The Real-Time Predictive Content of Money for Output*

Jeffery D. Amato
Bank for International Settlements

Norman R. Swanson
Texas A&M University

this revision: June 2000

Abstract

Data on monetary aggregates are subject to periodic *redefinitions*, presumably in part to improve their link to measures of output. Money data are also *revised* on a regular basis. Taking these data imperfections into account, we reassess the evidence on the marginal predictive content of M1 and M2 for real and nominal output. In particular, by first using the latest version of the data that is available, and then using sequences of historical time series that would have been available to forecasters in real-time, we are able to provide a comprehensive assessment of whether money is useful for predicting output. We conclude that the generally significant marginal predictive content of M1 or M2 for output that is found using a recently revised data set is not duplicated in a real-time setting, although M2 is shown to remain useful when 1-year ahead forecasts are constructed using fitted vector autoregressive models.

---

* Jeffery D. Amato, 4002 Basel, Switzerland (jeffery.amato@bis.org) and Norman R. Swanson, Department of Economics, Texas A&M University, College Station, TX 77843-4228 (nswanson@econ.tamu.edu). We gratefully acknowledge comments from Todd Clark, Dean Croushore, Athanasios Orphanides, and Todd Stark. Additionally, comments from seminar participants at the Federal Reserve Board, the 2000 Winter meeting of the American Economic Association, the editor, and, especially, an anonymous referee are appreciated. Excellent research assistance was provided by Myles Callan and ??????????. Swanson thanks the Private Enterprise Research Center and the Bush Program in the Economics of Public Policy at Texas A&M University for financial support. An earlier version of this paper was written while Amato was an economist at the Federal Reserve Bank of Kansas City. The views expressed herein are those of the authors and do not necessarily reflect those of the Bank for International Settlements, the Federal Reserve Bank of Kansas City, or the Federal Reserve System.

1 Introduction

The causal link between money and output is an ongoing issue of primary importance in macroeconomics. In particular, numerous studies over the last decade have attempted to assess whether money improves forecasts of output, beyond what can be achieved by using the history of other macroeconomic variables. Despite the immense amount of work carried out on this subject, the extent of the marginal predictive content of money for output remains largely uncertain. However, some recent studies (e.g. Becketti and Morris (1992), Feldstein and Stock (1994), Hafer and Kutan (1997), Stock and Watson (1989), and Swanson (1998)) have provided strong evidence to support the hypothesis that money is useful for predicting real and nominal output, at both monthly and quarterly frequencies.[1] These authors - and most contributors to this literature - focus their analysis on the properties of fitted residuals from regressions and, to a lesser extent, on out-of-sample prediction errors. In all cases, however, analyses are based on the use of substantially revised data.

As Diebold and Rudebusch (1991) point out for the case of the composite leading index (CLI), using revised data for investigating predictive ability can provide a distorted picture since the CLI is regularly subjected to *ex post* redefinitions to strengthen its historical link to output. The same criticism can be levelled at monetary aggregates, which are also occasionally subjected to redefinitions, presumably in part to improve their historical link with output. Redefinitions aside, monetary aggregates are also continually revised because of incomplete data collection and seasonal factor adjustments, for example. The combination of redefinitions and revisions suggests that it would be useful to reconsider the evidence on the predictive content of monetary aggregates for output. This is particularly true when out-of-sample prediction errors are used to examine the data, as these errors may change depending on the extent to which the revision process has been completed for each element of a data series.

Our approach is to provide a comprehensive analysis of the real-time marginal predictive content of M1 and M2 for both real and nominal output, at monthly and quarterly frequencies. The methoodology which we use to carry out our analysis follows closely that used by Diebold and Rudebusch (1991). To lay the groundwork for our real-time results, we first provide evidence based on the latest version of the data which is available. Some of these results confirm earlier findings in the literature, while others are new and therefore provide a fuller picture of the marginal significance

[1]Opposing evidence is discussed in Friedman and Kuttner (1992) and Friedman (1997), for example.

of money in output models. We then consider whether using real-time data in the same types of computations alters our conclusions about the usefulness of money. In so doing, we bridge the large literature that has investigated money-income causality with the growing literature that utilizes real-time data sets.[2]

Three of our primary findings can be summarized as follows. First, for in-sample analysis, both M1 and M2 appear to have marginal predictive ability for output when the latest version of data is used, but this weakens, or disappears, for real-time data. Second, neither M1 or M2 has marginal predictive content for output when correctly designed real-time out-of-sample experiments are carried out, with the noteable exception of M2, which is found to be useful when 1-year ahead forecasts based on fitted linear vector autoregressive (VAR) models are constructed. Indeed, VAR and vector error correction (VEC) models are found to perform better in most cases when money is *not* included, based on the application of Diebold and Mariano (1995) predictive ability tests. Finally, we find that, in many instances, VEC models outperform first-difference VAR models, in contrast to much of the literature which has found that VAR models typically produce more accurate forecasts of macroeconomic variables.

The remainder of the paper is organized as follows. In the next section, we briefly discuss the real-time data sets which we use, including the nature of the revisions and redefinitions to the output, price, and money series. In the third section, we present evidence on the in-sample marginal predictive content of M1 and M2 for output. In the fourth section, we carry out a truly *ex ante* forecasting experiment. This experiment is *ex ante* in the sense that only data available in real-time are used to construct sequences of forecasts. In addition, our predictive ability summary measures and tests based on *ex ante* forecasts can be viewed as out-of-sample Granger causality criteria, in the sense of Ashley, Granger, and Schmalensee (1980), where it is stated on page 1149 that: " *... a sound and natural approach to such tests [Granger causality tests] must rely primarily on the out-of-sample forecasting performance of models relating the original (non-prewhitened) series of interest*." Some concluding remarks are made in the final section.

[2]Other recent studies which utilize real-time data sets include Croushore and Stark (1999), Ghysels, Callan, and Swanson (1999), Orphanides (1998), and Rudebusch (1998).

2 Data

The variables used throughout this paper are nominal output, real output, the price level, M1, M2, and a short-term nominal interest rate. For series that are observed at a monthly frequency, we use industrial production (IP), the consumer price index - all items (CPI), M1, M2, and the secondary market rate on ninety-day United States Treasury bills. For series that are observed at a quarterly frequency, we use nominal GNP (NGNP), nominal GDP (NGDP), real GNP (RGNP), real GDP (RGDP), the implicit GNP deflator, and the GDP chain-weighted price index. Whether GNP or GDP is used depends on which series was favored by the Bureau of Economic Analysis in the National Income and Product Accounts at each point in time (hereafter, we simply refer to GDP, even if GNP is the relevant series). Data for the last month within each quarter is used to construct quarterly money and interest rates series. Except for the T-bill rate, all series were published seasonally adjusted.

In order to assess the impact of data revisions and redefinitions on measures of the predictive content of money for output, we have assembled a sequence of *real-time* data sets. A real-time data set for the period 1978:1, say, contains historical observations which were actually available to forecasters in 1978:1. In this way, a complete real-time data set contains a unique vector of real-time observations for each date in the sample. In principle, then, a real-time data set can be constructed for each period in the past, although the construction of such data sets is clearly tedious unless it is done as data become available.[3] For data at a monthly frequency, we compiled a vector of real-time observations available at each date from January 1978 until December 1997. Each of these real-time data vectors contains observations starting in 1959:1. Likewise, we use vectors of quarterly observations dating back to 1959:1 that would have been available to forecasters in each quarter from the first quarter of 1978 to the fourth quarter of 1997. Since new and revised observations are released at various points within the month (quarter), we chose roughly the 15th of the month (15th of the middle month of the quarter) as the cut-off date for data that was deemed to be available to forecasters in a particular period. In all cases, this means that in period t, the latest available observation is for period $t-1$ (i.e. there is a one period lag in data availability).

[3] Diebold and Rudebusch (1988, 1991) construct one of the first comprehensive real-time data sets. The data which they collect is for the composite leading index, and their papers detail completely the structure of real-time data sets.

Data on M1 and M2 are subject to three types of changes: near-term revisions, rebenchmarkings, and redefinitions. It is common for these series to undergo the first type of change, near-term revisions, in the weeks following their initial release. Revised versions of past data are published along with new observations in the Federal Reserve's weekly *H.6 - Money Stock, Liquid Assets & Debt Measures* statistical release. These near-term revisions are mainly a result of incomplete source data from depository institutions. Since most of these types of revisions occur within a month of the initial release date, new observations in our quarterly real-time data sets are based on already revised source data. As well, short-term revisions do not effect many changes in recently released data on a quarter-to-quarter basis. New observations in the monthly real-time data sets, however, are subjected to larger changes in the proceeding month. The second type of change, rebenchmarkings, are typically made each February, and published by the Federal Reserve in a volume called *Money Stock Revisions*. A rebenchmarking applies to data ranging from the fourth quarter two years prior to the fourth quarter of the preceding year, inclusive. The first part of a rebenchmarking involves revisions to the series due to additional source data obtained by the Federal Reserve after the near-term revisions have been completed. The second, more significant, part of a rebenchmarking involves updating the previous year's seasonal adjustment factors. Lastly, both M1 and M2 have been subjected to numerous redefinitions since they were introduced in 1960 and 1971, respectively.[4] With regard to M2, it is worth noting that there have been six redefinitions of the series since 1980. The redefinitions have largely been responses to financial innovation, presumably in order to maintain links between money and other macroeconomic variables. We refer the reader to Anderson and Kavjecz (1994) for a more detailed account of the revision and redefinition history of our monetary aggregates.[5]

To create the real-time money data sets, we used the full historical M1 and M2 series published in each yearly *Money Stock Revisions* as the basis for all of the quarterly data sets for that year, and for the monthly data sets from February of that year to January of the following year. Consistent

[4]The earliest version of a series that corresponds to present day M1 was the first monetary aggregate published by the Federal Reserve beginning in November 1960, simply called the *money supply*. It was subsequently redefined (at one time, two definitions existed, M1A and M1/M1B), and finally became M1 as it is currently defined in February 1988.

[5]Our real-time data sets for quarterly output and prices were obtained from the website of the Federal Reserve Bank of Philadelphia (http://www.phil.frb.org/econ/forecast/rtd.html). The data are discussed in detail in Croushore and Stark (1999).

with our timing convention, the monthly money series were supplemented by data provided in the *H.6* that was published closest to the middle of the month (i.e. between the 12th and 18th); the quarterly series were created analogously based on the appropriate *H.6* for the middle month of each quarter.[6]

In accord with our timing assumption, IP and CPI data are released on or around the 15th of each month, and a typical months' release of data for these variables is comprised of a first, or *preliminary*, release for the previous month, and one to five months of revisions to data previously released. In addition, more comprehensive rebenchmarking and base year revisions occur from time to time for each of these variables. Corresponding to our money data and our quarterly GDP data, we have collected monthly CPI and IP data beginning in 1959:1, and have constructed real-time data sets for the dates 1978:1-1997:12.

Turning first to the IP data, the following details are worth noting. Seasonally adjusted IP figures are compiled by the Federal Reserve Board. The primary source for seasonally adjusted IP data is the *Federal Reserve Bulletin*. Additional data for these series were obtained from Federal Reserve monthly statistical releases. Federal Reserve releases for IP are called *G.12.3* before May 1990, and *G.17* thereafter. Recent releases, and a partial real-time data set from 1972 onwards for seasonally adjusted data, can also be obtained from the Federal Reserve Board's website (http://www.bog.frb.fed.us/releases).

Aside from typical monthly revisions to recently released data, there have been various major updates to IP. Numerous updates include rebenchmarkings, for which at least 10 years of data were revised. For example, in the January 1997 rebenchmarking, it was announced that, henceforth, reformulation of indices based on updated weights would occur annually rather than every five years, as had been previously done. Also, for three of the major (rebenchmark and redefinition) revisions to IP, the Federal Reserve Board released separate publications - (1) *Industrial Production 1957-59 Base*, (2) *Industrial Production 1971,* and (3) *Industrial Production, 1976 Revision*.[7]

Our real-time data set for seasonally adjusted CPI was constructed from Federal Reserve Bank

[6]There is one significant exception to these timing rules. In 1981, *Money Stock Revisions* was not published until July. Consequently, for dates prior to this, in which otherwise we would have drawn upon the *Money Stock Revisions* published in February as a basis for historical observations, we used instead the full historical series already gathered for January of that year, and updated these with the appropriate *H.6* releases, as described in the text.

[7]As a result of a major revision, there is a missing entry in 1985:3. We replaced the missing observation with the first available data for that period (the second release).

of St. Louis publications. The main source of this data is *National Economic Trends*. However, recent releases of the CPI can be obtained from http://www.stls.frb.org/fred/dataindx.html. In general, benchmark revisions to the CPI occur every 12 months, at which time revisions to the data for the preceding 12 months are reported. There is one base year revision: January 1988 (see *National Economic Trends*, pp. iii - August 1988 for details).

The fact that many data series possibly face regular updates means that we can never claim to have a final record of historical data which is immune from potential future revision. Nonetheless, for the purpose of providing an *ex post* benchmark with which to compare the real-time forecasting properties of our models, we obtained a full sample of data that had already been subjected to one and a half years of revisions beyond the end date of our sample period.[8] We call this benchmark data set our *final (revised)* data. In closing our discussion of real-time data, it is worth stressing that observations on T-bill rates are not subject to revision, so that *final* T-bill rate data are also real-time.

Visual evidence of the extent to which revisions have impacted RGDP, NGDP, M1 and M2 is provided in Figures 1-2. Figure 1 shows plots of *preliminary* and *final* data (expressed as growth rates) of RGDP and M2. In Figure 2, each panel contains plots of the *differences* between preliminary and final data for all four quarterly series. In most periods, preliminary and final growth rates differ, and in some periods this difference is clearly substantial. For example, the average absolute revisions (in annualised percentage points) in these series are 1.35 (RGDP), 1.25 (NGDP), XX (M1), and 0.97 (M2).[9] The largest revisions in the plots are 7.78, 6.57, XX, and 4.79, respectively. In addition, note that the mean values of IP revisions, formed as $\ln(IP_{t+i}(t-1)) - \ln(IP_{t+i-1}(t-1))$, $i = 1, .., 12$ are significantly different from zero at the 5% nominal level in 8 of 12 cases, where the subscript refers to the release date of the data, and the bracketed index denotes the date to which the data pertains.[10] Similar results arise upon examination of our GDP, price index and monetary

[8]This final revised data set contains observations from the first period of 1959 through the last period of 1997. It was compiled in June 1999. (All final revised data used in this paper were obtained from the Federal Reserve Board's database.)

[9]For monthly data (not shown), the average absolute revisions (in annualised percentage points) are 4.91 (IP), XX (M1), and 1.72 (M2); the largest revisions are 27.46, XX, and 6.93, respectively.

[10]Test statistics are formed by dividing the mean revision by a heteroskedasticity and autocorrelation consistent (HAC) estimate of the standard error. The 12 mean values, begining with the value for $i = 1$, are: 0.076, 0.096, 0.023, 0.015, 0.014, 0.014, 0.014, 0.015, 0.015, 0.015, 0.005, and 0.014.

aggregate variables.

3 In-Sample Evidence

3.1 Methods

The first type of marginal predictive evidence which we consider is obtained from the fitted residuals of estimated linear time series equations. We estimate two types of equations: ones specified using logged first-differences of the variables (equations from a vector autoregression of order p - VAR(p) model) and ones specified using first-differences and error-correction terms (equations from a vector error correction - VEC(p) model). Our decision to use time series models with multiple variables reflects our desire to assess the marginal predictive content of money for output, after controlling for the influence of other common macroeconomic time series. In particular, lags of output and price series are commonly included as regressors in reduced-form exercises of the type considered here. In addition, many authors have reported that when a short-term interest rate, or the spread between the commercial paper and T-bill rates, is included in these reduced-form equations, money no longer has predictive ability in the output equation (e.g. Sims (1980) and Friedman and Kuttner (1992)). For this reason, we also include a short-term interest rate in all of the estimated equations to determine what money - a financial quantity - can explain beyond that explained by interest rates - a financial price. Including an interest rate has the additional advantage that it serves as a proxy for the opportunity cost of holding money in what may be long-run money demand relations in the VEC(p) models.[11]

Note that we specify equations in terms of the logged first-differences of all of the variables in view of the substantial evidence suggesting that these variables are characterized by the property that their first differences are second-order stationary, while their levels are not.[12] In addition, we allow for cointegration because many authors report evidence of cointegrating relationships among the variables which we are examining (e.g. Stock and Watson (1993)). In addition, allowing for the possibility of cointegration addresses Feldstein and Stock's (1994) critique that tests of Granger-

[11]The T-bill rate serves as a proxy for the opportunity cost of holding either M1 or M2 balances. A better measure of this opportunity cost would be the difference between the T-bill rate and the rate of return on holding M1 or M2 balances, respectively.

[12]An earlier version of this paper reported evidence on the integration properties of the series used here.

causality may be misleading if cointegrating relations are wrongly omitted from the model.

In total, we estimate 36 different models for any given data set (real-time or final). For the real output equations, the models are distinguished by: (i) the frequency of data - monthly or quarterly, (ii) the shortest time lag between the regressand and the regressors - 1-period or 1-year ahead., (iii) the exclusion (*small* model) or inclusion (*big* model) of M1 or M2 in the estimated models, and (iv) the choice of dependent (or *target)* variable - RGDP, NGDP, or IP. Our VAR models are estimated using least squares, and the equation of interest from the models can be written as follows:

$$\Delta y_t = \mu + \sum_{i=1}^{k} a_i \Delta y_{t-j-i} + \sum_{i=1}^{k} b_i \Delta p_{t-j-i} + \sum_{i=1}^{k} c_i \Delta R_{t-j-i} + \sum_{i=1}^{k} d_i \Delta m_{t-j-i} + \epsilon_t, \tag{1}$$

where j represents the prediction step (for quarterly data, $j = 1, 4$; for monthly data, $j = 1, 12$), Δy_t is the log first-difference of either real (RGDP, IP) or nominal (NGDP) output, Δp_t is the log first-difference of the price index, ΔR_t is the first-difference of the interest rate, Δm_t is the log first-difference of money, and ϵ_t is a residual. The lag length, k, is set equal to 9 (3) for monthly (quarterly) data.[13] The *small* models which exclude the money terms impose the restriction that $d_i = 0$ $(i = 1, ..., k)$.[14]

Similarly, our the equation of interest from our VEC models can be written as:

$$\Delta y_t = \mu + \sum_{i=1}^{k} a_i \Delta y_{t-j-i} + \sum_{i=1}^{k} b_i \Delta p_{t-j-i} + \sum_{i=1}^{k} c_i \Delta R_{t-j-i} + \sum_{i=1}^{k} d_i \Delta m_{t-j-i} + \sum_{r=1}^{h} \gamma_r \eta_{r,t-j} + \epsilon_t, \tag{2}$$

where the $\eta_{r,t}$ $(r = 1, ..., h)$ are the error-correction variables constructed using the maximum likelihood approach of Johansen (1988, 1991). Here, *small* models are characterized by the additional restriction that the cointegrating space of the variables is estimated after excluding the relevant money variable from the model.

To test for the in-sample marginal significance of money in the VAR models, we calculate standard Wald statistics from recursive estimation of 201 monthly models and 67 quarterly models. When $j = 1$, the monthly data samples start with 1959:1-1978:4 and end with 1959:1-1994:12; similarly, the quarterly data samples start with 1959:1-1978:2 and end with 1959:1-1994:4.[15] In addition, model parameters and cointegrating spaces are re-estimated before each new Granger

[13]These choices reflect common practice in the literature mentioned above. We obtain similar results if k is chosen based on the Schwarz Information Criterion, with a maximum allowable 12 (4) lags for monthly (quarterly) data.

[14]When Wald F-tests of the Granger non-causality null hypothesis are constructed, the lag lengths chosen for the *big* models are also applied to the *small* models.

[15]The timing of the samples used in our in-sample analysis can be explained by considering our quarterly sample periods, and noting that our last real-time data set was for the period 1997:4. At this time, data for the period

causality test statistic is constructed. In order to provide an alternative in-sample predictive ability measure, we also calculated sequences of Schwarz Information Criteria (SIC) values for the *small* and *big* VEC models. These criteria were compared in order to assess the number of times that the *big* model which includes money was *preferred* to the *small* model.[16]

3.2 Results

Table 1 summarizes our in-sample results. Panel A reports findings based on the use of *final* data, while those in Panel B correspond to the use of real-time data in all estimations.[17] The first column in the table states which measure of money is used in the *big* model; the second column reports the average cointegrating rank (i.e. h) selected across the sub-samples (with standard error in parentheses)[18]; columns three and four report the proportion of times that the null hypothesis that money does not Granger cause output is rejected at a 5% level in the VAR models; and columns five and six report the proportion of times that the *big* VEC model with money in it achieves a lower (i.e. better) SIC value than the *small* VEC model without money in it

Focusing first on Panel A, the average dimension of the cointegrating space is close to two at both frequencies, with substantial variability from sub-sample to sub-sample. Interestingly, the SIC

1997:3 were available. Thus, we could in principle extend our in-sample analysis beyond 1994:4. However, in order to facilitate comparison with our out-of-sample results, we need to take into account the fact that we compare our real-time forecasts with preliminary data, first revised data, etc., up to data which have been revised 12 times. In order to compare a forecast made in 1997:3, say, with the 12^{th} revision of actual realized data for the period 1997:4, we would need to have data available from calendar period 2000:4, as this is the calendar date at which the 12^{th} revision of data for period 1997:3 will be available. This is clearly not feasible. Hence, in order to line up the end points of the samples used in our in-sample analysis with the forecast period examined in our out-of-sample analysis, we end our in-sample period in 1994:4 (which is the calendar date of the last real-time data set used to construct forecasts in our out-of-sample analysis). For the 1-year ahead horizon ($j = 4$), end points of samples are three quarters (11 months for monthly data) earlier.

[16] See Granger, King, and White (1995) for a discussion of the use of model selection criteria to choose between competing models.

[17] Recall that all estimations are based on recursive samples, and that for *final* data, all recursive samples are drawn from a single vector of observations, while each real-time vector of observations (all such vectors taken together constitute a real-time data set) is by construction a recursive sample.

[18] Average cointegrating ranks are only reported for the *big* model, although different cointegrating ranks are estimated for the *small* model, and do differ from those estimated for the *big* model (especially if money enters the cointegrating relationships in the *big* model, for example).

selects VEC models with money 50 to 100% of the time in most cases, regardless of data frequency, money measure, or horizon. However, the evidence is generally less favorable for money when real-time data is used (compare the last two columns in Panels A and B). In 5 of the 12 cases, the money models are rejected more than 50% of the time; for quarterly data, the largest proportion of times a model with money is selected is only 58%. The in-sample evidence from F-test rejection frequencies based on VAR models is more mixed when comparing final and real-time data results. In Panel A, M1 and M2 appear useful for predicting output at a 1-period horizon based on final data. Given that most studies focus on the 1-step horizon, this result is not surprising, and agrees with much of the recent evidence presented in the literature, e.g., Feldstein and Stock (1994) for quarterly data and Swanson (1998) for monthly data. When real-time data is used, the results are less favourable for M1, but M2 still appears to have marginal predictive power. Finally, both M1 and M2 are uniformly useless at the 1-year horizon regardless of data type. As indicated by the average cointegrating ranks, the VAR models appear to be misspecified, and therefore more attention should be paid to results based on VEC models, both here and in the next section. Thus, overall, these results suggest that the significant marginal predictive content of money that is reported in the recent literature, and confirmed here for the most part, does not carry over to real-time data. However, we have yet to assess the predictive content of money for output in a truly *ex ante* forecasting scenario. This is done in the next section.

4 Out-of-Sample Evidence

4.1 Methods

If one's primary interest is to forecast future output, then it is natural to assess the marginal predictive content of money for output using some sort of tests of out-of-sample predictive ability (as discussed in Ashley, Granger and Schmalensee (1980), for example). If, in addition, one is interested in considering models which may be useful for real-time decision making, then one must also be careful not to use a *single* vector of time series data when constructing *sequences* of forecasts,[19] and instead use real-time data sets which were truly available at the point in time during which each of the forecasts were made. In short, one must go not only beyond the analysis

[19]For example, if one retrieves a series from CITIBASE, then that series is only truly real-time for the calendar date at which time the last observation of the sample was made public.

of in-sample predictability, but also beyond the naive approach of creating sequences of forecasts using data which are not real-time and erroneously calling such forecasts *ex ante.* Accordingly, in this section we analyze the properties of various truly *ex ante* sequences of 1-step and 1-year ahead forecasts of output.

The estimation procedures used here are the same as those used in our in-sample analysis, except that real-time vectors of observations available at each point in time are used to produce out-of-sample 1-step and 1-year ahead forecasts, using VAR and VEC models. These forecasts are compared with actual data in order to form sequences of forecast errors. However, note that it is not obvious which *actual* data our forecasts should be compared to. In particular, it is not clear what the true realized value is that should be used as a basis for comparing forecasts, since a given data point is potentially continually subject to revision. As pointed out by Robertson and Tallman (1998), the most relevant measure of the truth for judging forecasts may not be a value reported many years after the fact, but one that is available for forecast evaluation within reasonable proximity of the calendar date at which time a forecast is made. This is particularly true in financial markets, for example, where markets clearly react to macroeconomic announcements of preliminary figures, while they may not react as vigorously to announcements of revisions. On the other hand, policy setters, for example, may be interested in forming preliminary forecasts which are as close as possible to some *final* figure, hence suggesting that forecast errors be constructed using data which have been revised many times. However, in the face of radical redefinitions to the targeted series being forecasted, it may not be reasonable to compare real-time forecasts that are constructed to predict one definition of a series to a different definition of the series which has been adopted many years later, for example.

In response to the dilemma of how to choose the most appropriate actual realization of a variable when forming forecast errors with which to assess the usefulness of competing models, we compare forecasts to many different "vintages" of realized observations. Specifically, we extract four time series on which to base forecast comparisons from our sequences of real-time data sets, namely: values available after one period (*vint*1),[20] one year (*vint*4 or *vint*12), two years (*vint*8 or *vint*24), and three years (*vint*12 or *vin*t36). We also use the final revised (*final*) data in our comparisons, corresponding to the real-time forecast analysis of Diebold and Rudebusch (1991).

[20]These correspond to our *preliminary* releases of data.

The forecast errors are then used to construct mean square forecast error (MSFE) criteria. In addition, the forecast errors are used to form Diebold and Mariano (DM: 1995) predictive ability test statistics.[21] The MSFE criteria and DM statistics are constructed in order to compare *big* models which contain money with *small* models which do not contain money. This in turn allows us to choose between the models. When models with money are chosen, we have direct *ex ante* forecasting evidence that fluctuations in the money stock anticipate fluctuations in output. In order to facilitate comparison with what we will call our *erroneous method*, we also form sequences of forecasts using only *final* revised data, and compare these forecasts with *final* realizations. These forecasts are clearly not real-time, and are subject to the problems discussed above.

4.2 Results

Our out-of-sample forecasting results are contained in Tables 2 (GDP) and 3 (IP). The tables are broken into panels corresponding to money measure and forecast *target* variable combinations. For purposes of comparison, we include a first row in each panel which contains results based on fitted in-sample residuals (in-sample forecast errors) from equations estimated using the full sample of *final* data. The second row reports results based on sequences of out-of-sample forecasts constructed using sub-samples of the *final* data. Thus, the second row reports the results based on a method which we refer to above as our *erroneous method.* The remaining rows in each panel of the tables correspond to out-of-sample results based on forecasts constructed using real-time data. The entries in the tables are organized as follows. The first column of entries lists the forecast type (i.e. *in* for in-sample and *out* for out-of sample). The second and third columns describe the data sets used to estimate the forecast models and to form the forecast errors, respectively. Columns four and five report MSFEs based on 1-step ahead predictions from the VAR models with and without

[21] The DM tests which we construct correspond to our MSFE criterion, and are formed by first setting $d_t = e_{1t}^2 - e_{2t}^2$, where e_{1t} and e_{2t} are the forecast errors associated with our *big* and *small* forecast models, respectively. The null hypothesis of equal predictive ability is then tested by forming the statistic $dm = \overline{d}/\widehat{\sigma}_{\overline{d}}$, where $\overline{d} = \sum_{t=1}^{P} d_t$, $\widehat{\sigma}_{\overline{d}}$ is a HAC estimator of the standard error of $\overline{d}$, P is the out-of-sample forecast period, and dm has a nonstandard limiting distribution (see McCracken (1998) for complete details). Additionally, because critical values for the test depend on the rate at which the in-sample and out-of-sample periods grow with respect to each other (say π) as T increases, there is some question as to the appropriate critical values to use in our empirical exercises. For our puposes, unity was used as the 5% critical value (see McCracken (1998) for further discussion). Alternative predictive ability tests are discussed in Chao, Corradi, and Swanson (2000), for example.

money, while columns six and seven are the same, except that VEC models are used. Columns eight through eleven contain 1-year ahead prediction results which are analogous to the 1-step ahead results contained in columns four through seven. An asterisk beside an entry in the table denotes significantly better predictive ability (based on the application of the DM test at a 5% level). Corresponding to our in-sample results, Table 2 contains quarterly MSFEs from forecasts for the period 1978:2-1994:4, while Table 3 contains monthly results for the period 1978:4-1994:12.

A number of clear results emerge upon examination of the tables. For example, notice that when in-sample MSFEs are calculated from the full sample of *final* data[22] (first row of the panels in Tables 2 and 3), models with M1 (or M2) essentially always achieve lower MSFEs than models without M1 (or M2). This result essentially holds regardless of forecast horizon, data frequency, and model type (VAR or VEC), and is perhaps not surprising, since MSFEs in the first row of the three M2 panels in Tables 2 and 3 are somewhat reflective of the statistics reported in Table 1.[23]

Recall that the second row of entries in each panel of Tables 2 and 3 corresponds to our *erroneous method.* Note that for quarterly data (Table 2), use of this method suggests that M1 is not useful for predicting output, while M2 is, again regardless of forecast horizon, data frequency, and model type (VAR or VEC)of data frequency. However, although *big* model MSFEs are smaller when M2 is used, these differences are only significant at the annual forecast horizon. For example, in Panels B and D of Table 2, when VAR models are used, the *big* model MSFEs are 16.40% and 19.70%, respectively, while the corresponding *small* model MSFEs are 17.43% and 21.13%, and DM tests based on pairwise comparison of the alternative models reject the null hypothesis of equal predictive ability. Thus, even when out-of-sample forecasting is carried out, M2 still appears useful, as long as our *erroneous method* is used to construct the forecasts.[24] The findings differ markedly, however, when real-time data are used to form truly *ex ante* forecasts.

The main conclusion which is immediately apparent upon examination of the real-time results contained in the third through seventh rows of each panel in Tables 2 and 3 is that not only is M1 *not* useful for predicting output, but M2 is also *not* useful, with the exception that M2 is useful for

[22]Note that Diebold-Mariano predictive ability test statistics were not calculated for these cases, as the MSFEs for these cases are not based on out-of-sample predictions.

[23]Exact comparisons to Table 1 are not possible since these results are based on the *full* sample of final data, whereas the values in Panel A of Table 1 are based on averages across recursive samples.

[24]Results based on monthly data (Table 3) are less clear, although the general patterns of findings discussed above generally hold (see below for further details).

1-year ahead forecasting when VAR models are estimated. For example, note that when comparing MSFEs based on VEC models in Table 2, the *small* models always outperform the *big* models, regardless of forecast horizon, money measure, and output measure used, and the differences in forecasting ability are significant in many cases. This finding is immune to the data which are used to evaluate the forecasts, as the five different rows correspond to forecast errors constructed using different vintages of data, from *preliminary (vint1)* up to *final.* Interestingly, this finding does not hold up for M2 when VAR models are used instead of VEC models, and 1-year ahead forecasts are constructed. Indeed, in this particular case, models with M2 significantly (based on DM tests) outperform models without M2. Taken together with the fact that 1-year ahead VAR MSE values are lower than corresponding VEC MSE values, we have some evidence that M2 is in fact useful for predicting output, at least at the 1-year horizon.

Out-of-sample forecasts from monthly data paint a somewhat different picture. In particular, there appears nothing to choose between models with and without money, as there are only two cases for which one model outperforms the other, based on the DM test (and in both of these cases, the model without money "wins"). Note also how poorly the models forecast IP growth, as evidenced by the large MSFE values, particularly when compared with corresponding MSFEs from our quarterly results. This suggests that IP is not only difficult to forecast, but also that output predictions based on quarterly data may be preferred to those based on more noisy monthly data.

Turning again to Table 2, a final interesting observation can be made by comparing MSFEs for similar models and money measures *across* data vintage (i.e. data used for the forecast comparison). In particular, constructing forecast errors by subtracting real-time forecasts from *preliminary* actual data yields substantially lower MSFEs than in all other cases. In some sense, this is not surprising, given that much of the most recent data used in the construction of the forecast are almost preliminary, in the sense that they have been revised very little. Put another way, if the objective of the real-time forecaster is to construct a forecast as close as possible to some *final* value, then she/he should not expect to produce as accurate a picture of the future as he/she could have, had the objective been to predict *preliminary data,* highlighting the potential perils involved with basing forecasting assessments solely on data revised many years after the fact. However, notice that the *ranking* of the models does not change with vintage, as models without money generally outperform those with money regardless of vintage, although there is one noteable exception, as discussed above..

A visual check for the sub-period robustness of the results in Tables 2 and 3 is provided in Figure 2. The plots in this figure are of the differences between the absolute forecast errors of the *small* and *big* models used to predict RGDP and IP at a one-period horizon, when real-time data are used in forecast construction and *final* data are used for comparison purposes.[25] Our finding that money has little predictive content across the full range of samples is mirrored by the fact that the plots are not regularly above the zero line. There are periods when the forecast differences are small and periods when the differences are large. However, there is rarely a prolonged period in which one model dominates the other, even though the models sometimes greatly outperform one another, albeit by roughly equal magnitudes.

5 Conclusions

Our results show that one's perception of statistical relationships can be affected by the vintage of data used in forecast evaluations. Our most striking finding is that the significant marginal predictive content of M2 for output at the quarterly forecast horizon which is obtained using fully revised data disappears when data that were available in real-time are used, although M2 remains useful for constructing annula predictions in some cases. In addition, we present a substantial body of evidence suggesting that M1 has no marginal predictive ability for output, regardless of forecast horizon and of the data used to perform the evaluations. Indeed, we provide evidence that using money data in real-time forecasting may actually make matters worse.

Our conclusions suggest it may be useful to look at other (less broad) monetary aggregates, which may be subject to fewer revisions, e.g., the monetary base. Alternatively, in light of criticisms that have been levelled against simple-sum aggregates, a worthwhile topic for future research is to analyze the real-time properties of weighted money aggregates, like the Divisia index (Barnett, 1980).

[25]The figure provides evidence on *absolute* forecast errors, whereas the tables provide evidence on *squared* forecast errors.

References

[1] Anderson, R.G. and K.A. Kavajecz (1994): "A Historical Perspective on the Federal Reserve's Monetary Aggregates: Definition, Construction and Targeting", *Federal Reserve Bank of St. Louis Review*, March-April, 1-31.

[2] Ashley, R., C.W.J. Granger, and R. Schmalensee (1980): "Advertising and Aggregate Consumption: An Analysis of Causality", *Econometrica,* (48), 1149-1167.

[3] Barnett, W.A. (1980): "Economic Monetary Aggregates: An Application of Index Number and Aggregation Theory", *Journal of Econometrics*, (14) 11-48.

[4] Becketti, S. and C. Morris (1992): "Does Money Matter Anymore? A Comment on Friedman and Kuttner", Working Paper, Federal Reserve Bank of Kansas City.

[5] Chao, J., V. Corradi, and N.R. Swanson (2000), "An Out of Sample Test for Granger Causality", *Macroeconomic Dynamics*, forthcoming.

[6] Clark, T. and M.W. McCracken (1999): "Granger Causality and Tests of Equal Forecast Accuracy and Encompassing", Working Paper, Federal Reserve Bank of Kansas City.

[7] Croushore, D. and T. Stark (1999): "A Real-Time Data Set for Macroeconomists", Working Paper, Federal Reserve Bank of Philadelphia.

[8] Diebold, F.X. and R.S. Mariano (1995): "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, (13), 253-263.

[9] Diebold, F.X. and G.D. Rudebusch (1988): "Stochastic Properties of Revisions in Index of Leading Indicators", in: *Proceeding of the Business and Economic Statistics Section, American Statistical Association*, 712-717.

[10] 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.

[11] Feldstein, M. and J.H. Stock (1994): "The Use of a Monetary Aggregate to Target Nominal GDP", in G.N. Mankiw, (ed.), *Monetary Policy*, Chicago: Chicago University Press, 7-62.

[12] Friedman, B.M. (1997): "The Rise and Fall of Money Growth Targets as Guidelines for U.S. Monetary Policy", in I. Kuroda, (ed.), *Towards More Effective Monetary Policy*, New York: St. Martin's Press, 137-164.

[13] Friedman, B.M. and K.N. Kuttner (1992): "Money, Income, Prices and Interest Rates", *American Economic Review*, (82), 472-492.

[14] Ghysels, E., N.R. Swanson and M. Callan (1999): "Monetary Policy Rules with Model and Data Uncertainty", Working Paper, Texas A&M University.

[15] Granger, C.W.J., M.L. King and H. White (1995): "Comments on Testing Economic Theories and the Use of Model Selection Criteria", *Journal of Econometrics*, (67), 173-187.

[16] Hafer, R.W. and A.M. Kutan (1997): "More Evidence on the Money-Output Relationship", *Economic Inquiry*, (35), 48-58.

[17] Johansen, S. (1988): "Statistical Analysis of Cointegrating Vectors", *Journal of Economic Dynamics and Control*, (12), 231-254.

[18] Johansen, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica*, (59), 1551-1580.

[19] McCracken, M.W. (1998), "Asymptotics for Out of Sample Tests of Causality", Working Paper, Louisiana State University.

[20] Orphanides, A. (1998): "Monetary Policy Rules Based on Real-Time Data", Working Paper, Federal Reserve Board.

[21] Robertson, J.C. and E.W. Tallman (1998): "Data Vintages and Measuring Forecast Model Performance", *Federal Reserve Bank of Atlanta Economic Review*, Fourth Quarter, 4-20.

[22] Rudebusch, G.D. (1998): "Is the Fed Too Timid? Monetary Policy in an Uncertain World", Working Paper, Federal Reserve Bank of San Fransisco.

[23] Sims, C.A. (1980): "Macroeconomics and Reality", *Econometrica*, (48), 1-48.

[24] Stock, J.H. and M.W. Watson (1989): "Interpreting the Evidence on Money-Income Causality", *Journal of Econometrics*, (40) 161-182.

[25] Stock, J.H. and M.W. Watson (1993): "A Simple Estimator of Cointegrating Vectors in Higher-Order Integrated Systems", *Econometrica*, (61), 783-820.

[26] Swanson, N.R. (1998): "Money and Output Viewed Through a Rolling Window", *Journal of Monetary Economics*, (41), 455-473.

Table 1: Granger Causality via In-Sample Wald Testing and Model Selection *

Money Measure	Avg CI Rank (SE)	Wald Test Rejection Frequency horizon=1period	Wald Test Rejection Frequency horizon=1year	SIC Model Selection horizon=1period	SIC Model Selection horizon=1year
		Panel A: Results Based on Final Revised Data			
		Dependent Variable: RGDP			
M1	1.73 (0.45)	0.88	0.00	0.90	0.99
M2	2.15 (0.66)	0.91	0.00	0.48	1.00
		Dependent Variable: NGDP			
M1	1.73 (0.45)	0.87	0.00	0.57	0.66
M2	2.15 (0.66)	0.96	0.00	0.36	0.84
		Dependent Variable: IP			
M1	1.96 (0.36)	0.47	0.02	0.62	0.52
M2	2.19 (0.82)	1.00	0.18	0.74	0.86
		Panel B: Results Based on Real-Time Data			
		Dependent Variable: RGDP			
M1	1.69 (0.56)	0.36	0.09	0.58	0.57
M2	1.93 (0.80)	0.73	0.00	0.36	0.58
		Dependent Variable: NGDP			
M1	1.69 (0.56)	0.66	0.00	0.48	0.46
M2	1.93 (0.80)	1.00	0.00	0.42	0.58
		Dependent Variable: IP			
M1	1.92 (0.32)	0.36	0.79	0.63	0.60
M2	2.03 (0.77)	0.88	0.08	0.47	0.67

* Notes: The table summarizes results of in-sample tests and model selection procedures for assessing the marginal predictive content of money for output. The first observation of all datasets used is 1959:1. Data denoted as *final* were collected in June 1999. These data are not real-time in the sense that they are a single vector of observations which was available in June 1999. Data denoted as *real-time* correspond to the case where a new vector of truly real-time observations dating back to 1959 is constructed for each calendar date. Sequences of Wald F-tests of the Granger non-causality null hypothesis and sequences of Schwarz Information Criteria (SIC) used to select between models with and without money were constructed using sample periods starting with 1959:1-1978:4 and ending with 1959:1-1994:12, for the monthly variables. For the quarterly variables, sample periods start with 1959:1-1978:2, and end with 1959:1-1994:4. The first column in the table lists the money stock measure used in the *big* models (those including money), while the second column reports averages and standard errors of cointegrating ranks estimated across the different samples. Entries in the third and fourth columns denote rejection frequencies (for 5% nominal size F-tests) of the null hypothesis that the *smaller* model is preffered to the *bigger* model. Elements in the fifth and sixth columns are counts of the number of times the *bigger* models with money are selected (using the SIC criteria). In all cases, VARs (for the F-tests) and VECs (for the SIC criteria) are estimated with regressor sets which include 9 lags (monthly) or 3 lags (quarterly) of the dependent variable, prices, and interest rates (*small* models); in addition, the *big* models include the same number of lags of money terms. All models are estimated using maximum likelihood.

Table 2: Granger Causality via Ex-Ante Predictive Ability Testing - I*
Mean Squared Forecast Errors Based on Quarterly Data

			One-Quarter-Ahead Forecasts				One-Year-Ahead Forecasts			
Forecast	Data Used:		VAR		VEC		VAR		VEC	
Type	estimation	forecast comparison	money	no money	money	no money	money	no money	money	no money
Panel A: M1 Used in Models of RGDP										
in	final	final	11.56	11.60	8.79	10.00	12.37	12.36	10.52	11.22
out	final	final	14.07	13.52	15.59	13.75	18.32	17.43	18.56	16.86*
out	real-time	vint1	10.85	10.08	12.71	8.69*	12.69	12.70	12.28	12.75
out	real-time	vint4	12.41	11.05*	14.30	10.65*	14.37	14.09	13.90	13.70
out	real-time	vint8	12.97	11.69	14.87	10.94*	15.04	14.55	14.24	13.61
out	real-time	vint12	12.85	11.72	15.02	11.40*	14.46	14.19	13.66	13.59
out	real-time	final	15.69	14.87	17.08	12.93*	17.94	17.55	16.69	16.43
Panel B: M2 Used in Models of RGDP										
in	final	final	11.22	11.60	8.90	10.00	11.55	12.36	10.47	11.22
out	final	final	13.10	13.52	13.26	13.75	16.40*	17.43	16.33	16.86
out	real-time	vint1	9.44	10.08	10.07	8.69	11.32*	12.70	13.06	12.75
out	real-time	vint4	10.94	11.05	11.40	10.65	12.70*	14.09	13.87	13.70
out	real-time	vint8	11.71	11.69	11.44	10.94	13.08*	14.55	13.79	13.61
out	real-time	vint12	11.29	11.72	11.22	11.40	12.73*	14.19	13.54	13.59
out	real-time	final	14.38	14.87	14.42	12.93	16.01*	17.55	16.30	16.43
Panel C: M1 Used in Models of NGDP										
in	final	final	14.04	14.08	11.09	12.83	16.10	16.10	13.18	15.12
out	final	final	16.88	16.06	19.73	15.91*	21.88	21.13*	23.21	20.45*
out	real-time	vint1	12.43	10.73*	15.34	10.23*	16.21	16.29	17.64	16.12
out	real-time	vint4	15.47	12.86*	18.23	13.20*	18.16	18.26	19.82	17.50*
out	real-time	vint8	15.59	13.21*	18.22	13.42*	18.05	18.21	19.31	17.16
out	real-time	vint12	15.75	13.53*	19.15	14.27*	18.08	18.31	19.32	17.46
out	real-time	final	17.74	16.00*	20.23	15.43*	19.74	20.07	21.19	19.66
Panel D: M2 Used in Models of NGDP										
in	final	final	13.31	14.08	11.42	12.83	14.87	16.10	13.89	15.12
out	final	final	15.14*	16.06	15.45	15.91	19.70*	21.13	19.60	20.45
out	real-time	vint1	9.68*	10.73	11.89	10.23	14.29*	16.29	16.20	16.12
out	real-time	vint4	12.88	12.86	14.95	13.20	16.36*	18.26	17.91	17.50
out	real-time	vint8	13.14	13.21	14.75	13.42	16.12*	18.21	17.60	17.16
out	real-time	vint12	13.13	13.53	15.27	14.27	16.33*	18.31	17.64	17.46
out	real-time	final	15.44	16.00	17.46	15.43*	18.19*	20.07	20.29	19.66

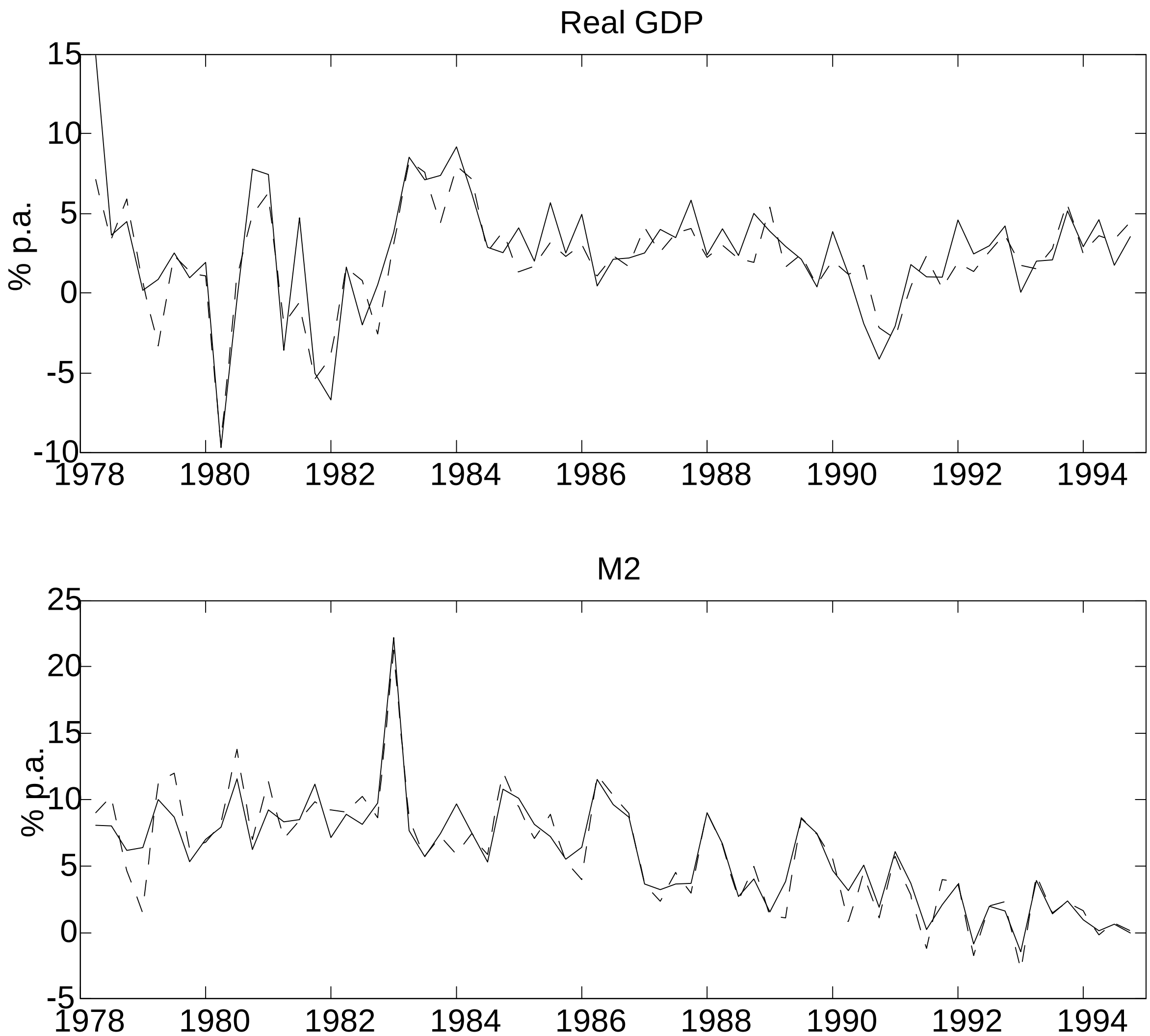
* Notes: See notes to Table 1. All entries are mean squared forecast errors (MSFEs) expressed in annualized percentage terms. Starred entries denote models which outperform their *bigger* (*smaller*) counterparts that include (exclude) money, and thus indicate that money does not have (has) marginal predictive content for output, based on the use of 5% nominal size DM tests (see Clark and McCracken (1999) and McCracken (1998) for a discussion of these tests in the current context). In column one, the forecast type is denoted "in" when out-of-sample forecasts are not constructed, and instead, in-sample residuals are used to construct the MSFEs. For those cases where the forecast type is denoted "out", out-of-sample forecasts are constructed. In columns two and three, the data sources for forecast construction and comparison are given, respectively. See the text for a description of *final* and *real-time* data sets. The terms *vint1* through *vint12* refer to the vintage of data release that is used to make the forecast comparison. The sample periods for the forecast sequences used to construct the MSFEs are the same as in Table 1.

Table 3: Granger Causality via Ex-Ante Predictive Ability Testing - II*
Mean Squared Forecast Errors Based on Monthly Data

			One-Month-Ahead Forecasts				One-Year-Ahead Forecasts			
Forecast	Data Used:		VAR		VEC		VAR		VEC	
Type	estimation	forecast comparison	money	no money	money	no money	money	no money	money	no money
Panel A: M1 Used in Models of IP										
in	final	final	66.25	68.82	64.87	67.42	76.85	76.34	72.53	73.59
out	final	final	72.54	75.28	77.86	79.79	83.72	81.61	84.50	81.79
out	real-time	vint1	53.92	53.35	60.60	56.43*	77.22	76.50	77.73	73.93*
out	real-time	vint12	64.28	65.07	69.07	66.98	85.99	85.75	88.75	84.53*
out	real-time	vint24	66.50	67.38	71.26	69.14	87.31	86.54	89.88	85.73*
out	real-time	vint36	65.97	66.64	70.06	67.97	84.16	83.28	86.50	81.85*
out	real-time	final	68.63	70.73	72.06	72.62	83.81	82.21	83.57	78.58*
Panel B: M2 Used in Models of IP										
in	final	final	67.60	68.82	64.84	67.42	75.92	76.34	71.90	73.59
out	final	final	73.21	75.28	83.31	79.79	82.68	81.61	81.49	81.79
out	real-time	vint1	51.88	53.35	54.79	56.43	76.95	76.50	72.96	73.93
out	real-time	vint12	64.21	65.07	67.65	66.98	85.00	85.75	83.74	84.53
out	real-time	vint24	66.42	67.38	70.19	69.14	85.34	86.54	84.68	85.73
out	real-time	vint36	66.41	66.64	70.18	67.97	82.02	83.28	81.79	81.85
out	real-time	final	69.61	70.73	73.34	72.62	81.33	82.21	78.88	78.58

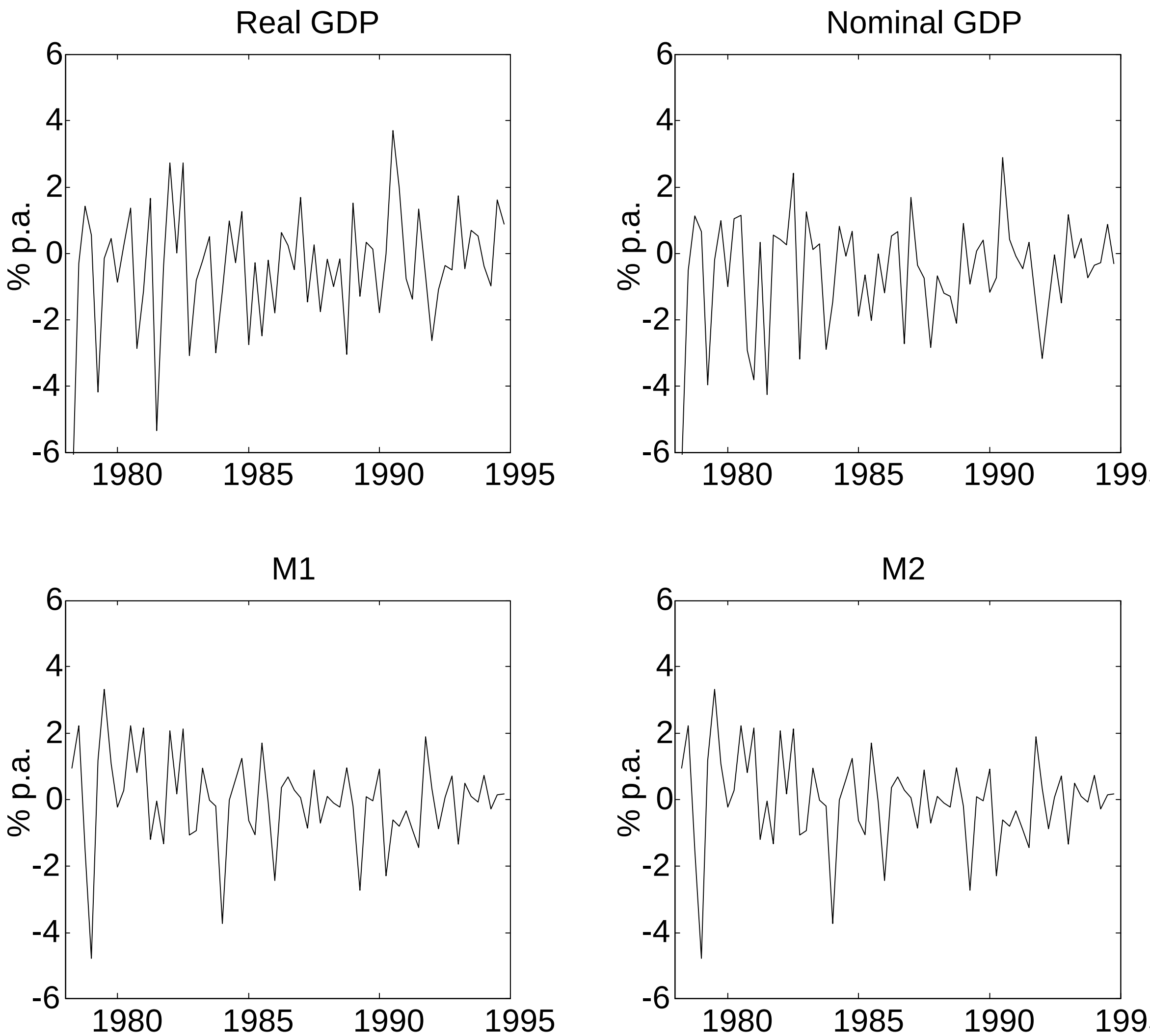
* Notes: See notes to Table 2.

Figure 1: Preliminary and Final Revised Output and Money Data



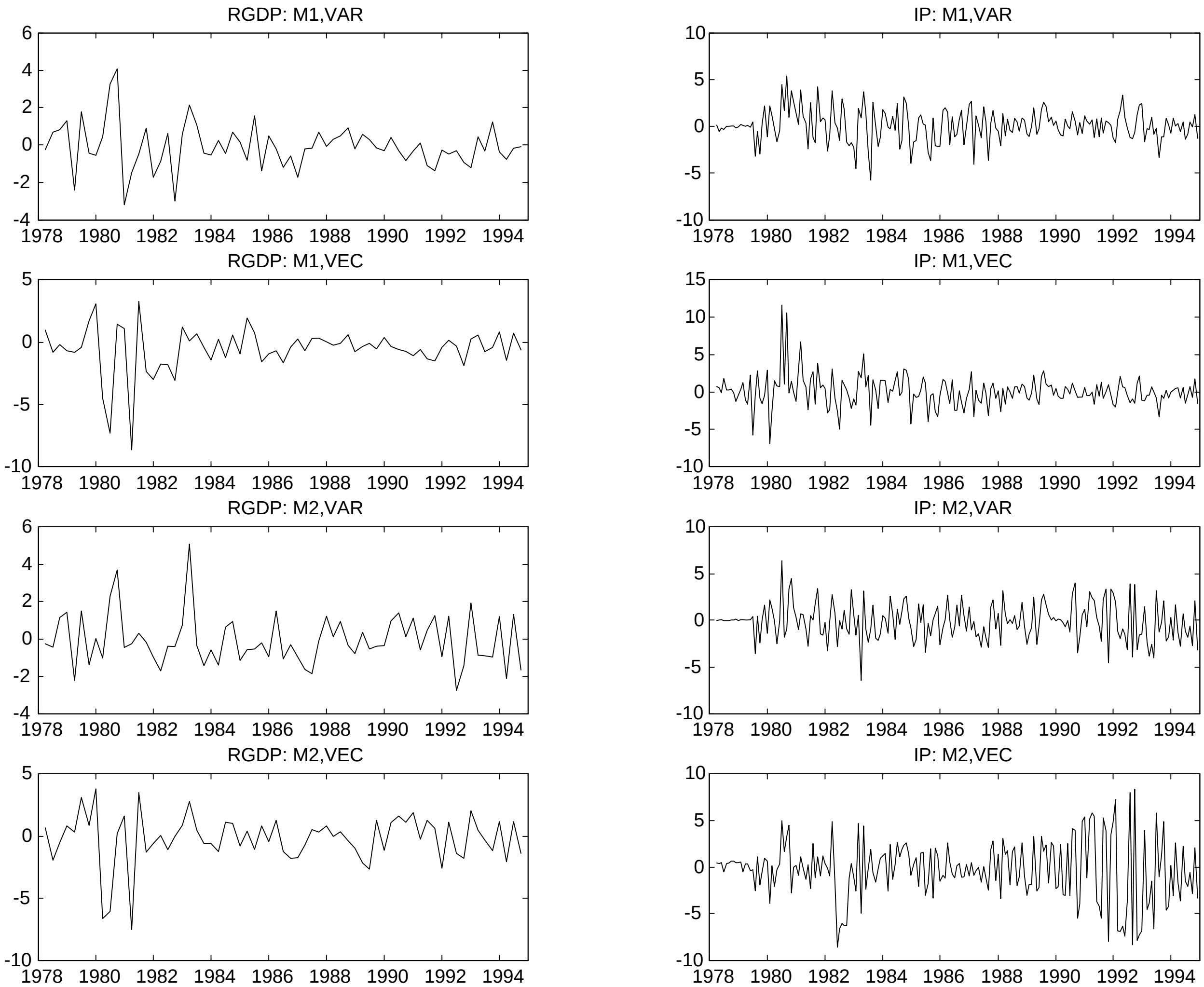
*Notes: The panels contain plots of the growth rates (expressed in annualized percentage terms) of RGDP and M2, respectively. The data are at a quarterly frequency and span the period 1978:2-1994:4. Two vintages of each time series are shown: *final* (solid line) and *preliminary* (dashed line). See the description in the text for data sources.

Figure 2: Output and Money Data Revisions



*Notes: The figure contains plots of the differences between *preliminary* and *final* data (expressed as growth rates in annualized percentage terms) of quarterly RGDP and M2 (upper panel) and monthly IP and M2 (lower panel). Quarterly data span the period 1978:2-1994:4; monthly data span the period 1978:4-1994:12. RGDP and IP are shown as solid lines, M2 as dashed lines . See the description in the text for data sources.

Figure 3: Differences in Absolute Forecast Errors of Models With and Without Money



*Notes: Each of the panels contains a plot of the absolute forecast error from the *small* model (i.e. excluding money) minus the absolute forecast error from the *big* model (i.e. including money). The left-hand side panels are based on forecasts of RGDP at a quarterly frequency and the right-hand side panels are based on forecasts of IP at a monthly frequency. The labels above the plots detail the *target* variable, the type of model used to make the forecasts (VAR or VEC), and which measure of money is used in the *big* models. All forecasts are constructed using real-time data sets, and forecast comparisons are based on $final$ actual data. The forecast errors for RGDP are for the the period 1978:2-1994:4, while those for IP are for the period 1978:4-1994:12. See the notes to Tables 2 and 3 for further details.