sub></i>	0.076	0.081	0.081	0.076	0.069	0.068
C	<i>u<sub>C2</sub></i>	0.077	0.081	0.082	0.080	0.074	0.069
C	<i>u<sub>C3</sub></i>	0.080	0.084	0.083	0.079	0.071	0.069
C	<i>u<sub>C4</sub></i>	0.078	0.082	0.081	0.077	0.070	0.067
<i>Begin Date of Forecast Period =1990:1</i>							
A	<i>u<sub>C1</sub></i>	0.057	0.046	0.050	0.049	0.041	0.042
A	<i>u<sub>C2</sub></i>	0.053	0.043	0.048	0.048	0.042	0.043
A	<i>u<sub>C3</sub></i>	0.058	0.047	0.051	0.050	0.041	0.043
A	<i>u<sub>C4</sub></i>	0.057	0.045	0.050	0.049	0.041	0.044
C	<i>u<sub>C1</sub></i>	0.063	0.053	0.055	0.055	0.045	0.045
C	<i>u<sub>C2</sub></i>	0.060	0.050	0.056	0.056	0.048	0.048
C	<i>u<sub>C3</sub></i>	0.066	0.055	0.057	0.057	0.046	0.046
C	<i>u<sub>C4</sub></i>	0.065	0.054	0.056	0.057	0.046	0.047
<i>Panel B: Diebold-Mariano Test Statistics Corresponding to Entries in Panel A</i>							
<i>Begin Date of Forecast Period =1970:1</i>							
A	<i>u<sub>C1</sub></i>	0.397	0.284	0.247	0.185	0.347	0.110
A	<i>u<sub>C2</sub></i>	0.592	0.567	0.478	0.279	0.378	0.486
A	<i>u<sub>C3</sub></i>	-1.086	-0.862	-0.846	-0.902	-0.472	0.132
A	<i>u<sub>C4</sub></i>	-0.127	0.033	-0.033	-0.241	0.087	0.292
C	<i>u<sub>C1</sub></i>	-0.480	-0.619	-0.702	-0.765	-0.618	-0.621
C	<i>u<sub>C2</sub></i>	-1.074	-1.141	-1.234	-1.387	-1.260	-0.904
C	<i>u<sub>C3</sub></i>	-0.792	-0.930	-0.961	-1.076	-1.007	-0.908
C	<i>u<sub>C4</sub></i>	-0.563	-0.670	-0.756	-0.857	-0.732	-0.676
<i>Begin Date of Forecast Period =1983:1</i>							
A	<i>u<sub>C1</sub></i>	0.070	0.455	0.317	0.219	0.064	0.653
A	<i>u<sub>C2</sub></i>	0.665	0.957	0.741	0.213	-0.427	0.822
A	<i>u<sub>C3</sub></i>	-0.440	-1.109	-1.249	-1.680	-1.528	-1.568
A	<i>u<sub>C4</sub></i>	-0.217	0.131	-0.020	-0.325	0.292	0.599
C	<i>u<sub>C1</sub></i>	-1.151	-1.301	-1.242	-1.610	-1.337	-0.976
C	<i>u<sub>C2</sub></i>	-1.062	-0.878	-1.454	-2.333	-2.420	-1.502
C	<i>u<sub>C3</sub></i>	-1.914	-1.866	-1.730	-2.235	-1.705	-1.158
C	<i>u<sub>C4</sub></i>	-1.474	-1.400	-1.355	-1.677	-1.529	-0.936
<i>Begin Date of Forecast Period =1990:1</i>							
A	<i>u<sub>C1</sub></i>	0.614	0.445	0.569	0.370	-0.176	1.601
A	<i>u<sub>C2</sub></i>	0.836	0.861	0.723	0.241	-0.414	-0.007
A	<i>u<sub>C3</sub></i>	-0.613	-0.991	-1.176	-1.751	-1.665	-1.380
A	<i>u<sub>C4</sub></i>	0.156	0.367	0.221	-0.109	-0.206	-0.339
C	<i>u<sub>C1</sub></i>	-1.316	-1.496	-0.902	-1.107	-0.794	-0.561
C	<i>u<sub>C2</sub></i>	-0.600	-0.640	-1.257	-1.503	-2.111	-1.646
C	<i>u<sub>C3</sub></i>	-1.856	-1.951	-1.279	-1.341	-0.962	-0.672
C	<i>u<sub>C4</sub></i>	-1.656	-1.768	-1.186	-1.409	-1.091	-0.870

(\*) See notes to Table 3.

Table A1.3: MSFEs Calculated Based on Simple Real-Time Autoregressions With Revision

Errors for Money <sup>(*)</sup>							
<i>Model</i>	<i>RevErr</i>	<i>k = 2</i>	<i>k = 3</i>	<i>k = 4</i>	<i>k = 6</i>	<i>k = 12</i>	<i>k = 24</i>
<i>Panel A: Mean Square Forecast Errors</i>							
<i>Begin Date of Forecast Period =1970:1</i>							
A	<i>u<sub>C1</sub></i>	1.035	1.030	1.013	0.937	0.982	0.992
A	<i>u<sub>C2</sub></i>	1.030	1.026	1.012	0.934	0.979	0.992
A	<i>u<sub>C3</sub></i>	1.030	1.024	1.008	0.935	0.983	1.001
A	<i>u<sub>C4</sub></i>	1.043	1.038	1.020	0.944	0.988	1.001
C	<i>u<sub>C1</sub></i>	1.037	1.029	1.014	0.935	0.961	0.979
C	<i>u<sub>C2</sub></i>	1.044	1.036	1.023	0.942	0.969	0.991
C	<i>u<sub>C3</sub></i>	1.040	1.030	1.012	0.933	0.963	0.983
C	<i>u<sub>C4</sub></i>	1.075	1.067	1.050	0.968	0.992	1.001
<i>Begin Date of Forecast Period =1983:1</i>							
A	<i>u<sub>C1</sub></i>	1.161	1.135	1.093	0.992	1.010	1.033
A	<i>u<sub>C2</sub></i>	1.147	1.121	1.080	0.984	1.002	1.031
A	<i>u<sub>C3</sub></i>	1.157	1.130	1.089	0.998	1.011	1.036
A	<i>u<sub>C4</sub></i>	1.163	1.137	1.095	0.994	1.012	1.035
C	<i>u<sub>C1</sub></i>	1.117	1.089	1.043	0.948	0.958	0.984
C	<i>u<sub>C2</sub></i>	1.112	1.084	1.039	0.947	0.956	0.987
C	<i>u<sub>C3</sub></i>	1.119	1.089	1.042	0.950	0.960	0.985
C	<i>u<sub>C4</sub></i>	1.121	1.093	1.046	0.951	0.960	0.987
<i>Begin Date of Forecast Period =1990:1</i>							
A	<i>u<sub>C1</sub></i>	1.064	1.065	1.042	0.905	0.897	0.950
A	<i>u<sub>C2</sub></i>	1.051	1.051	1.028	0.892	0.884	0.937
A	<i>u<sub>C3</sub></i>	1.050	1.049	1.025	0.907	0.896	0.948
A	<i>u<sub>C4</sub></i>	1.068	1.069	1.046	0.907	0.900	0.953
C	<i>u<sub>C1</sub></i>	1.028	1.028	1.002	0.875	0.861	0.916
C	<i>u<sub>C2</sub></i>	1.021	1.020	0.996	0.870	0.857	0.911
C	<i>u<sub>C3</sub></i>	1.028	1.025	0.997	0.877	0.864	0.918
C	<i>u<sub>C4</sub></i>	1.031	1.031	1.005	0.877	0.862	0.918
<i>Panel B: Diebold-Mariano Test Statistics Corresponding to Entries in Panel A</i>							
<i>Begin Date of Forecast Period =1970:1</i>							
A	<i>u<sub>C1</sub></i>	-0.282	-0.344	-0.287	0.033	0.165	0.413
A	<i>u<sub>C2</sub></i>	-0.028	-0.075	-0.121	0.122	0.192	0.254
A	<i>u<sub>C3</sub></i>	-0.031	-0.003	0.030	0.227	0.100	-0.272
A	<i>u<sub>C4</sub></i>	-0.613	-0.615	-0.533	-0.318	-0.242	-0.162
C	<i>u<sub>C1</sub></i>	-0.221	-0.171	-0.183	0.085	0.728	0.597
C	<i>u<sub>C2</sub></i>	-0.334	-0.300	-0.370	-0.119	0.387	0.174
C	<i>u<sub>C3</sub></i>	-0.325	-0.207	-0.133	0.140	0.689	0.461
C	<i>u<sub>C4</sub></i>	-1.051	-1.016	-0.972	-0.785	-0.194	-0.075
<i>Begin Date of Forecast Period =1983:1</i>							
A	<i>u<sub>C1</sub></i>	0.245	0.197	0.224	0.685	0.553	0.339
A	<i>u<sub>C2</sub></i>	0.441	0.410	0.408	0.562	0.471	0.231
A	<i>u<sub>C3</sub></i>	0.506	0.510	0.510	0.441	0.500	0.211
A	<i>u<sub>C4</sub></i>	0.169	0.123	0.145	0.610	0.462	0.256
C	<i>u<sub>C1</sub></i>	1.199	1.214	1.350	1.636	1.695	1.527
C	<i>u<sub>C2</sub></i>	1.035	1.046	1.137	1.284	1.308	1.154
C	<i>u<sub>C3</sub></i>	1.303	1.384	1.552	1.656	1.728	1.603
C	<i>u<sub>C4</sub></i>	1.106	1.121	1.258	1.548	1.627	1.441
<i>Begin Date of Forecast Period =1990:1</i>							
A	<i>u<sub>C1</sub></i>	-0.162	-0.188	-0.183	0.357	0.325	0.150
A	<i>u<sub>C2</sub></i>	0.185	0.175	0.162	0.758	0.703	0.523
A	<i>u<sub>C3</sub></i>	0.649	0.699	0.714	0.786	0.821	0.556
A	<i>u<sub>C4</sub></i>	-0.255	-0.280	-0.279	0.262	0.219	0.060
C	<i>u<sub>C1</sub></i>	0.767	0.743	0.801	1.353	1.429	1.111
C	<i>u<sub>C2</sub></i>	0.987	0.982	1.001	1.574	1.577	1.292
C	<i>u<sub>C3</sub></i>	0.946	1.032	1.158	1.328	1.424	1.199
C	<i>u<sub>C4</sub></i>	0.674	0.649	0.714	1.245	1.337	1.020

(\*) See notes to Table 3.

## Appendix 2 - Evidence from Extant Tests In the Literature on Rationality

Table A2.1: Rationality Test Results<sup>(\*)</sup>

Output	Linear Regression Test		
	F-Test Statistic	5% CV	10% CV
	2.893	3.053	2.336
	CCS Test & Chi Square Critical Values		
	Statistic	5% CV	10% CV
	0.000078	0.000044	0.000030
	CFS Test Statistic & Bootstrap Critical Values		
	Statistic	5%	10%
	0.023944	0.011842	0.009845
Prices	Linear Regression Test		
	F-Test Statistic	5% CV	10% CV
	1.851	3.053	2.336
	CCS Test & Chi Square Critical Values		
	Statistic	5% CV	10% CV
	0.000046	0.000030	0.000026
	CFS Test Statistic & Bootstrap Critical Values		
	Statistic	5% CV	10% CV
	0.056355	0.043220	0.034450
Money	Linear Regression Test		
	F-Test Statistic	5% CV	10% CV
	-0.52381	3.053	2.336
	CCS Test & Chi Square Critical Values		
	Statistic	5% CV	10% CV
	0.000021	0.000042	0.000032
	CFS Test Statistic & Bootstrap Critical Values		
	Statistic	5% CV	10% CV
	0.002549	0.015691	0.011907

<sup>(\*)</sup> For the linear regression test, we fit the regression  ${}_{t+2}u_t^{t+1} = \gamma_1 + \gamma_2 {}_{t+1}X_t + \gamma_3 {}_{t+1}u_{t-1}^t + \varepsilon_t$  and we report the F-statistic associated with the null hypothesis that  $\gamma_1 = \gamma_2 = \gamma_3 = 0$ . The CCS test reported on in this table is the Chao, Corradi and Swanson (2001) bivariate test with associated  $\chi_2^2$  critical values (see Corollary 2 of Chao et al. for details). The CFS test reported on in this table is  $M_{1,T}$ , where the null hypothesis of data rationality is defined as  $H_{0,1}$  (see Section 3 for further details).