Information in the Revision Process of Real-Time Datasets*

Valentina Corradi[1], Andres Fernandez[2], and Norman R. Swanson[2]
[1]University of Warwick
[2]Rutgers University

July 2007

Abstract

In this paper we first develop two statistical tests of the null hypothesis that early release data are rational. The tests are consistent against generic nonlinear alternatives, and are conditional moment type tests, in the spirit of Bierens (1982,1990), Chao, Corradi and Swanson (2001) and Corradi and Swanson (2002). We then use this test, in conjunction with standard regression analysis in order to individually and jointly analyze a real-time dataset for money, output, prices and interest rates. All of our empirical analysis is carried out using various variable/vintage combinations, allowing us to comment not only on rationality, but also on a number of other related issues. For example, we discuss and illustrate the importance of the choice between using first, later, or mixed vintages of data in prediction. Interestingly, it turns out that early release data are generally best predicted using first releases. The standard practice of using "mixed vintages" of data appears to always yield poorer predictions, regardless of what we term "definitional change problems" associated with using only first releases for prediction. Furthermore, we note that our tests of first release rationality based on ex ante prediction find no evidence that the data rationality null hypothesis is rejected for a variety of variables (i.e. we find strong evidence in favor of the "news" hypothesis). Thus, it appears that there is little benefit to using later releases of data for prediction and policy analysis, for example. Additionally, we argue that the notion of $final$ data is misleading, and that definitional and other methodological changes that pepper real-time datasets are important. Finally, we carry out an empirical example, where little evidence that money has marginal predictive content for output is found, regardless of whether various revision error variables are added to standard vector autoregression models of money, output, prices and interest rates.

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 . Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. The paper has been prepared for the conference entitled "Real-Time Data Analysis and Methods in Economics", held at the Federal Reserve Bank of Philadelphia, and the authors are grateful to seminar participants for many useful comments on early versions of this paper. Corradi gratefully acknowledges ESRC grant RES-062-23-0311, and Swanson acknowledges financial support from a Rutgers University Research Council grant.

1 Introduction

The literature on testing for and assessing the rationality of early release economic data is rich and deep. A very few of the most recent papers on the subject include Howrey (1978), Mankiw, Runkle, and Shapiro (1984), Mankiw and Shapiro (1986), Milbourne and Smith (1989), Keane and Runkle (1989,1990), Kennedy (1993), Kavajecz and Collins (1995), Faust, Rogers, and Wright (2005), Swanson and van Dijk (2006), and the papers cited therein.[1] In recent years, in large part due to the notable work of Diebold and Rudebusch (1991), many agencies have begun to collect and disseminate real-time data that are useful for examining data rationality, and more generally for examining the entire revision history of a variable. This has led to much renewed interest in the area of real-time data analysis. However, relatively few papers have examined issues such as data rationality using tests other than ones based on fairly simplistic models such linear regressions.[2]

In this paper, we add to the literature on whether preliminary (and later releases) of a given variable are rational by outlining two consistent out-of-sample tests for nonlinear rationality that are consistent against generic (non)linear alternatives. One test can be used to determine whether preliminary data are rational, in the sense that subsequent data releases simply take "news" into account that are not available at the time of initial release (see Mankiw and Shapiro (1986) and Faust, Rogers and Wright (2005) for further explanation of the "news" hypothesis). In this case, we have evidence that the first release data are those that should be used in out-of-sample prediction. If this test fails, then two additional testing procedures are proposed. The first involves simply using later release data in the same test to estimate the timing of data rationality (e.g. to estimate how many months it takes before data are rational). The second involves carrying out a different test which is designed to determine whether the data irrationality arises (i) simply because of a bias in the preliminary estimate, in which case the preliminary data should be adjusted by including an estimate of the bias prior to its use in out of sample prediction; or (ii) because available information has been used inefficiently when constructing first release data. The tests are related to the conditional moment tests of Bierens (1982,1990), de Jong (1996), and Corradi and Swanson (2002); and we provide asymptotic theory necessary for the valid implementation

[1]A small subset of other papers that examine data rationality, either explicitly, or implicitly via the analysis of real-time predictions, include Burns and Mitchell (1946), Muth (1961), Morgenstern (1963), Stekler (1967), Pierce (1981), Shapiro and Watson (1988), Diebold and Rudebusch (1991), Neftçi and Theodossiou (1991), Brodsky and Newbold (1994), Mariano and Tanizaki (1995), 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), and Bernanke and Boivin (2003).

[2]For a discussion of nonlinearity and data rationality, the reader is referred to Brodsky and Newbold (1994), Rathjens and Robins (1995), and Swanson and van Dijk (2006).

of the statistics using critical values constructed via a block (likelihood) bootstrap related to the bootstrap procedures discussed in Corradi and Iglesis (2007) and Corradi and Swanson (2007). Of final note is that in addition to being based on linear models, many of the regression based tests used to examine rationality in the literature are actually "in-sample", in the sense that "true" prediction experiments are not carried out when testing rationality. We feel that this is a crucial issue, and is analogous to the issue of whether the "correct" way to test for Granger causality is to use in-sample tests or ex ante out-of-sample predictive accuracy assessments. Our tests are truly out-of-sample; and hence quite different from prior tests of rationality such as those discussed and implemented in the keynote paper written by Mankiw and Shapiro (1986), and subsequent related papers.

In addition to outlining and implementing new rationality tests, we carry out a series of prediction experiments in this paper. One feature that ties together our prediction experiments and our data rationality tests is that we take no stand on what constitutes $final$ data, as we argue that the very notion of using so-called $final$ data in such analyses is fraught with problems. In particular, data are never truly $final$, as they are subject to an indefinite number of revisions. More importantly, it is not clear that financial market participants, government policy setters, and generic economic agents use final data in day to day decision making. For example, financial markets certainly react to first release data, and also to a small number of subsequent data revisions. However, they pretty clearly do not react to revisions to data from 10 or 20 years ago, say. Similarly, policy setters base their decisions to some extent on predictions of near-term early release target variables such as inflation, unemployment, and output growth. They clearly do not base their decisions upon predictions of $final$ versions of these data. The reason for this is that early release data are revised not only because of the collection of additional information relevant to the time period in question, but also because of "definitional changes". Indeed, distant revisions are due primarily to definitional and other structural data collection methodology changes; and it goes without saying that policy setters, for example, have very little premonition of definitional changes that will occur 10 or 20 years in the future, and indeed do not care about such. Of course, one might argue, and perhaps in certain contexts should argue that there exists some "true underlying" final definition of a variable; and definitional changes represent incremental efforts made by statistical agencies to converge to this definition. Moreover, it is not clear that empirical analysts should simply make level adjustments to series in order to allow for definitional changes - this is at best a crude solution to the problem (even if the raw data are transformed to growth rates). Instead, one might argue that a whole new series should be formed after each definitional change.

The above arguments, taken in concert, imply that the variable that one cares about predicting may in many cases be a near-term vintage; and if one cares about $final$ data, one may have a very difficult time, as one would need then to predict unknowable future definitional and other methodological changes implicit to the data construction process. Regardless of one's view on this matter, though, there are clearly instances where one may be interested in predicting first release data, and one of the objectives of this paper is to address how best to do this. Furthermore, it is pretty clear that much caution needs to be used when analyzing real-time data. For example, application of simple level shifts to datasets in order to deal with "definitional changes" may be grossly 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, say, may result in an entirely different time series that based on an earlier definition. This issue is examined in the sequel via discussion of a simple graphical method for determining "definitional changes", and via careful examination of the statistical properties of money, output, and prices for different sample periods corresponding to "definitional changes".

The preceding arguments are not meant to suggest that the standard approach in prediction exercises of using all available information when parameterizing models and subsequently when constructing predictions is invalid. Indeed, given that definitional changes may indeed make early releases of data from the distant past unreliable, when viewed in concert with early releases of very recent data, it still makes sense to use the most recently available data (as opposed to only using preliminary data, say) in prediction. However, it should nevertheless be stressed that the most recent data that are available for a particular variable consist of a series of observations, each of which is a different vintage. Thus, the standard econometric approach is to use a mixture of different vintages. This in turn poses a different sort of potential problem. Namely, if the objective is to construct predictions of early data vintages, then why use an arbitrarily large number of different data vintages when parameterizing prediction models? This is one of the issues addressed in the empirical part of this paper, where real-time prediction experiments are carried out using various vintages of data as the target variable to be predicted, and using various vintages and combinations of vintages, and revision errors in the information set from which the model is parameterized. For example, we investigate whether k^{th} available data are better predicted using all available information (i.e. using the latest real-time data), or using only past k^{th} available observations. This approach allows us to answer the following sort of question. If the objective is prediction of preliminary data, then do any advantages associated with using only past first release data to make such predictions outweigh the costs associated with the fact that distant first release data are clearly subject to the "definitional change" problems discussed above; when compared with predictions constructed using real-time data which are subject to the criticism that noise is added to the regression because many different vintages are used in model parameterization? As might be expected, the answer to the above question in part depends upon how severe the definitional changes are, and in part on how much history is used in our experiments; and hence how many definitional changes are allowed to contaminate our first release time series.

We also investigate whether revision errors can be used to improve predictions. Furthermore, we carry out a real-time assessment of the marginal predictive content of money for income using various vintage/variable combinations of output, money, prices, and interest rates.

Interestingly, it turns out that early release data are best predicted using first releases for all of the variables we examine, *even though we ignore definitional changes in our preliminary data series.* This suggests that the standard practice of using "mixed vintages" of data when constructing early release predictions is not optimal, in a predictive accuracy sense. Moreover, there appears to be no evidence of early release data irrationality, when our truly ex ante rationality tests are implemented; further supporting the notion that first release data should be used for forming predictions of early release data, and hence in subsequent policy analysis based upon such predictions. In support of the above findings, evidence is presented supporting the conclusion that adding revisions to simple autoregressive type prediction models does not lead to improved predictive performance. Finally, we find that there appears to be little marginal predictive content of money for output. This result holds for standard real-time vector autoregressions (VARs), and for VARs that are augmented with various revision error variables.

The rest of the paper is organized as follows. In Section 2, we outline some notation, and briefly explain the ideas behind testing for data rationality. In Section 3, we outline our nonlinear rationality tests, and in Section 4 we outline the empirical methodology used in our prediction experiments. Section 5 contains a discussion of the data used, and empirical results are gathered in Section 6. Concluding remarks are given in Section 7. All proofs are gathered in an appendix.

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 $k \to \infty$, 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.

The topic of testing rationality of preliminary data announcements is discussed in detail by Mankiw and Shapiro (1986), Keane and Runkle (1989,1990), Swanson and van Dijk (2006), and many others.[3] The notion of rationality can most easily be explained by considering Muth's (1961) definition of rational expectations, where the preliminary release ${}_{t+1}X_t$ is a rational forecast of the final data ${}_fX_t$ if and only if:

$$ {}_{t+1}X_t = E[{}_fX_t|\mathcal{F}_t^{t+1}], \tag{1} $$

where $\mathcal{F}_t^{t+1}$ contains all information available at the time of release of ${}_{t+1}X_t$ (see below for further discussion). This equation can be examined via use of the following regression model:

$$ {}_fX_t = \alpha + {}_{t+1}X_t\,\beta + {}_{t+1}W_t'\gamma + \varepsilon_{t+1}, \tag{2} $$

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 ε_{t+1} is an error term assumed to be uncorrelated with ${}_{t+1}X_t$ and ${}_{t+1}W_t$. The null hypothesis of interest (i.e. that of rationality) in this model is that $\alpha = 0$, $\beta = 1$, and $\gamma = 0$, and corresponds to the idea of testing for the rationality of ${}_{t+1}X_t$ for ${}_fX_t$ by finding out whether the conditioning information in ${}_{t+1}W_t$, available in real-time to the data issuing agency could have been used to construct better conditional predictions of final data.

Based on regressions in the spirit of the above model, and on an examination of preliminary and final money stock data, Mankiw, Runkle, and Shapiro (1984) find evidence against the null that $\alpha = 0$, $\beta = 1$, and $\gamma = 0$ in (2), suggesting that preliminary money stock announcements are not rational. On the other hand, Kavajecz and Collins (1995) find that seasonally unadjusted money announcements are rational while adjusted ones are not. For GDP data, Mankiw and Shapiro (1986) find little evidence against the null hypothesis of rationality, while Mork (1987) and Rathjens and Robins (1995) find evidence of irrationality, particularly in the form of prediction bias (i.e. $\alpha \neq 0$ in (2)). Keane and Runkle (1990) examine the rationality of survey price forecasts rather than preliminary (or real-time) data, using the novel approach of constructing panels of real-time survey predictions. This allows them to avoid aggregation bias, for example, and may be one of the

[3] For a very clear discussion of some approaches used to test for rationality, see also Faust, Rogers, and Wright (2005), where the errors-in-variables and rational forecast models are used to discuss the notions of "noise" and "news", respectively.

reasons why they find evidence supporting rationality, even though previous studies focusing on price forecasts had found evidence to the contrary. Swanson and van Dijk (2006) consider the entire revision history for each variable, and hence discuss the "timing" of data rationality by generalizing (2) as follows:

$$ {}_{f}X_t - {}_{t+k}X_t = \alpha + {}_{t+k}X_t\,\beta + {}_{t+k}W'_{t+k-1}\gamma + \varepsilon_{t+k} \tag{3} $$

and

$$ {}_{t+k}X_t - {}_{t+1}X_t = \alpha + {}_{t+1}X_t\,\beta + {}_{t+1}W'_{t}\gamma + \varepsilon_{t+k}, \tag{4} $$

where $k = 1, 2, \ldots$ defines the release (or vintage) of data (that is, for $k = 1$ we are looking at preliminary data, for $k = 2$ the data have been revised once, etc.). The primary objective of fitting the second of these two regression models is to assess whether there is information in the revision error between periods $t + k$ and $t + 1$ that could have been predicted when the initial estimate, ${}_{t+1}X_t$, was formed. Using this approach, Swanson and van Dijk find that data rationality is most prevalent after 3 to 4 months, for unadjusted industrial production and producer prices.

In the sequel, we consider questions of data rationality similar to those discussed above. However, we do not assume linearity, as is implicit in the tests carried out in all of the papers discussed above. Additionally, we examine a number of empirical prediction models in order to assess the relative merits of using various vintages of real-time data, in the context of predicting a variety of different variable/vintage combinations (see above for further discussion).

3 Consistent Out of Sample Tests for Rationality

3.1 The Framework

As mentioned above, a common approach in the literature has been to use linear regressions in order to test for rationality. Therefore, failure to reject the null equates with an absence of linear correlation between the revision error and information available at the time of the first data release. It follows that these tests do not necessarily detect nonlinear dependence. Our objective in this section is to provide a test for rationality which is consistent against generic nonlinear alternatives. In other words, we propose a test that is able to detect any form of dependence between the revision error and information available at the time of the first data release. This is accomplished by constructing conditional moment tests which employ an infinite number of moment conditions (see e.g. Bierens (1982,1990), Bierens and Ploberger (1997), de Jong (1996), Hansen (1996a), Lee, Granger and White (1993) and Stinchcombe and White (1998), Corradi and Swanson (2002)).

To set notation, let ${}_{t+2}u_t^{t+1} = {}_{t+2}X_t - {}_{t+1}X_t$, and ${}_{t+1}W_t = \left({}_{t+1}X_t, {}_{t+1}u_t^{t+1}\right)$. Furthermore, let $\mathcal{F}_t^{t+1} = \sigma\left({}_{s+1}W_s;\ 1 \leq s \leq t\right)$. Thus, ${}_{t+2}u_t^{t+1}$ denotes the error between the 2^{nd} and the 1^{st} releases, and $\mathcal{F}_t^{t+1}$ contains information available at the time of the first release, assuming a one month lag before the first datum becomes available. All of our results generalize immediately to the case where ${}_{t+1}W_t$ contains information other than ${}_{t+1}X_t$ and ${}_{t+1}u_t^{t+1}$. Furthermore, even though our discussion focusses primarily on ${}_{t+2}u_t^{t+1}$, our results additionally generalize immediately to ${}_{t+k}u_t^{t+j}$ for $k \geq 2$, $j \geq 1$, and $k > j$.

In the sequel, consider testing the following hypotheses, against their respective negations:
$H_{0,1} : E\left({}_{t+2}u_t^{t+1}|\mathcal{F}_t^{t+1}\right) = 0$, and
$H_{0,2} : E\left({}_{t+2}u_t^{t+1}|\mathcal{F}_t^{t+1}\right) = E\left({}_{t+2}u_t^{t+1}\right)$.
Hypothesis $H_{0,1}$ is the null hypothesis that the first release is rational, as the revision error in this case is a martingale difference process, 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 news that were not available at the time of the first release into account. Thus, if we fail to reject $H_{0,1}$, it means that the first data release already incorporates all available information at the current time. Furthermore, in this case we might expect that first release data are best predicted using historical first release data rather than using mixed vintages of data (as is currently the standard practice). See Section 5.2 for further discussion of this point.

In principle, one might imagine forming a joint test for the null hypothesis that $E\left({}_{t+k}u_t^{t+1}|\mathcal{F}_t^{t+1}\right) = 0$, for $k = 2, ..., \overline{k}$, where ${}_{t+k}u_t^{t+1}$ denotes the revision error between the k^{th} and the 1^{st} releases. However, under the null of rationality, for $k > 2$, ${}_{t+k}u_t^{t+1}$ is perfectly correlated with ${}_{t+2}u_t^{t+1}$, as:

$$E\left({}_{t+k}u_t^{t+1}\ {}_{t+2}u_t^{t+1}\right) = E\left(\left({}_{t+2}u_t^{t+1}\right)^2\right) + \sum_{j=3}^{k} E\left({}_{t+j}u_t^{t+1}\ {}_{t+2}u_t^{t+1}\right) = E\left(\left({}_{t+2}u_t^{t+1}\right)^2\right),$$

which in turn follows because the revision error is uncorrelated, under the null hypothesis. In this sense, by considering additional revision errors, one gains no further information. Therefore, a test statistic for the joint null $E\left({}_{t+k}u_t^{t+1}|\mathcal{F}_t^{t+1}\right) = 0$, for $k = 2, ..., \overline{k}$ will be characterized by a degenerate limiting distribution. On the other hand, one can certainly use ${}_{t+k}u_t^{t+k-1}$, $k \geq 1$ in place of ${}_{t+2}u_t^{t+1}$ in $H_{0,1}$. Indeed, by sequentially testing $H_{0,1}$ using increasing values of k, one can estimate which release of the variable of interest is the first one that fails to reject, and is hence rational.

Hypothesis $H_{0,2}$ also forms the basis for a rationality test, because rationality entails that the revision error is indeed independent of any function which is measurable in terms of information available at time $t+1$. Nevertheless, the first release may be a biased estimator of the second release. In this sense, the first release would be unconditionally biased. Unconditional bias may arise due to the fact that the statistical reporting agency produces releases according to an asymmetric loss function. For example, there may be a preference for a pessimistic release in the first stage, followed by a more optimistic one in the later stage (see e.g. Swanson and Van Dijk (2006) for further discussion). Intuitively, one might argue that the cost of a downward readjustment of the preliminary data is higher than the cost of an upward adjustment, say.

Our first objective is to provide a test for $H_{0,1}$, which is consistent against all possible deviations from the null. Now, failure to reject $H_{0,1}$ would clearly suggest that one should use first release data for out-of-sample prediction On the other hand, if $H_{0,1}$ is rejected, one remains with the problem of ascertaining the cause of the rejection. A logical next step would the be to construct a statistic for testing $H_{0,2}$ against its negation. If the null hypothesis fails to reject, then there is unconditional bias, but there is no issue of the inefficient use of available information. In this case, then, one should use the preliminary release plus the estimated mean of ${}_{t+2}u_t^{t+1}$ as an appropriately *adjusted* preliminary release in prediction and policy applications.

3.2 Test Statistics and Assumptions

Bierens (Theorem 1, (1990)) shows that if $E\left({}_{t+2}u_t^{t+1}|{}_{t+1}W_t\right) \neq 0$, then $E\left({}_{t+2}u_t^{t+1}\exp_{t+1} W_t'\gamma\right) \neq 0$, for all $\gamma \in \Gamma$, except a subset of zero Lebesgue measure. Stinchcombe and White (1998) show that if $w\left({}_{t+1}W_t', \gamma\right)$ is a generically comprehensive function, then whenever $\Pr\left(E\left({}_{t+2}u_t^{t+1}|{}_{t+1}W_t\right) = 0\right) < 1$, $E\left({}_{t+2}u_t^{t+1}w\left({}_{t+1}W_t', \gamma\right)\right) \neq 0$ for all $\gamma \in \Gamma$, except a subset of zero Lebesgue measure. In addition to the Bierens exponential function, the class of generically comprehensive functions includes the logistic function, and cumulative distribution functions in general. Suppose that ${}_{t+1}W_t$ is a $q-$dimensional vector (whenever ${}_{t+1}W_t$ does not have bounded support in $\mathcal{R}^q$, then it is customary to map the q elements of W_t into bounded subsets of $\mathcal{R}$). Examples of $w\left({}_{t+1}W_t', \gamma\right)$ include: $w\left({}_{t+1}W_t', \gamma\right) = \exp(\sum_{i=1}^q \gamma_i \Phi(W_{i,t}))$ and $w\left({}_{t+1}W_t', \gamma\right) = 1/(1+\exp(c - \sum_{i=1}^q \gamma_i \Phi(W_{i,t})))$, with $c \neq 0$ and Φ a measurable one to one mapping from $\Re$ to a bounded subset of $\Re$.

In our context, we want to test whether the revision error is independent of the entire history. Thus, we need to ensure that if:

$\Pr\left(E_{t+2}\left({}_{t+2}u_t^{t+1}|{}_{t+1}W_{t,t-1}W_t, \ldots, {}_2W_1\right) = 0\right) < 1$, then $E\left({}_{t+2}u_t^{t+1}w\left(\sum_{i=1}^t {}_{t+1-i}W_{t-i}'\gamma_i\right)\right) \neq 0$, for all $\gamma_i \in \Gamma$. In order to test

$$H_{0,1} : E\left({}_{t+2}u_t^{t+1}|\mathcal{F}_t^{t+1}\right) = 0, \text{ a.s.}$$

versus

$$H_{A,1} : \Pr\left(E\left({}_{t+2}u_t^{t+1}|\mathcal{F}_t^{t+1}\right) = 0\right) < 1,$$

we shall rely on the following statistic suggested by de Jong (1996):

$$M_{1,T} = \sup_{\gamma \in \Gamma} |m_{1,T}(\gamma)|, \tag{5}$$

where in our context:

$$m_{1,T}(\gamma) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T-2} {}_{t+2}u_t^{t+1} w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right),$$

and where Φ a measurable one to one mapping from $\Re^{t-1}$ to a bounded subset of $\Re$, and

$$\Gamma = \left\{\gamma_j : a_j \leq \gamma_j \leq b_j,\ i = 1, 2;\ |a_j|, |b_j| \leq Bj^{-k},\ k \geq 2\right\}. \tag{6}$$

As shown in Lemma 1 of de Jong, $(\Gamma, \|\gamma - \gamma'\|)$ is a compact metric space, with $\|\gamma - \gamma'\| = \left(\sum_{j=1}^{\infty} j^k |\gamma_j - \gamma'|^2\right)^{1/2}$, where $|\cdot|$ denotes the Euclidean norm in $\mathcal{R}^2$. In practice, one can allow for $k = 2$ and choose $a_j = aj^{-2}$ and $b_j = bj^{-2}$, where a and b belong to some compact set in $\mathcal{R}^2$. It is immediate to see the weight attached to past observations decreases over time. Indeed as stated in the assumptions below, the revision error is a short memory process, and therefore it is "independent" of its distant past, under both hypotheses.

As mentioned above, if $H_{0,1}$ is not rejected, then one can conclude that the revision error is unpredictable, and thus the first release data already incorporates available information in an efficient way. Thus, we can rely on the use of first release data both for forecasting and for policy evaluation. On the other hand, if $H_{0,1}$ is rejected, then it is important to distinguish between the case of inefficiency and (unconditional) bias. Thus, whenever $H_{0,1}$ is rejected, it remains to

$$H_{0,2} : E\left({}_{t+2}u_t^{t+1}|\mathcal{F}_t^{t+1}\right) = E\left({}_{t+2}u_t^{t+1}\right), \text{ a.s.}$$

versus

$$H_{A,2} : \Pr\left(E\left({}_{t+2}u_t^{t+1}|\mathcal{F}_t^{t+1}\right) = E\left({}_{t+2}u_t^{t+1}\right)\right) < 1.$$

Now, note that $m_{1,T}(\gamma)$ does not contain estimated parameters, so that there is no difference between in-sample and out-of-sample tests, when considering our test of $H_{0,1}$. This is no longer true when testing $H_{0,2}$ versus $H_{A,2}$, as implementation of the test requires the computation of the deviation from zero of the of revision error. In this case, we thus propose splitting the sample T , such that $T=R+P$, where only the last P observations are used for testing rationality. The mean is estimated recursively as:.

$$\widehat{\mu}_t=\frac{1}{t}\sum_{j=1}^{t}\ {}_{j+2}u_j^{j+1},\ \text{ for } t=R,...,R+P-2$$

It follows that the statistic of interest is:

$$M_{2,P}=\sup_{\gamma\in\Gamma}\left|m_{2,P}(\gamma)\right|,$$

where

$$m_{2,P}\left(\gamma\right)=\frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\ \left({}_{t+2}u_t^{t+1}-\widehat{\mu}_t\right)w\left(\sum_{j=1}^{t-1}\gamma_j'\Phi\left({}_{t+1-j}W_{t-j}\right)\right).$$

In the sequel, we require the following assumptions. We state the assumptions for the case in which ${}_{t+1}W_t\in\mathcal{R}^k$, so that if we set ${}_{t+1}W_t=\left({}_{t+1}X_{t},{}_{t+1}u_t^{t+1}\right)$, $k=2$.

Assumption A1: (i) The weights γ_j be defined as in (6). (ii) $\sum_{j=1}^{t-1}\gamma_j'\Phi\left({}_{t+1-j}W_{t-j}\right)$ is near epoch dependent (NED) of size -1 on ${}_{t+1-j}W_{t-j}$, where ${}_{t+1-j}W_{t-j}$ is strictly stationary and strong mixing sequence with size $-2r/(r-2)$, $r>2$, (iii) $E\left(\left({}_{t+2}u_t^{t+1}\right)^{2r}\right)<\infty$, and $E\left(w\left(\sum_{j=1}^{t-1}\gamma_j'\Phi\left({}_{t+1-j}W_{t-j}\right)\right)^{2r}\right)<\infty$, uniformly in γ, for $r>2$, and $E\left(\left({}_{t+2}u_t^{t+1}w\left(\sum_{j=1}^{t-1}\gamma_j'\Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right)\right)^2>0$, (iv) $\gamma_j'\Phi\left({}_{t+1-j}W_{t-j}\right)\in\mathcal{U}=(\underline{u},\overline{u})$, with $-\infty<\underline{u}<\overline{u}<\infty$, (v) $w\left(\cdot,\gamma\right)$ is a generically comprehensive function, such that $\sup_{\gamma\in\Gamma}\left\|w\left(\cdot,\gamma\right)\right\|_{q,2,\mathcal{U}}<\infty$, for $q>(k+1)/2$, where for each $\gamma\in\Gamma$, $\left\|w\left(\cdot,\gamma\right)\right\|_{q,2,\mathcal{U}}=\left(\sum_{|\alpha|\leq q}\int_{\mathcal{U}}\left|D^\alpha w\left(x,\gamma\right)\right|^2dx\right)^{1/2}$, with $D^\alpha w\left(x,\gamma\right)=\frac{\partial^{|\alpha|}w(x,\gamma)}{\partial x_1\times\ldots\times\partial x_k}$ and $|\alpha|=\sum_{l=1}^k\alpha_l$. (iv).

Remarks:

(i) Given **A1(i)**, it is immediate to see that $\sum_{j=1}^{t-1}\gamma_j'\Phi\left({}_{t+1-j}W_{t-j}\right)$ is NED of size -1 on ${}_{t+1-j}W_{t-j}$, and so **A1(ii)** follows. In fact, assuming for simplicity ${}_{t+1}W_t$ to be a scalar, letting $\mathcal{F}_s=\sigma\left({}_{i+1}W_i,\ i\leq s-1\right)$, $s<t$,

$$\begin{aligned}&\sup_{s<t\leq T-s}\left(E\left(\sum_{j=1}^{t-1}j^{-2}\Phi\left({}_{t+1-j}W_{t-j}\right)-E\left(\sum_{j=1}^{t-1}j^{-2}\Phi\left({}_{t+1-j}W_{t-j}\right)|\mathcal{F}_s\right)\right)^2\right)^{1/2}\\&\leq\sum_{j=s}^{T-1}j^{-2}\sup_t\left(E\left(\left|\Phi\left({}_{t+1}W_t\right)\right|^2\right)\right)^{1/2}\simeq O\left(s^{-1-\varepsilon}\right),\ \varepsilon>0,\end{aligned}$$

and the claim follows straigthforwardly from the definition of NED mapping, see e.g. Gallant and White (1988, Ch.4).

(ii) It is immediate to see that **A1(iv)** is satisfied if we choose Φ to be the atan function, as is customary in the consistent test literature (see e.g. Bierens (1982, 1990)).

(iii) The norm $\|w(\cdot)\|_{q,2,\mathcal{U}}$ defined in **A1(v)** is known as the Sobolev norm, and it ensures that the function has derivatives up to order q which are square integrable. This condition is satisfied for all q, by most of the test functions used in the consistent test literature, such as the exponential, the logistic, and cumulative distribution functions in general.

3.3 Asymptotics

We now state the limiting behavior of the two statistic suggested above.

Theorem 1: *Let Assumption A1 hold. Then: (i) Under* $H_{0,1}$, $M_{1,T} \xrightarrow{d} \sup_\gamma |m_1(\gamma)|$, *where* $m_1(\gamma)$ *is a zero mean Gaussian process with covariance kernel given by:*

$$\begin{aligned} & C(\gamma_1, \gamma_2) \\ &= E\left(\left({}_{t+2}u_t^{t+1}\right)^2 w\left(\sum_{j=1}^{t-1} \gamma_{1,j}' \Phi\left({}_{t+1-j}W_{t-j}\right)\right) w\left(\sum_{j=1}^{t-1} \gamma_{2,j}' \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right). \end{aligned}$$

(ii) Under $H_{A,1}$, *there exist an* $\varepsilon > 0$, *such that* $\Pr\left(\frac{1}{\sqrt{T}} M_{1,T} > \varepsilon\right) \to 1$.

From the statement in part (i) of Theorem 1, it is immediate to see that the covariance kernel does not contain cross terms, capturing the correlation between ${}_{t+2}u_t^{t+1} w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right)$ and ${}_{s+2}u_s^{s+1} w\left(\sum_{j=1}^{s-1} \gamma_j' \Phi\left({}_{s+1-j}W_{s-j}\right)\right)$ for all $s \neq t$. This is because the revision error is a martingale difference sequence under the null.

Whenever $w(\cdot)$ is an exponential function, the statement in Theorem 1 corresponds to that in Theorem 4 of de Jong (1996), for the case of no parameter estimation error. Given A1, the proof of the theorem above follows easily from the empirical process central limit theorem of Andrews (1991), for heterogeneous NED arrays.[4] On the other hand, de Jong (1996) requires $\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)$ to be $v-$stable on ${}_{t+1-j}W_{t-j}$ instead of near epoch dependent, where $v-$stability is a slightly weaker concept than NED; indeed his proof is much more involved, though he confines attention to the exponential function.[5]

As mentioned above, if we reject $H_{0,1}$, then we want to be able to distinguish between the case of unconditional biasedness and the case of inefficient use of information. The theorem below establishes the asymptotic properties of $M_{2,P}$.

Theorem 2: *Let Assumptions A1 hold. Then, if as* $R, P \to \infty$, $P/R \to \pi$, $0 < \pi < \infty$,

[4] Hansen (1996b) provides an empirical process central limit theorem for the case in which A1(iv) fails to hold and $\gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)$ is unbounded.

[5] For the relationship between NED and $v-$stability, see e.g. Pötscher and Prucha (1997, Ch.6.2).

(i) Under $H_{0,2}$, $M_{2,P} \xrightarrow{d} \sup_\gamma |m_2(\gamma)|$, where $m_2(\gamma)$ is a zero mean Gaussian process with covariance kernel given by:

$$
\begin{aligned}
&C(\gamma_1, \gamma_2) \\
&= E\left(\left(\left({}_{t+2}u_t^{t+1}\right) - \mu_u\right)^2 w\left(\sum_{j=1}^{t-1} \gamma'_{1,j} \Phi\left({}_{t+1-j}W_{t-j}\right)\right) w\left(\sum_{j=1}^{t-1} \gamma'_{2,j} \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right) \\
&+ 2\Pi\sigma_u^2 E\left(w\left(\sum_{j=1}^{t-1} \gamma'_{1,j} \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right) E\left(w\left(\sum_{j=1}^{t-1} \gamma'_{2,j} \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right) \\
&- \Pi E\left(\left(\left({}_{t+2}u_t^{t+1}\right) - \mu_u\right)^2 w\left(\sum_{j=1}^{t-1} \gamma'_{1,j} \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right) E\left(w\left(\sum_{j=1}^{t-1} \gamma'_{2,j} \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right) \\
&- \Pi E\left(\left(\left({}_{t+2}u_t^{t+1}\right) - \mu_u\right)^2 w\left(\sum_{j=1}^{t-1} \gamma'_{2,j} \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right) E\left(w\left(\sum_{j=1}^{t-1} \gamma'_{1,j} \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right) \qquad (7)
\end{aligned}
$$

where $\mu_u = E\left({}_{t+2}u_t^{t+1}\right)$, and $\sigma_u^2 = Var\left({}_{t+2}u_t^{t+1}\right)$, and $\Pi = 1 - \pi^{-1}\ln(1+\pi)$.

(ii) Under $H_{A,2}$, there exist an $\varepsilon > 0$, such that $\Pr\left(\frac{1}{\sqrt{P}} M_{1,P} > \varepsilon\right) \to 1$.

The difference between the limiting distribution in Theorems 1 and 2 is apparent in the last two lines of (7), which reflect the contribution of the recursively estimated sample mean. Furthermore, the limiting distributions in both theorems depend on the nuisance parameter $\gamma \in \Gamma$, and thus standard critical values are not available. Corradi and Swanson (2002) provide valid asymptotic critical values via the use of simulated conditional P-values. One possible drawback of this approach is that it requires direct estimation of the contribution of parameter estimation error to the covariance kernel. One the other hand, the construction of bootstrap critical value does not require direct estimation of parameter estimation components. Hence, in the next section we suggest an easy to implement bootstrap procedure, and we establish first order validity thereof.

3.4 Bootstrap Critical Values

First order validity of the block bootstrap in the context of recursive estimation is established in Corradi and Swanson (2007), for the case in which the test function, w, depends only on a finite number of lags. Furthermore, it follows intuitively that if we resample the data, say ${}_{t+2}u_t^{t+1}$ and ${}_{t+1}W_t$, then the statistic computed using the resampled observations cannot mimic the behavior of the original statistic, simply because the correct temporal ordering is no longer preserved. A scenario analogous to this one arises in the context of the conditional variance of a GARCH model, which is a near epoch dependent map on all of the history of the process. White and Goncalves (2004), in order to obtain QMLE estimation of GARCH parameters, suggest resampling (blocks of) the likelihood, and more recently Corradi and Iglesias (2007) show higher order refinement for QMLE GARCH estimation based on similar resampling. In the sequel, we use the same idea and jointly resample $\left({}_{t+2}u_t^{t+1}, w_t\left(\gamma\right)\right)$, where $w_t(\gamma) = w\left(\sum_{j=1}^{t-1}\gamma_j'\Phi\left({}_{t+1-j}W_{t-j}\right)\right)$. Under the null, ${}_{t+2}u_t^{t+1}w_t\left(\gamma\right)$ is a martingale difference sequence. Therefore, first order asymptotic validity of the bootstrap statistic can be achieved by simply resampling blocks of length one, as in the *iid* case. On the other hand, in order to achieve higher order refinement, one has to use the block bootstrap with a block size increasing with the sample, even in the case of martingale difference sequences. This is because for refinement it no longer suffices to merely mimic the first two sample moments when showing asymptotic validity; and the martingale difference assumption does not help for higher moments (see e.g. Andrews (2002)). However, our statistics are not pivotal, because of the presence of the nuisance parameters, γ, that are unidentified under the null, and thus we cannot obtain higher order bootstrap refinements. For this reason, when considering $M_{1,T}$ it suffices to make $T-2$ independent draws from $\left({}_3u_1^2, w_1\left(\gamma\right), \ldots, {}_T u_{T-2}^{T-1}, w_{T-2}\left(\gamma\right)\right)$. This can be simply accomplished by drawing $T-2$ independent discrete uniform variates on $1, \ldots, T-2$, say I_i, $i = 1, \ldots, T-2$. Hereafter, let $\left({}_3u_1^{*2}w_1^*\left(\gamma\right), \ldots, {}_T u_{T-2}^{*T-1}, w_{T-2}^*\left(\gamma\right)\right) = \left({}_{I_1+2}u_{I_1}^{*I_1+1}, w_{I_1}^*\left(\gamma\right), \ldots, {}_{I_T} u_{I_{T-2}}^{*I_{T-1}}, w_{I_{T-2}}^*\left(\gamma\right)\right)$. Thus, for all i, $w_{I_i}\left(\gamma\right) = \left(\sum_{j=1}^{I_i-1}\gamma_j'\Phi\left({}_{I_i+1-j}W_{I_i-j}\right)\right)$ which is equal to $w\left(\sum_{j=1}^{t-1}\gamma_j'\Phi\left({}_{t+1-j}W_{t-j}\right)\right)$ for $t = 1, \ldots, T-2$, with equal probability $1/(T-2)$. Note that, for any bootstrap replication, we use the same set of resampled values across all $\gamma \in \Gamma$. Hereafter, E^* and Var^* denote the mean and the variance with respect to the law governing the boostrap samples.

The bootstrap analog of $M_{1,T}$, say $M_{1,T}^*$, is then defined to be:

$$M_{1,T}^* = \sup_{\gamma\in\Gamma}\left|m_{1,T}^*(\gamma)\right|,$$

where

$$m_{1,T}^*\left(\gamma\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-2} {}_{t+2}u_t^{*t+1}w_t^*(\gamma).$$

We next turn to constructing the bootstrap analog of $M_{2,P}$, say $M_{2,P}^*$. First, construct the bootstrap analog of $\widehat{\mu}_t$, say $\widehat{\mu}_t^*$, which is defined as:

$$\widehat{\mu}_t^* = \arg\min_{\mu}\left(\frac{1}{t}\sum_{j=1}^{t}\left({}_{j+2}u_j^{*j+1} - \mu\right)^2 - 2\mu\frac{1}{T}\sum_{i=1}^{T-2}\left({}_{i+2}u_i^{i+1} - \widehat{\mu}_t\right)\right). \tag{8}$$

$$\widehat{\mu}_t^* = \frac{1}{t}\sum_{j=1}^{t}\left({}_{j+2}u_j^{*j+1} - \frac{1}{T}\sum_{i=1}^{T-2} {}_{i+2}u_i^{i+1}\right) + \widehat{\mu}_t \text{ for } t = R, \ldots, R+P-2. \tag{9}$$

From (9), it is immediate to see that $\widehat{\mu}_t^*$ is not the exact counterpart of $\widehat{\mu}_t$, this is due to the recentering term in (8), which is necessary to ensure that $E^*\left(\widehat{\mu}_t^*\right) = \widehat{\mu}_t$.

Next, construct:

$$M^*_{2,P} = \sup_{\gamma\in\Gamma} \left| m^*_{2,P}(\gamma) \right|,$$

where

$$m^*_{2,P}(\gamma) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} \left(\left({}_{t+2}u_t^{*t+1} - \widehat{\mu}^*_t \right) w^*_t - \frac{1}{T} \sum_{i=1}^{T-2} \left({}_{i+2}u_t^{i+1} - \widehat{\mu}_t \right) w_t \right).$$

Note that we need to recenter the above bootstrap statistic around the full sample mean, as observations have been resampled from the full sample. In fact, given the resampling scheme described above, it is immediate to see that $E^* \left(\left({}_{t+2}u_t^{*t+1} - \widehat{\mu}^*_t \right) w^*_t \right) = \frac{1}{T} \sum_{t=1}^{T-2} \left({}_{t+2}u_t^{t+1} - \widehat{\mu}_t \right) w_t$.

Theorem 3: *Let Assumption A1 hold. Then, if as* $R, P \to \infty$, $P/R \to \pi$, $0 < \pi < \infty$, *then:*

(i)

$$P \left[\omega : \sup_{x\in\Re} \left| P^* \left(\sup_{\gamma\in\Gamma} \left| m^*_{1,T}(\gamma) \right| \leq x \right) - P \left(\sup_{\gamma\in\Gamma} \left| m^{\mu}_{1,T}(\gamma) \right| \leq x \right) \right| > \varepsilon \right] \to 0,$$

where $m^{\mu}_{1,T}(\gamma) = \sqrt{T} \left(m_{1,T}(\gamma) - E\left(m_{1,T}(\gamma) \right) \right).$

(ii)

$$P \left[\omega : \sup_{x\in\Re} \left| P^* \left(\sup_{\gamma\in\Gamma} \left| m^*_{2,P}(\gamma) \right| \leq x \right) - P \left(\sup_{\gamma\in\Gamma} \left| m^{\mu}_{2,P}(\gamma) \right| \leq x \right) \right| > \varepsilon \right] \to 0,$$

where $m^{\mu}_{2,P}(\gamma) = \sqrt{P} \left(m_{2,P}(\gamma) - E\left(m_{2,P}(\gamma) \right) \right).$

The above results suggest proceeding in the following manner. For any bootstrap replication, compute the bootstrap statistic, $M^*_{1,T}$ ($M^*_{2,P}$). Perform B bootstrap replications (B large) and compute the quantiles of the empirical distribution of the B bootstrap statistics. Reject $H_{0,1}$ ($H_{0,2}$) if $M_{1,T}$ ($M_{2,P}$) is greater than the $(1-\alpha)^{th}$-percentile of the corresponding bootstrap distribution. Otherwise, do not reject. Now, for all samples except a set with probability measure approaching zero, $M_{1,T}$ ($M_{2,P}$) has the same limiting distribution as the corresponding bootstrapped statistic, ensuring asymptotic size equal to α. Under the alternative, $M_{1,T}$ ($M_{2,P}$) diverges to (plus) infinity, while the corresponding bootstrap statistic has a well defined limiting distribution, ensuring unit asymptotic power.

4 Empirical Methodology

In addition to the nonlinear rationality tests outlined above, we consider the issue of prediction using various variable/vintage combinations. In particular, we construct sequences of predictions for various $ex-ante$ sample periods using recursively estimated models of the following variety:

Model A (First Available Data): ${}_{t+k}X_{t+1} = \alpha + \sum_{i=1}^{p} \beta_i \; {}_{t+2-i}X_{t+1-i} + \varepsilon_{t+k}$;

Model B (k^{th} Available Data) : ${}_{t+k}X_{t+1} = \alpha + \sum_{i=1}^{p} \beta_i \; {}_{t+2-i}X_{t+3-k-i} + \varepsilon_{t+k}$; and

Model C (Real-Time Data) : ${}_{t+k}X_{t+1} = \alpha + \sum_{i=1}^{p} \beta_i \; {}_{t+1}X_{t+1-i} + \varepsilon_{t+k}$

By comparing the $ex-ante$ predictive performance of the above three models for various values of k, we can directly examine various properties of the revision process and hence various features pertaining to the rationality of early data releases. The first regression model has explanatory variables that are formed using only 1^{st} available data, while the second regression model has explanatory variables that are available $k-1$ months ahead, corresponding $(k-1)^{st}$ available data. Thus, the first model corresponds to the approach of simply using first available data and ignoring all later vintages, regardless of which vintage of data is being forecasted. On the other hand, the second model only uses data that have been revised $k-1$ times in order to predict data that likewise have been revised $k-1$ times. This sort of prediction model might be useful, for example, if it turns out that data are found to be "rational" after 3 releases using the above rationality tests, and if it is thus decided that only 3^{rd} release data is "trustworthy" enough to be used in policy analysis, say. It follows immediately upon inspection of the models that Models A and B are the same for $k = 2$, regardless of p. In Model C, the latest vintage 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. Note also that it follows that for $p = 1$, Models A and C are the same for all k, while Models A, B, and C are the same for $k = 2$. Finally, Models A and B are identical for all p if $k = 2$.

Now, 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 B), for example. Of course, the last statement has a caveat. Namely, it is possible that 1^{st} vintage data are best predicted using only 1^{st} vintage regressors, 2^{nd} vintage using 2^{nd} vintage 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 vintage, and only one of these vintages can possibly correspond to the vintage being predicted at any point in time (see discussion in the introduction for further details). This eventuality would be consistent with a finding that Model B yields the "best" predictions. However, the problem with using Model B is that we are ignoring more recent vintages of some variables, and so the model is in some sense "out of date". This is a substantial disadvantage, and may well be expected to result in Model C yielding the best predictions for values of k great than 2. Our approach in the empirical application is to set: (i) $p = 1$; (ii) $p = SIC$; (iii) $p = AIC$; (iv) $p = 0$.

Additionally, we set k ={1, 2, 3, 6, 12, 24}. Furthermore, for Type C regressions, in addition to the basic regression model, we consider models where we include additional regressors of the form: (i) ${}_{t+1}W_t' = {}_{t+1}u_{t-k}^t$, for each of $k = 1, 2, ..., 24$; (ii) ${}_{t+1}W_t' = ({}_{t+1}u_{t-1}^t, {}_{t+1}u_{t-2}^t, {}_{t+1}u_{t-3}^t)'$; and (iii) ${}_{t+1}W_t' = {}_{t+1}u_{t+1-k}^{t+2-k}$, for each of $k = 3, 4, ..., 24$.

Additionally, we 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 which consist of various releases of output growth.

All of our prediction experiments are based on the examination of the mean square forecast errors associated with predictions constructed using recursively estimated models; and MSFEs are in turn examined via the use of Diebold and Mariano (1995) and Clark and McCracken (2001) type predictive accuracy tests (see also Clark and McCracken (2005), McCracken (2006), and West (1996)).

5 Empirical Results

5.1 Data

The variables used in the empirical part of this paper are real GDP (seasonally adjusted), the GDP chain-weighted price index (seasonally adjusted), the money stock (measured as M1, seasonally adjusted) and the real interest rate (measured as the rate on 3-month Treasury bills).[6] All series have a quarterly frequency and our real time data sets for each of the four variables were obtained from the Philadelphia Reserve System's real time data set for Macroeconomists (RTDSM).[7]

The first vintage in our sample is 1965.Q4, for which the first calendar observation is 1959.Q3. 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.Q4 vintage and the corresponding 2006.Q3 calendar date, allowing us to keep track of the exact data that were available at each vintage for every pos-

[6] Results based on the use of M2 in place of M1 in all empirical exercises are available upon request, and are qualitatively the same as those reported for M1.

[7] 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. The data we use 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).

sible 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. We use log-differences throughout our analysis (except for interest rates); and the log-differences of all the variables, except the interest rate, are plotted in Figures 1-3. The figures also exhibit 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. 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 in the zero interval, 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 offer little useful information, other than signalling the presence of definition 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 is illustrated in Figure 4 where we have plotted four early calendar dates (1959.Q4; 1960.Q4; 1961.Q4; and 1962.Q4) across all available vintages in our sample, for output (GDP) and money (both in log-differences). Of note is that the data varies significantly across the 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 late 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). This can be graphically illustrated by noticing that in Figure 4, for output, the four calendar dates plotted 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. This can be observed in the lower graph in Figure 4. This suggests a need for great caution when analyzing real time data, particularly money stock.

In the introduction of this paper, we argued that the variable that one cares about predicting is likely to be a near-term vintage, and if one cares about *final* data, one may have a very difficult time, as one would need then to predict unknowable future definitional and other methodological changes implicit to the data construction process. The above discussion supports this argument. Namely, "pure revision" appears to occur in the near-term, "definitional change" occurs in the long term, and little occurs between. Furthermore, and as argued in the introduction, application of simple level shifts to datasets in order to address the "definitional change" issue may be grossly 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. However, it should be noted that although our 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, our other empirical evidence suggests that the "definitional change" issue may not be very damaging. This issue is discussed in the next sub-section.

5.2 Basic Predictive Regression Results

As discussed in Section 4, we carried out three types of simple autoregressive prediction experiments, where the objective was to forecast output. The methods involved fitting regression Models A, B, and C. Given the discussion of the previous sub-section, it is quite apparent that Models A and B are, strictly speaking, invalid. This follows because the models are estimated recursively, using all data available prior to the construction of each new prediction (see footnote to Table 3 for further details). This in turn means that the structure of Models A and B ensures that data from different "definitional periods" will be used in the estimation of the models, for many of the predictions made; and hence parameter estimates will in principle be corrupted. However, Model C does not have this feature, as one always uses currently available data for model estimation and prediction construction. Indeed, Model C is arguably the type of model that the Federal Reserve uses to construct predictions. Namely, data are updated at each point in time, as they become available, and prior to re-estimation of models and prediction construction. The other models use what might be called "stale data". However, the other models might have a certain advantage over Model C. For example, Model A *only* uses first available data to predict first available data. Indeed, Model A also uses first available data for cases where we are interested in predicting second and later vintages of data (i.e. when we are predicting ${}_{t+k}X_{t+1}$, for $k > 2$). Model B, on the other hand, uses only k^{th} vintage data for predicting k^{th} vintages. Thus, Models B and C do not suffer from the "mixed vintages" problem that Model A does. They do not use explanatory datasets for which each and every observation corresponds to a different vintage. One aspect of our prediction experiments is that we are able to shed light upon the following trade-off: What is worse: using "mixed vintage" datasets that are not subject to "definitional change" problems, or using datasets containing only the vintage corresponding to the desired prediction vintage, but with "definitional change" problems?

The answer to the preceding question is immediately apparent upon inspection of Table 3, where results for basic autoregressive versions of Models A, B, and C are presented for various prediction periods, various values of k, and for models both with and without included revision error regressors.[8] Model C *never* yield the lowest mean square forecast error (MSFE), *regardless of what vintage of data is being predicted, and regardless of the variable being predicted*.[9] In particular, for output, prices, and money, Models A and B always "win", and for most cases where Model B "wins", it is simply because it is for a case where Models A and B yield identical results (see Section 4 for a discussion of these cases). Thus, it indeed appears to be the case that Model A "wins" in virtually every case considered across all variables, and regardless of whether revision errors are added as additional regressors or not. Thus, we have surprising evidence that preliminary data are useful even for predicting later data releases. While this is pretty strong evidence against Model C, one must keep in mind that Model B uses "stale data" when predicting vintages for $k > 2$, and hence it might not be viewed as too surprising that Model A dominates Model B. In summary, we have interesting evidence suggesting that real-time datasets are crucial, but perhaps not always in the way people have though, as the "mixed vintage" problem appears to be sufficiently important as to cause Model C to "lose" every one of our prediction competitions. This findings holds *even though* Model A is subject to the "definitional change" problem, suggesting that the "definitional

[8] In Model C, additional "revision error" regressors include: (i) ${}_{t+1}W_t' = {}_{t+1}u_{t-k}^{t}$; (ii) ${}_{t+1}W_t' = ({}_{t+1}u_{t-1}^{t}, {}_{t+1}u_{t-2}^{t})'$; and (iii) ${}_{t+1}W_t' = {}_{t+1}u_{t+1-k}^{t+2-k}$. Note that results are only reported for $k = 2, 3$, and 4. Other values of k were considered, but results are not reported here as they are qualitatively the same as those that are reported, and because "pure revisions" appear to die out extremely quickly in the data (see above discussion). Finally, note that although models were estimated using lags selected with the AIC and the SIC, and for lags set equal to unity, results are only reported for lags selected using the SIC. The reason for this is that models with lags selected using the SIC uniformly yielded the lowest MSFEs across all cases reported.

[9] Note also that the random walk model does not yield the lowest MSFE for any cases considered. The only exception is output for $k = 2$, when the forecasting period starting date is1983:01, in which case Model C beats Models A and B, and the random walk model beats Models A, B, and C.

change" problem is not too severe.[10]

It is also apparent from inspection of Table 3 that adding revision errors to the prediction models does not result in lower MSFEs, at least for the cases that we have considered. This finding, coupled with our finding that Model A yields the lowest MSFEs across all cases, suggests that all of our data are to some extent efficient. Of course, this result is based upon autoregression type models, and one would need to include additional regressors, both linearly and nonlinearly, to properly examine the issue of efficiency. In the next section this is done via examination of vector autoregressive predictive models for output. In the subsequent section, we examine the issue using the tests discussed in Section 3 of this paper.

5.3 Rationality Tests

We now turn to an illustration of how one might empirically implement the rationality tests discussed above. Note that in-depth empirical research of the series, however, is left to future research.

We begin by testing the following hypothesis:

$$H_{0,1} : E\left({}_{t+2}u_t^{t+1} | \mathcal{F}_t^{t+1}\right) = 0, \text{ a.s.}$$

This is done for the three variables ($X's$) in our dataset (i.e. real GDP (seasonally adjusted), the GDP chain-weighted price index (seasonally adjusted), and the money stock (measured as M1, seasonally adjusted). Recall that the relevant test statistic is:

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

where

$$m_{1,T}(\gamma) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T-2} {}_{t+2}u_t^{t+1} w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right),$$

We will distinguish two cases for ${}_{t+1-j}W_{t-j}$: (i) ${}_{t+1-j}W_{t-j}$ is a scalar,

$${}_{t+1-j}W_{t-j} = {}_{t+1-j}X_{t-j}; \quad j = 0, 1, ..., t-1,$$

[10] Diebold Mariano (1995) test statistics were constructed for all of the results reported in the tables discussed in this subsection, and all test statistic values were far below standard normal critical values, in absolute value, suggesting that there is actually little to choose between the models (tabulated values are available upon request from the authors). However, we note that it remains an interesting result of our experiments that Model A virtually always yields lower MSFEs, when comparing point values; whereas in a truly experimental setting one might expect an equal number of "wins" based on point MSFEs for each of the different models, under the null hypothesis. For this reason, we conjecture that more refined testig with bigger samples of data, for example, might be expected to yield test statistics that are significantly different from zero.

For further discussion of valid test critical value construction when applying Diebold Mariano statistics in the current context, refer to Clark and McCracken (2007).

and, (ii) ${}_{t+1-j}W_{t-j}$ is a vector,

$$ {}_{t+1-j}W_{t-j} = \left[{}_{t+1-j}X_{t-j},\ {}_{t+1-j}u_{t-1-j}^{t-j} \right]' \ j = 0, 1, ..., t-1. $$

Following Corradi and Swanson (2002) we set w as the exponential function, and Φ the inverse tangent function. Finally, we set $\gamma_j \equiv \gamma \cdot j^{-k}$, where γ is defined over a fine grid; $\gamma \in [0, 3]$, for the scalar case; and

$$ \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \in [0, 3]x[0, 3] $$

for the vector case. We report results for $k = 2, 3$ and 4. To sum-up, the test statistic, under the scalar case (i), is computed as the supremum of

$$ m_{1,T}(\gamma) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T-2} {}_{t+2}u_t^{t+1} \exp\left(\sum_{j=1}^{t-1} \left(\gamma_1 j^{-k} \tan^{-1}(X_{t-j}) \right) \right), $$

and, under the vector case (ii), is the supremum of

$$ m_{1,T}(\gamma) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T-2} {}_{t+2}u_t^{t+1} \exp\left(\sum_{j=1}^{t-1} \left(\gamma_1 j^{-k} \tan^{-1}({}_{t+1-j}X_{t-j}) + \gamma_2 j^{-k} \tan^{-1}\left({}_{t+1-j}u_{t-1-j}^{t-j} \right) \right) \right), $$

When forming bootstrap samples for $M_{1,T}$, it suffices to make $T-2$ independent draws from $\left({}_3u_1^2, w_1(\gamma), ..., {}_T u_{T-2}^{T-1}, w_{T-2}(\gamma) \right)$. This can be simply accomplished by drawing $T-2$ independent discrete uniform variates on $1, ..., T-2$, say I_i, $i = 1, ..., T-2$. Note that for implementation of the bootstrap in the current context, we select observations in which we can "back-fill" as much as $T-3$ observations in order to construct $w(\cdot)$. In our case, since our *full sample* consists of 164 observations, we choose the first observation of the summation to be the 81st. Finally, note that for any k/variable combination, we also form a test of $H_{0,2}$, if $H_{0,1}$ is rejected.

Results are gathered in Table 5, were it is immediately apparent that the null hypothesis fails to reject for all k/variable combinations, using critical values taken from the 90th and 95th percentiles of the bootstrap distributions. This suggests two things. First, when viewed in a truly ex ante context, there is no evidence of data irrationality for any of the series examined here. In particular, recall that $H_{0,1}$ is the null hypothesis that the first release is rational, as the revision error in this case is a martingale difference process, at least when adapted to the filtrations used for the results reported in Panels A and B of the table. This is consistent with the "news" version of rationality, according to which subsequent data revisions only take news that were not available at the time of the first release into account. Thus, we have evidence that first, second and third data releases already incorporate "all" available information at their time of release. This also implies, for example, that first release data might be expected to be best predicted using historical first release data rather than using mixed vintages of data (as is currently the standard practice). Indeed, this is what was found based upon our simple regression experiments reported in the previous subsection. Second, bias adjustment to preliminary releases may not be useful, in an ex ante forecasting context. Indeed, given that $H_{0,1}$ fails to reject, we did not even test $H_{0,2}$.

5.4 Marginal Predictive Content of Money for Output

In this subsection, we implement Models A,B, and C to examine whether additional variables and their respective revision errors improve predictive performance. 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 are included, so that we can additionally assess the marginal predictive content of money for output.[11] As a benchmark, the autoregressive and random walk prediction model results from Table 3 are included.

Interestingly, the vector autoregression models yield lower MSFEs than their autoregressive counterparts for prediction periods beginning in 1970:1 and 1983:1; but not for the recent very stable period beginning in 1990:1. Furthermore, 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. This result holds regardless of whether we add revision errors as additional regressors or not. Thus, we have evidence that including money in linear output prediction models does not improve predictive performance. Additionally, models with revision errors never outperform their counterparts that do not contain revision errors. This is further evidence that our preliminary data are efficient.

6 Concluding Remarks

We outline two new tests for data rationality, both of which are designed to assess rationality from an ex ante forecasting perspective. An illustrative empirical implementation of the tests yields support for the "news" hypothesis, in the sense that early data revisions to U.S. output, price, and money variables appear to only take news that were not available at the time of data release into account. Furthermore, and consistent with this finding, we carry out a series of prediction experiments lending support to the notion that first available data are best predicted with datasets

[11] Amato and Swanson (2001) also use real-time data to assess the marginal preditive content of money for output. However, they do not consider models that additionally include various types of revision errors as predictors, as is done in this paper.

constructed using only past first available data, and not using mixed vintages of data, as is the usual approach. Finally, we find little real-time marginal predictive content of money for output, both when past historical data as well as past revision errors are used as explanatory variables.

Many problems in this literature reamin unsolved. For example, from an empirical perspective it remains to extend the analysis that we carry out to later vintages (only the first three vintages were examined in our empirical analysis), and to further examine the important problem of definitional change that is addressed in the paper. From a theoretical perspective, it remains to extend many of the standard predictive accuracy and related tests to the case of real-time data, including those that have been hithertofor carried out using “in-sample” regression approaches.

7 References

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

Andrews, D.W.K. (1991), An Empirical Process Central Limit Theorem for Dependent Non-Identically Distributed Random Variables, *Journal of Multivariate Analysis,* 38, 187-203.

Andrews, D. W. K. (2002), Higher-Order Improvements of a Computationally Attractive $k-$Step Bootstrap for Extremum Estimators, *Econometrica* 70, 119-162.

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

Bierens, H.B. (1982), Consistent Model Specification Tests, *Journal of Econometrics*, **20**, 105-134.

Bierens, H.B. (1990), A Conditional Moment Test of Functional Form, *Econometrica* , **58**, 1443-1458

Bierens, H.J. and W. Ploberger, (1997): Asymptotic Theory of Integrated Conditional Moment Tests, *Econometrica*, **65**, 1129-1152.

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 Multi-Step Forecasts, *Econometric Reviews* **24**, 369-404.

Clark, T. E. and M.W. McCracken (2007), Tests of Equal Predictive Ability with Real-Time Data, Working Paper, Federal Reserve Board.

Corradi, V. and E.M. Iglesias (2007), Bootstrap Refinements for QML Estimators of the GARCH(1,1) Parameters, Working Paper, Warwick University.

Corradi, V. and N.R. Swanson (2002), A Consistent Test for Nonlinear Out of Sample Predictive Accuracy, *Journal of Econometrics* **110**, 353-381.

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

Corradi, V. and N.R. Swanson (2006b), Predictive Density and Conditional Confidence Interval Accuracy Tests, *Journal of Econometrics* **135**, 187-228.

Corradi, V. and N.R. Swanson (2007), Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes, *International Economic Review*, **48**, 67-109.

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.

de Jong, R.M., (1996), The Bierens Test Under Data Dependence, *Journal of Econometrics*, **72**, 1-32.

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., J.H. Rogers and J.H. Wright (2003), Exchange Rate Forecasting: The Errors We've Really Made, *Journal of International Economics* **60**, 35–59.

Faust, J., J.H. Rogers and J.H. Wright (2005), News and Noise in G-7 GDP Announcements, *Journal of Money, Credit and Banking*, 37, 403-417

Gallant, A.R., and H. White, (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models,* Blackwell, Oxford.

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

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

Gonçalves, S. and H. White (2004), Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models, *Journal of Econometrics* 119, 199-219.

Granger, C.W.J. (2001), An Overview of Nonlinear Macroeconometric Empirical Models, *Macroeconomic Dynamics* **5**, 466–481.

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

Hansen, B.E. (1996a), Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis, *Econometrica*, **64**, 413-430.

Hansen, B.E., (1996b), Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays, *Econometric Theory,* 12, 347-359.

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.

Lee, T.-H., H. White and C.W.J. Granger, (1993), Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests, *Journal of Econometrics*, **56**, 269-290.

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. (2006), Asymptotics for Out–of–Sample Tests of Causality, *Journal of Econometrics*, forthcoming.

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.

Pötscher, B.M., and I.R. Prucha, (1997), *Dynamic Nonlinear Econometric Models,* Springer, Berlin.

Ratjens, P. and R.P Robins (1995), Do Goverment 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.

Stock, J.H. and M.W. Watson (1996), Evidence on Structural Instability in Macroeconomic Time Series Relations, *Journal of Business & Economic Statistics* **14**, 11–30.

Stock, J.H. and M.W. Watson (1999), Business Cycle Fluctuations in US Macroeconomic Time Series, in J.B. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics* Vol. 1A, Amsterdam: Elsevier Science, pp. 3–64.

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.

Zarnowitz, V. (1978), Accuracy and Properties of Recent Macroeconomic Forecasts, *American Economic Review* **68**, 313–319.

Appendix

Proof of Theorem 1:

(i) Given A1(iv)-(v), for each $\gamma \in \Gamma$, ${}_{t+2}u_t^{t+1} w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right)$ is Lipschitz of order 1 on $\mathcal{U}$. Thus, given A1(ii), from Theorem 4.2 in Gallant and White (1988), it follows that ${}_{t+2}u_t^{t+1} w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right)$ is NED of size -1 on the strong mixing base $\Phi\left({}_{t+1-j}W_{t-j}\right)$ of size $-2r/(r-2)$. Given A1(ii)-(iii), it follows from Lemma 2 in Andrews (1991) that for any $\gamma, \gamma' \in \Gamma$, $Cov\left(m_T(\gamma), m_T(\gamma')\right)$ exists. Straightforward calculation, show that $Cov\left(m_T(\gamma), m_T(\gamma')\right)$ is as defined in the statement of the Theorem. Thus, Assumption C in Andrews (1991) is satisfied, and the statement follows from his Theorem 4, recalling that given A1(i), $\Gamma \subseteq \mathcal{R}^k$.

(ii)

$$
\begin{aligned}
& m_{1,T}(\gamma) \\
& = \frac{1}{\sqrt{T}} \sum_{t=1}^{T-2} \left({}_{t+2}u_t^{t+1} w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right) - E\left({}_{t+2}u_t^{t+1} w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right)\right) \\
& + \frac{1}{\sqrt{T}} \sum_{t=1}^{T-2} E\left({}_{t+2}u_t^{t+1} w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right).
\end{aligned}
$$

The first term on the RHS above is $O_P(1)$ as it convergence in distribution by part (i). As $w(\cdot, \gamma)$ is generically comprehensive, $E\left({}_{t+2}u_t^{t+1} w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right) \neq 0$ for all $\gamma \in \Gamma$. The statement then follows.

Proof of Theorem 2:

(i) For any given γ, let $\mu_u = E\left({}_{t+2}u_t^{t+1}\right)$ and $\mu(\gamma) = E\left(w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right)$,

$$
\begin{aligned}
m_{2,P}(\gamma) &= \frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} \left({}_{t+2}u_t^{t+1} - \mu\right) w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right) \\
&\quad - \frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} \left(\frac{1}{t} \sum_{i=1}^{t} \left({}_{i+2}u_i^{i+1} - \mu_u\right)\right) w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right) \\
&= \frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} \left({}_{t+2}u_t^{t+1} - \mu_u\right) w\left(\sum_{j=1}^{t-1} \gamma_j' \Phi\left({}_{t+1-j}W_{t-j}\right)\right) \\
&\quad - \frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} \left(\frac{1}{t} \sum_{i=1}^{t} \left({}_{i+2}u_i^{i+1} - \mu_u\right)\right) \mu(\gamma) + o_P(1) \qquad (10)
\end{aligned}
$$

where the $o_P(1)$ term holds uniformly in γ, because of the uniform law of large numbers for NED processes on a strong mixing base (e.g. Gallant and White, 1988, Ch.3). As for first term on the RHS of (10), $(m_{2,P}(\gamma), m_{2,P}(\gamma'))$ converges in distribution for each for any $\gamma, \gamma' \in \Gamma$ and is stochastic equicontinuous on Γ, by Theorem 4 in Andrews (2001). As for second term on the RHS of (10), it can be treated by a similar argument as in the proof of Theorem 1 in Corradi and Swanson (2002).

(ii) As in Part (ii) of Theorem 1.

Proof of Theorem 3:

(i) First note that, as I_i are identically and independently distributed, and thus $w_t^*(\gamma)$ is *iid* conditionally on the sample. Now,

$$E^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-2} {}_{t+2}u_t^{*t+1} w_t^*\right) = \sqrt{T} E^*\left({}_{t+2}u_t^{*t+1} w_t^*\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T-2} {}_{t+2}u_t^{t+1} w_t$$

and thus for all γ, $E^*\left(m_{1,T}^*(\gamma)\right) = m_{1,T}(\gamma)$. Also,

$$\begin{aligned}
&Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-2} {}_{t+2}u_t^{*t+1} w_t^*\right) \\
&= E^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-2}\left({}_{t+2}u_t^{*t+1} w_t^* - \frac{1}{T}\sum_{t=1}^{T-2} {}_{t+2}u_t^{t+1} w_t\right)\right)^2 \\
&= \frac{1}{T}\sum_{t=1}^{T-2}\left({}_{t+2}u_t^{t+1} w_t - \frac{1}{T}\sum_{t=1}^{T-2} {}_{t+2}u_t^{t+1} w_t\right)^2 \\
&= E\left(\left({}_{t+2}u_t^{t+1}\right)^2 w\left(\sum_{j=1}^{t-1}\gamma_j'\Phi\left({}_{t+1-j}W_{t-j}\right)\right)^2\right) + o_p(1)
\end{aligned}$$

Then, by Theorem 3.5 in Künsch (1989), pointwise in γ,

$$\sqrt{T}\left(m_{1,T}^*(\gamma) - E^*\left(m_{1,T}^*(\gamma)\right)\right) \xrightarrow{d^*} N\left(0, Var^*\left(m_{1,T}^*(\gamma)\right)\right),$$

conditional on the sample, and for all samples except for a subset of probability measure approaching zero. Joint convergence of $\left(\sqrt{T}\left(m_{1,T}^*(\gamma) - E^*\left(m_{1,T}^*(\gamma)\right)\right), \sqrt{T}\left(m_{1,T}^*(\gamma') - E^*\left(m_{1,T}^*(\gamma')\right)\right)\right)$ follows immediately as a consequence of the Wold device. Given A1(iv)-(v), $m_{1,T}^*(\gamma) - E^*\left(m_{1,T}^*(\gamma)\right)$ has a series expansion with smooth (first order Lipschitz) coefficients, as described in equations (3.3) and (3.4) in Andrews (1991). Thus, stochastic equicontinuity on Γ follows from Theorem 1 in Andrews (1991). The statement in part (i), then follows by noting that for each $\gamma_1, \gamma_2 \in \Gamma$,

$$Cov^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-2} {}_{t+2}u_t^{*t+1} w_t^*(\gamma_1), \frac{1}{\sqrt{T}}\sum_{t=1}^{T-2} {}_{t+2}u_t^{*t+1} w_t^*(\gamma_1)\right)$$

$$= E\left(\left({}_{t+2}u_t^{t+1}\right)^2 w\left(\sum_{j=1}^{t-1}\gamma_{1,j}'\Phi\left({}_{t+1-j}W_{t-j}\right)\right) w\left(\sum_{j=1}^{t-1}\gamma_{2,j}'\Phi\left({}_{t+1-j}W_{t-j}\right)\right)\right) + o_p(1).$$

(ii) Let $w_t(\gamma) = w\left(\sum_{j=1}^{t-1}\gamma_j'\Phi\left({}_{t+1-j}W_{t-j}\right)\right)$, $\mu(\gamma) = E(w_t(\gamma))$, and $\mu_u = E\left({}_{t+2}u_t^{t+1}\right)$.

$$\begin{aligned}
m_{2,P}^*(\gamma) &= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\left(\left({}_{t+2}u_t^{*t+1} - \widehat{\mu}_t^*\right) w_t^*(\gamma) - \frac{1}{T}\sum_{i=1}^{T-2}\left({}_{i+2}u_i^{i+1} - \widehat{\mu}_t\right) w_i(\gamma)\right) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\left(\left({}_{t+2}u_t^{*t+1} - \widehat{\mu}_t\right) w_t^*(\gamma) - \frac{1}{T}\sum_{i=1}^{T-2}\left({}_{i+2}u_i^{i+1} - \widehat{\mu}_t\right) w_i(\gamma)\right) \\
&- \frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\left(\widehat{\mu}_t^* - \widehat{\mu}_t\right) w_t^*(\gamma) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\left(\left({}_{t+2}u_t^{*t+1} - \widehat{\mu}_t\right) w_t^*(\gamma) - \frac{1}{T}\sum_{i=1}^{T-2}\left({}_{i+2}u_i^{i+1} - \widehat{\mu}_t\right) w_i(\gamma)\right) \\
&- \frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\left(\widehat{\mu}_t^* - \widehat{\mu}_t\right)\mu(\gamma) - \frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\left(\widehat{\mu}_t^* - \mu_u\right)\left(w_t^*(\gamma) - \mu(\gamma)\right) \\
&= I_P(\gamma) - II_P(\gamma) - III_P(\gamma)
\end{aligned}$$

We need to show that, conditional on the sample, and for all samples except for a subset of probability measure approaching zero, $I_P(\gamma)$ has the same limiting distribution as in the statement in (i), $II_P(\gamma)$ has the same limiting distribution as $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\left(\frac{1}{t}\sum_{i=1}^{t}\left({}_{i+2}u_i^{i+1} - \mu_u\right)\right)\mu(\gamma)$, and that $III_P(\gamma)$ is $o_P(1)$. Now,

$$\begin{aligned}
&E^*\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\left(\left({}_{t+2}u_t^{*t+1} - \widehat{\mu}_t\right) w_t^*(\gamma)\right)\right) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\frac{1}{T}\sum_{i=1}^{T-2}\left({}_{i+2}u_t^{i+1} w_t(\gamma) - \widehat{\mu}_t\right) w_i(\gamma),
\end{aligned}$$

and

$$\begin{aligned}
&Var^*\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\left({}_{t+2}u_t^{*t+1} - \widehat{\mu}_t\right) w_t^*(\gamma)\right) \\
&= \frac{1}{P}\sum_{t=R}^{T-2}\frac{1}{T}\sum_{i=1}^{T-2}\left(\left({}_{i+2}u_i^{t+1} - \widehat{\mu}_t\right) w_i(\gamma) - \frac{1}{T}\sum_{i=1}^{T-2}\left({}_{i+2}u_t^{i+1} w_t(\gamma) - \widehat{\mu}_t\right) w_i(\gamma)\right)^2
\end{aligned}$$

Thus, $I_P(\gamma)$ can be treated along the same lines as in Part (i).

Recalling the definition of $\widehat{\mu}_t^*$, given in (9),

$$-II_P(\gamma) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-2}\left(\widehat{\mu}_t^* - \widehat{\mu}_t\right)\mu(\gamma)$$

$$= \frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} \left(\frac{1}{t} \sum_{j=1}^{t} \left({}_{j+2}u_j^{*j+1} - \frac{1}{T} \sum_{i=1}^{T-2} {}_{i+2}u_i^{i+1} \right) \right) \mu(\gamma)$$

$$= \mu(\gamma) \frac{a_{R,0}}{\sqrt{P}} \left(\sum_{j=1}^{R} \left({}_{j+2}u_j^{*j+1} - \frac{1}{T} \sum_{i=1}^{T-2} {}_{i+2}u_i^{i+1} \right) \right.$$

$$\left. + \frac{1}{\sqrt{P}} \sum_{i=1}^{P-2} a_{R,i} \left({}_{R+2+i}u_{R+i}^{*R+1+i} - \frac{1}{T} \sum_{i=1}^{T-2} {}_{i+2}u_i^{i+1} \right) \right)$$

From Lemma A5 and Lemma A6 in West (1996), it follows that

$$Var^* \left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} (\widehat{\mu}_t^* - \widehat{\mu}_t)\, \mu(\gamma) \right)$$
$$= 2\Pi\mu(\gamma)^2 \left(E^* \left({}_{j+2}u_j^{*j+1} - \frac{1}{T} \sum_{i=1}^{T-2} {}_{i+2}u_i^{i+1} \right)^2 \right)$$
$$= 2\Pi\mu(\gamma)^2 \sigma_u^2 + o_P(1),$$

with $\Pi = 1 - \pi^{-1} \ln(1 + \pi)$. So, for all $\gamma_1, \gamma_2 \in \Gamma$,

$$Cov^* \left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} (\widehat{\mu}_t^* - \widehat{\mu}_t)\, \mu(\gamma_1), \frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} (\widehat{\mu}_t^* - \widehat{\mu}_t)\, \mu(\gamma_2) \right)$$
$$= 2\Pi\mu(\gamma_1)\mu(\gamma_2)\sigma_u^2.$$

Thus, $II_P(\gamma)$ properly captures the contribution of parameter estimation error. Finally, recalling that $E^*(w_t^*(\gamma)) = \frac{1}{T} \sum_{i=1}^{T} w_i(\gamma)$,

$$III_P(\gamma)$$
$$= \frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} (\widehat{\mu}_t^* - \mu_u)\, (w_t^*(\gamma) - E^*(w_t^*(\gamma)))$$
$$+ \frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} (\widehat{\mu}_t^* - \mu_u) \left(\frac{1}{T} \sum_{i=1}^{T} (w_i(\gamma) - \mu(\gamma)) \right)$$
$$= \frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} (\widehat{\mu}_t^* - \mu_u)\, (w_t^*(\gamma) - E^*(w_t^*(\gamma))) + o_p(1),$$

uniformly in γ. As, $w_j^*(\gamma)$ and ${}_{j+2}u_j^{*j+1}$ are iid conditionally on the sample, for all $t \neq j$,

$$E^* \left(\left({}_{j+2}u_j^{*j+1} - E^* \left({}_{j+2}u_j^{*j+1} \right) \right) (w_t^*(\gamma) - E^*(w_t^*(\gamma))) \right) = 0,$$

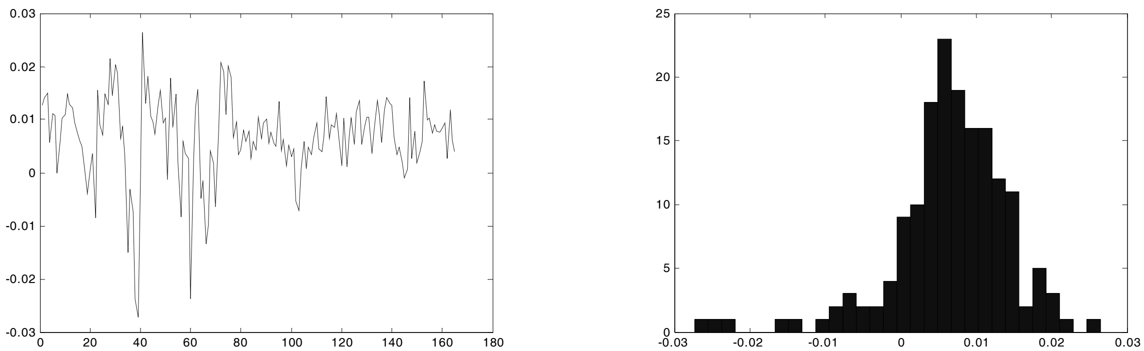
and thus

$$Var^* \left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} (\widehat{\mu}_t^* - \mu_u)\, (w_t^*(\gamma) - E^*(w_t^*(\gamma))) \right),$$

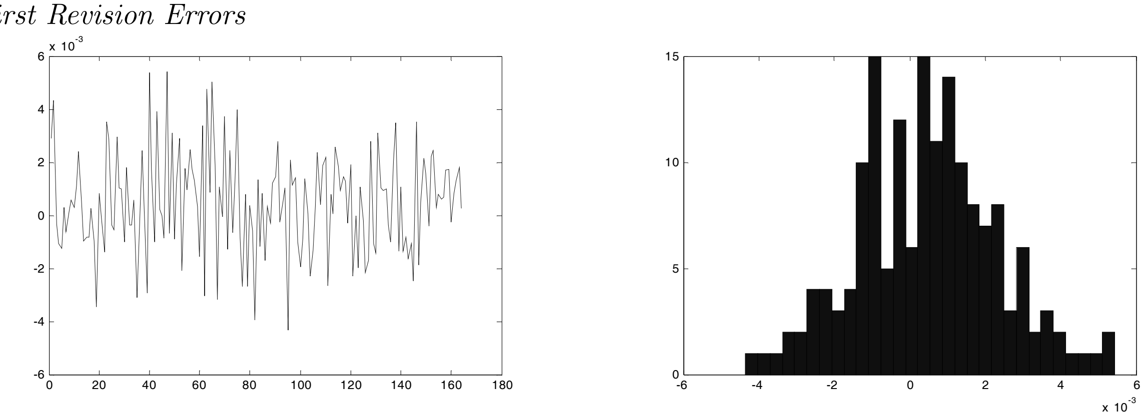
which ensures that $\frac{1}{\sqrt{P}} \sum_{t=R}^{T-2} ((\widehat{\mu}_t^* - \mu_u)\, (w_t^*(\gamma) - \mu(\gamma))) = o_p(1)$, where the $o_p(1)$ term does not depend on γ. Hence, $III_P(\gamma) = o_p(1)$, uniformly in γ.

Figure 1: Output Growth Rates, First, and Second Revision Errors — 1965:4 - 2006:4

Growth Rates



First Revision Errors



Second Revision Errors

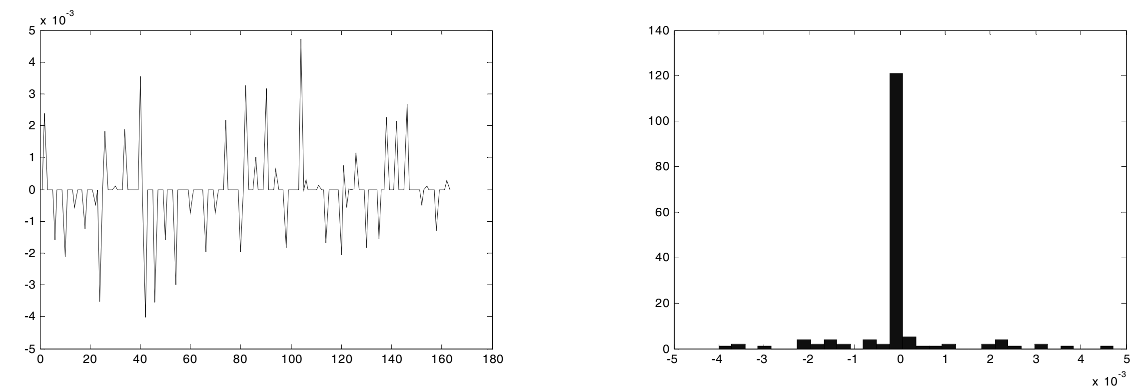
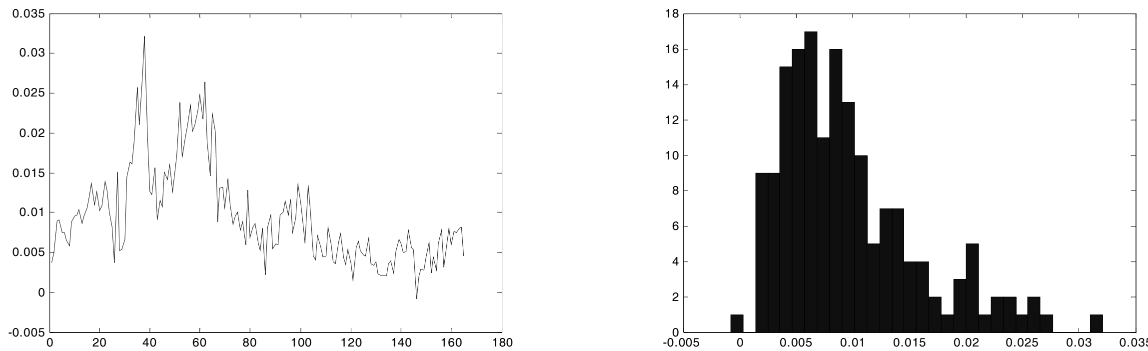
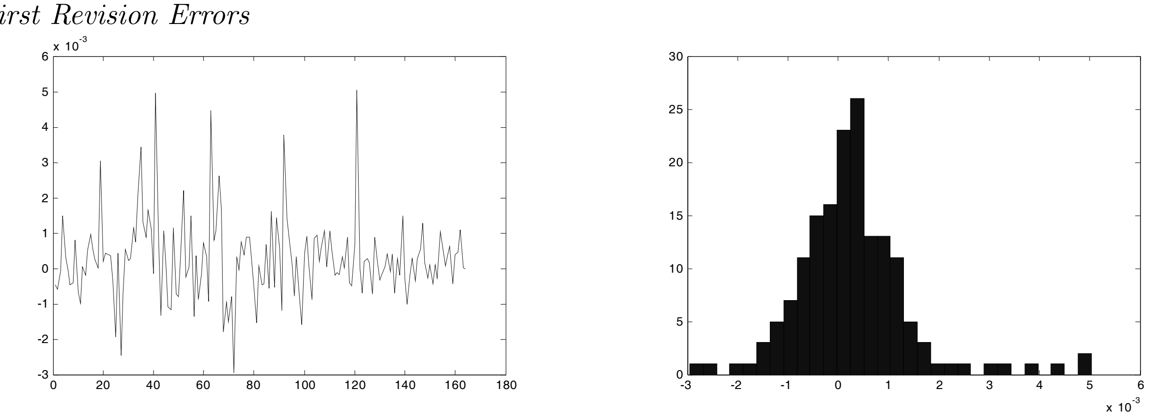


Figure 2: Price Growth Rates, First, and Second Revision Errors — 1965:4 - 2006:4

Growth Rates



First Revision Errors



Second Revision Errors

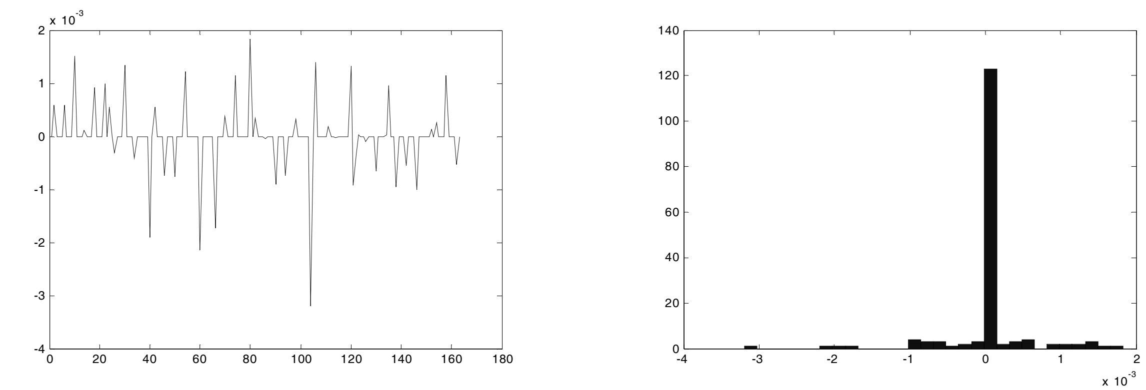
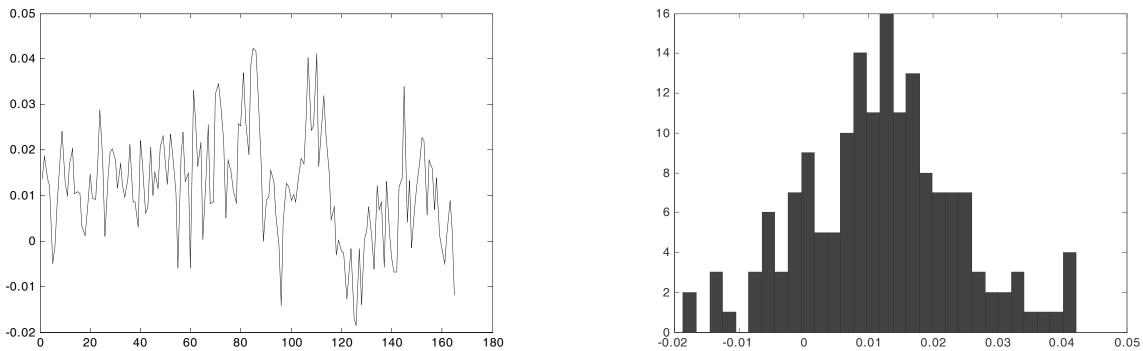
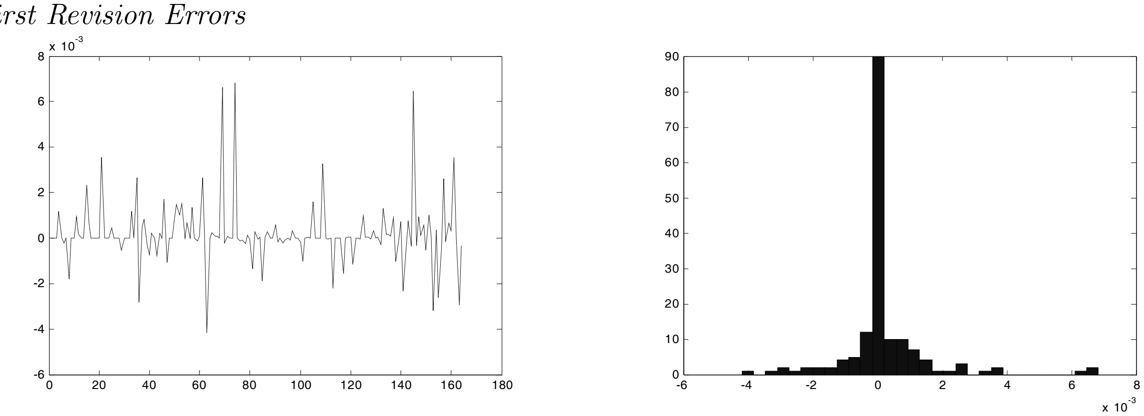


Figure 3: Money Growth Rates, First, and Second Revision Errors — 1965:4 - 2006:4

Growth Rates



First Revision Errors



Second Revision Errors

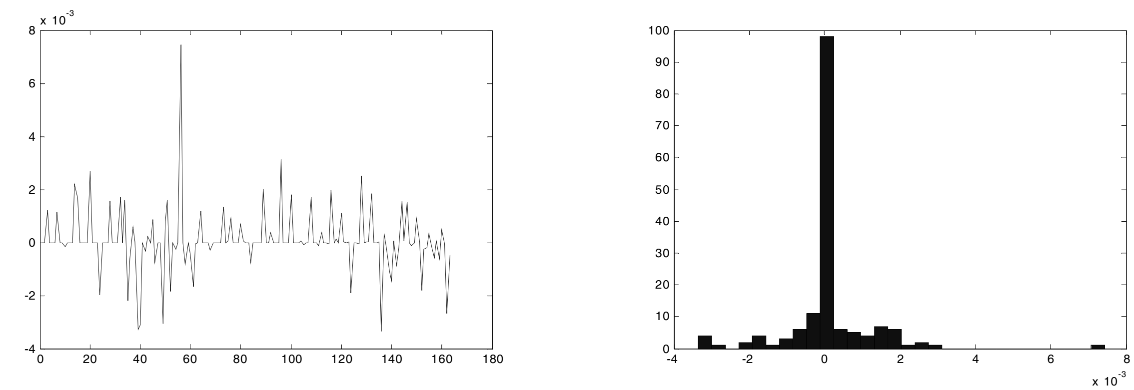
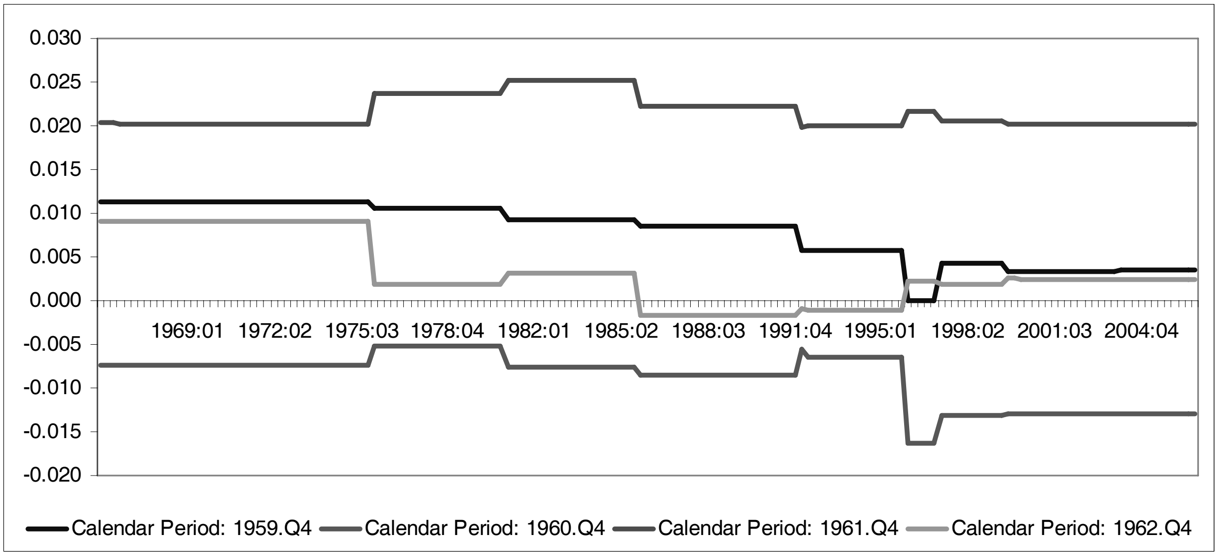
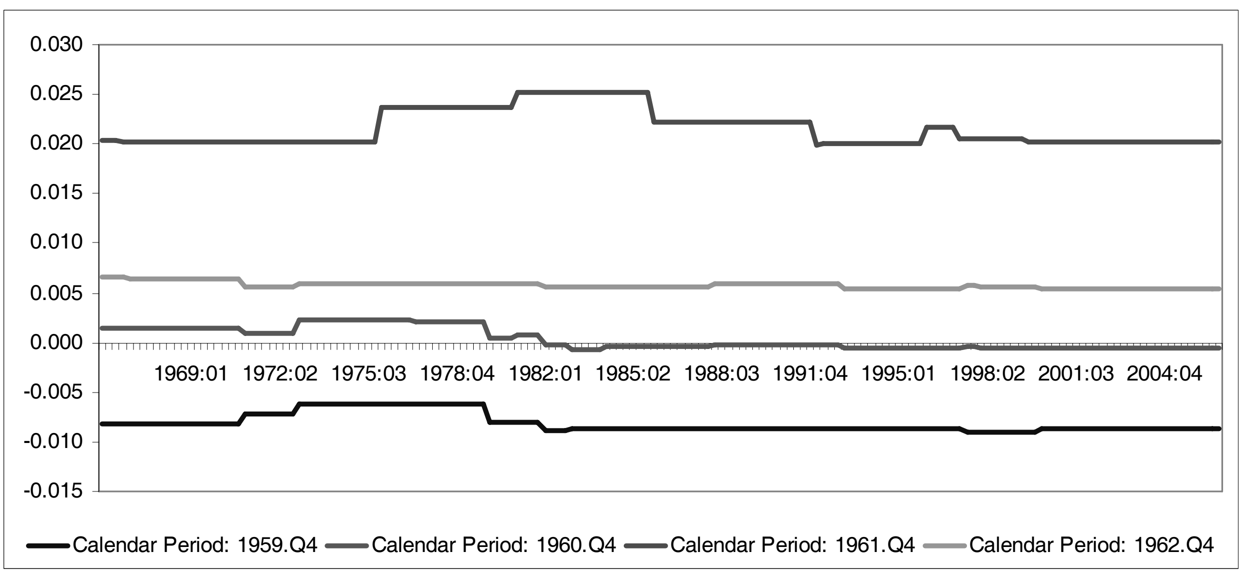


Figure 4: Data From Four Different Calendar Dates Plotted Across All Data Vintages(*)

Output



Money



(*) See Table 2 for summary statistics based on the subsamples denoted by structural breaks in the calendar dates depicted in the above plot for output. All lines denote the values for a given calendar date as reported across all vintages of the real-time dataset. See above for further details.

Table 1: Growth Rate and Revision Error Summary Statistics – Output, Prices and Money$^{(*)}$

Vrbl	*Vint*	*R-Err*	*smpl*	$\overline{y}$	$\widehat{\sigma}_y$	$\widehat{\sigma}_{\overline{y}}$	*skew*	*kurt*	*LB*	*JB*	*ADF*
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, $\overline{y}$ is the mean of the series, $\widehat{\sigma}_y$ is the standard error of the series, $\widehat{\sigma}_{\overline{y}}$ is the standard error of $\overline{y}$, *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(*)

$Vrbl$	$Vint$	R-Err	$smpl$	$\overline{y}$	$\widehat{\sigma}_y$	$\widehat{\sigma}_{\overline{y}}$	$skew$	$kurt$	LB	JB	ADF
	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 With and Without Revision Errors(*)

Model	*RevErr*	*Begin Date=1970:1* $k=2$	$k=3$	$k=4$	*Begin Date=1983:1* $k=2$	$k=3$	$k=4$	*Begin Date=1990:1* $k=2$	$k=3$	$k=4$
					Panel A: Output					
Model A	none	*0.00978*	0.01086	0.01088	0.00246	0.00281	0.00285	0.00141	0.00163	0.00162
Model B	none	*0.00978*	0.01036	0.01134	0.00246	0.00258	0.00280	0.00141	0.00162	0.00171
Model C	none	0.01007	0.01134	0.01157	0.00241	0.00277	0.00282	0.00141	0.00164	0.00164
RWd	none	0.00990	0.01123	0.01134	0.00238	0.00282	0.00288	0.00142	0.00171	0.00167
Model C	u_{C1}	0.00972	0.01118	0.01105	0.00241	0.00277	0.00283	0.00141	0.00163	0.00164
Model C	u_{C2}	0.00969	0.01110	0.01096	0.00241	0.00277	0.00284	0.00141	0.00163	0.00164
Model C	u_{C3}	0.01059	0.01191	0.01205	0.00259	0.00296	0.00302	0.00154	0.00176	0.00177
					Panel B: Prices					
Model A	none	0.00206	0.00228	0.00228	0.00063	0.00066	0.00066	0.00041	0.00036	0.00036
Model B	none	0.00206	0.00248	0.00296	0.00063	0.00071	0.00083	0.00041	0.00038	0.00046
Model C	none	0.00217	0.00245	0.00253	0.00067	0.00069	0.00069	0.00042	0.00037	0.00037
RWd	none	0.00601	0.00651	0.00642	0.00341	0.00350	0.00352	0.00257	0.00257	0.00260
Model C	u_{C1}	0.00212	0.00241	0.00246	0.00070	0.00068	0.00071	0.00043	0.00037	0.00038
Model C	u_{C2}	0.00208	0.00230	0.00239	0.00071	0.00070	0.00072	0.00043	0.00037	0.00038
Model C	u_{C3}	0.00235	0.00274	0.00273	0.00069	0.00077	0.00077	0.00043	0.00043	0.00043
					Panel C: Money					
Model A	none	0.01705	0.01695	0.01673	0.01151	0.01140	0.01121	0.00718	0.00730	0.00734
Model B	none	0.01705	0.01883	0.01929	0.01151	0.01419	0.01442	0.00718	0.00922	0.00988
Model C	none	0.01763	0.01739	0.01723	0.01165	0.01153	0.01137	0.00723	0.00739	0.00745
RWd	none	0.02220	0.02228	0.02214	0.01791	0.01796	0.01782	0.01259	0.01290	0.01300
Model C	u_{C1}	0.01766	0.01746	0.01734	0.01181	0.01167	0.01151	0.00739	0.00753	0.00760
Model C	u_{C2}	0.01759	0.01751	0.01727	0.01184	0.01169	0.01153	0.00740	0.00755	0.00760
Model C	u_{C3}	0.01759	0.01763	0.01735	0.01168	0.01157	0.01141	0.00723	0.00739	0.00745

(*) 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. 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*): ${}_{t+k}X_{t+1} = \alpha + \sum_{i=1}^{p} \beta_{i\,t+2-i}X_{t+1-i} + \varepsilon_{t+k}$ Model B *(k^{th} Available Data)* : ${}_{t+k}X_{t+1} = \alpha + \sum_{i=1}^{p} \beta_{i\,t+2-i}X_{t+3-k-i} + \varepsilon_{t+k}$ Model C *(Real-Time Data)* : ${}_{t+k}X_{t+1} = \alpha + \sum_{i=1}^{p} \beta_{i\,t+1}X_{t+1-i}$ $+\varepsilon_{t+k}$ In the above models, X denoes the growth rate of either output, prices, or money. Also, RWd is the random walk with drift model in log levels, $u_{C1} = {}_{t+1}u_{t-k}^{t}$
,k=1;u$_{C2} = {}_{t+1}u_{t-k}^{t}$, $k = 1, 2$; and $u_{C3} = {}_{t+1}u_{t+1-k}^{t+2-k}$, $k = 3$. Further details are contained in Sections 4 and 5.

Table 4: MSFEs Calculated Based on Real-Time Autoregressions and Vector Autoregressions With and Without Money and Revision Errors$^{(*)}$

		Begin Date=1970:1			*Begin Date=1983:1*			*Begin Date=1990:1*		
Model	*Rev Err*	$k=2$	$k=3$	$k=4$	$k=2$	$k=3$	$k=4$	$k=2$	$k=3$	$k=4$
AR - Mod A	none	0.00978	0.01086	0.01088	0.00246	0.00281	0.00285	0.00141	0.00163	0.00162
AR - Mod B	none	0.00978	0.01036	0.01134	0.00246	0.00258	0.00280	0.00141	0.00162	0.00171
AR - Mod C	none	0.01007	0.01134	0.01157	0.00241	0.00277	0.00282	0.00141	0.00164	0.00164
VAR noM - Mod A	none	0.00836	0.00991	0.01041	0.00202	0.00239	0.00243	0.00141	0.00148	0.00151
VAR noM - Mod B	none	0.00836	0.01008	0.01120	0.00202	0.00246	0.00264	0.00141	0.00162	0.00193
VAR noM - Mod C	none	0.00793	0.01014	0.01057	0.00222	0.00276	0.00280	0.00148	0.00159	0.00158
VAR M - Mod A	none	0.00849	0.01170	0.01096	0.00213	0.00256	0.00254	0.00145	0.00158	0.00160
VAR M - Mod B	none	0.00849	0.01034	0.01137	0.00213	0.00286	0.00308	0.00145	0.00187	0.00208
VAR M - Mod C	none	0.00880	0.01004	0.01063	0.00249	0.00276	0.00279	0.00149	0.00161	0.00164
RWd	none	0.00990	0.01123	0.01134	0.00238	0.00282	0.00288	0.00142	0.00171	0.00167
VAR noM - Mod C	u_{C1}	0.00861	0.01059	0.01100	0.00222	0.00287	0.00291	0.00147	0.00158	0.00156
VAR noM - Mod C	u_{C2}	0.00901	0.01079	0.01135	0.00226	0.00269	0.00283	0.00149	0.00166	0.00166
VAR noM - Mod C	u_{C3}	0.00910	0.01052	0.01113	0.00238	0.00277	0.00280	0.00195	0.00216	0.00214
VAR M - Mod C	u_{C1}	0.00933	0.01107	0.01152	0.00245	0.00276	0.00275	0.00160	0.00167	0.00168
VAR M - Mod C	u_{C2}	0.00953	0.01128	0.01203	0.00247	0.00276	0.00281	0.00155	0.00163	0.00169
VAR M - Mod C	u_{C3}	0.00900	0.01065	0.01145	0.00235	0.00272	0.00283	0.00192	0.00214	0.00220

$^{(*)}$ See notes to Table 3. Forecast mean square errors (MSFEs) are reported based on predictions of output 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. All estimated models are either pure autoregressions (denoted AR), vector autoregressions without money (denoted VAR noM), vector autoregressions with money (denoted VAR M), or a random walk model in log levels with drift. Vector autoregression models with money include output, prices, money, and interest rates, whereas those without money simply exclude money and all revisions thereof. Models with revision errors include additional explanatory variables (denoted u_{C1}, u_{C2}, or u_{C3}). In such models, revision errors from each equation in the system are used in all other equations in the system, and lags for each variety of revision error are chosen using the Schwarz Information Criterion (SIC). Lags are also selected using the SIC for all other variables. Definitions of Mod A, Mod B, and Mod C are given in the footnote to Table 3, where they are denoted Models A, B, and C. Revision error definitions are also given in the footnote to Table 3. Further details are contained in Sections 4 and 5.

Table 5: Rationality Test Results[(*)]

Vrbl	k	$M_{1,T}$	*Bootstrap Crit. Val.* 5%	10%
		W Scalar Valued		
Output	2	0.7243	1.0808	1.1926
	3	0.3812	0.5701	0.6030
	4	0.3045	0.4117	0.4410
Prices	2	0.0506	0.0870	0.0970
	3	0.0499	0.0824	0.0868
	4	0.0495	0.0759	0.0862
Money	2	0.0366	0.7539	0.9446
	3	0.0283	0.2893	0.3478
	4	0.0381	0.1894	0.2058
		W Vector Valued		
Output	2	156.84	354.01	388.58
	3	31.431	58.352	63.634
	4	19.863	31.277	32.487
Prices	2	3.1020	4.1520	4.9397
	3	3.2314	1.9002	2.0049
	4	3.1644	2.2276	2.5439
Money	2	0.1436	7.2841	8.2426
	3	0.3421	11.533	13.396
	4	0.3817	11.581	13.464

[(*)] The null hypothesis of data rationality (see discussion of $H_{0,1}$ above) is testsed for two cases: (i) W Scalar Valued (${}_{t+1-j}W_{t-j} ={}_{t+1-j}\ X_{t-j},\ j = 0, 1, ..., t-1$); and (ii) W Vector Valued (${}_{t+1-j}W_{t-j} = \left[{}_{t+1-j}X_{t-j}, {}_{t+1-j}\, u^{t-j}_{t-1-j}\right]'$ $j = 0, 1, ..., t-1$). The third column of the table gives test statistic values. Bootstrap critical values are reported in the 4th and 5th columns. See Section 5.3 for complete details.