Bachmeier, Lance J.; Swanson, Norman R.

Working Paper

Predicting Inflation : Does The Quantity Theory Help?

Working Paper, No. 2003-17

Provided in Cooperation with:
Department of Economics, Rutgers University

Suggested Citation: Bachmeier, Lance J.; Swanson, Norman R. (2003) : Predicting Inflation : Does The Quantity Theory Help?, Working Paper, No. 2003-17, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
https://hdl.handle.net/10419/23174

Predicting Inflation: Does The Quantity Theory Help? *

Lance J. Bachmeier[1] and Norman R. Swanson[2]

[1] East Carolina University
[2] Rutgers University

August 2003

Abstract

Various inflation forecasting models are compared using a simulated out-of-sample forecasting framework. We focus on the question of whether monetary aggregates are useful for forecasting inflation, but unlike previous work we examine a wide range of forecast horizons and allow for estimated as well as theoretically specified cointegrating relationships in some of our models. Our findings indicate that there are forecasting gains from allowing monetary aggregates to enter into prediction models via cointegrating restrictions among money, prices, and output derived from a simple version of the quantity theory, but only when the cointegrating relations are specified *a priori* based on economic theory. When *estimated* cointegrating relations are used in a vector error correction (VEC) model, a vector autoregression (VAR) model in differences predicts better. These results hold, even during the 1990s, and evidence is presented suggesting that previous findings of a breakdown in the cointegrating relationship among prices, money, and output is the result of a failure of M2 as a measure of the money stock, and is not due to money demand instabilities. Two Monte Carlo experiments that lend credence to our findings are also reported on. The first shows that cointegration vector parameter estimation error is crucial when using VEC models for forecasting, and helps to explain previous findings of the failure of VEC models to forecast better than VAR models. The second shows that random walk and other atheoretical models routinely forecast better than *correctly* specified alternative models, due to parameter estimation error, indicating that caution needs to be exercised when interpreting the results of such comparisons, particularly when making statements concerning the usefulness of empirical models for use in policy-setting.

JEL classification: E31, C32.
Keywords: Inflation, Phillips curve, forecast evaluation, cointegration.

---

* Bachmeier: Department of Economics, East Carolina University, Greenville, NC 27858, BachmeierL@mail.ecu.edu. Swanson: Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, nswanson@econ.rutgers.edu. The authors wish to thank Graham Elliott, Clive W.J. Granger, Dennis Jansen, Allan Timmerman, and participants at the Texas Camp Econometrics Conference, the Southern Economic Association Meetings, and departmental seminars at Texas A&M University and East Carolina University for useful comments and suggestions. The first author thanks the Private Enterprise Research Center at Texas A&M University for support through a Bradley Dissertation Fellowship.

1 Introduction

Inflation forecasting has played a key role in recent U.S. monetary policy, and this has led to a renewed search for variables which serve as good indicators of future inflation. One frequently used indicator, based on the "Phillips curve," is the unemployment rate or a similar measure of the output gap (see e.g. Gerlach and Svensson (2003), Clark and McCracken (2003), Mankiw (2001), Gali and Gertler (1999), and the references contained therein). The Phillips curve is believed by many to be the preferred tool for forecasting inflation (see e.g. Mankiw (2001), Stock and Watson (1999a) and Blinder (1997)), although its use in formulating monetary policy is not without controversy (see e.g. Sargent (2000)). Another approach, based on the quantity theory of money, uses monetary aggregates to predict inflation. Despite the strong theoretical motivation for this approach, though, there is little evidence that measures of the nominal money supply are useful for predicting inflation relative to a conventional unemployment rate Phillips curve model (see e.g. Stock and Watson (1999a)). Stock and Watson's results indicate that even simple univariate time series models generally forecast about as well as models which include measures of the money supply, so that it is hard to make the case that nominal money supply data have any predictive content for inflation.[1]

This paper evaluates inflation forecasts made by models which allow for prices, money, and output to be cointegrated, and in the process re-examines the question of whether monetary aggregates have marginal predictive content for inflation. Our work is motivated in part by economic theory, as the presence of a cointegrating relationship among the series we look at corresponds to an implicit assumption that prices, the money supply, and output "hang together" in the long run, an implicit feature of most analyses based on the quantity theory.

From a statistical point of view, a system with cointegrated regressors does not have a finite-order VAR representation, so that a VAR in differences will be misspecified and may not forecast well regardless of the relevance of the included variables. Our analysis is therefore focused on the questions, 'Are there gains, in terms of forecast accuracy, from imposing the restriction that

[1]Similar evidence can be found in Leeper and Roush (2002, 2003). Inflation forecasts using many other variables, such as commodity prices, interest rates, exchange rates, and wages have also been studied in the literature (see e.g. Stock and Watson (1999a, 2003) and the references contained therein.) This paper does not consider forecasts made using these other variables. We also do not consider the quality of forecasts made by the private sector (see e.g. Croushore (1998)).

prices, money, and output are cointegrated?', 'Does it matter whether cointegrating restrictions are imposed *a priori* based on economic theory, or can they be estimated?', and 'Do models imposing cointegration among prices, money, and output forecast inflation as well as the Phillips curve and other alternative models?'

The econometric framework that we employ is similar to that of Stock and Watson (1999a), but differs from theirs in two ways. First, Stock and Watson (1999a) consider 1-year horizon inflation forecasts, while we consider forecast horizons of up to 10 years. This is potentially important in our context, as we include versions of the quantity theory of money in our analysis, a theory which arguably may not yield substantive gains to forecasting in the "short-run". Additionally, future inflation at many horizons is in general of interest to policymakers (even if the weight attached to inflation at different horizons is a matter of individual preference), so that "long-run" predictions are only unuseful if and when they fail to have marginal predictive content for inflation.[2] A second difference between our work and that of Stock and Watson (1999a) is that some of our models differ from theirs, including those which impose quantity theory based cointegrating restrictions, for example. In these types of models we: (i) impose a cointegration restriction derived from the assumption of stationary velocity, and (ii) estimate cointegrating restrictions. We also examine a fairly broad variety of (linear) models, including: simple autoregressive models in price levels and differences; conventional unemployment rate Phillips curve models; and VAR models in levels and differences with money, prices, and output. As a strawman model with which to compare our "best" models, we use various random walk models, and all models are evaluated using standard loss criteria such as mean square forecast error as well as via Diebold and Mariano (1995) tests of equal predictive accurcy.

Our approach is to consider alternative h-quarter ahead inflation predictions from the models mentioned above. We analyze two different periods, one from 1979:4-1992:4, and one from 1989:4-1999:4. These periods are analyzed separately because, as is shown below, Johansen (1988,1991) trace tests based on the full sample find cointegrating ranks of at least one through 1993 and zero thereafter, so that it is reasonable to allow for the possibility of a structural break around 1993.[3]

[2]The argument that short run inflation stabilization is not a feasible objective, and therefore that monetary policy should primarily be concerned with inflation at long horizons, goes back at least to Friedman (1959). See Amato and Laubach (2000) for one approach to determining the forecast horizon(s) of interest to a central bank.

[3]For a comprehensive and interesting discussion of the cointegration properties of our data in the 1990s, see Carlson, Hoffman, Keen, and Rasche (2000).

Sequences of 1-quarter to 5-year ahead predictions are made for the period 1979:4–1992:4, with one sequence of predictions constructed for each model and for each forecast horizon. This is done by re-estimating each model in a recursive fashion, using observations through 1979:4-h for the first forecast and observations through 1992:4-h for the last forecast, for $h = 1, ..., 20$ quarters. By focusing part of our attention on the period 1979-1992, we are able to assess whether predictions made using cointegrating restrictions estimated over a period for which it is well accepted that cointegration was present dominate prediction made without imposing cointegration. Furthermore, if estimated cointegrating restrictions over this period fail to yield predictive performance improvements, while restrictions imposed *a priori* based on economic theory do yield improvements, then we have direct evidence that the lack of success of cointegration type models in forecasting noted widely in the literature may be due in large part to parameter estimation error. We also carry out a more extensive version of the above exercise for the period from 1989:4-1999:4. Here, we consider predictions of up to 10 years ahead. Before describing our findings, it is worth stressing that there is much evidence that pre-existing cointegrating relations broke down in the 1990's (see e.g. Carlson et al. (2000)). However, we are interested in a wide range of forecast horizons, and cointegration tests are implicitly based on one-step ahead forecasting models. Thus, the failure of empirical cointegration tests does not imply that there are not long-run restrictions among the variables which will not yield improved long-run predictions (for more on this see Christoffersen and Diebold (1998)).

Our findings are clear-cut and can be summarized as follows. First, by allowing for prices, money, and output to be cointegrated, and by considering a variety of forecast horizons, there is evidence that M2 has marginal predictive content for inflation, and this holds true even during the 1990s, when cointegration tests are unable to find cointegration. In fact, for the longest horizons, the forecasts of our benchmark VEC model continue to dominate the forecasts from a VAR in differences over the entire 1989:4-1999:4 period. Further analysis reveals that the only time the VAR in differences forecasts inflation better than the VEC model is when data observed *after* 1993 are used in the *construction* of the predictions.[4] This is consistent with the hypothesis that M2 is not

[4]Thus, long-run predictions of inflation for the 1990s based on VEC models are better than those made with VARs in differences, even though the 1990s is supposed to be a period where cointegrating relations break down. The reason that the long-run VEC predictions remain the "best" is that data beyond 1993 are not used in their construction.

a good measure of the nominal money stock after 1993, and if it had remained an accurate measure of the nominal money stock, there would have been no observed breakdown in the relationship among prices, money and output. Thus, while much of the evidence in the literature points to a breakdown of cointegrating relationships during the 1990s as evidence of "unstable" money demand, our results indicate that this may not be the case.[5] However, even though we instead argue in favor of defining an appropriate monetary aggregate, this task is by no means trivial (see e.g. Feldstein and Stock (1996) and Barnett and Serletis (2000)).

Second, our findings supporting the usefulness of imposing cointegration are limited to the case where velocity is restricted to be stationary. For the period 1979:4-1992:4, we find that: (i) a VEC model that imposes stationary velocity typically forecasts better than a VAR in differences; (ii) forecasts from the VEC with stationary velocity also dominate, at all forecast horizons, those made using a VEC for which the cointegrating rank and vectors are estimated; and (iii) forecasts from a VAR in differences dominate the forecasts, at all horizons, made using a VEC with estimated cointegrating rank and vectors. These findings are suggestive. For example, we thus have evidence that when the cointegrating restriction(s) are estimated, we do better by simply using a VAR in differences. This corresponds to the common finding in the applied econometrics literature that VEC models do not usually predict better than VAR models (see e.g. Clements and Hendry (1996), Hoffman and Rasche (1996), and Lin and Tsay (1996)). What is interesting, though, is that when we impose the parameter (cointegration) restriction directly, based on theory, the VEC model does outperform the VAR model. This in turn suggests that one reason for VEC failure in practical applications may be imprecise estimation of cointegration vector(s) and/or cointegration space ranks, rather than incorrect model specification. Put another way, theory is important, and should be incorporated whenever possible. Given this finding, we perform a series of Monte Carlo experiments to investigate the importance of cointegration vector rank and parameter estimation error on VEC model forecasts. Using simulated data calibrated to be consistent with the historical U.S. record, we find that, for some configurations, the impact of cointegration vector rank and parameter estimation error on VEC model forecasts is substantial.

A second set of Monte Carlo experiments is also run because we find that a random walk model

[5]See e.g. Barnett and Serletis (2000), Estrella and Mishkin (1997), and Feldstein and Stock (1996) for further discussion of these alternative reasons for the breakdown in the cointegrating relation connecting prices, money and ouput.

is the *only* model (including the Phillips curve model) which the VEC model does not dominate at long horizons. Although such a finding is not important for our analysis, as we are interested in determining which variables have marginal predictive content for inflation, and comparison with the random walk model cannot answer this question, it is common to use a random walk model as a benchmark in out-of-sample forecast comparisons. A common interpretation of the failure of a model based on economic theory to forecast better than a random walk model is that the theory-based model is incorrectly specified. This interpretation is investigated using simulated data, calibrated to be consistent with the historical U.S. data, for two data generating processes, a second-order autoregressive (AR(2)) model, and a VAR model. We show that, for samples as large as 500, it is difficult to reject the null hypothesis that an AR(1) model forecasts as well as an AR(2) model, even when the data are generated according to an AR(2) process, and a random walk model usually forecasts better than a VAR model, even when the data are generated according to a VAR model. This serves to point out that one needs to be cautious when interpreting the results of out-of-sample forecast comparisons with atheoretical time series models, as parameter estimation error can cause correctly specified econometric models to forecast poorly. In particular, results from this experiment suggest that failure of an estimated version of a particular theoretical model to outperform a strawman random walk model in forecasting should not be taken as evidence that the theoretical model is not useful.

The remainder of the paper is organized as follows. Section 2 discusses the data used in our empirical investigation, while Section 3 outlines the methodology used. Quantitative findings are presented in Section 4, and Section 5 discusses the results of our Monte Carlo experiments. In Section 6, concluding remarks and directions for future research are given.

2 The Data

All data were downloaded from the Federal Reserve Economic Database on the Federal Reserve Bank of St. Louis web site. All data are quarterly U.S. figures for the period 1959:1 to 1999:4. For the price level, P_t, we use the gross domestic product implicit price deflator. In accordance with this choice of price index, we also use gross domestic product, Q_t, in chained 1996 dollars as our measure of real output. The money supply, M_t, data which we use are seasonally adjusted $M2$ figures. This choice of monetary aggregate is obviously not without its drawbacks. Barnett

and Serletis (2000), for example, contains a number of contributions which show the importance of using a monetary services index rather than simple sum $M2$. While the points made in these papers are important and valid, Diewert (2000) points out that divisia money supply measures require arbitrary choices in their construction, and these arbitrary choices can have a significant impact on empirical analysis. It is therefore natural to expect policy makers, at least in principle, to be interested in findings based on simple sum $M2$ measures[6]. Finally, unemployment, U_t, is the seasonally adjusted civilian unemployment rate.

3 Methodology

To begin, consider the equation of exchange, namely

$$M_t V_t = P_t Q_t \tag{1}$$

where P_t, M_t, and Q_t are defined above, and V_t is the velocity of money with respect to nominal output. Now, assume that P_t, M_t, and Q_t are I(1), using the terminology of Engle and Granger (1987). This assumption is standard in the literature testing for a cointegrating relationship between prices, money, and output, although there has been some debate as to whether prices are I(1) or I(2), (see e.g. Culver and Papell (1997) and the references contained therein). Unit root tests show that a unit root can be rejected for the first difference, but not the level, or the logged level for all three series.[7] In addition, for the time being, assume that $v_t = \log(V_t)$ is I(0) (see e.g. Feldstein and Stock (1994) or Estrella and Mishkin (1997) for a discussion of this assumption). Now, rearranging (1),

$$v_t = p_t - m_t + q_t, \tag{2}$$

where lowercase letters signify the use of natural logarithms. Assuming that there exists a vector autoregressive (VAR) representation of p_t, m_t, and q_t, the assumption of stationary velocity implies: (i) that there exists a cointegrating restriction among p_t, m_t, and q_t; and (ii) that the cointegrating

[6]It should in general be of interest to carry out empirical investigations with both varieties of monetary aggregates. Swanson (1998), for example, does this and finds little difference between empirical findings based on the two different types of aggregates, when using monetary services index data available on the St. Louis Federal Reserve Bank website.

[7]Augmented Dickey Fuller unit root tests with covariates, according to the procedure outlined in Elliot and Janssen (2003), were run on the natural logarithms of all variables, with lags selected according to the approach outlined in Ng and Perron (1995), and all were found to be I(1).

vector linking the variables is $(1, -1, 1)$, up to a scalar multiple. The Granger Representation Theorem (Engle and Granger (1987)) then states that the VAR in levels can be written as a vector error correction (VEC) model with price equation:

$$\Delta p_t = \beta_0 + \sum_{i=1}^{l} \beta_{pi} \Delta p_{t-i} + \sum_{i=1}^{l} \beta_{mi} \Delta m_{t-i} + \sum_{i=1}^{l} \beta_{qi} \Delta q_{t-i} + \phi z_{t-1} + \varepsilon_t,$$

where $z_{t-1} = \alpha'(p_{t-1}, m_{t-1}, q_{t-1})'$, α is a 3x1 vector of constants (i.e. the cointegration vector), ε_t is an error term, and l denotes the number of lags included in the VEC model. This is our benchmark model, is used to predict inflation, where we assume that $\alpha = (1, -1, 1)'$, and is referred to as the *"quantity theory VEC model"*. Alternatively, rather than fixing α (and assuming that v_t is stationary), we estimate α using the methodology of Johansen (1988,1991), allowing for the possibility that there may be no cointegration, so that $\alpha = 0$, and z_{t-1} is not included in the above model.[8] This model is referred to as our *"estimated VEC model"*.[9]

The above two models, as well as all of our other models, can be written as restricted versions of the following VEC model:

$$y_{t+h} = \beta_0 + \beta(L)x_t + \phi z_t + \varepsilon_{t+h}, \quad h = 1, ..., h_{max}, \tag{3}$$

where y_{t+h} is a scalar equal to either p_{t+h} or Δp_{t+h}, β_0 is a constant, $\beta(L)$ is a matrix polynomial in the lag operator L, x_t is a vector of explanatory variables, $h_{max} = 20$ for 1979:4-1992:4, and $h_{max} = 40$ for 1989:4-1999:4. When computing forecasts for the VEC models, we set $y_{t+h} = \Delta p_{t+h}$ and $x_t = (\Delta p_t, \Delta m_t, \Delta q_t)'$. Note also that ϕ is restricted to be equal to zero for models specified in levels or models specified in differences which do not allow for cointegration. For a given value of h, (3) is re-estimated recursively (i.e. re-estimated before each new forecast is constructed) in order to yield a sequence of 53 real-time rolling inflation forecasts for 1979:4-1992:4 and 41 rolling real-time inflation forecasts for 1989:4-1999:4. This procedure is then carried out again for a new value of h, and is repeated until h_{max} sequences of real-time $h-$step ahead forecasts are constructed.

[8]The approach of estimating the cointegrating restriction is standard in the literatures on stable money demand and on money income causality, for example.

[9]It is also standard in this literature to include a nominal interest rate among the variables in the cointegrating relationship. We do not include an interest rate variable because a strong theoretical argument can be made that real interest rates should be stationary (see e.g. Watson (1994)). Given our assumption that inflation is stationary, this implies that nominal interest rates are stationary. Experimentation revealed that our results are affected very little by the inclusion of an interest rate variable in the cointegrating relationship.

All estimation in this paper is based on the principle of maximum likelihood, and all lags (l) are re-estimated at each point in time (i.e. before each new prediction is constructed) and for each h, using the Schwarz (1978) Information Criterion (SIC), which is widely known to dominate other lag selection criteria (such as the Akaike (1973,1974) Information Criterion) when the objective is to produce optimal forecasting models (see e.g. Engle and Brown (1986) and Swanson and White (1995,1997)). Based on the above model, we define our prediction of cumulative inflation at period $t+h$ as $\widehat{\pi}_{t+h} = \sum_{k=1}^{h} \Delta\widehat{p}_{t+k|t}$, which implies that $\widehat{p}_{t+h|t} = p_t + \widehat{\pi}_{t+h}$, where in all cases the "$|t$" symbol denotes conditioning on information available at time t.

It remains to specify the rest of the models that will serve as "competitors" for the *quantity theory VEC model.*[10] One clear candidate is a VAR model with $y_{t+h} = \Delta p_{t+h}$, $x_t = (\Delta p_t, \Delta m_t, \Delta q_t)'$, and $\phi = 0$, which is the same as the above VEC model except for the restriction that $\phi = 0$. This model, thus, does not allow for prices, money and output to be cointegrated. It is henceforth called our *"VAR in differences model"*. An alternative to the *VAR in differences model* shall be called the *"VAR in levels"* model, and sets $y_{t+h} = p_{t+h}$, $x_t = (p_t, m_t, q_t)'$, and $\phi = 0$. As this model involves regression with I(1) variables, inference based on the estimated coefficients is not standard (see e.g. Sims, Stock and Watson (1990)). However, as our objective is prediction and not inference, this does not pose a problem for us. In addition, note that by estimating the VAR model in levels, we are allowing for cointegration among the variables, although the model is inefficient in the sense that we are not imposing the cointegrating restriction.

In addition to the VEC and VAR models, we estimate a conventional unemployment rate Phillips curve, which is shown in Stock and Watson (1999a) to be quite robust and is seldom beaten in their forecasting experiments except when their new index of aggregate activity based on 168 economic indicators is used. This model is called our "*differences Phillips curve model*", and sets $y_{t+h} = \Delta p_{t+h}$, $x_t = (\Delta p_t, U_t)'$, and $\phi = 0$. A levels version of this model, for which $y_{t+h} = p_{t+h}$, $x_t = (p_t, U_t)'$, and $\phi = 0$, is called the "*levels Phillips curve model*". We follow Stock and Watson (1999a) in assuming the NAIRU is constant and omitting supply shock variables, and therefore use just the unemployment rate when making Phillips curve forecasts. Finally, we also estimate differences and levels versions of a "*simple autoregressive model*", with $y_{t+h} = \Delta p_{t+h}$, $x_t = \Delta p_t$, and $\phi = 0$ for the differences version and $y_{t+h} = p_{t+h}$, $x_t = p_t$, and $\phi = 0$ for the levels version, and

[10] Table 1 presents a summary of the alternative models.

various random walk models including: $\Delta p_{t+h} = \Delta p_t + \varepsilon_{t+h}$, $h = 1, ..., h_{max}$ ("*differences random walk model*") and $p_{t+h} = \beta_0 + p_t + \varepsilon_{t+h}$, $h = 1, ..., h_{max}$ ("*random walk with drift model*").

Given forecast and actual inflation values, we are able to formally compare the predictive accuracy of the models. This is done by first forming real-time prediction errors as $e_{t+h|t} = \pi_{t+h} - \widehat{\pi}_{t+h}$, for each model, and for each value of h, $t + h$ = 1979:4-1992:4 for the first period and $t + h$ =1989:4-1999:4 for the second time period.[11] Then, predictions from the alternative models are compared using forecast mean square error (*MSE*), forecast mean absolute deviation (*MAD*), and forecast mean absolute percentage error (*MAPE*) criteria (see e.g. Swanson and White (1995,1997) for a discussion of these criteria). As these criteria are only point estimates, we additionally construct predictive accuracy tests based on each of the above three loss functions, along the lines of Diebold and Mariano (DM: 1995). These tests allow us to statistically assess whether our benchmark model outperforms each of the other models.[12] To illustrate, consider the following example. Let $l(e_{1t})$ denote the loss associated with the model 1 forecast at time t, and $l(e_{2t})$ denote the loss associated with the model 2 forecast at time t, say. If the two models have equal predictive accuracy, then $l(e_{1t}) = l(e_{2t})$. For the *MSE* loss function, we are interested in testing $H_0 : E[d_t] = E[e_{1t}^2 - e_{2t}^2] = 0$ against its negation, where $d_t = e_{1t}^2 - e_{2t}^2$. When model 1 and model 2 are nonnested, d_t is an asymptotically *N(0,1)* random variable. Thus, the DM test can be constructed, for example, by simply regressing d_t on an intercept and forming the heteroskedasticity and autocorrelation consistent t-statistic associated with the intercept. In our context, however, the models are sometimes nested, and McCracken (1999, 2000) shows that the usual DM t-statistic has a nonstandard distribution, for which he tabulates appropriate critical values. In summary, if d_t is significantly greater than zero, model 2 is "preferred", while model 1 is "preferred" if it is significantly less than zero. Otherwise, there is nothing to choose between the models. As our benchmark model is the *quantity theory VEC model*, we compare it against each of the other models to find out which one "wins" our prediction contest.

[11] Given that we have 10 different models, h_{max} = 20 for 1979:4-1992:4, and h_{max} = 40 for 1989:4-1999:4, a total of 27,000 different predictions and prediction errors are calculated.

[12] For a discussion of DM type tests in the context of multiple alternative models see Corradi, Swanson and Olivetti (2001) and White (2000).

4 Quantitative Findings

4.1 Lag length estimates and cointegration tests

Panel (a) of Figure 1 is a graph of the average number of lags chosen by the SIC for the differences autoregressive model at each forecast horizon for the period 1989:4-1999:4. The wide range of lag lengths chosen suggests that it is preferable to estimate the lag length each time a forecast is made, as opposed to holding the lag length fixed through time and across horizons, and underscores the difficulty in finding stable VAR and VEC models using standard estimation strategies (see e.g. Boivin (1999)). Panel (b) shows recursive estimates of the cointegration rank among p_t, m_t, and q_t. The cointegrating rank is estimated to be one or two until 1994, then falls to zero and remains there through the end of the sample, consistent with the recent literature (see e.g. Estrella and Mishkin (1997)). Panels (c) and (d) show the estimated coefficients on M2 and real GDP, respectively, for the cointegration vector associated with the largest eigenvalue in the model. In 1975 the estimated coefficient on M2 begins to fluctuate (even though the cointegration vector remains statistically significant), but by 1978 the estimated coefficient is again close to -1. The coefficient on output does not fluctuate as much as the coefficient on M2, though it does show large fluctuations in 1975, 1976 and 1982. Overall, all large deviations of the estimated coefficients from the restriction associated with the quantity theory VEC model appear to have been transitory.

4.2 Forecast evaluation results for 1979:4-1992:4

Figure 2 shows results from comparison of the quantity theory VEC and the VAR in differences model for the period 1979:4-1992:4, and answers our first question, namely, 'Are there gains, in terms of forecast accuracy, from imposing the restriction that prices, money, and output are cointegrated?' Panels (a)-(c) are graphs of *MSE, MAD* and *MAPE* for the two models at each of the 20 different forecast horizons, with the quantity theory VEC model performing better at all horizons regardless of the loss function. Panels (d)-(f) in Figure 2 are graphs of the DM statistics for tests of the null hypothesis that the two models predict equally well, according to *MSE, MAD*, and *MAPE* loss functions, respectively, for each of the 20 different horizons. In this case the only difference between the two models is that the quantity theory VEC model includes an error correction term, so that the two models are nested. A significantly positive DM statistic implies that imposing cointegration improves forecasts relative to a VAR in differences model, while a significantly negative DM statistic

implies the opposite. Notice that the DM statistics are almost always positive, and are often significant (using critical values which are in the 1-2 range in our framework - see McCracken (1999)). Thus, our benchmark model, in which the quantity theory restriction is imposed *without* estimating a cointegrating vector, performs better than the VAR in differences.

Nevertheless, estimation of the cointegration rank and vector has two advantages. Namely, it allows for cases in which the cointegrating relationship cannot be identified *a priori*, and it allows for the possibility that the system is evolving over time. Panel (a) of Figure 3 shows DM statistics for the comparison of the estimated VEC model and quantity theory VEC model, with positive statistics implying the superiority of the quantity theory VEC model at a particular forecast horizon. The quantity theory VEC model dominates at all horizons, meaning that cointegration vector parameter and/or cointegrating rank estimation error is very important for forecasts in samples of the size we have. Notice also that the quantity theory VEC does increasingly better as the horizon increases, suggesting that the quantity theory is particularly useful for long-run prediction. Furthermore, note that in panel (b) of Figure 3, the VAR in differences dominates the estimated VEC model at all horizons, and by a sizable amount (in this panel, a significantly positive statistic implies that the estimated VEC forecasts better than the VAR in differences). In other words, if we had estimated the cointegration vector and rank each time a forecast was made, rather than imposing stationary velocity and a cointegrating rank of unity, we would have reached a conclusion that the quantity theory VEC was not useful, and would have concluded that imposing cointegration never improves out-of-sample prediction in our context! These findings at least partially explain previous findings that imposing cointegration often does not result in improvement over forecasts constructed using VAR models. In short, we find that there can be large gains from *a priori* knowledge of the cointegrating vector and rank, and that economic theory plays an important role, at least when our objective is prediction. One reason why this is the case appears to be that parameter and cointegration rank estimation error is large in our framework, as is shown via a series of Monte Carlo experiments in the next section.

Figure 4 presents graphs of the DM statistics from comparison of the quantity theory VEC model with various other alternative models under the assumption of *MSE* loss. The dashed lines in each graph are 90% and 95% critical values, so that a DM statistic above the lower dashed line indicates that the quantity theory VEC model forecasts better for that horizon at a significance level of 10%, and a DM statistic above both dashed lines indicates that the quantity theory VEC

model forecast better at a significance level of 5%. When the models are nested, the critical values depend on the length of the out-of-sample prediction period, the number of observations used to estimate the model parameters, and the number of additional parameters in the larger model, and therefore vary across horizons, and reported critical values are again taken from McCracken (1999). Comparison with the VAR in levels (Panel (b)) is of some interest here, as this model allows for cointegration of unknown form (see e.g. Sims, Stock and Watson (1990)). Note that in this case, the quantity theory VEC model dominates at all horizons, so that failure to impose the correct cointegrating restriction, which leads to estimation inefficiency in the levels VAR model, also leads to poorer predictions. Similar results are obtained for the other levels models. As noted in the introduction, Phillips curve models are generally believed to provide the best inflation forecasts. As might be expected, Panel (e) shows that the differences Phillips curve forecasts better at short horizons, while the quantity theory VEC model forecasts better at the longer horizons. It is interesting to note that the quantity theory VEC begins to dominate the Phillips curve model after around three years, suggesting that the long-run (at which time the quantity theory begins to be useful) is perhaps not very long! Similar results are obtained when comparing the quantity theory VEC model with the differences autoregressive model. At horizons of two through ten quarters, a simple autoregressive model forecasts about as well as the quantity theory VEC model, a result consistent with the findings of Stock and Watson (1999a). For the longer horizons, though, the quantity theory VEC model performs much better, and we have evidence that the variables in the quantity theory VEC model, including M2, have marginal predictive content for inflation beyond that in the autoregressive model.

Figure 4 also shows the results of comparing the quantity theory VEC model with the differences random walk model, and we observe that even at the longest forecast horizons, the differences random walk model forecasts about as well as the quantity theory VEC. There are several reasons, though, that this is not evidence that the quantity theory VEC model is useless, at least for purposes of monetary policy. First, the differences random walk model is not a reasonable policy model, because it contains no control variables, and merely summarizes the historical time series properties of the inflation series. Further, the relevant question for policy is whether the variables in the quantity theory VEC model contain information about future inflation, and comparison with the differences random walk model cannot answer this question. Finally, failure of the quantity theory VEC model to forecast better than the differences random walk model does not necessarily

imply that the quantity theory VEC model is incorrectly specified. In fact, the Monte Carlo experiments discussed in the next section show that a parsimonious, but misspecified, time series model may forecast better than a correctly specified model due to parameter estimation error.

4.3 Forecast evaluation results for 1989:4-1999:4

Panels (a)-(c) in Figure 5 show MSE, MAD and MAPE for the quantity theory VEC and the VAR in differences models for the period 1989:4-1999:4. For low horizons (less than 5 years), the loss associated with the quantity theory VEC model is greater than that for the VAR in differences, regardless of loss function. The test statistics in panels (d)-(f) confirm that the quantity theory VEC model generally forecasts much worse than the VAR in differences model at low horizons. Figure 6 shows that the estimated VEC model does even worse than the quantity theory VEC model, and Figure 7 shows that the alternative models typically also outperform the quantity theory VEC model at low horizons. Imposing cointegration on low-horizon forecasts in the 1990's would have been harmful, which is not surprising given the previous literature documenting the breakdown in the stability of M2 demand. Surprisingly, though, the quantity theory VEC model generally forecasts much better than the VAR in differences model and the alternative models at longer horizons. This result suggests that failure to find cointegration after 1993 may not be due to the fact that "money demand became unstable", as is the usual interpretation, since in that case the VEC model would be misspecified, and this would affect inflation forecasts at all horizons.

To make our argument concrete, assume that the inflation series is generated as

$$\Delta p_{t+h} = \beta_0 + \sum_{i=0}^{l-1} \beta_{pi} \Delta p_{t-i} + \sum_{i=0}^{l-1} \beta_{mi} \Delta m_{t-i} + \sum_{i=0}^{l-1} \beta_{qi} \Delta q_{t-i} + \phi z_t + \varepsilon_{t+h}, \qquad (4)$$

where $z_{t-1} = \alpha'(p_{t-1}, m_{t-1}, q_{t-1})'$, and $\alpha = (1, -1, 1)'$ is the cointegration vector. Consider a permanent money demand shock which causes the cointegrating relationship to break down. Once the cointegrating relationship has broken down, the coefficient on z_t changes (to zero), and this will cause the quantity theory VEC model to be misspecified, at least until there are enough new observations to enable our recursive estimation methodology to account for the change in ϕ. Now consider the case where there are no money demand instabilities and M2 is initially an accurate measure of the money stock, m_t, but at some point in time M2 begins to underestimate the nominal money stock, so that $M2_t = m_t + \nu_t$. This would be classified as a measurement error problem and would, of course, only affect forecasts made using data affected by the measurement error,

hence allowing for our finding that longer horizons predictions are always better when formed with the quantity theory restriction, even in the 1990s. The above discussion also suggests a test for distinguishing between different explanations for the breakdown in cointegrating relations during the 1990s. If the breakdown is due to instabilities in money demand, a VEC model will be misspecified and all forecasts made after the breakdown will be affected. If, however, the breakdown is the result of measurement error, the VEC model should forecast poorly only for those periods where the M2 data used to make the forecasts were observed after the instabilities began.

Table 2 breaks down the mean absolute forecast errors into three time periods. "Period 1" refers to forecasts from 1989:4-1992:4, "period 2" refers to forecasts after 1993:1 which did not use M2 data from 1993:1 or later,[13] and "period 3" refers to the remaining forecasts. Table 2 shows that the mean absolute forecast error for the quantity theory VEC model is slightly higher for time period 2 than time period 1, but is substantially higher for time period 3 than for either time period 1 or time period 2. The five-year forecast horizon MAE is almost five times higher for time period 3 than for time period 2. This indicates that the breakdown of cointegrating relationships was likely not a result of what has conventionally been termed an "unstable money demand function". Rather, it seems that M2 needs to be redefined, as done for instance in Feldstein and Stock (1996). Carlson, et al (2000), for instance, have argued that by allowing for a one-time structural change in the early 1990's, an M2 cointegrating relation continues to have relevance in the context of a nominal income policy rule. In that sense, it is possible that the quantity theory VEC model may still be relevant in spite of the problems with M2, but it is too early to test this possibility in a framework consistent with the real-time forecasting methodology employed in this paper[14].

4.4 Additional Issues

This section evaluates the robustness of our findings to the criticism made by a number of authors (see e.g. Woodford (1994)) that indicator variables, when used correctly by policymakers, will fail to be useful for reduced-form forecasting purposes. Woodford (1994) illustrates the point with the

[13]For instance, "period 2" covers 1993:1-1993:4 for one-year horizon forecasts, 1993:1-1994:4 for two-year horizon forecasts, and so on.

[14]A question we do not address in this paper is detection of the instabilities in practice, but there are reasons to believe it would not have been difficult. First, Johansen's (1988,1991) cointegration test finds a cointegrating rank of zero after 1993:4. Second, Greenspan (1993) states that policymakers knew about the problems with M2 in the first half of 1993.

following simple example. Assume that inflation is determined by

$$\pi_{t+1} = s_t + u_t + \varepsilon_{t+1},$$

where π_t is the inflation rate at time t, s_t is an information variable, u_t is a control variable, and ε_{t+1} cannot be forecast at time t. If the goal is to minimize the variance of inflation, the optimal policy will be to set $u_t = -s_t$, meaning $\pi_{t+1} = \varepsilon_{t+1}$, and s_t will then not be useful for forecasting time $t+1$ inflation in a reduced-form VAR model. Admittedly, this result requires strong assumptions, but it is plausible that money supply data have been used as indicator variables in the past, and if the corresponding policy led to changes in M2, it would weaken the evidence that monetary aggregates are useful for forecasting inflation.[15] This section provides some evidence on the importance of the criticism, with a thorough analysis of the issue left to future research. We compare forecasts over the earlier period, 1979:4-1992:4, for the "simple autoregressive model" in differences and the quantity theory VEC model, with an additional term included in the latter model to account for growth in m_t between period t and period $t+k$. The benchmark model is the autoregressive model

$$\Delta p_{t+h} = \beta_0 + \sum_{i=0}^{l-1} \beta_{pi} \Delta p_{t-i} + \varepsilon_{t+h},$$

where all variables are defined as before. This autoregressive model is compared to the following modified version of the quantity theory VEC model

$$\Delta p_{t+h} = \beta_0 + \sum_{i=0}^{l-1} \beta_{pi} \Delta p_{t-i} + \sum_{i=0}^{l-1} \beta_{mi} \Delta m_{t-i} + \sum_{i=0}^{l-1} \beta_{qi} \Delta q_{t-i} + \phi z_t + \gamma \Delta m_{t+h} + \varepsilon_{t+h}, \quad (5)$$

where $\Delta m_{t+h} = m_{t+h} - m_t$ is the change in M2 from period t to period $t+h$, and all other variables defined as before. Inclusion of Δm_{t+h} accounts for the possibility that M2 was used by the Fed as an indicator variable for future inflation, and that the corresponding monetary policy influenced the path of M2. As γ cannot be estimated consistently due to the potential that it is endogenous, we present results for values of $\gamma = (0, 0.25, 0.5, 1)$, a range that includes all reasonable values of γ. Recall that we found in Section 4 that the quantity theory VEC model did not forecast better at low horizons than the autoregressive model in differences, implying that lagged values of m_t did not help forecast inflation.

[15] Romer and Romer (2000) and Bernanke and Boivin (2003) find that Federal Reserve Greenbook forecasts are very accurate relative to other forecasts, and Bernanke and Boivin (2003) suggest that this may result from the fact that Greenbook forecasts are made conditional on future policy actions.

Figure 8 shows the DM statistics at horizons of 1 through 20 quarters associated with the comparison of the above modified quantity theory VEC model with the autoregressive model, for different values of γ. As before, positive DM statistics imply that the quantity theory VEC is preferred. Figure 8 shows that forecasts from the quantity theory VEC model tend to be slightly worse when a correction is made for the possibility that M2 has been used as an indicator variable in the past, implying that this explanation cannot explain the lack of marginal predictive content of M2 for inflation at low horizons.

It should be noted that we have presented only a crude estimate of the potential effects of the Fed (correctly) using monetary aggregates as indicator variables in the past, and a more thorough analysis would require the work in this section to be extended to include formal structural policy models, including identified VAR models. Such an analysis of identified VAR models would also be of independent interest, as the results of structural VAR analysis based on different identifying assumptions often leads to markedly different conclusions, yet there is typically no way to distinguish among the various identifying assumptions. Asymptotically, a correctly identified VAR model should be expected to forecast better than any other identified VAR model, but verifying this would require a formal analysis of the properties of such a comparison, which is left to future research.

5 Monte Carlo Experiments

In this section, we investigate the importance of parameter estimation error for the forecasts of several of the models considered above.

The first set of comparisons is designed to study the importance of cointegration vector rank and parameter estimation error on forecasts from VEC models. In particular, 5000 samples of data were generated using the following data generating process (DGP):

$$\Delta Y_t = a_3 + b_3 \Delta Y_{t-1} + c_3 Z_{t-1} + \epsilon_{3t}, \tag{6}$$

where $Y_t = (p_t, m_t, q_t)'$, with p_t, m_t, and q_t defined as above, Δ is the first difference operator, $\epsilon_{3t} \sim IN(0, \Sigma_3)$, with Σ_3 a $3x3$ matrix, and $Z_{t-1} = dY_{t-1}$, with d is an $rx3$ matrix of cointegration vectors, r is the rank of the cointegrating space (which is either 0, 1, or 2), and a_3, b_3, c_3, and Σ_3 are parameters estimated using historical U.S. data. In all of our comparisons, data are generated

with one lag of Y_t and cointegrating rank, r, equal to unity, and d either estimated from the historical U.S. data or set equal to $(1, -1, 1)$. We estimate the parameters of (6) using four different sample periods: the entire sample, covering 1959:1-1999:4; the period prior to the well known "monetarist experiment", covering 1959:1-1979:3; the period 1979:4-1989:3; and the period 1989:4-1999:4, which corresponds to the empirical work in the previous section.[16] Given data generated according to (6), 2 prediction models are estimated, including: (i) versions of (6) where r and d are estimated, corresponding to the "estimated VEC" model; (iii) versions of (6) where r=0 is imposed, corresponding to the "VAR in differences" model. Note that we have generated the data according to a VEC model in all cases, so that we should expect the estimated VEC prediction model to perform well, assuming that coefficients are estimated with sufficiently little parameter estimation error, for example. Results from this experiment are gathered in Table 3. The results vary across the different DGP's, but two patterns emerge. First, for small samples ($T = 100$), imprecise estimates of the cointegrating vector parameters and rank generally prevent the VEC model forecasts from dominating the VAR in differences forecasts, and in many cases the VAR in differences model even forecasts more accurately. Second, as the sample size grows, the VEC model forecasts begin to dominate more often, and for some DGP's the VEC model almost always forecasts better for $T = 500$.

Our second Monte Carlo experiment is designed to show that parsimonious time series models will often forecast better than more heavily parameterized, but correctly specified rival models, likely due to parameter estimation error. Specifically, we generate data according to two DGP's. The first DGP is a second-order autoregressive (AR(2)) process:

$$\Delta p_t = a_1 + b_1 \Delta p_{t-1} + c_1 \Delta p_{t-2} + \epsilon_{1t}, \tag{7}$$

where p_t and Δ are defined as above, so that Δp_t is the percentage change in the price level from period t-1 to period t, $\epsilon_{1t} \sim IN(0, \sigma_1^2)$, and a_1, b_1, c_1, and σ_1^2 are parameters estimated using historical U.S. data for the period 1959:1-1999:4. The second DGP is a first-order vector autoregressive (VAR(1)) process:

$$\Delta Y_t = a_2 + b_2 \Delta Y_{t-1} + \epsilon_{2t}, \tag{8}$$

[16] See Section 2 for a description of the historical U.S. data used to estimate the parameters of the DGP's in this section.

where $Y_t = (p_t, m_t, q_t)'$, with p_t, m_t, q_t and Δ defined as above, $\epsilon_{2t} \sim IN(0, \Sigma_2)$, with Σ_2 a $3x3$ matrix, and a_2, b_2, and Σ_2 are parameters estimated using historical U.S. data for the period 1959:1-1999:4. Given these data generating processes, 5000 samples of varying lengths ($T = 164$, which corresponds to the actual sample size used in the empirical work above, and $T = 300, 500$) were generated. For each sample generated from the DGP given in equation (7), both AR(1) and AR(2) models were fitted, and one-step ahead forecasts were compared using the DM test. Although the AR(2) model is correctly specified, it requires the estimation of an additional parameter beyond that of the AR(1) model, so that it is not clear which model will forecast better, out-of-sample. For each sample generated according to DGP (8), one-step ahead forecasts are compared for the "differences random walk" and "VAR in differences" models analyzed in the previous section. Again, even though the VAR in differences model is correctly specified, there is no reason to expect that it will forecast better than the differences random walk model, as the lag length and several other parameters need to be estimated for the VAR in differences model. As a final metric for assessing the importance of parameter estimation error, "true" model forecasts, for which the model parameters are imposed *a priori* to be equal to their true values, rather than estimated, are also included for all of the comparisons.

Table 4 shows the percentage of times the DM test was able to reject the null hypothesis that the AR(1) and AR(2) models forecast equally well, given that the data generating process is an AR(2) model. The figure shows results for two comparisons, where the AR(1) model forecasts are compared to those of an AR(2) model for which the coefficients are estimated (the "AR(2) Model" comparisons), and also where the AR(1) model forecasts are compared to those of an AR(2) model where the true coefficients are imposed rather than estimated (the "True Model") comparisons. We see that for samples of size 164, which is the sample size used in the empirical work, the power is never more than 20%. This means that in practice we would have mistakenly concluded that an AR(1) model is the correct specification 80% of the time. As expected, the power of the test increases with the sample size, but is never more than 51% when the AR(2) model parameters are estimated, even for samples of size 500.

Table 5 has related results for two comparisons. In the first (the "VAR Model" comparisons), differences random walk model forecasts are compared to VAR in differences forecasts, with estimated lag lengths and coefficients. In the second (the "True Model" comparisons), differences random walk model forecasts are compared to forecasts from a first-order VAR model where the

coefficients are imposed to be equal to their true values rather than estimated. The results depend on the specification, but when the VAR parameters are estimated, the random walk model almost always does better. In fact, for all of the configurations, the DM statistics are never greater than 1.96, but are often less than -1.96, with negative DM statistics implying that the random walk model forecasts better. On the other hand, for the "True Model" comparisons, very few of the DM statistics are negative, and the percentage of DM statistics greater than 1.96 is greater than 80% for all but three cases. In nearly all cases, then, a VAR model where the lag length and coefficients are known *a priori* will forecast better than a random walk model, but when the lag length and coefficients need to be estimated, the random walk model forecasts better.

6 Concluding Remarks

In this paper, we show that M2 has marginal predictive content for inflation, and that imposing a cointegrating restriction among prices, money and output that is implied by the quantity theory yields predictions that are superior to those from a variety of other models, including a VAR in differences and a version of the Phillips curve. This result is seen to be particularly robust at longer forecast horizons, and at longer horizons even holds throughout the 1990s, when cointegrating relationships are believed to have broken down. We additionally provide some evidence suggesting that the reported breakdown in the money demand cointegrating relationship may be the result of a failure of M2 as a measure of the money stock, and not the result of an actual breakdown in cointegrating relations, a result that agrees with Carlson, et al (2000).

Our finding that imposing cointegration is useful for forecasting inflation is, however, limited to the case where the cointegration vector is imposed rather than estimated. When the cointegration vector is estimated, the corresponding VEC model always forecasts much worse than a VAR in differences model. This suggests that previous work which has found that VEC models do not forecast better than VAR models may be in some part due to the presence of cointegration vector parameter estimation error. We support this notion by presenting Monte Carlo evidence showing that the effect of parameter estimation error on VEC model forecasts is so substantive that it results in many cases to VAR models forecast-dominating VEC models, even the true model is a VEC, and when the cointegrating rank is known. We additionally present evidence that failure to beat a random walk model is not in itself a useful yardstick for measuring the validity of a theoretical

model, at least if the objective is forecasting. This is done in part by using data simulated to be consistent with the historical U.S. record over the 1959:1-1999:4 period, and showing that a random walk model usually forecasts better than a VAR model for which the lag length and coefficients are estimated, even when the true data generating process is a first-order VAR model. Given this result, and other related arguments, we conclude that use of a random walk as a strawman model in analyses such as ours is not warranted.

Some limitations of the current paper and directions for future research are the following. First, all of the forecasting models in this paper are simple linear models. Nonlinear models may offer forecasting gains (see e.g. Stock and Watson (1999b)). Second, although we have considered only one long-run relationship, it might be of interest to consider some of the many other cointegrating relationships that have been proposed in the literature, both domestic (see e.g. Ahmed and Rogers (2000)) and international (see e.g. Ahmed, et al (1993)). Third, the analysis of Section 4.4 should be extended to include identified structural VAR's and other models which take account of the effects of actual monetary policy actions. Finally, more work needs to be done on the definition of an appropriate monetary aggregate. Attempts to exploit forecasting relationships between monetary aggregates and policy objectives have been subject to criticism, because in practice it takes too long to detect flaws in the monetary aggregate or parameter instability. Although there has been important work done that deals with the problem of instability (see e.g. Carlson, et al (2000)), much remains to be done in this area, particularly in the area of ex ante analysis of instability.

References

[1] Ahmed, S., B.W. Ickes, P. Wang, and B.S. Yoo (1993): "International Business Cycles", *American Economic Review*, (83), 335-359.

[2] Ahmed, S. and J.H. Rogers (2000): "Inflation and the Great Ratios: Long Term Evidence from the U.S.", *Journal of Monetary Economics*, (45), 3-35.

[3] Akaike, H. (1973): "Information Theory and an Extension of the Maximum Likelihood Principle", in B.N. Petrov and F. Csaki, (eds.), *2nd International Symposium on Information Theory,* Budapest: Akademiai Kiado, 267-281.

[4] Akaike, H. (1974): "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control,* (AC-19), 716-723.

[5] Amato, J.D. and T. Laubach (2000): "Forecast-based Monetary Policy", Working Paper 89, Bank for International Settlements.

[6] Barnett, W.A. and A. Serletis, eds. (2000): *The Theory of Monetary Aggregation,* New York: Elsevier.

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

[8] Blinder, A.S. (1997): "Is There A Core of Practical Macroeconomics That We Should All Believe?", *American Economic Review*, (87), 240-243.

[9] Boivin, J. (1999): "Revisiting the Evidence on the Stability of Monetary VAR's ", Working Paper, Columbia University.

[10] Carlson, J.B., D.L. Hoffman, B.D. Keen, and R.H. Rasche (2000): "Results of a Study of the Stability of Cointegrating Relations Comprised of Broad Monetary Aggregates", *Journal of Monetary Economics*, (46), 345-383.

[11] Christoffersen, P. and F.X. Diebold (1998): "Cointegration and Long-Horizon Forecasting", *Journal of Business and Economic Statistics*, (16), 450-458.

[12] Clark, T.E. and M.W. McCracken (2003): "The Predictive Content of the Output Gap for Inflation: Resolving In-Sample and Out-of-Sample Evidence", Working Paper 03-06, Federal Reserve Bank of Kansas City.

[13] Clements, M.P. and D.F. Hendry (1996): "Intercept Corrections and Structural Change", *Journal of Applied Econometrics*, (11), 475-494.

[14] Corradi, V., N.R. Swanson, and C. Olivetti (2001): "Predictive Ability with Cointegrated Variables", *Journal of Econometrics*, (104), 315-358.

[15] Croushore, D. (1998): "Evaluating Inflation Forecasts", Working Paper 98-14, Federal Reserve Bank of Philadelphia.

[16] Culver, S.E. and D.H. Papell (1997): "Is There a Unit Root In the Inflation Rate? Evidence From Sequential Break and Panel Data Models", *Journal of Applied Econometrics*, (12), 435-444.

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

[18] Diewert, W.E. (2000): "Preface to the Theory of Monetary Aggregation", in W.A. Barnett and A. Serletis, (eds.), *The Theory of Monetary Aggregation*, New York: Elsevier.

[19] Elliott, G. and M. Jansson (2003): "Testing for Unit Roots with Stationary Covariates", *Journal of Econometrics*, (115), 75-89.

[20] Engle, R.F. and S.J. Brown (1986): "Model Selection for Forecasting", *Applied Mathematics and Computation*, (20).

[21] Engle, R.F. and C.W.J. Granger (1987): "Co-Integration and Error Correction: Representation, Estimation, and Testing", *Econometrica*, (55), 251-276.

[22] Estrella, A. and F.S. Mishkin (1997): "Is There a Role for Monetary Aggregates in the Conduct of Monetary Policy?", *Journal of Monetary Economics*, (40), 279-304.

[23] 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.

[24] Feldstein, M. and J.H. Stock (1996): "Measuring Money Growth When Financial Markets are Changing", *Journal of Monetary Economics,* (37), 3-27.

[25] Friedman, M. (1959): *A Program for Monetary Stability,* New York: Fordham University Press.

[26] Gali, J. and M. Gertler (1999): "Inflation Dynamics: A Structural Econometric Analysis", *Journal Of Monetary Economics,* (44), 195-222.

[27] Gerlach, S. and L.E.O. Svensson (2003): "Money and Inflation in the Euro Area: A Case for Monetary Indicators?", *Journal of Monetary Economics*, forthcoming.

[28] Greenspan, A. (1993): "Statement to the Congress", *Federal Reserve Bulletin,* (79), 849-55.

[29] Hoffman, D.L. and R.H. Rasche (1996): "Assessing Forecast Performance in a Cointegrated System", *Journal of Applied Econometrics,* (11), 495-517.

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

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

[32] Leeper, E.M. and J.E. Roush (2002): "Putting 'M' Back into Monetary Policy", NBER Working Paper 9552.

[33] Leeper, E.M. and J.E. Roush (2003): "Putting 'M' Back into Monetary Policy", *Journal of Money, Credit, and Banking,* forthcoming.

[34] Lin, J.L. and R.S. Tsay (1996): "Co-Integration Constraint and Forecasting: An Empirical Examination", *Journal of Applied Econometrics*, (11), 519-538.

[35] Mankiw, N.G. (2001): "The Inexorable and Mysterious Tradeoff Between Inflation and Unemployment", *Economic Journal,* (111), C45-C61.

[36] McCracken, M.W. (1999): "Asymptotics for Out of Sample Tests of Causality", Working Paper, Louisiana State University.

[37] McCracken, M.W. (2000): "Robust Out-of-Sample Inference", *Journal of Econometrics*, (99), 195-223.

[38] Ng, S. and P. Perron (1995): "Unit Root Tests in ARMA Models with Data Dependent Methods for the Truncation Lag", *Journal of the American Statistical Association*, (90), 268-281.

[39] Romer, C.D. and D.H. Romer (2000): "Federal Reserve Information and the Behavior of Interest Rates", *American Economic Review*, (90), 429-457.

[40] Said, S.E. and D.A. Dickey (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order", *Biometrika*, (71), 599-607.

[41] Sargent, T.J. (2000): *The Conquest of American Inflation*, Princeton, NJ: Princeton University Press.

[42] Schwarz, G. (1978): "Estimating the Dimension of a Model", *The Annals of Statistics*, (6), 461-464.

[43] Sims, C.A., J.H. Stock, and M.W. Watson (1990): "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica*, (58), 113-144.

[44] Stock, J.H. and M.W. Watson (1999a): "Forecasting Inflation", *Journal of Monetary Economics*, (44), 293-335.

[45] Stock, J.H. and M.W. Watson (1999b): "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series", in R. Engle and H. White (eds.), *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*, Oxford: Oxford University Press 1-44.

[46] Stock, J.H. and M.W. Watson (2003): "Forecasting Output and Inflation: The Role of Asset Prices", *Journal of Economic Literature*, forthcoming.

[47] Swanson, N.R. (1998): "Money and Output Viewed Through a Rolling Window", *Journal of Monetary Economics*, (41), 455-474.

[48] Swanson, N.R. and H. White (1995): "A Model Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks", *Journal of Business and Economic Statistics*, (13), 265-279.

[49] Swanson, N.R. and H. White (1997): "A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks", *Review of Economics and Statistics*, (79), 540-550.

[50] Watson, M.W. (1994): "Vector Autoregressions and Cointegration", in R.F. Engle and D.L. McFadden, (eds.), *Handbook of Econometrics, Volume IV*, New York: North-Holland, 2843-2915.

[51] White, H. (2000): "A Reality Check for Data Snooping", *Econometrica*, (68), 1097-1126.

[52] Woodford, M. (1994): "Non-Standard Indicators for Monetary Policy: Can Their Usefulness be Judged from Forecasting Regressions?", in N.G. Mankiw, (ed.), *Monetary Policy*, Chicago: University of Chicago Press.

Table 1: Summary of Forecasting Models

Benchmark Model

1. *Quantity Theory VEC model:*

$$\Delta p_{t+h} = \beta_0 + \sum_{i=0}^{l-1} \beta_{pi} \Delta p_{t-i} + \sum_{i=0}^{l-1} \beta_{mi} \Delta m_{t-i} + \sum_{i=0}^{l-1} \beta_{qi} \Delta q_{t-i} + \phi z_t + \varepsilon_{t+h}.$$

Alternative Models

2. VAR in differences model:

$$\Delta p_{t+h} = \beta_0 + \sum_{i=0}^{l-1} \beta_{pi} \Delta p_{t-i} + \sum_{i=0}^{l-1} \beta_{mi} \Delta m_{t-i} + \sum_{i=0}^{l-1} \beta_{qi} \Delta q_{t-i} + \varepsilon_{t+h}.$$

3. VAR in levels model:

$$p_{t+h} = \beta_0 + \sum_{i=0}^{l-1} \beta_{pi} p_{t-i} + \sum_{i=0}^{l-1} \beta_{mi} m_{t-i} + \sum_{i=0}^{l-1} \beta_{qi} q_{t-i} + \varepsilon_{t+h}.$$

4. Simple autoregressive model (differences):

$$\Delta p_{t+h} = \beta_0 + \sum_{i=0}^{l-1} \beta_{pi} \Delta p_{t-i} + \varepsilon_{t+h}.$$

5. Simple autoregressive model (levels):

$$p_{t+h} = \beta_0 + \sum_{i=0}^{l-1} \beta_{pi} p_{t-i} + \varepsilon_{t+h}.$$

6. Differences Phillips curve model:

$$\Delta p_{t+h} = \beta_0 + \sum_{i=0}^{l-1} \beta_{pi} \Delta p_{t-i} + \sum_{i=0}^{l-1} \beta_{ui} U_{t-i} + \varepsilon_{t+h}.$$

7. Levels Phillips curve model:

$$p_{t+h} = \beta_0 + \sum_{i=0}^{l-1} \beta_{pi} p_{t-i} + \sum_{i=0}^{l-1} \beta_{ui} U_{t-i} + \varepsilon_{t+h}.$$

8. Differences random walk model:

$$\Delta p_{t+h} = \Delta p_t + \varepsilon_{t+h}.$$

9. Random walk with drift model:

$$p_{t+h} = \beta_0 + p_t + \varepsilon_{t+h}.$$

10. Estimated VEC model:

$$\Delta p_{t+h} = \beta_0 + \sum_{i=0}^{l-1} \beta_{pi} \Delta p_{t-i} + \sum_{i=0}^{l-1} \beta_{mi} \Delta m_{t-i} + \sum_{i=0}^{l-1} \beta_{qi} \Delta q_{t-i} + \phi \widehat{z}_t + \varepsilon_{t+h}.$$

Table 2: Mean absolute forecast errors for the quantity theory VEC model

Horizon	Period 1	Period 2	Period 3
1 Year	0.51	0.63	0.61
2 Years	1.21	1.33	2.34
3 Years	1.60	2.31	6.33
4 Years	2.22	3.10	11.41
5 Years	2.63	3.07	14.12

Notes: "Period 1" forecasts were for the period 1989:4-1992:4, "period 2" forecasts were those from 1993:1 and later which only used M2 data from 1992:4 and earlier, and "period 3" forecasts were those from 1993:1 and later which used M2 data from 1993:1 or later.

Table 3: Monte Carlo Results

DM Statistics for Comparison of Estimated VEC and Differences VAR

I. Cointegration Vector (1, -1, 1) Used In True DGP

		$P=(1/3)T$			$P=(1/2)T$			$P=(2/3)T$		
Sample		A	B	C	A	B	C	A	B	C
1959:1-1999:4	T=100	0.11	0.34	0.68	0.10	0.33	0.68	0.09	0.32	0.68
	T=250	0.01	0.08	0.36	0.01	0.06	0.29	0.01	0.06	0.27
	T=500	0.00	0.01	0.11	0.00	0.00	0.05	0.00	0.00	0.03
1959:1-1979:3	T=100	0.15	0.47	0.80	0.15	0.46	0.80	0.13	0.44	0.79
	T=250	0.12	0.46	0.83	0.10	0.46	0.83	0.09	0.44	0.83
	T=500	0.11	0.49	0.84	0.10	0.49	0.84	0.10	0.49	0.85
1979:4-1989:3	T=100	0.16	0.44	0.77	0.14	0.43	0.78	0.12	0.41	0.77
	T=250	0.12	0.46	0.82	0.11	0.44	0.81	0.10	0.43	0.81
	T=500	0.10	0.46	0.83	0.09	0.45	0.82	0.09	0.44	0.82
1989:4-1999:4	T=100	0.21	0.54	0.83	0.21	0.55	0.85	0.21	0.57	0.87
	T=250	0.16	0.44	0.77	0.14	0.43	0.77	0.12	0.44	0.79
	T=500	0.13	0.37	0.67	0.11	0.35	0.67	0.09	0.34	0.66

II. Estimated Cointegration Vector Used In True DGP

		$P=(1/3)T$			$P=(1/2)T$			$P=(2/3)T$		
Sample		A	B	C	A	B	C	A	B	C
1959:1-1999:4	T=100	0.07	0.26	0.60	0.07	0.25	0.59	0.07	0.24	0.59
	T=250	0.00	0.04	0.26	0.00	0.03	0.18	0.00	0.02	0.14
	T=500	0.00	0.01	0.09	0.00	0.00	0.04	0.00	0.00	0.02
1959:1-1979:3	T=100	0.04	0.16	0.45	0.03	0.13	0.41	0.03	0.12	0.41
	T=250	0.01	0.06	0.21	0.01	0.03	0.12	0.00	0.01	0.07
	T=500	0.06	0.16	0.31	0.03	0.09	0.23	0.01	0.04	0.13
1979:4-1989:3	T=100	0.20	0.52	0.81	0.18	0.51	0.81	0.16	0.49	0.81
	T=250	0.17	0.54	0.84	0.18	0.54	0.85	0.17	0.56	0.86
	T=500	0.15	0.48	0.82	0.14	0.51	0.83	0.13	0.51	0.84
1989:4-1999:4	T=100	0.26	0.68	0.94	0.29	0.73	0.95	0.34	0.77	0.97
	T=250	0.19	0.61	0.91	0.21	0.65	0.93	0.23	0.71	0.96
	T=500	0.18	0.53	0.84	0.17	0.55	0.86	0.18	0.60	0.89

Notes: "A" refers to percentage of cases in 5000 replications where the DM statistic was less than or equal to -1, assuming an MSE loss function. "B" refers to percentage of cases where the DM statistic was less than or equal to 0. "C" refers to the percentage of cases where the DM statistic was less than or equal to 1. A negative DM statistic implies the VAR in differences model performed better.

Table 4: Monte Carlo Results

Power of test of H_0:AR(1) model forecasts as well as AR(2) model

(DGP is an AR(2) model)

Comparison	OOS Period	Sample Size	Power
AR(2) Model	$P = (2/3)T$	$T = 164$	0.20
		$T = 300$	0.34
		$T = 500$	0.51
True Model		$T = 164$	0.38
		$T = 300$	0.47
		$T = 500$	0.60
AR(2) Model	$P = (1/2)T$	$T = 164$	0.18
		$T = 300$	0.29
		$T = 500$	0.42
True Model		$T = 164$	0.31
		$T = 300$	0.38
		$T = 500$	0.50
AR(2) Model	$P = (1/3)T$	$T = 164$	0.20
		$T = 300$	0.28
		$T = 500$	0.38
True Model		$T = 164$	0.29
		$T = 300$	0.34
		$T = 500$	0.43

Notes: The last column of numerical entries shows the power of the Diebold and Mariano (1995) predictive ability test to determine whether an AR(2) model forecasts significantly better, one-step ahead, than an AR(1) model, under MSE loss. The data generating process is an AR(2) model, with parameters estimated using historical US data for the period 1959:1-1999:4. Power of the test indicates the percentage of times in 5000 replications that the predictive ability test rejected equal forecast accuracy of AR(1) and AR(2) models at a significance level of 95 percent, where critical values are taken from McCracken (1999). "AR(2) Model" refers to comparison of the AR(1) model forecasts with AR(2) model forecasts, where the parameters of both models are estimated. "True Model" refers to comparison of the AR(1) model forecasts with AR(2) model forecasts, where the parameters of the AR(2) model are imposed to be equal to their true values.

Table 5: Monte Carlo Results

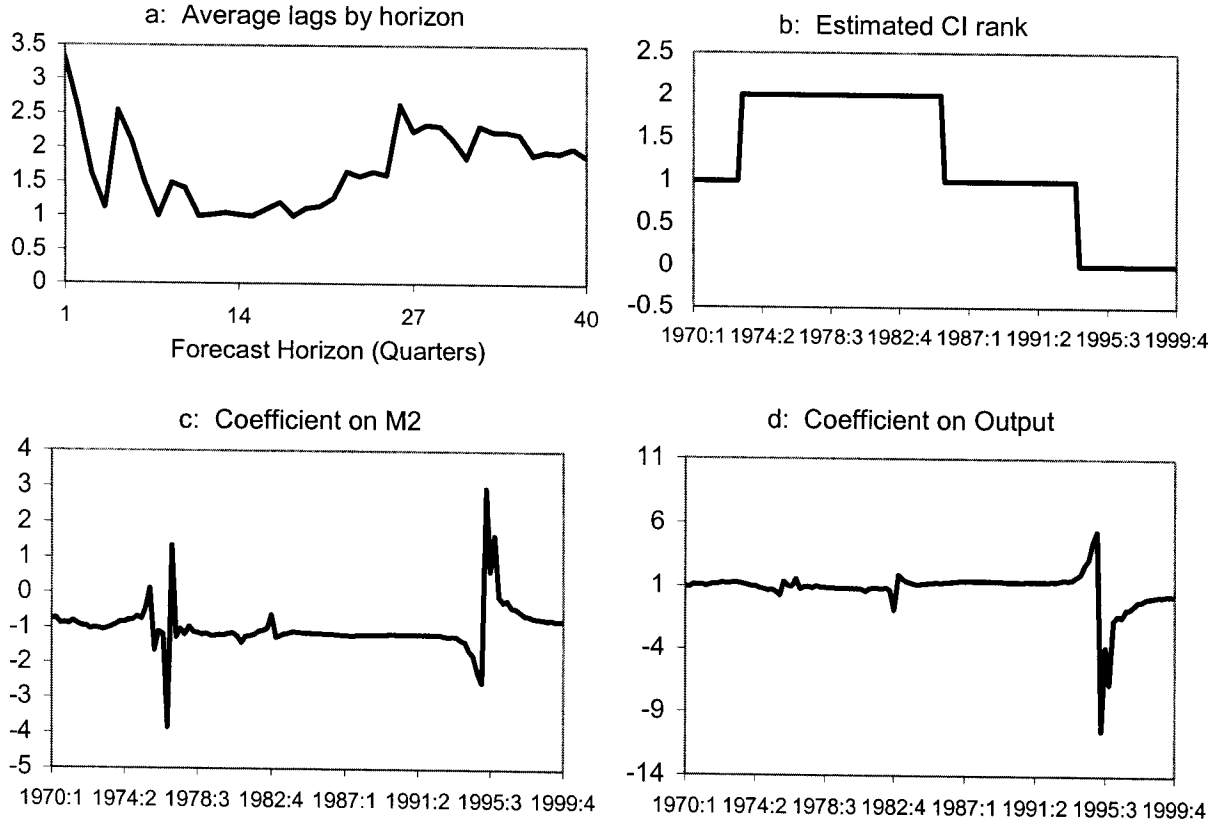
Forecast comparison of VAR and random walk models

(DGP is a VAR model)

Comparison	OOS Period	Sample Size	$DM \leq -1.96$	$DM \leq 0$	$DM \leq 1.96$
VAR Model	$P = (2/3)T$	$T = 164$	0.60	0.98	1.00
		$T = 300$	0.86	1.00	1.00
		$T = 500$	0.97	1.00	1.00
True Model		$T = 164$	0.00	0.00	0.17
		$T = 300$	0.00	0.00	0.03
		$T = 500$	0.00	0.00	0.00
VAR Model	$P = (1/2)T$	$T = 164$	0.47	0.97	1.00
		$T = 300$	0.76	1.00	1.00
		$T = 500$	0.93	1.00	1.00
True Model		$T = 164$	0.00	0.00	0.32
		$T = 300$	0.00	0.00	0.08
		$T = 500$	0.00	0.00	0.01
VAR Model	$P = (1/3)T$	$T = 164$	0.36	0.95	1.00
		$T = 300$	0.57	0.99	1.00
		$T = 500$	0.82	1.00	1.00
True Model		$T = 164$	0.00	0.01	0.47
		$T = 300$	0.00	0.00	0.24
		$T = 500$	0.00	0.00	0.06

Notes: See notes to Table 4.

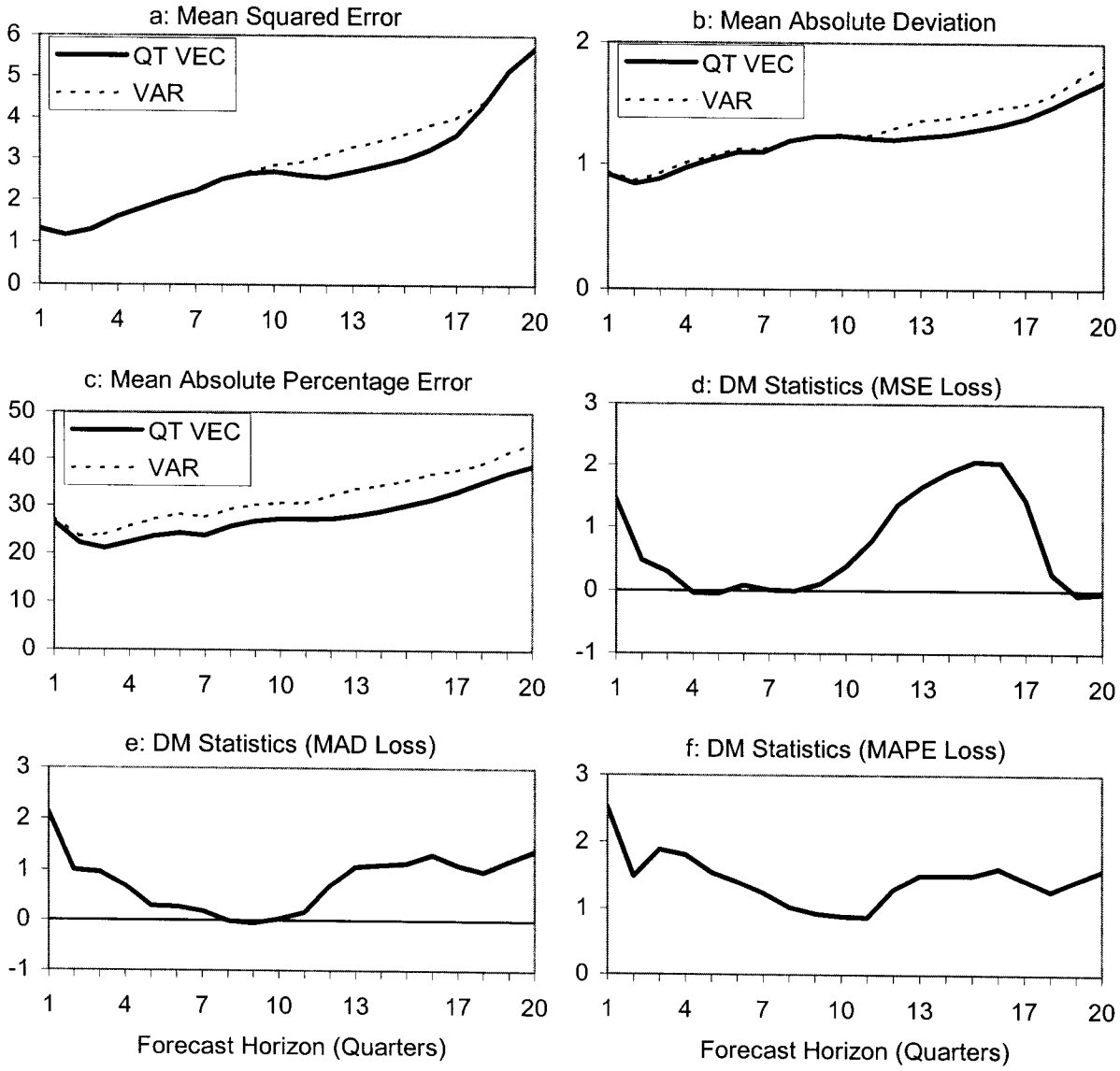
Figure 1: Lag and Cointegration Estimates



Notes: (a) Average number of lags for the autoregressive model of the inflation rate at each forecast horizon, chosen by the SIC. (b) Recursive cointegrating rank estimates. (c) Recursive estimates of the coefficient on M2 in the cointegration vector with largest eigenvalue. (d) Recursive estimates of the coefficient on real GDP in the cointegration vector with largest eigenvalue, using the method of Johansen (1988, 1991).

Figure 2: Inflation Forecasts for 1979:4-1992:4

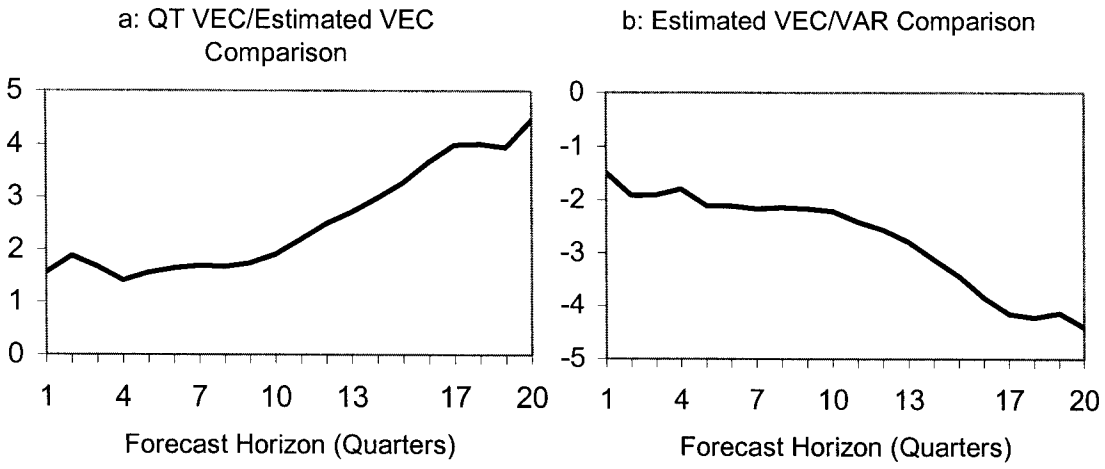
Forecast comparison of quantity theory VEC and VAR in differences models



Notes: (a)-(c) contain plots of mean squared error, mean absolute deviation, and mean absolute percentage error, respectively, for each forecast horizon. (d)-(f) show test statistics for the Diebold and Mariano (1995) predictive accuracy test. Positive statistics imply that the quantity theory VEC model performs better at a particular horizon. The horizontal axis is the forecast horizon, in quarters.

Figure 3: Inflation Forecasts for 1979:4-1992:4

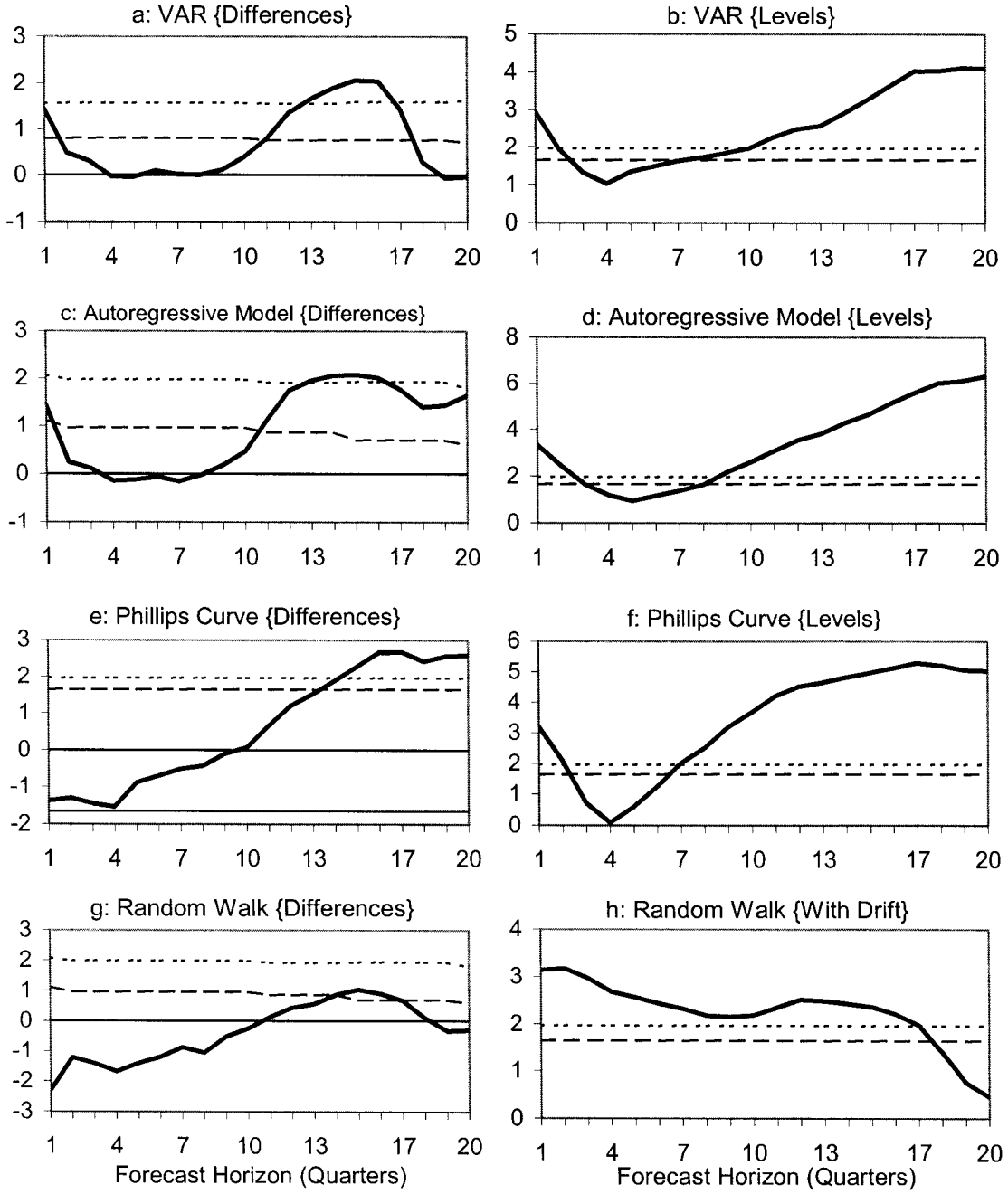
Effects of Estimating the Cointegration Vector and Rank



Notes: (a) Test statistics for the Diebold and Mariano (1995) test of the null hypothesis that the quantity theory VEC and estimated VEC models have equal forecast accuracy at a particular forecast horizon, assuming MSE loss. Positive statistics imply the quantity theory VEC model dominates at a particular horizon. (b) Test statistics for the Diebold and Mariano (1995) test of the null hypothesis that the estimated VEC and the VAR in differences models have equal forecast accuracy at a particular horizon, assuming MSE loss. Positive statistics imply that the quantity theory VEC model dominates at a particular horizon. The horizontal axis in both graphs is the forecast horizon, in quarters.

Figure 4: Inflation Forecasts for 1979:4-1992:4

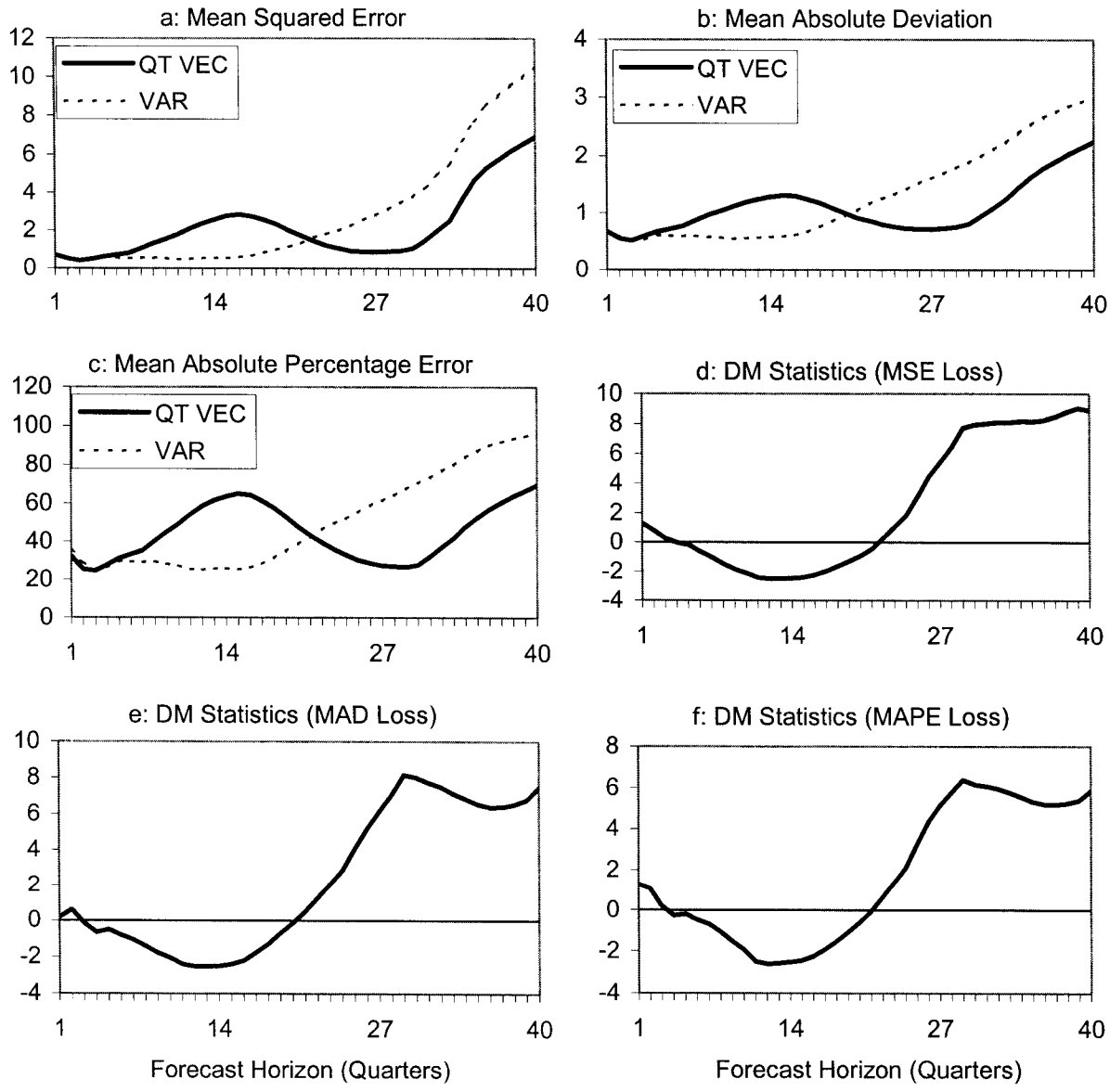
Comparison of Quantity Theory VEC and Alternative Model Forecasts



Notes: Figures show test statistics and 95% and 90% critical values for the Diebold and Mariano (1995) test of the null hypothesis that the quantity theory VEC model and alternative model have equal forecast accuracy at that horizon, assuming an MSE loss function. Positive statistics imply the quantity theory VEC model forecast better at that horizon. The horizontal axis is the forecast horizon, in quarters.

Figure 5: Inflation Forecasts for 1989:4-1999:4

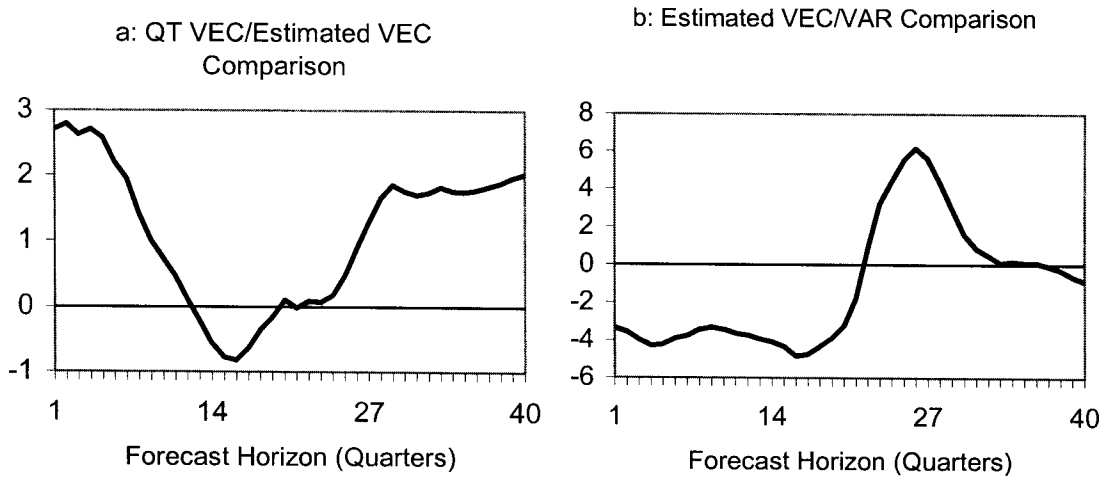
Forecast comparison of quantity theory VEC and VAR in differences models



Notes: See notes to Figure 2.

Figure 6: Inflation Forecasts for 1989:4-1999:4

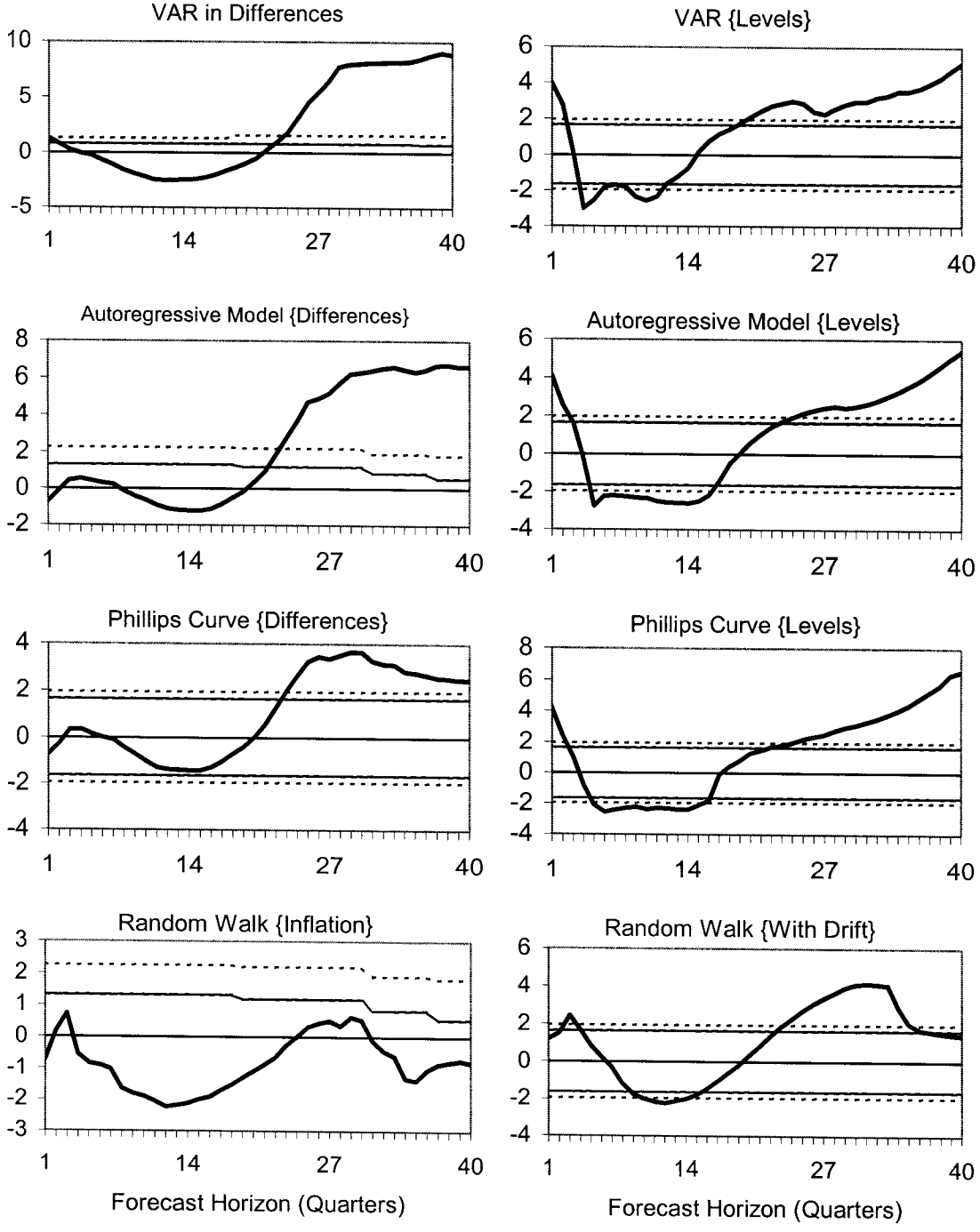
Effects of Estimating the Cointegration Vector and Rank



Notes: See notes to Figure 3.

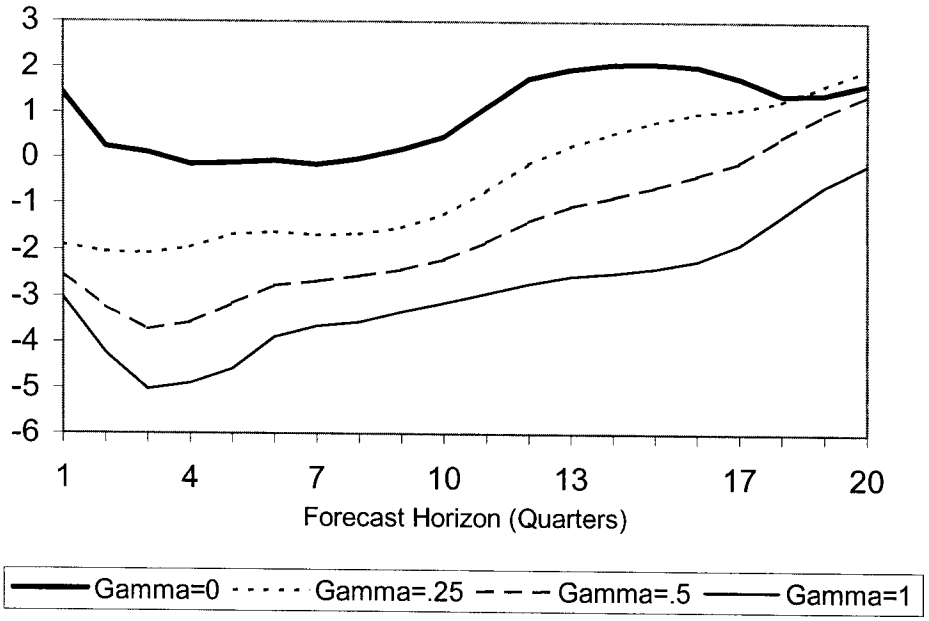
Figure 7: Inflation Forecasts for 1989:4-1999:4

Comparison of Quantity Theory VEC and Alternative Model Forecasts



Notes: See notes to Figure 4.

Figure 8: Inflation Forecasts for 1979:4-1992:4



Notes: Plots are of Diebold and Mariano (1995) test statistics for the null hypothesis that the differences autoregressive model and the model given by equation (5) above have equal predictive accuracy, under MSE loss, for different values of gamma. Negative statistics imply that the autoregressive model forecasts better at a particular horizon. The horizontal axis is the forecast horizon, in quarters.