

The Real-Time (In)Significance of M2

Jeffery D. Amato*
Federal Reserve Bank of Kansas City

September 1998

JEL Classification: E52

Keywords: Granger causality, variance decomposition, money rule, data revision

Abstract: We examine the relationships between output, prices, interest rates, and M2, using data sets which were available in real time from 1973:1 to 1997:4. Our purpose is threefold. First, we delineate a potential role for M2 in policymaking. Second, we provide a more accurate basis for interpreting historical policymaking. Third, we evaluate the cause and effect of the historical redefinitions of M2. Our latter two objectives make it necessary for us to use data which was available to policymakers at the time decisions were made. In regard to our first objective, our approach is both novel and complementary to the existing literature.

*Correspondance: 925 Grand Boulevard, Kansas City, MO, 64198; tel: 816-881-2761; fax: 816-881-2199; email: jamato@frbkc.org. Charmaine Buskas and Hank Hui provided excellent research assistance. The author is grateful to Dean Croushore and Tom Stark for providing their real-time data on nominal and real output. The views expressed herein are solely those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

“Overall, annual revisions to the monetary aggregates due to revisions to source data, seasonal factors and definitions render treacherous any attempt by a researcher to update or extend previous studies by mixing differing vintages of monetary aggregates data.”

“...to an unknown but perhaps considerable extent, selection of the definitions of the monetary aggregates has been based on the relative ability of alternate aggregates to predict economic activity.”

Anderson and Kavajecz (1994)

1 Introduction

We examine the relationships between output, prices, interest rates, and M2, using data sets which were available in real time from 1973:1 to 1997:4. Our analysis is focused on three objectives. The first is to evaluate a potential role for M2 in future policymaking; the second is to provide an interpretation of historical policymaking which is logistically accurate; the third is to determine the cause and effect of the historical redefinitions of M2.

Most authors who have examined reduced-form relationships among the variables of interest here have been concerned with our first goal; namely, to determine whether money aggregates should play a role in the conduct of monetary policy, either as an intermediate target, or simply as an information variable.¹ Researchers have employed standard time-series techniques to investigate the marginal predictive power of money for real and nominal output, and prices, and whether money targeting rules suggest welfare improvements over historical policies. Results have been reported for various sampling methods (samples of fixed length over a long period of time, rolling samples, and recursive samples),² but always based on one vintage of data. By undertaking similar tasks with real-time data sets we provide a novel

¹See, among others, Stock and Watson, 1989; Friedman and Kuttner, 1992; Beckett and Morris, 1992; Feldstein and Stock, 1994; Miyao, 1996; Friedman and Kuttner, 1996; Friedman, 1997; Swanson, 1998. Friedman (1997) was also concerned with the varying use of money aggregates in the historical setting of policy.

²Rolling samples have a fixed window length (typically 10 to 15 years), with moving startpoints and endpoints. Recursive samples expand in size, with a fixed starting point and a moving endpoint.

perspective on the link between M2 and the variables of ultimate interest to policymakers.

To be specific, we construct various statistics (e.g. measures of predictability, performance measures of a policy rule) for each quarter from 1973 to 1997, using data that would have been available in that quarter. Viewed as a whole, our results represent the collective experience economists would have obtained across the range of our samples in evaluating the properties of M2. By contrast, studies which provide sub-sample evidence based on one vintage of data are providing a “snapshot” of existent relationships.³ In regard to our first objective, our approach should be viewed as complementary to studies such as those mentioned above. However, to the extent that the usefulness of a variable in real-time decision making can only be evaluated using data which is available in real time, our approach improves upon existing methods. For example, it may be preferable to evaluate out-of-sample forecasts in an environment which simulates that in which the policymaker exists, in particular, one where forecasts are constructed from real-time data. By extension, if we view Granger-causality tests also as tests of predictability then it may be more useful to construct these tests with real-time data sets. Furthermore, by the very nature of counterfactual experiments, it may be more appropriate to perform them using data from the time period for which the alternative policy regime is supposed to have been adopted.^{4,5}

Our second and third objectives make it necessary for us to use data which was available to policymakers and the Federal Reserve Board’s staff at the time decisions were made. By abstracting from the presence of revisions, results based on one vintage of data may lead to a misinterpretation

³Papers that include tests of structural stability which use only the current vintage of data (e.g. Feldstein and Stock, 1994) fall into this category.

⁴An additional reason to use real-time data is that our reduced-form models implicitly incorporate expectations. Of course, agents’ expectations are formed based only on data available in real time.

⁵Since we report various statistics for each quarter from 1973 to 1997, one may also view our results as filling in gaps in a literature which contains results for only a small subset of sample periods, although this description is not entirely accurate. As suggested by Anderson and Kavajecz (1994), it is very difficult to replicate the exact data vintage of most studies even with full knowledge of data sources, sample periods, etc. Not all authors use all of the data that is available in the quarter that they create their data sets and instead may trim their data in some arbitrary manner. Possible explanations for this practice include a desire to match their sample period to previous studies or to avoid effects stemming from mismeasurement of recent observations.

of the history of policymaking. For instance, consider this description of the monetary policymaking process: each quarter, new data becomes available on many variables, mostly with a lag, that policymakers use to estimate the current and future states of the economy. These estimates aid policymakers in adjusting their instruments, based on a view of the economy. The arrival of new data, however, poses two challenges. As new data becomes available, policymakers have the option to assess whether their preferred models remain empirically valid. Secondly, in light of revisions and rebenchmarkings to key series, it may be necessary to change the course of policy. If a series is frequently redefined, it may become impractical as an input to policy.⁶ To the extent that this description is accurate, it seems crucial that historical analysis use both models and data which were relevant for past policymaking. In fact, Federal Reserve (Fed) policymakers' preferred measure of output and its deflator, as well as M2, have all been subject to substantial revisions, rebenchmarkings, and redefinitions.⁷ By using real-time data, our approach allows the reader to achieve a more accurate view of the evolving perceptions of economic relationships among policymakers.

Perhaps one reason M2 undergoes continual redefinition is that, in the face of financial innovation and other structural change, the Fed desires to have one measure of money which corresponds to the abstract notion of money represented in economic models. More specifically, as Anderson and Kavajecz (1994) point out, redefinitions are likely the result of the Fed wishing to have a series which has a strong link to output and prices and admits a stable long-run demand function. Our methods and data sets allow us to address when, by these criteria, there was a need for a redefinition of M2. Furthermore, we can provide a partial assessment, *ex post*, of whether the redefinitions of M2 achieved either of these goals.

Our models, and the statistics we compute from them, allow us to achieve our three objectives simultaneously in a convenient manner and, for purposes of comparability, reflect techniques employed in the literature cited above. We construct three measures of the marginal predictive power of M2, including Granger-causality tests, variance decompositions, and mean squared errors of out-of-sample forecasts. Each statistic captures a different aspect of the link between M2 and the target variables, although the out-of-sample di-

⁶Frequent redefinitions of a variable may represent both a symptom and a cause for omitting it in policy deliberations.

⁷Herein, all changes to a series will be referred to simply as revisions, except in places where the discussion is specifically about a revision, rebenchmarking, or redefinition.

agnostics are traditionally viewed as the most stringent test of predictability. Each of these statistics are derived from unrestricted vector autoregressions (VARs) in first-differences. Our analysis of predictability allows us to address whether M2 may be useful as an information variable, how the evolving perception of the marginal content in M2 may have affected Fed official's attention to money targets in the past, and whether - by the criteria set forth by Anderson and Kavajecz (1994) - the redefinitions in M2 were both necessary and effective. Since our data are assumed to be integrated of order one, we then apply cointegration tests to determine whether a long-run demand function exists. A long-run demand function must exist if M2 is to be used to anchor the price level in the long run. We will argue later, as well, that the redefinitions in M2 were likely undertaken in light of evidence against cointegration. Lastly, in parallel to recent work by Orphanides (1998), we consider the effects of data vintage on a simple money rule analyzed by Feldstein and Stock (1994). We calculate statistics under this counterfactual rule to further our analysis of a possible role for M2 in policymaking.

The main results in the paper are: (i) statistics computed using the real-time data sets differ considerably from those obtained using sub-samples from a single vintage of data, e.g. the most recently available data set; (ii) for the real-time data sets, the in-sample link from M2 to output and prices in the VAR appears remarkably stable; (iii) there is strong evidence that M2 Granger-causes real output, but it holds little marginal predictive content on the basis of results from variance decompositions and out-of-sample forecasts; (iv) in many periods, a case can be made that, *a priori*, redefinitions of M2 were necessary, however there is only minimal evidence, except in one important case, that the redefinitions achieved their presumed goals; and, (v) policymakers would have been led to grossly varying conclusions at different points in time about the properties of the simple rule we consider.

The remainder of the paper is organized as follows. In the next section, we briefly discuss the data sets, including the nature of the revisions, rebenchmarkings, and redefinitions. In the third section, we present evidence on the predictive power of M2 in the context of a VAR. In the fourth section, we test the possibility that there exists a long-run money demand relationship, as expressed through cointegration, between real money balances, real income, and a measure of opportunity costs. Thereafter, since tests for cointegration are known to perform poorly in small samples, we assume cointegration exists and we provide estimates of elasticities, as well as results of Granger-causality tests in vector error correction models (VECMs).

In the fifth section, we analyze the properties of a simple feedback rule for policymaking. Some concluding remarks are given in the final section.

2 Data

The variables used throughout this paper are the natural logarithms of nominal output (x_t), real output (y_t), price deflator (p_t), and M2 (m_t), and the level of the secondary market rate on ninety-day United States Treasury bills (R_t), expressed in percentages.⁸ The series for output (and prices) correspond to the series favored in NIPA at each point in time (see discussion below). All data are at a quarterly frequency, where the within quarter monthly observations on M2 and the interest rate have been averaged.

Data on T-bill rates is not subject to revision, so all sub-samples for this variable are drawn from the full sample, covering the period from 1959:1 to 1997:4. We make the assumption that policymakers (and others) obtain the first estimate of a variable in the quarter proceeding the date of the variable. That is, data becomes available with a one quarter lag. For each of the series other than the T-bill rate, a separate “real-time” data set exists for each quarter from 73:1 to 97:4. By “real-time”, we mean data that was available to policymakers in that quarter. The label of the data set corresponds to the date in which the policymaker used the data. Thus, a data set with a label of quarter t contains data from the common starting point, 59:1, to quarter $t - 1$. This structure allows for revisions to historical observations in each new data set.⁹

Both the output/price and M2 series have been subject to revisions, rebenchmarkings, and redefinitions during the full sample period under consideration. A brief summary of the nature and timing of these changes is given in Boxes 1 to 4.¹⁰ In regards to M2, it is worth noting that there have

⁸Upper case letters refer to levels; lower case letters are the natural logarithms of the level.

⁹The data on T-bill rates were obtained from the Federal Reserve Board's database for the United States. The real-time data sets for output and prices were constructed by Dean Croushore and Tom Stark, to be discussed in the forthcoming Federal Reserve Bank of Philadelphia Working Paper, “A Real-Time Data Set for Use by Macroeconomists.” The real-time data sets for M2 were constructed by the author and Charmaine Buskas, and are to be discussed in a manuscript in preparation.

¹⁰For more detailed discussions, please see the aforementioned papers which describe the construction of the data sets. An excellent description of the changes to the money

been six redefinitions of the series since 1980, with the major redefinition occurring in February 1980, a time marked by substantial financial innovations. The description of release/revision dates in Boxes 2 and 4 conform to our assumption on the first availability of data. In Table 1, we give a taste of the magnitude of the effect of revisions on the series, in the forms of both log-levels and growth rates.¹¹ The numbers in this table are the means of the absolute differences between the variable as it appears in the full-sample data set (i.e., the one labelled 97:4) and the final observation on the variable from the corresponding real-time data sets.¹² The size of the revisions to M2, real output, and the price deflator are non-trivial, and warrant the investigations to follow.

Finally, to specify the models we use in this paper requires knowledge of the order of integration of the variables. To assess these properties, we performed unit root tests on x_t , y_t , p_t , m_t , and R_t for each real-time data set. Using augmented Dickey-Fuller (ADF) t -statistics, we could not reject the null hypothesis of a unit root in any of the variables for the vast majority of data sets.¹³

aggregates published by the Federal Reserve can be found in Anderson and Kavajecz (1994).

¹¹Throughout, growth rates are constructed as the first-difference of log-levels, expressed as annualized percentages.

¹²The final observation (ie. the one dated $t - 1$) is missing from the 94:1 and 96:1 data sets for output and prices; these data points were omitted in these calculations. For output and prices, the first two observations are missing from the 96:1 through 97:1 data sets. All of the calculations in this paper were performed using trimmed data sets when these missing values were encountered.

¹³Results of the tests are not shown; they may be obtained from the author upon request. All tests were constructed with four lags of first-differences of the variables. The tests for the T-bill rate included a constant, while the tests for all other variables included a constant and a trend. At conventional levels of significance, a unit root could be rejected in the T-bill rate for a few data sets in the mid-seventies and in nominal output for a few data sets in the early seventies. Strictly speaking, we should also test whether the series are integrated of order two (I(2)) versus I(1). Evidence in this regard can be found by computing the size of the largest eigenvalue in the companion form matrix of the systems estimated below. Except for the data sets from 74:1 to 75:1, the largest eigenvalue is less than one and typically between 0.9 and 0.95.

3 VARs

In this section we provide evidence on the predictive properties of M2 for output and prices in the context of first-differenced VARs.¹⁴ This section is divided into two parts. In the first sub-section, we construct tests of the hypothesis that M2 does not Granger-cause either nominal output, real output, or the price deflator.¹⁵ Secondly, we use variance decompositions to determine what percentage of output and price fluctuations can be attributed to the innovation in the money equation. Both of these exercises provide evidence on in-sample relationships. In contrast, the second sub-section contains results on the out-of-sample marginal predictive power of M2.

Before proceeding, it should be recognized that our results may be difficult to interpret if, during certain periods in the range of our samples, the Fed had indeed successfully targeted M2. In this case, targeting of M2 would considerably reduce its volatility, necessarily making it appear that it has little explanatory power.¹⁶ But during the episode which the Fed arguably paid closest attention to the money aggregates, the volatility of M2 actually increased.¹⁷

3.1 In-Sample Results

As in Friedman (1997), we test the hypothesis that all of the coefficients on M2 terms are statistically insignificant in the equations for real output and prices from a VAR with four lags. We also run tests in equations for nominal output, as in Feldstein and Stock (1994). Our decision to use a VAR reflects our desire to determine the *marginal* predictive power of M2 beyond what can be explained by output and prices alone. Many authors have reported that when a short-term interest rate, or the spread between the commercial paper rate and the T-bill rate, is also included in these reduced-form equations,

¹⁴By using first-differenced VARs, we are implicitly assuming that cointegration does not exist among the variables.

¹⁵For the systems with a root greater than one, the distribution of the F-statistic is still valid if cointegration exists among the *growth rates* of the variables and we have included at least one lag more than the “true” order of the VAR (see Dolado and Lutkepohl, 1996).

¹⁶Friedman (1997) makes this point in a similar context. It is commonly referred to as Goodhardt’s Law.

¹⁷This claim is evident in plots of the sample standard deviation of M2 growth computed with rolling windows ranging from 3 to 5 years in length (available from the author upon request).

money loses its predictive power (e.g. Sims, 1980, and Friedman and Kuttner, 1992). For this reason, we also include a short-term interest rate in all of the equations to see if M2 - a financial quantity - contains information beyond what is in the T-bill rate - a financial price. By fixing the number of lags, we avoid statistical problems associated with pre-testing. Besides, some lag selection criteria are designed to choose the model with the best fit, which would clearly bias the results of causality tests in favor of M2.¹⁸

The top two panels of Figure 1 show p-values of F-statistics computed from the system including real output and prices. Neither real output nor prices are included in the system with nominal output, with the p-values in this case shown in the bottom panel of Figure 1. The solid line in each of the panels corresponds to p-values computed from using all of the data in each of the one hundred real-time data sets. Values are plotted against the date of the data set so, for a given date, one can see how significant money appeared to policymakers at that time. The dashed line reports values computed by taking recursive samples from the most recent vintage of data (i.e. the 97:4 data set). A comparison of the solid and dashed lines indicates the sensitivity of retrospective inference when statistics are constructed from a recent vintage of the data. The dash-dot line gives the p-value computed from the full sample of data, and gives the reader a sense of what inferences would be drawn if we took the standard approach of reporting results based on only the full sample of currently available data. To ease the interpretation of the figure, in Table 2 we report the number of times each of the curves falls below conventional levels of significance. In the table, columns labelled "Real-Time" and "Full-Sample" correspond, respectively, to the solid and dashed lines.

For the output series, the obvious implication to be drawn from this figure is that our view of money is highly dependent on the vintage of data we use. In real-time, we would have always inferred that M2 Granger-caused output at a 5% level of significance, and for the most part at the 1% level. In contrast, using the 97:4 data set, we would not have rejected the null over half of the time at 5%, and never at 1%. Our full-sample recursive results even contrast with Friedman's (1997) recursive/rolling results, the paper which is most directly comparable to ours. In his paper, he always accepts the

¹⁸For an example of a paper in which lag selection criteria are used to specify models before conducting Granger-causality tests, see Swanson (1998).

non-Granger-causal null, even at a 10% level of significance.¹⁹ Alternately, there appears to be little evidence that M2 growth Granger-causes inflation, consistent with the results in Friedman (1997).

If the economic system evolves with time, then we may wish to discount “older” information. In the absence of a rigorous theory of structural change, one method for discounting early observations is to use rolling windows. In Figure 2, we report the same statistics as in Figure 1, but computed with windows fifteen years in length.²⁰ Although the curves are more variable (which is to be expected when using shorter windows of data), the results for the real-time data sets change little. On the other hand, the patterns based on the full-sample of data are much more pronounced. An example of the danger in using results computed from sub-samples of one recent vintage of data can be seen in the panels for output. One period of special interest is the time leading up to, and including, the adoption of money growth targets in 1975, and the widely heralded change in Fed operating procedure in October, 1979. From anecdotal evidence at the time,²¹ one might argue that the Fed began to look at M2 seriously upon the adoption of the new operating procedure. It is tempting to interpret the Fed’s increased attention towards M2 in the late seventies, and subsequent downgrading of M2 in the late eighties, as a response to the increased, and subsequently decreased, significance of M2. But in real-time, M2 always appeared to Granger-cause output.²² Besides, if the shifted focus towards M2 was to control inflation,

¹⁹Friedman (1997) uses the Federal Funds rate in place of the 90-day T-bill rate, and a rolling window twenty years in length. (He actually uses recursive samples until his window size reaches eighty quarters. In this case, his results for before 80:1 are more comparable to our recursive case.)

²⁰In terms of sampling windows, a true comparison of our results to Friedman’s (1997) would lie somewhere between our recursive and rolling cases (see the description of Friedman’s method given previously).

²¹See the announcement of the change in operating procedure in the *Federal Reserve Bulletin*, October, 1979. The popular press at the time seemed to believe that the Fed had entered a new regime, with a stronger focus on targeting money growth to tame inflation.

²²Swanson (1998) attempts to resolve the mixed evidence on the money-real output relationship by using lag selection criteria to trim the order of his VAR equations before constructing F-tests. Using both recursive and 15-year rolling samples from a single vintage of data, he finds results similar to those reported in our Table 2 for the real-time data sets. It is difficult to compare our respective results, however, since he uses data at a monthly frequency, substituting industrial production for GNP/GDP and wholesale prices for the implicit price deflator.

the decision surely was not based on the evidence offered in the middle panels of either Figure 1 or 2. In the first figure, M2 never appeared significant; in the latter figure, the relationship was more volatile, although most p-values are greater than 0.1.

Granger-causality tests are attractive because they do not require potentially controversial identification assumptions. However, they provide only one piece of evidence of the link between M2 and output and prices. A second method which sheds light on the properties of M2 is the decomposition of the variance of forecast errors, which has been employed most notably by Friedman and Kuttner (1992, 1996) and Friedman (1997). Through identification of “structural” contemporaneous shocks, variance decompositions allow us to determine the percentage of variation in a variable (over some horizon) accounted for by each of the shocks; in particular, we can determine the portions of output and price variation which are caused by innovations in M2.

Figures 3 and 4 are plots of variance decompositions at a two-year horizon, computed using recursive and 15-year rolling samples, respectively. The solid lines give values computed from sub-samples from the 97:4 data set, with one-standard-error bands shown in dots.²³ Real-time values are shown in dashes. In all of the computations, we make the same identifying assumptions as in the work by Friedman and Kuttner (1992); namely, we use a Cholesky decomposition of the residual covariance matrix, putting output first, prices second, money third, and interest rates last.^{24,25} From Figure 3, it is clear that for our real-time data, M2 accounts for basically none of the two-year-ahead variance in any of the series and none of the point estimates are statistically significant (not shown). In contrast, the “full-sample” recursive estimates are often statistically significant. Turning to Figure 4, the paths of the real-time estimates exhibit more movement. The estimates become statistically significant at times, but they rarely exceed 15%. Friedman (1997) offers a different picture for real output. He finds that M2 can account for as much

²³We chose to plot one-standard-error bands instead of the conventional 95% confidence intervals. Standard errors of variance decompositions are notoriously large, leaving 95% confidence intervals relatively meaningless.

²⁴Friedman and Kuttner (1992) show that ordering M2 before the interest rate leads to slightly more variation being attributed to M2 than in the case when M2 is ordered last.

²⁵The systems underlying these calculations are the same as for the Granger-causality tests, with the exception that they are specified in (log-)levels form (as is conventional), rather than in first-differences.

as 30% of the variation in output during the eighties.²⁶

The advantage of reporting confidence intervals in the full-sample case is that we can determine whether inferences drawn from simply using subsamples of a recent vintage of data give an accurate portrait of what could be determined in real-time. In the lower panel of Table 2, we report the number of times the real-time values fall within the 97:4-vintage confidence intervals. Under recursive sampling, the real-time value typically falls outside the band, indicating that inferences from a single vintage of data may not be robust. The results are better aligned under rolling windows - this is mainly due to the fact that the full-sample confidence intervals now include zero more often and because the real-time series is more choppy.

As mentioned in the introduction, our approach to using real-time data sets allows us to assess whether breakdowns in statistical relations involving M2 induced changes in its definition. Based on Figures 1 and 2, there is little evidence to suggest this was the case in regard to the output series. Furthermore, it is not apparent the redefinitions affected the significance of M2 in these relations, although it very well may be the case that the relationships would have broken down if the redefinitions did not occur. On the other hand, the results for prices in these figures - and the variance decompositions for all of the series - imply that the redefinitions were necessary, but that none of them achieved their presumed goal of restoring marginal predictive power to M2.^{27,28}

3.2 Out-of-Sample Results

The statistics reported in the previous section only shed light on the in-sample predictive power of M2. A stronger test of predictability is whether information on M2 can reduce the magnitude of out-of-sample forecast errors. One difficulty in judging out-of-sample real-time forecasts, however, is that it is not clear which “realized” values should be used as a basis for comparison, since a given data point is subject to revision, rebenchmarking, and redefinition. The convention we adopt here is to compare forecasts to

²⁶During 1982, his confidence intervals do not even include the recursive estimates from our 97:4 data set.

²⁷For example, the redefinition of 1996 had little impact on any of the statistics, a result consistent with the evidence in Whitesell and Collins (1996).

²⁸It may be the case that Fed staff focused their redefinitions of M2 on restoring a long-run relationship between money and prices. This issue is taken up below.

the first released estimate of an observation.²⁹

In Table 3, we report root mean squared errors (RMSEs) of one-step-ahead forecasts of quarterly variables (panel A) and of two-year-ahead forecasts of annual variables (panel B). In both panels, for each data set a VAR system was estimated and forecasts made for two periods, and six to nine periods, beyond the last observation, respectively.³⁰ The errors in forecasts were then averaged across data sets. The VAR(1) system refers to an autoregression in the variable being forecasted; the VAR(2) systems include an output series, real or nominal, and the price series; the VAR(3) systems add the T-bill rate and M2 to the VAR(2) systems, respectively; the VAR(4) systems add both variables.³¹ The reason for considering systems without M2 is to see if M2 had *marginal* predictive power for output and prices.

Overall, the results for M2 are relatively uninspiring. In every case, the system which includes M2 but not interest rates never dominates all other systems, although in almost half of the cases the four-variable system has the lowest RMSE. Nonetheless, in predicting nominal output and the price deflator, there is little to choose among the multivariable systems, although variables beyond lags of the series being predicted had marginal predictive power. Not so for real output - the autoregressions performed best! These results contrast with those reported in Feldstein and Stock (1994) using one vintage of data. For recursive one-year-ahead forecasts, they found adding M2 to the model resulted in over a ten-percent reduction in the RMSE.³² As foreshadowed in the introduction, the results for inflation are not surprising since comparisons of out-of-sample forecasts are more stringent tests than Granger-causality tests. On the other hand, the mediocre results for M2 in

²⁹Alternately, we could extract the final revised value of an observation under an unchanged definition of the variable.

³⁰We call the forecasts in the upper panel "one-step-ahead" forecasts because they are made for the period immediately preceding the date of the data set, although formally they are two-step-ahead forecasts because of the lag in receiving data. Likewise, the forecasts in the lower panel are the average of quarterly forecasts six to nine periods in the future instead of five to eight periods.

³¹Each VAR was specified in first-differences, and contained four lags of the variables.

³²Another interesting observation can be drawn from the tables. One argument for performing estimation with rolling samples is that the economic environment may be changing, in which case out-of-sample forecasts may be more accurate from systems estimated using only recent observations. At a close horizon, however, the recursive-based forecasts outperformed the rolling-based forecasts by a margin of twelve to three; at the longer, two-year horizon, the rolling forecasts were better ten of fifteen times.

the case of nominal output, and the relatively poor forecasts of real output, deflate the significance of our almost uniform finding that M2 Granger-causes both nominal and real output.

4 Long-Run Money Demand and VECMs

In the model of the previous section it was implicitly assumed that there did not exist cointegration among any of the variables. Since each of the variables appear to have a unit root, there must exist a cointegrating vector if there is to be a stable long-run demand function for M2. From a theoretical perspective, we would expect that a stable long-run demand function would exist for our abstract notion of money. Furthermore, for the P* model of inflation (e.g. Hallman, Porter, and Small, 1991) to have content, velocity must be stationary.³³ The existing evidence on cointegration is mixed. The results in Miyao (1996) are generally negative, whereas a number of other authors have obtained more positive results (e.g. Stock and Watson, 1993). The use of real-time data sets thus provides an additional opportunity to test for the existence of cointegration.

In Figures 5 and 6, we report ADF t-statistics constructed using residuals of six special cases of the following regression of real money balances on real income and the T-bill rate:³⁴

$$m_t - p_t = \mu_o + \mu_1 t + \beta_y y_t + \beta_R R_t + u_t \quad (1)$$

In the top panel of each figure, no restrictions are imposed on the regression; in the middle panels, we set the long-run income elasticity to one ($\beta_y \equiv 1$); in the bottom panels, we also impose $\beta_R \equiv 0$. (The latter case is a test of the stationarity of velocity.) Figures 5 and 6 report statistics in the

³³If the distribution of velocity contained breaks which could be modelled in a plausible *a priori* manner, then one could still employ the P* model to forecast inflation. Orphanides and Porter (1996) provide an *ex post* correction to the supposed break in M2-velocity in the early nineties.

³⁴The T-bill rate serves as a proxy for the opportunity cost of holding M2 balances. A better measure would be the difference between the T-bill rate and the rate of return on holding M2 balances. As of the time of writing, however, real-time data sets did not exist on M2's rate of return. Anyhow, one might expect the spread to be a cointegrating relation (i.e. stationary), in which case tests for cointegration between velocity and the spread would be inappropriate.

demeaned ($\mu_1 \equiv 0$) and detrended (μ_1 estimated) cases, respectively. The null hypothesis in each case is that cointegration does not exist.

It is clear from both figures that we cannot reject the null for most time periods when the interest semi-elasticity is estimated.³⁵ As can be seen in the middle panels, one may be inclined to accept the alternative of cointegration using a 1992 data set. Since 1993, the evidence has gotten worse, which likely reflects what has come to be called the “missing money” episode of the early nineties, a period characterized by a large gap between M2-velocity and standard measures of opportunity costs. Again, it is interesting to note that in the period just before the change in operating procedure, the evidence was most strongly against cointegration; only with the redefinition in 80:1 can the no-cointegration null be rejected. For M2-velocity (lower panels), the evidence is more favorable towards cointegration from 80:1 to 93. Unlike in the previous section, the cointegration tests strongly support the argument that the redefinition of M2 in 80:1 achieved the goal of reestablishing stationary velocity, but it is also apparent that velocity is now a higher order of integration.^{36,37}

It is well-known that tests for cointegration have poor power properties, especially in small samples. It also may have been the case that policymakers held strong priors in favor of the existence of a long-run money demand function, in which case it would be of interest to obtain estimates of the income and interest (semi-)elasticities. Figure 7 shows estimates of these elasticities obtained using the DOLS method of Stock and Watson (1993).³⁸ In the top two panels are estimates from the unrestricted regression; the bottom panel reports estimates of the interest semi-elasticity when the income elasticity is restricted to equal one. This latter restriction is supported by the results in the top panel. The null that $\beta_y = 1$ cannot be rejected at any point in time.

Two features of the graph immediately stand out. Firstly, the hypothesis, $\beta_R = 0$, can never be rejected, and most estimates are very close to zero

³⁵The null is rejected more in the demeaned case, perhaps reflecting the greater power of the test when a trend is not estimated.

³⁶The apparent lack of cointegration in recent years has led some authors to propose a further redefinition of M2, one that includes bond and stock mutual funds (e.g. Collins and Edwards, 1994).

³⁷In light of Miyao’s (1996) warnings about robustness, these results should be taken to be suggestive.

³⁸Each regression was estimated with a constant and four leads and lags of the first-differences of the right hand side contemporaneous variables.

and often of the wrong sign. Secondly, there was considerable uncertainty about estimates of β_y and β_R leading up to the change in Fed procedures.³⁹ Interestingly, using data to 1993, both Stock and Watson (1993) and Miyao (1996) can reject the null, $\beta_R = 0$.⁴⁰ It is perhaps not surprising, though, that we cannot reject a zero interest semi-elasticity in light of the pattern of results in the cointegration tests.⁴¹

Feldstein and Stock (1994) argue that VARs specified in first-differences may be subject to specification error, and that the results of Granger-causality tests may be overturned if allowance is made for cointegration among the variables. In particular, they construct Granger-causality tests of M2 from equations which can be considered to be part of a VECM, and they find money enters significantly.⁴² Our findings above suggest that the best possible specification of a VECM would include only M2-velocity as a cointegrating residual. In Figures 8 and 9, we report p-values of F-tests of the null that the coefficients on all the money terms - the lagged first-differences and the cointegrating residual - are zero, again using recursive and 15-year rolling samples, respectively. As before, we report results for both real-time data sets and the most recent vintage of data. Looking at Figure 8, the conclusion for the real-time data sets is that our inferences about Granger-causality are not affected by our assumption about cointegration. Again, the two sampling methods provide different pictures, although the rolling sample results from the 97:4 data set are somewhat different than before. In particular, the relationships falsely look more volatile, when in fact they would have been seen as stable. In light of the similar outcomes of the tests with those from the VAR, our conclusions about the necessity and effect of the redefinitions still hold.

³⁹The large standard errors may be due to short data samples.

⁴⁰One source of the difference between our results and Stock and Watson's (1993) could be that we computed standard errors using a Newey-West correction with window length equal to six, whereas they modelled the errors as an AR(2) process.

⁴¹Our results do not necessarily imply values for the coefficient on a better measure of opportunity cost and, as such, should be taken as suggestive. Nonetheless, many studies do not contain results for regressions which include the rate of return on M2 balances.

⁴²This argument is only valid if the VAR does not encompass the "true" VECM. For instance, if the number of lags in the VAR is identical to the order of the VECM (see Dolado and Lutkepohl, 1996). It is highly unlikely, however, that the true order of a VECM is ever known. Besides, Feldstein and Stock (1994) assume the systems have order three, whereas we include four lags, so that our VAR systems meet the criterion in Dolado and Lutkepohl (1996), even if they may be misspecified.

5 A Simple Policy Rule

Although the evidence in favor of incorporating M2 into policymaking appears to have been mixed at best over the range of our data sets, it nonetheless may be worthwhile to analyze the real-time properties of rules involving M2 to gain further perspective on the perceived attractiveness of incorporating it into ongoing policy deliberations. A number of authors have argued that the Fed should adopt a simple feedback rule for conducting monetary policy. In fact, it has become fashionable recently to model the Fed as acting according to a rule. The most popular form of such a rule, which has come to be known as “Taylor’s rule” (originating in the work of Taylor, 1993), specifies that the Fed should move the Federal Funds rate in response to deviations of inflation from a target and deviations of real output from its potential, or natural, level. In this case, the Federal Funds rate is taken to be the instrument of policy, an assumption which has received considerable empirical and anecdotal support in recent years. An alternative rule which has also been examined is a simple partial adjustment rule for a money aggregate (such as M2), which is used to target nominal output, or some combination of real output and inflation.⁴³ Feldstein and Stock (1994) analyze the properties of the following rule in systems like those considered so far in this paper:

$$\Delta m_t = \mu_m + \lambda(\mu_x - \Delta x_{t-1}) + (1 - \lambda)(\Delta m_{t-1} - \mu_m) \quad (2)$$

where, in their case, μ_m and μ_x are the target growth rates of M2 and nominal GDP, respectively. This rule has the Fed responding to deviations of last period’s nominal income growth from target, with some smoothing of the policy variable, captured by $1 - \lambda$. The parameter λ governs the strength of the Fed’s response to developments in the economy, where a high value could be interpreted as an aggressive targeting regime. In this sense, it is similar in nature to Taylor’s rule with interest rate smoothing. In fact, a hybrid version of the rule can be expressed as:

$$\Delta m_t = \mu_m + \lambda\{2\theta(\pi^* - \Delta p_{t-1}) + 2(1 - \theta)(\mu_y - \Delta y_{t-1})\} + (1 - \lambda)(\Delta m_{t-1} - \mu_m) \quad (3)$$

where π^* is the Fed’s target inflation rate, and μ_y is the long-run potential growth rate of real output. The Fed’s relative preference for controlling inflation fluctuations is captured in the parameter θ .

⁴³Taylor (1985) has called the latter case, in which the weights on output and inflation may differ, to be a *modified* nominal-income rule.

5.1 Revisions to the Rule

In recent work, Orphanides (1998) studies the effect of data revisions on a simple form of Taylor's rule. He finds that the prescriptions of the rule are quite different when real-time data are used, as compared to historical calculations of the rule using a recent vintage of data. As a consequence, a simple form of Taylor's rule does not appear to describe the Fed's behavior as well as Taylor (1993) claimed it did. In parallel fashion, Table 4 shows the effects of data revisions on the hybrid rule in (3). The elements of the table were calculated by first constructing what the implied value of the rule would be under both the real-time data sets and the full-sample data set (i.e., the 97:4 data set). Next, we averaged over the absolute percentage deviations of the last elements implied by the real-time data sets from the corresponding elements of the full-sample data set. Values are shown for various special cases of the parameters λ and θ . We assumed that $\mu_m = \mu_x = \pi^* + \mu_y$, and that the Fed's inflation target is held constant (in which case, it drops out of the calculations).

For the "moderate" regime ($\lambda = 0.4$, the case studied by Feldstein and Stock, 1994), the size of the revisions are very large, reaching almost sixty-percent in the case that the Fed has a relative distaste for real output variability. Overall, the effects of the revisions are highly dependent on the choice of λ and θ ; in all cases, they are non-trivial. The consequences of substantial revisions to the rule are that they make it difficult to compare, using a recent vintage of data, how the rule would have performed in certain episodes versus the historical record.

5.2 Performance Measures

Feldstein and Stock (1994) provide an extensive analysis of what they call the performance measure, r , of their rule, defined as:

$$r = \left(\frac{\text{var}(Z_{it}^*)}{\text{var}(Z_{it})} \right)^{\frac{1}{2}}$$

where $\text{var}(\cdot)$ is an unconditional variance and, in their case, Z_{it}^* and Z_{it} refer to nominal GDP growth when the system is controlled and uncontrolled, respectively. Since they analyze their rule only in the context of unrestricted VARs, the interpretation of their results is fraught with difficulties, which is well-argued by Taylor (1994) in the discussion of their paper. Namely,

their methodology fails to address the Lucas critique and it is difficult to interpret their results since their model lacks structure. Nonetheless, since the positive results of their counterfactual exercises suggest that the Fed could have achieved more efficient outcomes by using the rule in (2) (as measured by estimates of r less than one), it is useful to consider what the path of the performance measure would have looked like in real time.⁴⁴

In Figures 10 and 11, we plot performance measures calculated using real-time data sets versus values from the full sample, for various combinations of λ and θ .⁴⁵ All of the panels in Figure 10 correspond to nominal income targeting ($\theta = 0.5$). The panels on the left-hand and right-hand sides report results for “moderate” and “aggressive” regimes, respectively. For the moderate regime, the real-time results are not as positive as what would be determined using the full sample today. In some cases, the rule actually performs worse than the uncontrolled system ($r > 1$). Under the aggressive regime, the results are more varied. Using data sets from the nineties, the variance of nominal income would be reduced under the rule, but at the expense of increased variability in real output. Most of the efficiency gain would lie in inflation. However, historically, there are some periods in which the rule would have worsened outcomes for both real output and inflation. It appears that evaluation of the rule is sensitive to choice of the smoothing parameter, λ .

In Figure 11, we show estimates of r under moderate hybrid regimes. If the Fed has a relative distaste for real output variability (e.g. $\theta = 0.2$), then it would have appeared through time that small reductions in output variability could have been achieved. Historically, the picture generally looked more pessimistic for inflation than it does today. Alternatively, suppose the Fed had a strong preference for reducing inflation fluctuations (e.g. $\theta = 1.0$) – what might be called a pure inflation targeting regime. The gains for inflation would have been perceived to have changed substantially through time; output variance would have been higher, almost uniformly.

⁴⁴Furthermore, the stability of the paths of F-statistics in our tests of Granger-causality imply that the reduced-form system was stable over the full sample, even in the face of the supposed change in operating procedures in 79:4.

⁴⁵Results are for variances of quarterly variables. Feldstein and Stock (1994) also report results for variables over a semi-annual and annual horizon. The dates where no value is plotted correspond to cases where the largest eigenvalue in the uncontrolled system is greater than one - that is, the unconditional variance does not exist.

6 Conclusions

Our results show strongly that one's perception of statistical relationships between M2, output, prices, and interest rates is affected by the vintage of data used. In general, we urge researchers to use caution when interpreting historical episodes under the guise of a recent vintage of data. Secondly, we have demonstrated that M2 has a dubious marginal role to play in ongoing policy deliberations. Throughout the range of our samples, M2 Granger-caused real output but not prices, and it does not provide an improvement in out-of-sample forecasts over what can be obtained from the series themselves. Velocity appeared to be stable during the eighties, otherwise we could not reject a unit root; for the most part, there did not exist a cointegrating relation between velocity and the 90-day T-bill rate. Preliminary calculations for a simple feedback rule imply that its properties were highly variable over time.

Our conclusions, of course, are conditional on the models and techniques we employed. Investigations using the real-time data sets with structural models could shed further light on the effects of data vintage. In light of criticisms that have been levelled against simple-sum aggregates, a second topic for future research is to analyze the real-time properties of alternative money aggregates, like the Divisia index (see Barnett, 1980).

References

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*, March-April, 1-31.

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

Bernanke, B.S. and I. Mihov (1995): "Measuring Monetary Policy", NBER Working Paper, no. 5145.

Beckett, S. and C. Morris (1992): "Does Money Matter Anymore? A Comment on Friedman and Kuttner", Federal Reserve Bank of Kansas City Working Paper, (7).

Collins, S. and C.L. Edwards (1994): "An Alternative Monetary Aggregate: M2 Plus Household Holdings of Bond and Equity Mutual Funds", *Federal Reserve Bank of St. Louis*, November-December, 7-29.

Dolado, J.J. and H. Lutkepohl (1996): "Making Wald Tests Work for Cointegrated VAR Systems", *Econometric Reviews*, (15), 369-386.

Federal Reserve Bulletin (1979): Washington, D.C.: Federal Reserve Board, October.

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.

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.

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

Friedman, B.M. and K.N. Kuttner (1993): "Why Does the Paper Bill Spread Predict Real Economic Activity", in J.H. Stock and M.W. Watson. (eds.), *Business Cycles, Indicators and Forecasting*, Chicago: University of Chicago Press, 213-249.

Friedman, B.M. and K.N. Kuttner (1996): "A Price Target for U.S. Monetary Policy? Lessons from the Experience with Money Growth Targets", *Brookings Papers on Economic Activity*, (1), 77-125.

Hallman, J.J., R.D. Porter and D.H. Small (1991): "Is the Price Level Tied to the M2 Monetary Aggregate in the Long Run?", *American Economic Review*, (81), 841-858.

Miyao, R. (1996): "Does a Cointegrating M2 Demand Relation Really Exist in the United States?", *Journal of Money, Credit, and Banking*, (28), 365-380.

Orphanides, A. (1998): "Monetary Policy Rules Based on Real-Time Data", Finance and Economics Discussion Series Working Paper, Federal Reserve Board, (3).

Orphanides, A. and R. Porter (1996): "P* Revisited: Money-Based Inflation Forecasts with a Changing Equilibrium Velocity", manuscript, Federal Reserve Board.

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

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

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

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

Taylor, J.B. (1985): "What Would Nominal GDP Targeting do to the Business Cycle?", *Carnegie-Rochester Conference Series on Public Policy*, (22), 61-84.

Taylor, J.B. (1993): "Discretion versus Policy Rules in Practice", *Carnegie-Rochester Conference Series on Public Policy*, (39), 195-214.

Taylor, J.B. (1994): "Comment", in G.N. Mankiw, (ed.), *Monetary Policy*, Chicago: University of Chicago Press, 63-66.

Whitesell, W. and S. Collins (1996): "A Minor Redefinition of M2", Finance and Economics Discussion Series Working Paper, Federal Reserve Board, (7).

Box 1. Redefinitions of M2

Date	Redefinition
April 1971	Federal Reserve begins to publish M2. M2 includes currency, demand deposits, time deposits at commercial banks other than certificates of deposits.
February 1980	M2 adds overnight and (continuing contract) RP's that are issued by commercial banks to the non-bank public, overnight Eurodollars issued by Caribbean branches of member banks to US non-bank customers, MMMF shares.
July 1981	M2 adds travelers checks.
January 1982	M2 adds retail RP's, excludes institution-only MMMF's.
February 1983	M2 adds money market deposit accounts.
February 1990	M2 adds overnight RP's issued by thrift institutions move from term RP's (non-M2 component of M3) to overnight RP's.
February 1996	M2 excludes overnight wholesale RP's, overnight eurodollars.

Box 2. Revisions of M2

Document	Release Date	Revision Type
H6 - Money Stock, Liquid Assets & Debt Measures	Thursday - weekly	Revisions to data dated two months and earlier
	Thursday - weekly (mid-month)	Preliminary estimate for previous month
	Thursday - weekly (end of month)	First Revision for previous month
Money Stock Revision	Released in February - annual	Rebenchmarked and seasonally readjusted; data up to December of previous year

Box 3. Redefinitions of Output and Deflator

Date	Redefinition
1947:1 - 1991:4	GNP; deflator and real GDP constructed using a fixed weight measure (ie. weights of a single year for all periods).
1992:1 - 1995:4	GDP; deflator and real GDP constructed using a fixed weight measure (ie. weights of a single year for all periods). Base Period is 1987.
1996:1 - 1997:4	GDP; deflator and real GDP constructed using a chain weight measure (ie. weights of adjacent years). Base period is 1992.

Box 4. Revisions of Output and Deflator

Release Date	Revision Type
First month after end of quarter	Advance; Preliminary and Incomplete
Second month after end of quarter	Preliminary; Updated and Revised
Third month after end of quarter	Final; Final Revised Data

Table 1. Size of Revisions to Data Series

	Log Levels	Growth Rate
M2	14.39	24.21
Nominal Output	3.84	12.03
Real Output	78.77	45.21
Price Deflator	74.95	16.84

Note: The log levels are calculated as the mean absolute difference (expressed in percentages) between the last element of the real-time datasets and the full-sample data set. The growth rate is calculated as the mean absolute percentage difference between the last element of the real-time data sets and the full-sample data set.

Table 2. Comparison of Real-Time versus "Full-Sample" Data Vintages

A. Frequency of Rejections of Granger Non-Causality

Model (Sampling Method)	Figure	Level of Test	Nominal Output		Real Output		Price Deflator	
			Real-Time	Full-Sample	Real-Time	Full-Sample	Real-Time	Full-Sample
VAR (Recursive)	1	0.05	100	46	100	54	1	11
		0.01	90	0	90	0	0	0
VAR (15 Yr. Rolling)	2	0.05	91	36	99	52	15	23
		0.01	75	23	81	22	3	11
VECM (Recursive)	8	0.05	100	51	100	100	0	9
		0.01	89	1	91	76	0	0
VECM (15 Yr. Rolling)	9	0.05	97	46	95	76	18	53
		0.01	79	28	83	46	11	33

**B. Percentage of Times Real-Time Value Falls
Within One Standard Error of "Full-Sample" Value**

Sampling Method	Figure	Nominal Output	Real Output	Price Deflator
Recursive	3	31	44	44
15 Yr. Rolling	4	70	83	83

Table 3. RMSEs of Real-Time Forecasts

A. One-Quarter-Ahead, Quarterly Value

Model	Nominal Output		Real Output		Price Deflator	
	Recursive	15 Yr. Rolling	Recursive	15 Yr. Rolling	Recursive	15 Yr. Rolling
VAR(1)	3.8957	4.247	3.9925	4.0371	3.4823	4.5634
VAR(2)	3.6536	3.9201	4.0936	4.3381	2.2613	2.3331
VAR(3) with R_t	3.6353	3.9542	4.0066	4.5486	2.2415	2.3901
VAR(3) with m_t	3.6893	3.7473	4.2176	4.2174	2.2723	2.2671
VAR(4)	3.7118	3.7411	4.1843	4.1954	2.2463	2.2164

B. Two-Years-Ahead, Annual Value

Model	Nominal Output		Real Output		Price Deflator	
	Recursive	15 Yr. Rolling	Recursive	15 Yr. Rolling	Recursive	15 Yr. Rolling
VAR(1)	5.0693	4.9476	2.7799	2.8405	4.0752	3.8467
VAR(2)	4.9207	4.8841	2.9221	2.8991	3.5304	3.5059
VAR(3) with R_t	4.9289	4.8776	2.9246	2.8883	3.5182	3.4912
VAR(3) with m_t	4.8598	4.9144	2.8666	2.8819	3.5178	3.5101
VAR(4)	4.8605	4.9159	2.8815	2.9176	3.4724	3.4583

Note: Each entry gives the root mean squared error of forecasts. In both panels, for each variable, the sub-column "Recursive" refers to systems estimated using the full sample of each real-time data set; the sub-column "15 Yr. Rolling" refers to systems estimated using only the last 15 years of data from each real-time data set. In the top panel, forecasts are for the first quarter proceeding the dating of the data sets, so in effect, they are two-step-ahead forecasts. In the lower panel, forecasts are for annual observations two years proceeding the date of the data set, so in effect, they are averages of quarterly forecasts for six to nine periods into the future. All values are expressed in annualized percentages.

Table 4. Size of Revisions to Partial Adjustment Rule for M2

		λ		
		0.1	0.4	0.9
θ	0.2	19.16	57.39	23.98
	0.5	7.75	30.47	63.48
	0.8	18.62	45.42	12.57
	1	19.29	42.13	26.34

Note: First, the implied value for the rule is constructed using the real-time and full-sample data sets. Next, the mean absolute revision, in percentages, is calculated by taking the last implied value from each of the real-time data sets and comparing it to the corresponding value from the full-sample data set.

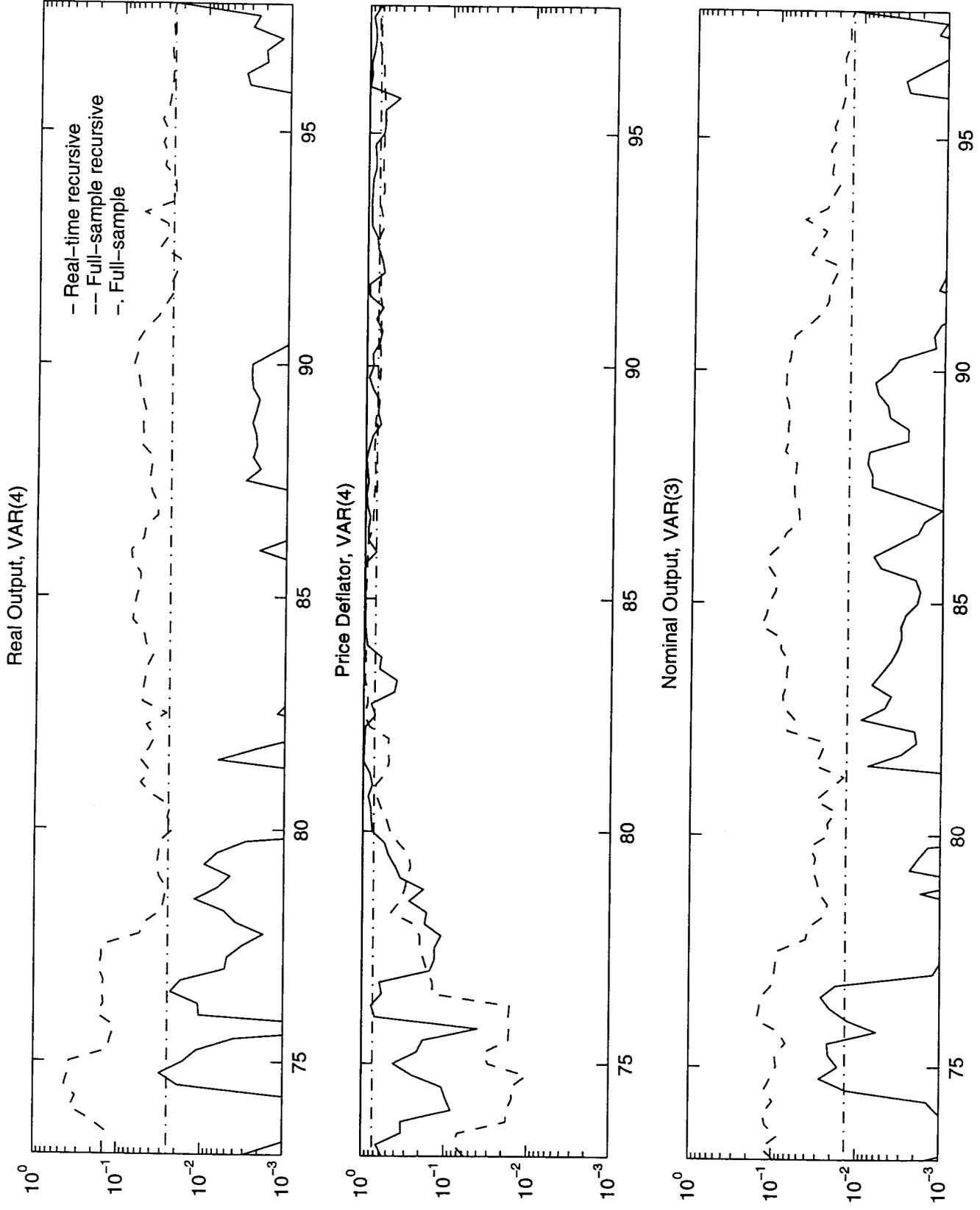


FIGURE 1. p-values Testing Significance of M2 (Recursive Samples)

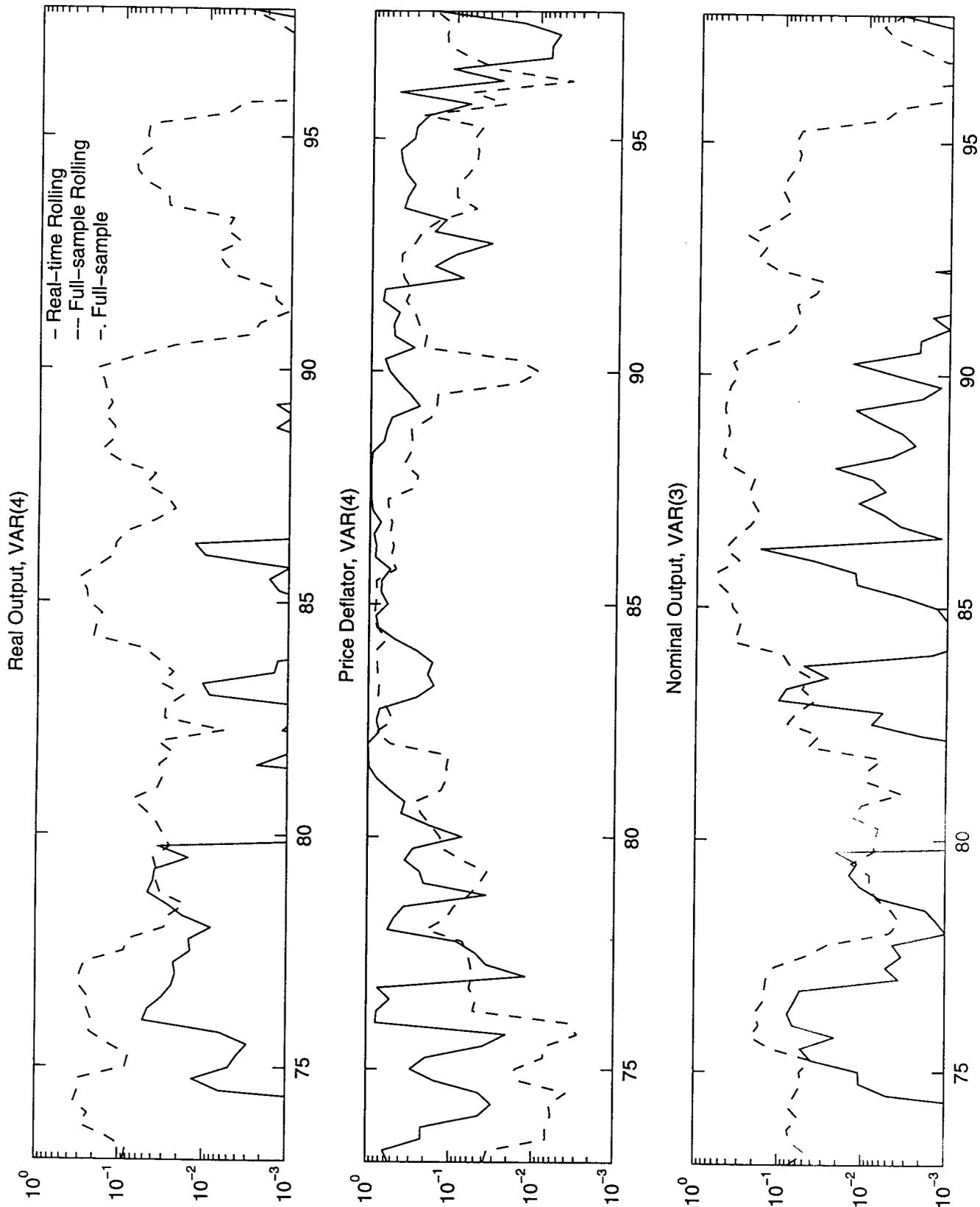


FIGURE 2. p-values Testing Significance of M2 (15 Yr. Rolling Samples)

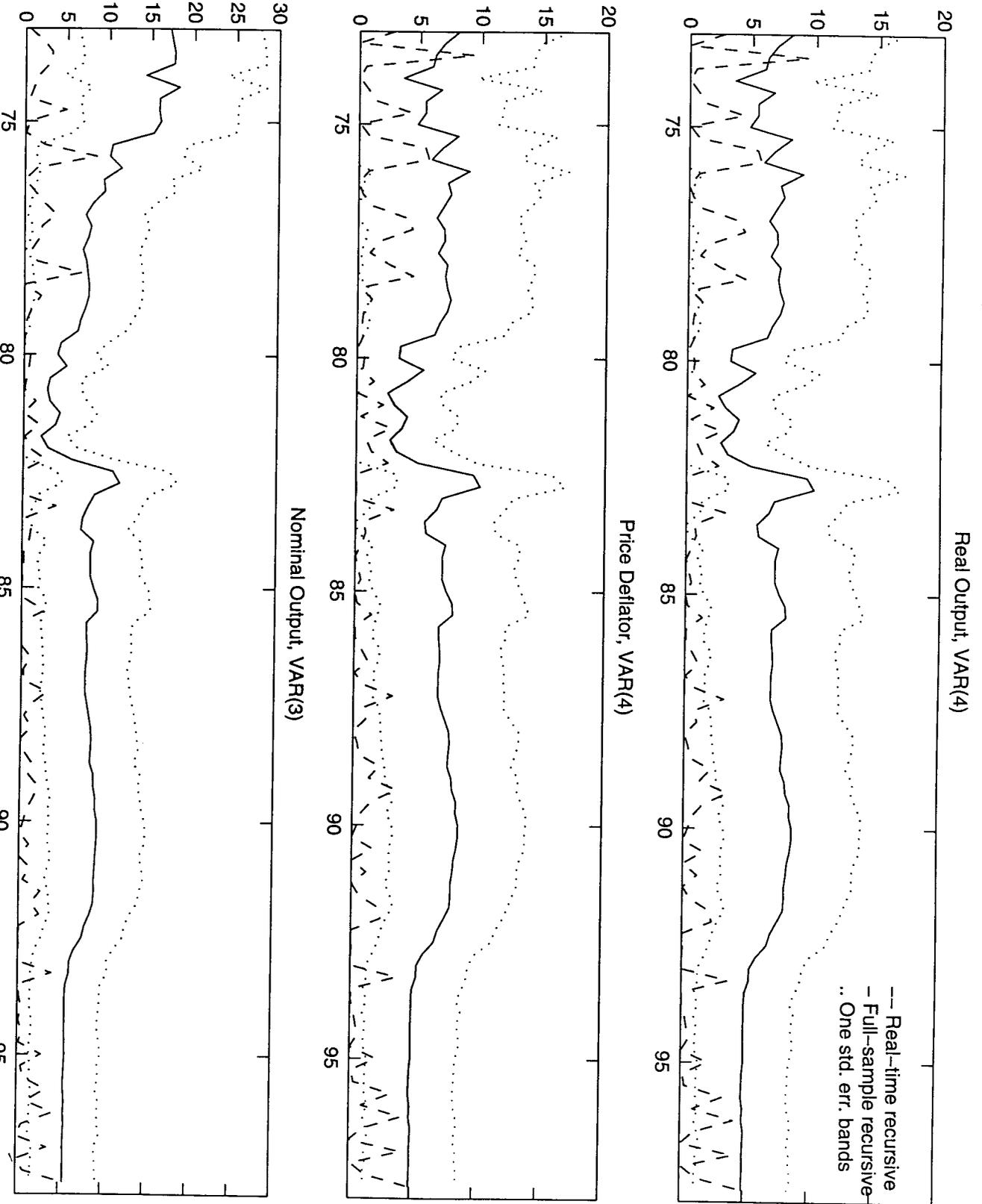


FIGURE 3. Percentages of Variation Accounted for by M2 at a Two-Year Horizon (Recursive Samples)

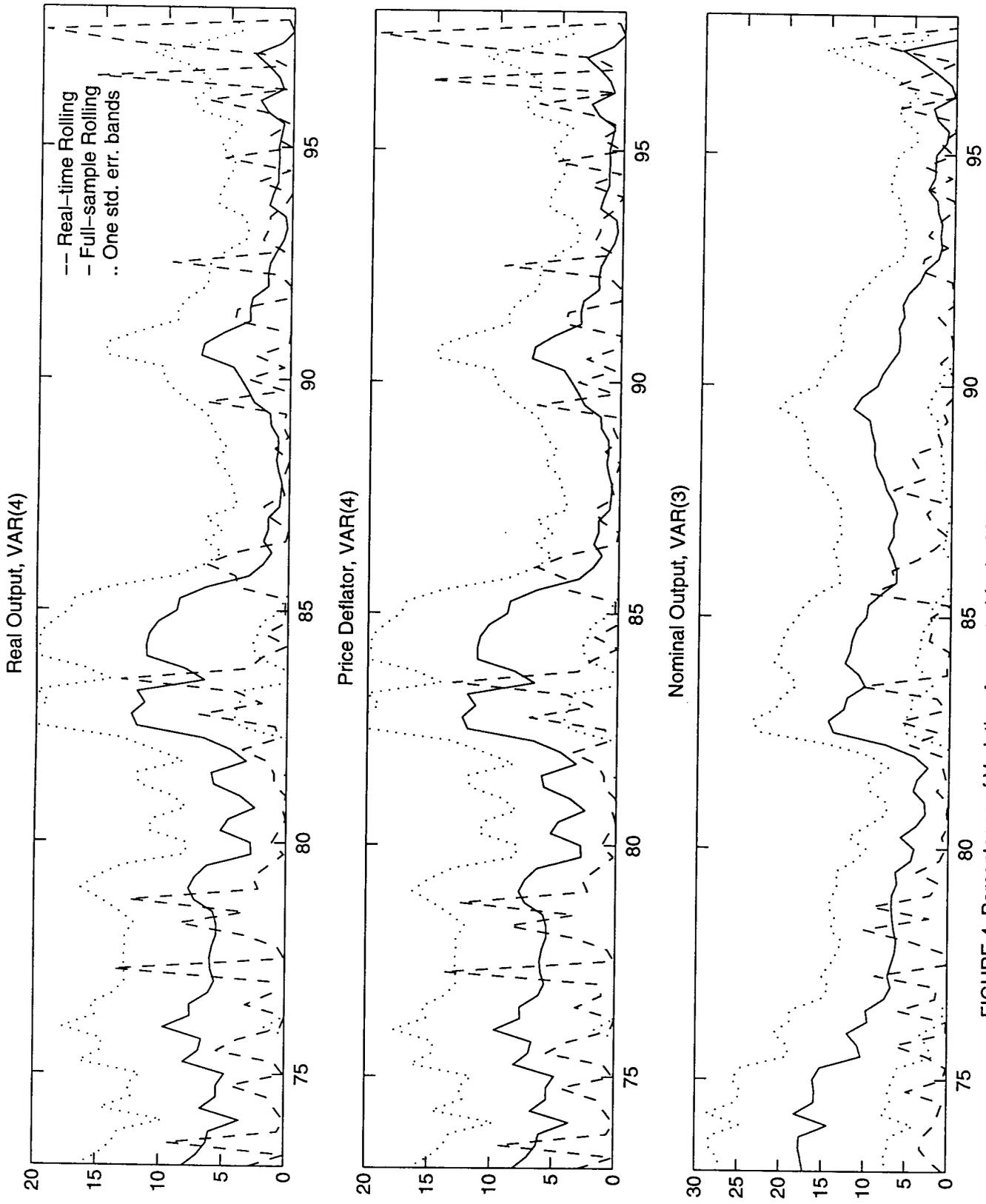


FIGURE 4. Percentages of Variation Accounted for by M2 at a Two-Year Horizon (15 Yr. Rolling Samples)

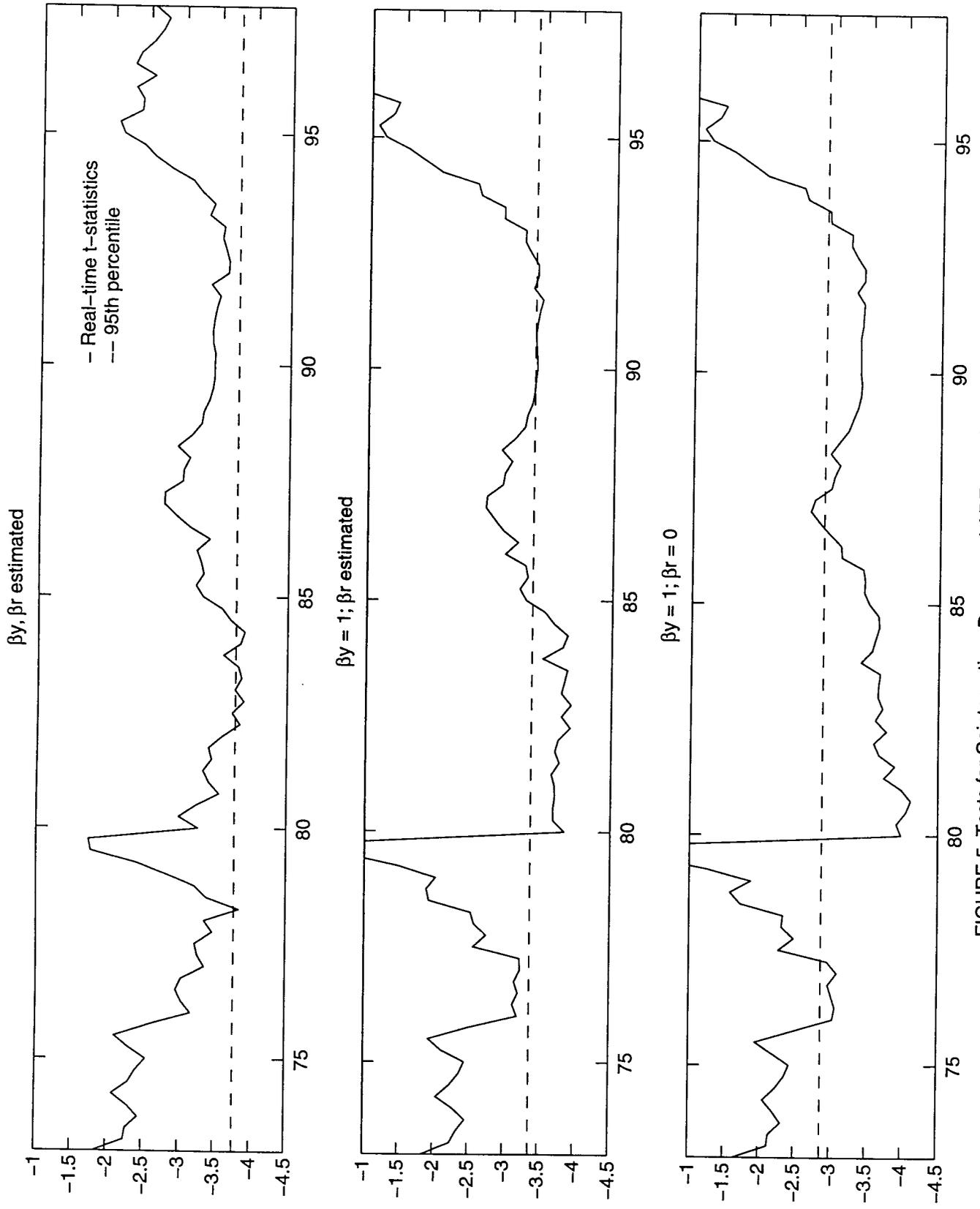


FIGURE 5. Tests for Cointegration: Demeaned ADF t-statistics (Recursive Samples)

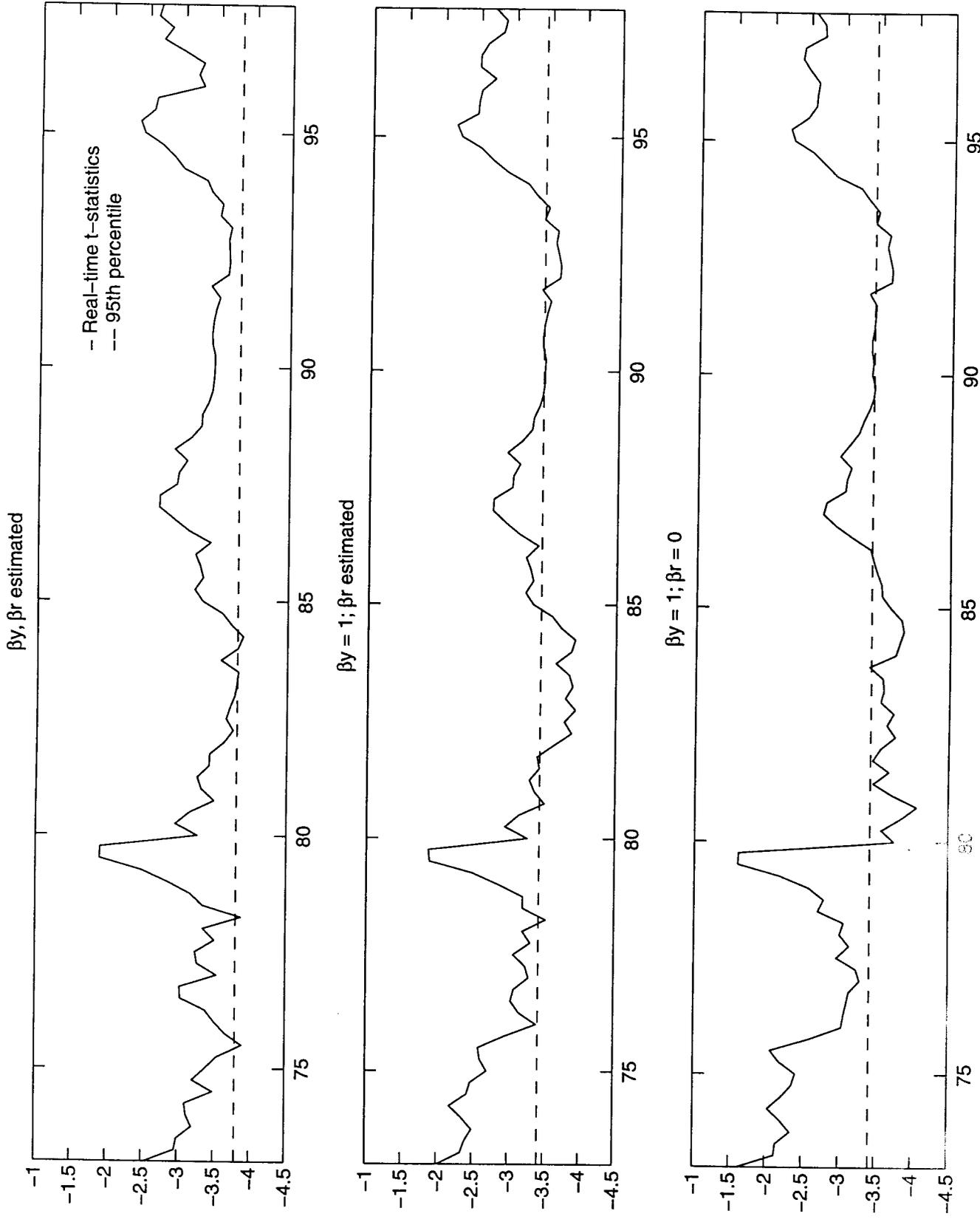


FIGURE 6. Tests for Cointegration: Detrended ADF t-statistics (Recursive Samples)

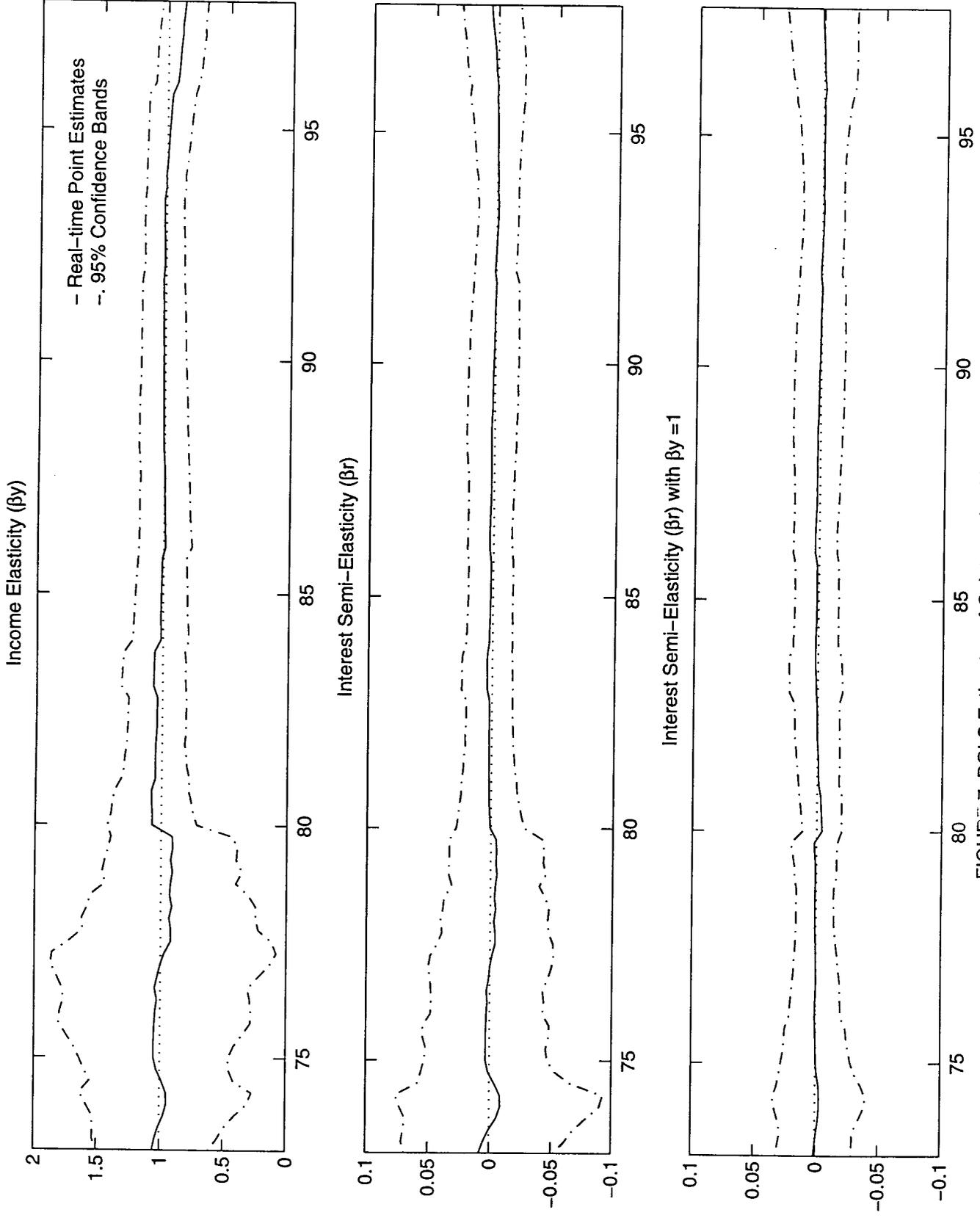


FIGURE 7. DOLS Estimates of Cointegrating Vector (Recursive Samples)

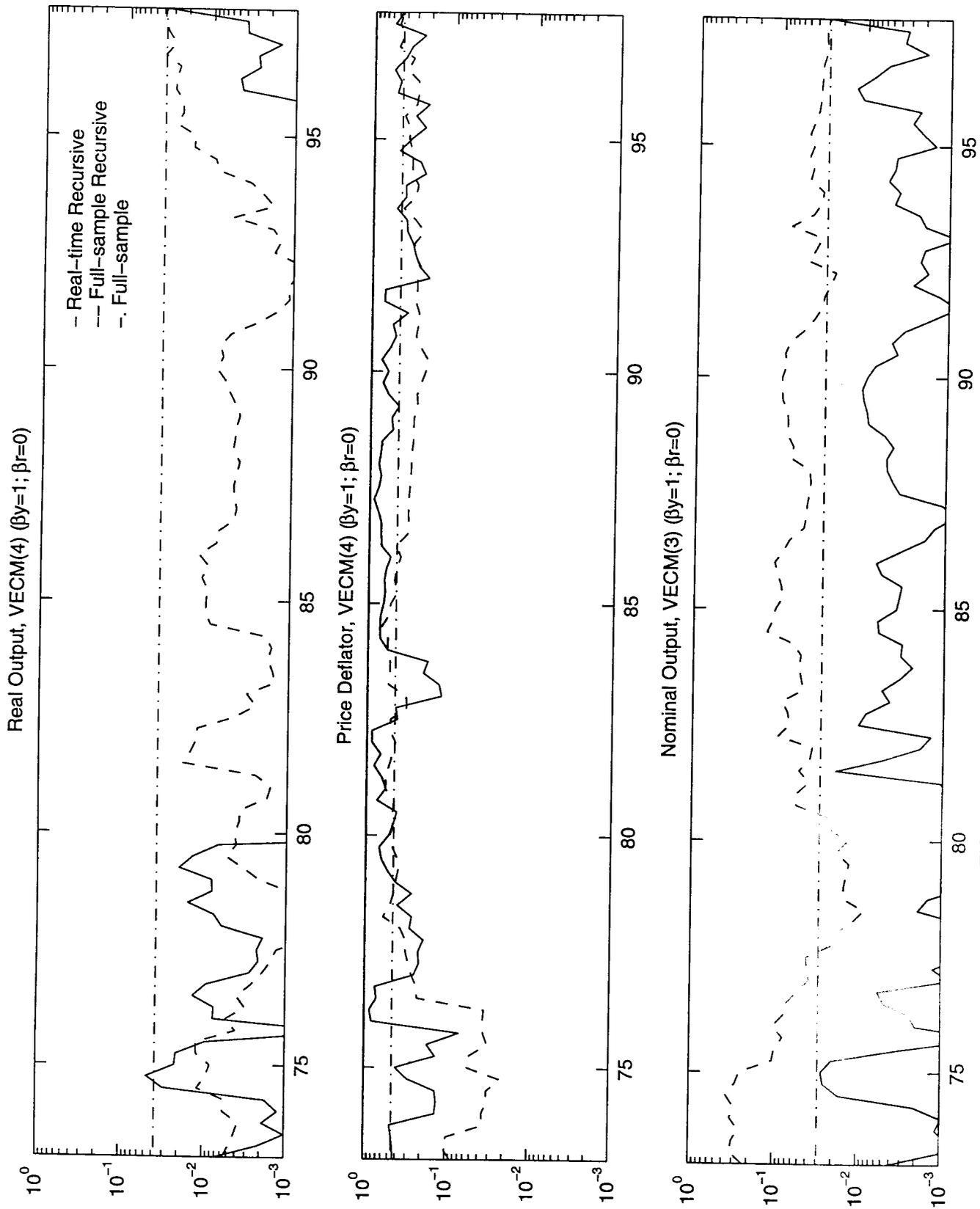


FIGURE 8. p-values Testing Significance of M2 (Recursive Samples)

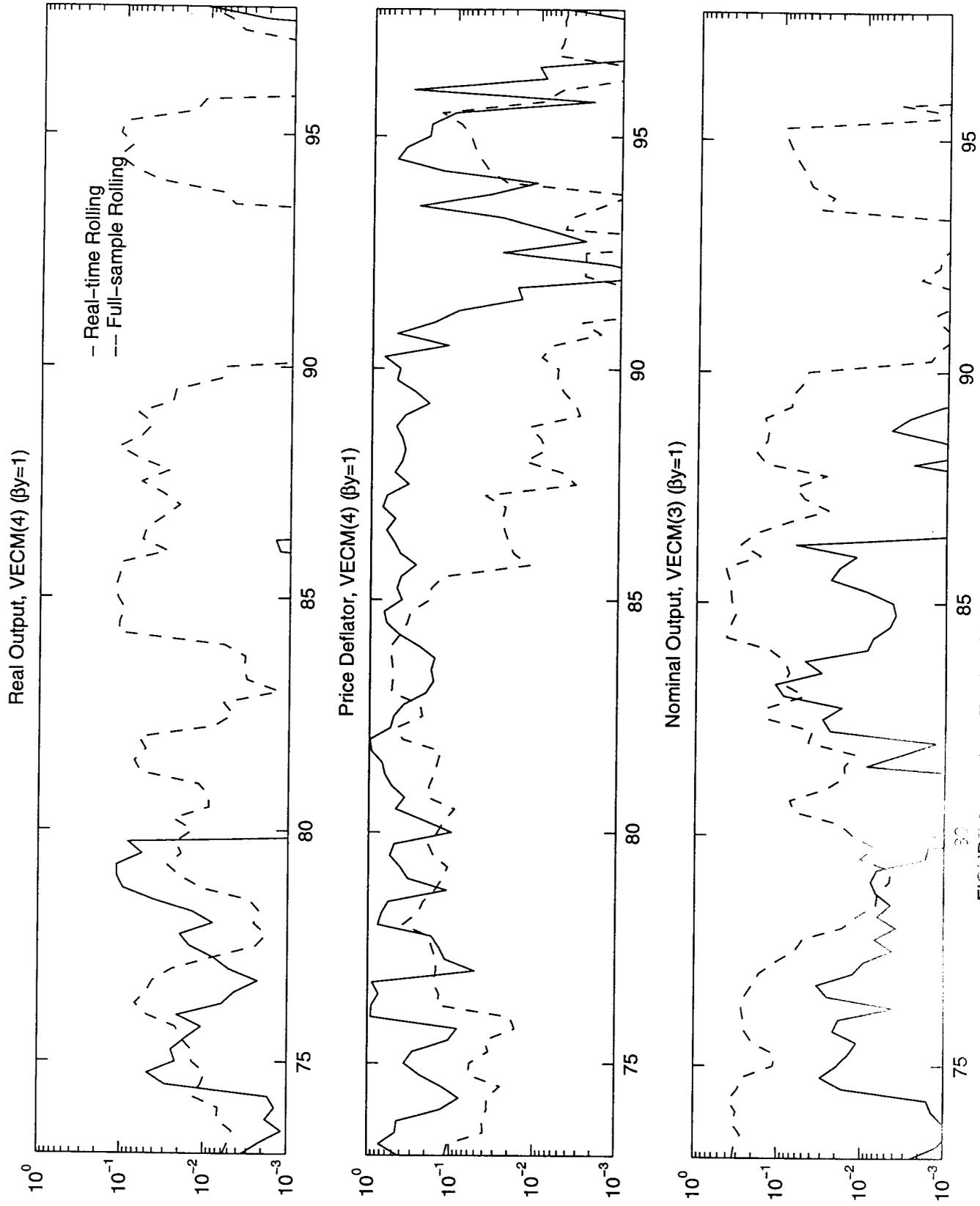


FIGURE 1. p-values Testing Significance of M2 (15 Yr. Rolling Samples)

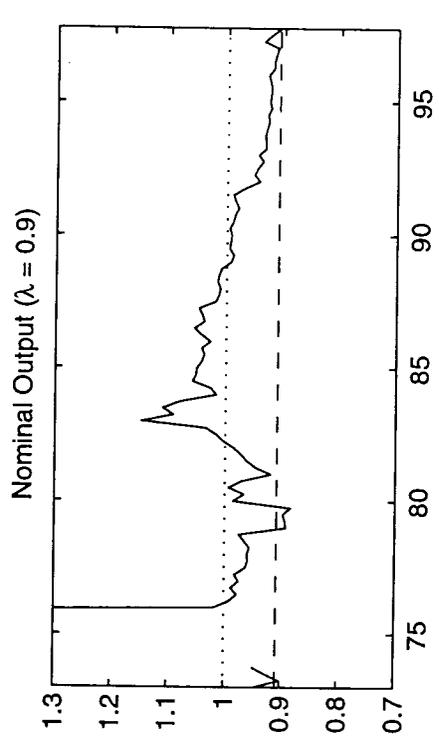
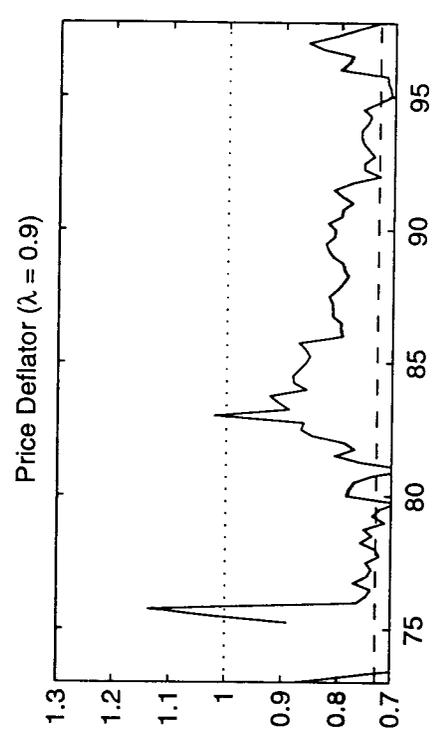
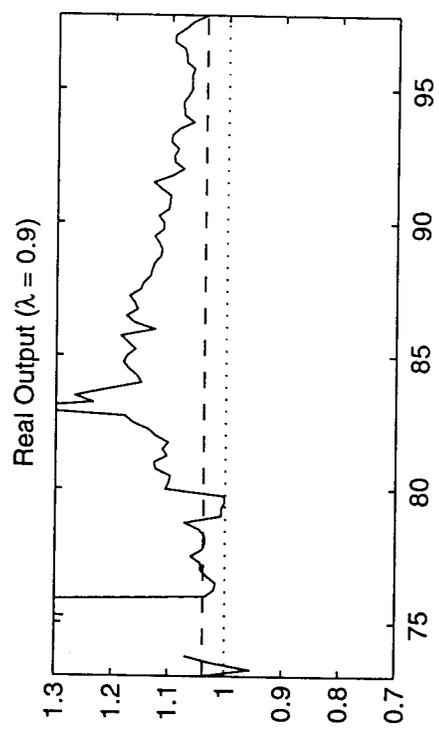
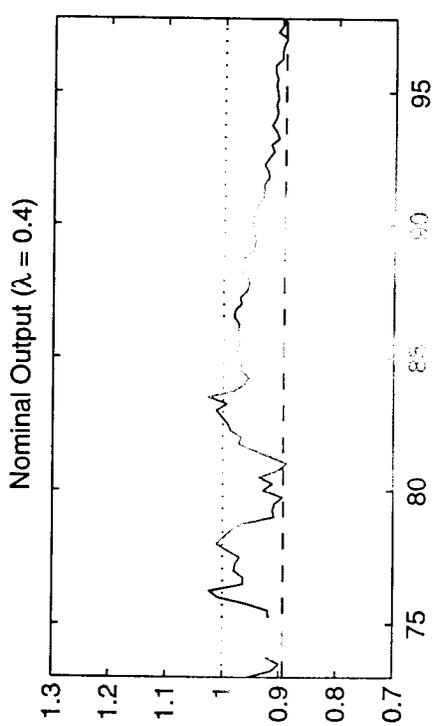
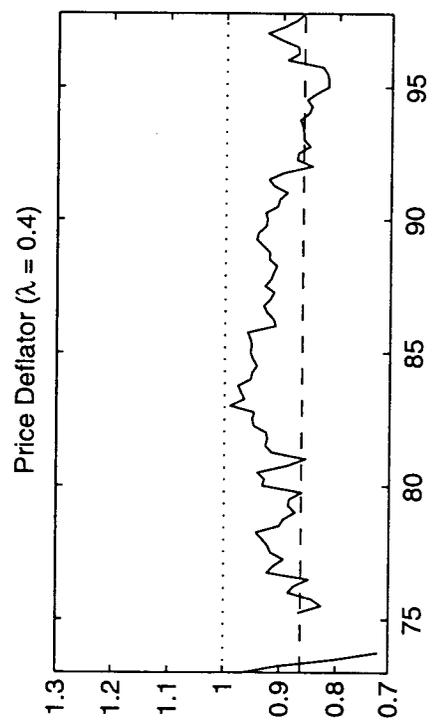
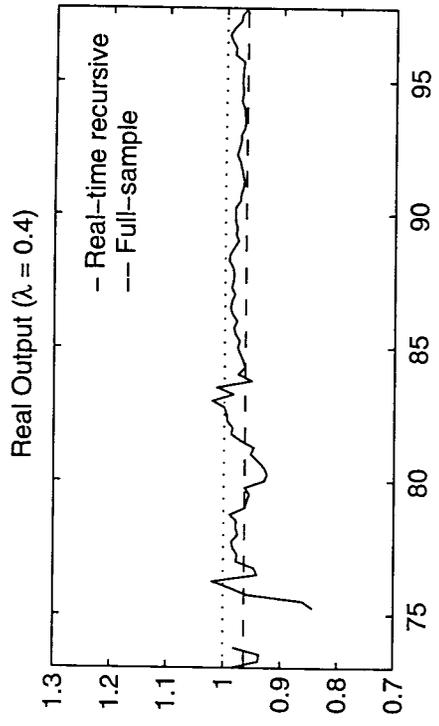


FIGURE 10. Feldstein-Stock Performance Measure for Nominal Income Targeting ($\theta = 0.5$)

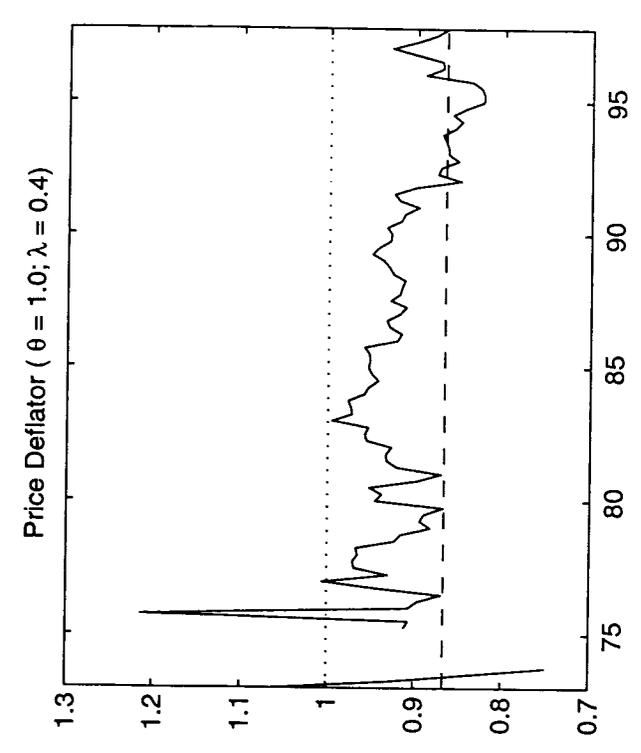
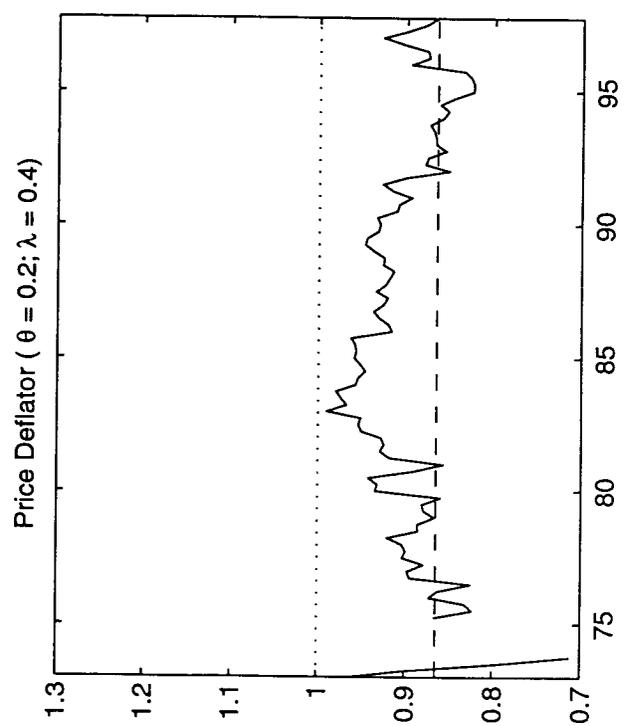
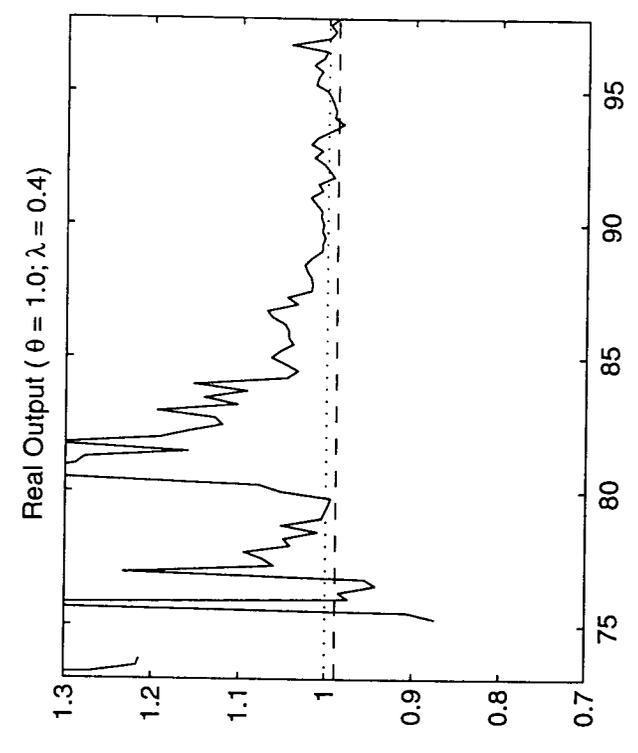
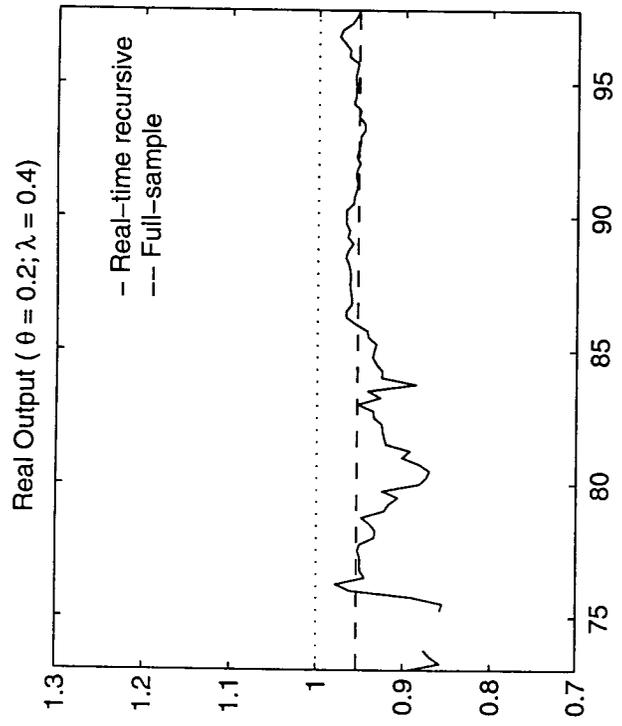


FIGURE 11. Feldstein-Stock Performance Measure for Hybrid Targeting