Econometric Institute Research Report EI-9945/A

A Multivariate STAR Analysis of the Relationship Between Money and Output*

Philip Rothman[†]
Department of Economics, East Carolina University

Dick van Dijk[‡]
Tinbergen Institute and Econometric Institute, Erasmus University Rotterdam

Philip Hans Franses[§]
Econometric Institute, Erasmus University Rotterdam

November 12, 1999

Abstract

Using a standard 4-variable linear vector error correction model (VECM), we first show that the null hypothesis of linearity can be strongly rejected against the alternative of smooth transition autoregressive nonlinearity. An important result from this stage of the analysis is that the quarterly growth rate of money is identified as the transition variable, the variable which governs the smooth switching between regimes. This implies there is a nonlinear causal relationship between money and output. A smooth transition VECM (STVECM) is then used to examine whether money nonlinearly Granger causes output in the sense that lagged values of money enter the model's output equation as regressors. We evaluate this type of nonlinear Granger causality with both in-sample and out-of-sample analysis. For the in-sample analysis we compare alternative models using predictive accuracy tests. These results vary strongly across use of the AIC and SIC. Our use of an out-of-sample forecasting exercise to study money-income Granger causality, both linear and nonlinear, we believe is new to the literature. The forecasting results do not suggest that money is nonlinearly Granger causal for output. In fact, they show that by allowing money to nonlinearly Granger cause output, the forecasting performance of the STVECM is significantly worsened.

---

*We thank participants of the Tinbergen Institute/Macroeconomic Dynamics Conference on "Nonlinear Modeling of Multivariate Macroeconomic Relations" for their helpful comments. Rothman's research was supported by an East Carolina University Faculty Senate Summer Research Grant and van Dijk wishes to acknowledge the hospitality of the Department of Economics at the University of Western Australia.

[†]Department of Economics, Brewster Building, East Carolina University, Greenville, NC 27858 USA, email: rothmanp@mail.ecu.edu (corresponding author)

[‡]Econometric Institute, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, email: djvandijk@few.eur.nl

[§]Econometric Institute, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, email: franses@few.eur.nl

1 Introduction

Macroeconomists have long been interested in whether 'money matters'. Over the past two decades this question has frequently been investigated by testing if various measures of money Granger cause output, that is, by examining whether movements in money have predictive content for fluctuations in output. Using U.S. postwar data, results across several important papers tend to conflict with one another as the sample period is changed and/or different variables are included in the underlying vector autoregressions [VARs]; see, for example, Christiano and Ljungqvist (1988), Stock and Watson (1989), and Friedman and Kuttner (1993).

Swanson (1998) offers a useful brief survey of this literature and points out that the standard practice has been to make use of estimated VARs which exclude long-run cointegrating restrictions and are specified with a priori fixed lag lengths. In his analysis, then, Swanson (1998) accounts for cointegration among the variables considered and identifies the models' lag lengths by using the Schwarz and Akaike information criteria [SIC and AIC]. In addition, he uses rolling fixed-length windows of data, to allow for the possibility that the system is evolving over time. With this approach Swanson (1998) reports robust evidence which strongly rejects the null hypothesis that money does not Granger cause output for the U.S. economy over the 1959:01-1996:03 period.

While Swanson (1998) investigates the Granger causal relationship between money and output using linear VARs and vector error-correction models [VECMs], he also notes that various sorts of nonlinear models have been receiving increasing attention in the econometric time series literature. This is a key point of departure for our paper, in that we propose to investigate the question of money-output causality with a multivariate nonlinear time series model of the smooth transition autoregressive [STAR] type. A good deal of evidence in favor of nonlinear dynamical structure for macroeconomic data has been reported in the literature, though most of these findings are based on univariate analysis; see Granger (1999) for a survey of the recent literature. Such evidence of nonlinearity implies that the standard linear models are misspecified. Our use of STAR models allows us to consider such effects of model misspecification and also enables us to ask whether money is nonlinearly Granger causal for output. With the STAR models we use it is interesting to distinguish between two possible sources of nonlinear Granger causality between money and output. First, money may be identified as the transition variable which governs the smooth switching between regimes in the STAR model. Second, lagged money variables may enter as regressors in the output equation of the STAR model. The problem of nonlinear Granger causality within a STAR model is also studied by Skalin and Teräsvirta (1999), but our treatment differs somewhat from theirs. Their analysis is univariate and they cast the question of nonlinear Granger causality within the framework of additive nonlinearity as developed by Eitrheim and Teräsvirta (1996). Likewise, our parametric approach to nonlinear Granger causality is different from, yet complementary to, the nonparametric procedure of Hiemstra and Jones (1994).

Though we differ from Swanson (1998) in adopting a STAR approach, we do follow his decision to use rolling fixed-length windows and include cointegrating restrictions. While we do not a priori set fixed lag lengths as several earlier studies did, we select the order of the model using an information criterion for practically the full sample period, and then impose this order on the models estimated for each of the rolling windows. Further, our analysis is not quite as general as Swanson's (1998) with respect to the breadth of money measures examined, use of both linear and quadratic deterministic trends, and pre-specification as opposed to estimation of the cointegrating vectors. We make several simplifying assumptions to allow us to focus on the possible role played by nonlinearity of the STAR type in quantifying the relationship between money and output.

The second main contribution of our paper is that we consider out-of-sample forecasting-based tests of Granger causality. Heretofore all studies of the Granger causal relationship between money and output have been based on in- sample fits. Accordingly, we are the first in the literature to focus on comparison of out-of-sample forecasting performance as a test of whether money Granger causes output. While in-sample comparisons of models with and without money indeed are consistent with what has become known as a test of 'Granger causality', it is important to note Granger's argument that the notion of Granger causality is inherently a statement about out-of-sample predictability; see, for example, his interview in Phillips (1997). In our post-sample forecasting exercise, we compare the performance of the various models across a relatively long range of forecast horizons. Use of multiple forecast steps in this exercise is important since Dufour and Renault (1998) show that, for linear projections, non-causality at one forecast horizon does not imply non-causality at all horizons in the presence of auxiliary variables.

The paper closest to ours is Weise (1999), who uses a multivariate STAR model to study an empirical question initially addressed by Cover (1992), that is, whether the effects of money supply shocks on output are asymmetric. There are several technical differences between our paper and Weise's (1999), with respect to linearity testing, estimation of the model parameters, accounting for cointegration, determination of model lag lengths, use of the full-sample versus analysis of a sequence of rolling windows of fixed length, dimension of the baseline linear model, variables used to measure both output and money, and frequency of the data. But the primary contrast between the two papers is that our main focus is on causality testing, while, through use of generalized impulse response analysis of the type introduced by Koop *et al.* (1996), Weise's (1999) chief concern is with monetary shock asymmetry. On this point we note that, at least for linear models, it is known that a zero impulse response is not necessarily equivalent to non-causality; see Dufour and Tessier (1993).

Another paper that is similar in spirit to ours is Thoma (1994), who reports the important result that the p-values of conventional money-income causality tests across increasing windows of data appear to be strongly correlated with the level of real activity. Swanson (1998) makes the interesting observation that Thoma's (1994) use of growing windows of data implies an assumption that the system being modeled is converging to some final state, whereas use of rolling windows of fixed size allows for the possibility that the system evolves over time. While Thoma (1994) does include an out-of-sample forecasting exercise in his study, these results have no bearing on the issue of money-income causality since, in all of the models he uses to generate out-of-sample forecasts, money Granger causes income. Further, it is useful to note that some of the state-dependent models estimated by Thoma (1994) can be interpreted as first-order approximations to STAR models.

The paper proceeds as follows. In Section 2 we discuss linearity testing against STAR alternatives within a multivariate context and present the results of such testing for our set of data. We compare the in-sample fits across the sequence of fixed windows of data for the various models considered in Section 3. Our out-of-sample forecasting results are examined in Section 4 and Section 5 concludes the paper.

2 Multivariate STAR Models and Linearity Testing

Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{kt})'$ be a $(k \times 1)$ vector time series. In our case we have $\mathbf{x}_t = (y_t, m_t, p_t, r_t)'$,with y_t the log of industrial production, m_t the log of nominal M2, p_t the log of the producer price index, and r_t the 90-day Treasury bill rate; the data are for the U.S. economy and cover the sample period 1959:01-1997:06. All data are taken from the Citibase data bank and the specific Citibase series used are: IP for y_t, FM2 for m_t, PW for p_t,and FYGM3 for r_t. A k-dimensional smooth transition vector error-correction model [STVECM] then can be specified as

$$\Delta\mathbf{x}_t = \left(\boldsymbol{\mu}_1 + \boldsymbol{\alpha}_1\mathbf{z}_{t-1} + \sum_{j=1}^{p-1}\boldsymbol{\Phi}_{1,j}\Delta\mathbf{x}_{t-j}\right)(1 - G(s_t;\gamma,c)) + \left(\boldsymbol{\mu}_2 + \boldsymbol{\alpha}_2\mathbf{z}_{t-1} + \sum_{j=1}^{p-1}\boldsymbol{\Phi}_{2,j}\Delta\mathbf{x}_{t-j}\right)G(s_t;\gamma,c) + \boldsymbol{\varepsilon}_t, \quad (1)$$

where $\boldsymbol{\mu}_i$, $i = 1, 2$, are $(k \times 1)$ vectors, $\boldsymbol{\alpha}_i$, $i = 1, 2$, are $(k \times r)$ matrices, $\mathbf{z}_t = \boldsymbol{\beta}'\mathbf{x}_t$ for some $(k \times r)$ matrix $\boldsymbol{\beta}$ denoting the error-correction terms, $\boldsymbol{\Phi}_{i,j}$, $i = 1, 2$, $j = 1, \ldots, p-1$, are $(k \times k)$ matrices, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{kt})$ is a k-dimensional vector white noise process with mean zero and $(k \times k)$ covariance matrix $\boldsymbol{\Sigma}$. The transition function $G(s_t;\gamma,c)$ is assumed to be a continuous function bounded between zero and one, with parameters γ and c determining the smoothness and location of the change in the value of $G(s_t;\gamma,c)$. In this paper we restrict attention to the logistic transition function

$$G(s_t;\gamma,c) = \frac{1}{1 + \exp\{-\gamma(s_t - c)/\hat{\sigma}(s)\}}, \qquad \gamma > 0, \quad (2)$$

where $\hat{\sigma}(s)$ is the sample standard deviation of s_t.

While we compare the in-sample and out-of-sample performance of the various models across a long sequence of rolling windows of fixed length, we carry out linearity testing for the sample period 1960:01-1994:12; we decided not to use the last 30 observations for linearity testing and model estimation, reserving them solely for out-of-sample forecasting. Use of the sample period 1960:01-1994:12 for linearity testing reflects our decision to impose a common transition variable for all windows, that is, to keep the variable identified as s_t the same for all windows, in contrast to letting s_t vary across all windows. Since the STVECM can be considered a type of regime-switching model, our decision assumes that the indicator for the smooth shifting between regimes is stable across all windows. We feel this is reasonable if the regimes are to be given a sensible economic interpretation.

The procedure we follow for specifying STVECMs is a straightforward modification of the specification procedure for univariate STAR models put forward by Teräsvirta (1994). We start by specifying a linear VECM for $\mathbf{x}_t$, that is,

$$\Delta \mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\alpha} \mathbf{z}_{t-1} + \sum_{j=1}^{p-1} \boldsymbol{\Phi}_j \Delta \mathbf{x}_{t-j} + \boldsymbol{\varepsilon}_t, \tag{3}$$

where the order p should be such that the residuals $\hat{\boldsymbol{\varepsilon}}_t$ have zero autocorrelations at all lags. The choice of p is based on applying the SIC in a linear VAR model for $\mathbf{x}_t$ with a deterministic linear trend, which selects $p = 2$ as the appropriate lag order. Through bootstrap simulations we have estimated the size of the linearity test described below conditional on using both the AIC and SIC. These simulations show that the test appears to be conservative in both cases at conventional nominal significance levels, but that use of the AIC leads to far more severe downward size distortion.

We set the cointegrating rank $r = 1$ and pre-specify the cointegrating vector β as $\beta = (1, -1, 1, 0)'$, that is, z_t is the (log) velocity of M2, a time series plot for which appears in Figure 1. Since our measure of output is industrial production, which is not also a measure of income, in contrast to GDP, it is perhaps more accurate to refer to this as 'quasi-velocity'. Our choice of r and the pre-specification scheme to impose for the cointegration vector is in the 'Hendry-style', in that we appeal to economic theory to set these; see, for example, Hendry and Mizon (1993). We abstain from testing the value of r for simplicity. While such testing can be done, it perhaps is not so easy. If the STVECM is the true data generating process, then quasi-velocity is likely to be nonlinear too, so that the standard cointegration tests would need to be adapted. Further, we note that in a study of a similar base line linear system, Swanson (1999) also set the cointegrating rank to $r = 1$.

The next step in the specification procedure consists of testing linearity against the alternative of a STVECM as given in (1). Testing linearity is hampered by the fact that the STVECM contains nuisance parameters that are not identified under the null hypothesis. This can be understood by noting that the null hypothesis of linearity can be expressed in multiple ways, either as $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2$, and $\boldsymbol{\Phi}_{1,j} = \boldsymbol{\Phi}_{2,j}$ for $j = 1, \ldots, p-1$, or as $H_0' : \gamma = 0$ in (1). We follow the approach of Luukkonen *et al.* (1988) and replace the transition function $G(s_t;\gamma,c)$ with a suitable Taylor approximation to circumvent the identification problem. For example, a first-order Taylor expansion of $G(s_t;\gamma,c)$ yields the reparameterized model

$$\Delta\mathbf{x}_t = \boldsymbol{M}_0 + \boldsymbol{A}_0\mathbf{z}_{t-1} + \sum_{j=1}^{p-1}\boldsymbol{B}_{0,j}\Delta\mathbf{x}_{t-j} + \boldsymbol{M}_1 s_t + \boldsymbol{A}_1\mathbf{z}_{t-1}s_t + \sum_{j=1}^{p-1}\boldsymbol{B}_{1,j}\Delta\mathbf{x}_{t-j}s_t + \mathbf{e}_t, \quad (4)$$

where $\mathbf{e}_t$ consists of the original shocks $\boldsymbol{\varepsilon}_t$ and the error arising from the Taylor approximation. Note that in (4) it is assumed that s_t is not one of the variables in $\mathbf{x}_{t-j}$, $j = 1,\ldots,p-1$ or a linear combination thereof. If this is not the case, the term $\boldsymbol{M}_1 s_t$ should not be included. The parameters in $\boldsymbol{M}_i$, $\boldsymbol{A}_i$, and $\boldsymbol{B}_{i,j}$, $i = 0,1$, $j = 1,\ldots,p-1$, are functions of the parameters in the original STVECM (1) such that the original null hypothesis of linearity is equivalent to the null hypothesis that the parameters associated with the auxiliary regressors, s_t, $\mathbf{z}_{t-1}s_t$, and $\Delta\mathbf{x}_{t-j}s_t$, $j = 1,\ldots,p-1$ are equal to zero, that is, $H_0'': \boldsymbol{M}_1 = \boldsymbol{A}_1 = \boldsymbol{B}_{1,j} = 0$, $j = 1,\ldots,p-1$. This hypothesis can be tested by a standard variable addition test. The resulting Lagrange Multiplier [LM] statistic has an asymptotic χ^2 distribution with $k(r+1)+(p-1)k^2$ degrees of freedom under the null hypothesis. The statistic, which will be denoted as S_1, can easily be computed from an auxiliary regression of the residuals from the linear VECM under the null hypothesis on a constant, $\mathbf{z}_{t-1}$, $\Delta\mathbf{x}_{t-j}$, s_t, $\mathbf{z}_{t-1}s_t$, and $\Delta\mathbf{x}_{t-j}s_t$, $j = 1,\ldots,p-1$, whereas an F version of the test can be used as well. We also consider multivariate analogues of the S_2 and S_3 statistics of Luukkonen *et al.* (1988). The S_2 statistic is based on a third-order Taylor approximation of the logistic transition function. This higher-order expansion results in the reparameterized model (4) with s_t^i, $\mathbf{z}_{t-1}s_t^i$, and $\Delta\mathbf{x}_{t-j}s_t^i$, $i = 2,3$, $j = 1,\ldots,p-1$ as additional auxiliary regressors. The corresponding parameters are equal to zero under the null hypothesis of linearity. The S_2 statistic is designed to have power against alternatives in which only the constant in the VECM changes, that is, $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ but $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2$ and $\boldsymbol{\Phi}_{1,j} = \boldsymbol{\Phi}_{2,j}$ for $j = 0,1,\ldots,p-1$ in (1). It turns out that only the parameters corresponding to s_t^2 and s_t^3 are functions of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. To save degrees of freedom, a parsimonious version of the S_2 statistic can be obtained by augmenting (4) with additional auxiliary regressors s_t^2 and s_t^3 (or $\Delta x_{i,t-j}^3$ and $\Delta x_{i,t-j}^4$ for the case in which $s_t = \Delta x_{i,t-j}$). The resultant statistic is the S_3 statistic.

It is well-known that neglected heteroskedasticity may lead to spurious rejection of the null hypothesis of linearity. Specification tests have been developed, by Davidson and

MacKinnon (1985) and Wooldridge (1990, 1991), which can be used in the presence of heteroskedasticity without the need to specify the form of the heteroskedasticity (which often is unknown) explicitly. Their procedures can be readily applied to make robust the linearity tests against STAR-type nonlinearity; see also Granger and Teräsvirta (1993, pp. 69-70). As our time series appear to be quite heteroskedastic, we use robust versions of the LM statistics according to the procedures outlined in Wooldridge (1991) to guard against spurious rejection of linearity. In particular, we use 'Procedure 3.1' of Wooldridge (1991).

To identify an appropriate transition variable s_t, the LM statistics can be computed for several candidates $s_{1t}, \ldots, s_{mt}$, say, and the one for which the p-value of the test statistic is smallest can be selected. Here we consider 84 different candidate transition variables: lagged growth rates in output (Δy_{t-j}), lagged growth rates in M2 (Δm_{t-j}), lagged inflation rates (Δp_{t-j}), lagged changes in the 90-day Treasury bill rate (Δr_{t-j}), lagged changes in the inflation rate ($\Delta \pi_{t-j}$, with $\pi_t = p_t - p_{t-1}$), lagged changes in the federal funds rate (Δff_{t-j}), and lagged changes in the relative price of oil (Δo_{t-j}, with $o_t = p_t^{\mathrm{OIL}}/p_t$ and p_t^{OIL} the crude petroleum producer price index); we use the Citibase series FYFF and PW561 for ff_t and p_t^{OIL}, respectively. We follow Weise (1999) in testing linearity with lagged changes in the inflation rate. Our inclusion of the federal funds rate and the relative price of oil as candidate transition variables is motivated by work which has demonstrated that these variables have played key roles in U.S. business cycle fluctuations in the post-World War II era; see, for example, Bernanke and Blinder (1992) and Hamilton (1983). Since all of these monthly time series contain a substantial amount of short-run fluctuations which do not necessarily represent changes in regime, the monthly changes may not be the most suitable transition variables. Therefore we also consider quarterly changes in the above-mentioned variables, for example $\Delta_3 y_{t-j} = y_{t-j} - y_{t-j-3}$, etc. In all cases we set $j = 1, \ldots, 6$.

The results of our linearity testing appear in Table 1, which reports the top 10 ranked candidate transition variables for the 1960:01-1994:12 sample period. F versions of the test statistics and associated p-values for robust versions of the S_1, S_2, and S_3 linearity tests, for both multivariate (system-wide) tests and single equation tests using just the output equation, are presented. Three features stand out in the table. First, the linearity null hypothesis is strongly rejected using all three tests, for both the multivariate and univariate tests. In most cases the p-values are less than or equal to 0.01. This is so even though our bootstrap simulations have found the test to be quite conservative. Second, in 51 out of 60 cases, a lag of either the growth rate or quarterly change in M2 is selected as one of the top 10 ranked candidate transition variables. In all cases some such lag of money is ranked among the top 3 candidate transition variables. Third, it is interesting to note the absence of lags of both the change in the federal funds rate and the relative price of oil in Table 1. With respect to the federal funds rate, this result may reflect both the well-known reduced reliance of the Federal Reserve on the funds rate as an intermediate target during the 'Volcker experiment' and the possibility, noted by Bernanke and Blinder (1992), that before 1966 this interest rate was a less important monetary instrument, since it was below the discount rate during most of that period. With respect to the relative price of oil, our results perhaps mirror Hooker's (1996) finding that oil prices do not Granger cause many key U.S. macroeconomic variables in the post-1973 era. Likewise, this may also stem from our decision not to use Hamilton's (1996) 'net increase' in oil prices measure.

Since fluctuations in s_t govern the shifting between regimes in the STVECM framework, these test results can be interpreted as implying that money indeed does Granger cause output. Further, they suggest that the nature of this causality is nonlinear. Here our argument is similar to that used in the time-varying transition probability Markov switching literature. Ravn and Sola (1999), for example, use such Markov switching models to study whether various policy variables lead to transitions in aggregate activity.

Given these results it is necessary to select a transition variable to use in estimation of the STVECMs. The S_1 and S_3 tests select $\Delta_3 m_{t-4}$ as the top-ranked candidate transition variable, while the S_2 test selects $\Delta_3 m_{t-1}$. Since S_3 is the parsimonious version of S_2, the absence of $\Delta_3 m_{t-4}$ among the top 10 ranked candidate transition variables using S_2 may possibly reflect this test's use of far more degrees of freedom. To allow for this possibility, we use $\Delta_3 m_{t-4}$ as the system-wide transition variable in estimation of the STVECMs.

3 Model Estimation and In-Sample Model Evaluation

For our rolling window in-sample analysis of Granger causality, we estimate five types of models for each window. In all we use 241 windows of data, each with a fixed length of 15 years. The first window covers the 1960:01-1974:12 sample, and the last is for the 1980:01-1994:12 period. Swanson (1998) uses both 10-year and 15-year windows in his study, but finds that results with the 15-year fixed windows are more reliable. The extra degrees of freedom yielded by these longer windows are particularly important for estimation of the highly parameterized STVECMs we use.

To provide two important benchmarks, for each window we estimate two linear models. ‘Model 1’ is the unrestricted linear VECM and ‘Model 2’ is the linear VECM restricted so that money does not Granger cause output. That is, in the equation for Δy_t in Model 2, neither Δm_{t-1} nor z_{t-1} appear, while these two variables are included as regressors in the output growth rate equation for Model 1. Given our pre-specification of (the log of) M2 quasi-velocity as the error-correction term, Model 1 is estimated by ordinary least squares [OLS] and the parameters of Model 2 are obtained by seemingly unrelated regressions estimation (since the regressors are not identical in all equations).

For each window we estimate three STVECMs, using $\Delta_3 m_{t-4}$ as the transition variable s_t in each case. ‘Model 3’ is the unrestricted STVECM and ‘Model 4’ is the STVECM restricted so that money does not Granger cause output. In Model 4, then, neither Δm_{t-1} nor z_{t-1} appear in the equation for Δy_t. ‘Model 5’ imposes a weaker restriction, that is, the parameters associated with these two variables in $\mathbf{\Phi}_{1,j}$ and $\mathbf{\Phi}_{2,j}$, $j = 1, \ldots, p-1$, and α_1 and α_2 do not vary across regimes in the output growth rate equation, so that in Model 5 money only linearly Granger causes output in a nonlinear multivariate model. In Table 2 we provide a list of all model names and definitions.

Estimation of the STVECM parameters is a relatively straightforward application of nonlinear least squares [NLS], which is equivalent to quasi maximum likelihood based on a normal distribution. Under certain (weak) regularity conditions, the resulting estimates are consistent and asymptotically normal; see White and Domowitz (1984) and Pötscher and Prucha (1997), among others.

To facilitate the nonlinear optimization, we make use of the fact that for fixed values of the parameters in the transition function, γ and c, estimates of $\boldsymbol{\mu}_i$, $\boldsymbol{\alpha}_i$, $\mathbf{\Phi}_{i,j}$, $i = 1, 2$, $j = 1, \ldots, p-1$, can be obtained by OLS. A convenient method to obtain sensible starting values for the nonlinear optimization algorithm then is to perform a two-dimensional grid search over γ and c. Furthermore, the objective function (the log of the determinant of the residual covariance matrix) can be concentrated with respect to $\boldsymbol{\mu}_i$, $\boldsymbol{\alpha}_i$, $\mathbf{\Phi}_{i,j}$, $i = 1, 2$, $j = 1, \ldots, p-1$. This reduces the dimensionality of the NLS estimation problem considerably as the objective function needs to be minimized with respect to the two parameters γ and

c only.

Figure 2 presents a time series plot of the NLS point estimates of γ across all 241 windows for Model 3; results for Models 4 and 5 are very similar. This graph reveals that, with the exception of a few blips, $\hat{\gamma}$ starts out being roughly constant and small and then tends to increase relatively steadily across the first half of the windows of data. Throughout most of the second half of the windows, $\hat{\gamma}$ hits the upper bound allowed by the estimation program ($\hat{\gamma} = 500$). At this value of the smoothness parameter, the switching in the STVECM is of the far more discrete threshold type.

Noting several advantages in doing so, Swanson (1998) uses comparison of AIC and SIC across models with and without money as tests of Granger causality, calling such comparisons 'predictive accuracy' tests. In this approach, these complexity based likelihood measures are calculated for models both with and without money. If the 'best' model contains any money variables, Swanson (1998) argues that this can be interpreted as money Granger causing output.

Table 3 reports the ranking of these five models across all 241 windows as determined by both the AIC and SIC and Table 4 presents pair-wise comparisons. The results differ considerably across use of these two information criteria and we first discuss those obtained using the AIC. With this criterion the STVECMs 5, 3, and 4 are ranked 1, 2, and 3, implying that use of the STAR approach provides a substantial improvement in modeling this vector of time series data. The choice between Models 3 and 5 via the AIC does not seem terribly sharp, though. Slightly more than half the time Model 5 dominates Model 3, suggesting that money only linearly Granger causes output in the STVECM framework, while for a bit less than half of the windows Model 3 is ranked higher than Model 5, implying that money Granger causes output in a nonlinear fashion. It is revealing, however, to contrast the Model 3 to Model 5 comparison with that between Model 4 and Model 5. For nearly every window Model 5 dominates Model 4, which, along with the near tie between Models 3 and 5, demonstrates that allowing for either or both linear and nonlinear money-output Granger causality significantly improves the performance of the STVECM. Finally, Models 1 and 2 are ranked 4 and 5, a result which is consistent with Swanson's (1998) linear analysis.

Tables 3 and 4 also show that the SIC penalizes the much more heavily parameterized STVECMs a good deal more than does the AIC. Using the SIC the unrestricted linear

VECM Model 1 is the top-ranked model, not only beating out the restricted linear VECM Model 2 but also all of the STVECMs. Comparison across the three STVECMs with the SIC does not provide terribly strong evidence in favor of money nonlinearly Granger causing output. For half of the windows, Model 3 is ranked higher than Model 4, and vice versa. Using the SIC Model 5 dominates Model 3 roughly 90% of the time and Model 4 for about 80% of the windows of data, suggesting that money is only linearly Granger causal for output. We note, however, that the SIC is often found to be unreliable in selecting a nonlinear time series model for out-of-sample forecasting; see, for example, Swanson and White (1995).

In Figures 3 and 4 we present time series plots of differences in AIC and SIC, respectively, across models and data windows. These do not suggest very strong variation in the predictive accuracy test results across time.

4 Out-of-Sample Forecasting

For each of the five models discussed above we generate a sequence of 241 out-of-sample forecast profiles as follows. Starting with the estimated model for the 1960:01-1974:12 window, we compute a set of out-of-sample forecasts for forecasts steps $k = 1, \ldots, 30$, so that the out-of-sample forecast period is 1975:01-1977:06. Then we roll the fixed 15-year window by one observation and compute the next set of out-of-sample forecasts using the model estimated for the second window. This is continued until the sample is exhausted, with the last out-of-sample period being 1995:01-1997:06.

It is important to recall that we use the same variable as the transition variable in all windows. This is based on the linearity testing done in Section 3 using the sample period 1960:01-1994:12, so that by way of the procedure we use to identify the transition variable across data windows, there is in this sense some out-of-sample information that is being used. Accordingly, it may be more accurate to say that we carry out a 'quasi' out-of-sample forecasting exercise.

Generating out-of-sample forecasts for Models 1 and 2 is a straightforward exercise, since they are linear. It is well-known that computing multi-step ahead forecasts for nonlinear models is not so simple, since the expected value of a nonlinear function is not equal to the function evaluated at the expected value of its argument; see Brown and

Mariano (1989) and Granger and Teräsvirta (1993, pp. 130-135). We use the bootstrap method to compute multi-step ahead forecasts for the nonlinear Models 3, 4, and 5. The bootstrap approach is favored over the Monte Carlo method since no assumptions are required about the distribution of the underlying stochastic error terms. Also, Lin and Granger (1994) and Clements and Smith (1997) show that it compares favorably well to other methods with respect to computing multi-step ahead forecasts for STAR models.

We present our out-of-sample forecasting results in Tables 5 and 6. In Table 5 we use several different criteria to compute the ranks of the forecasting performance of these models, where these ranks are computed across all 30 forecast horizons. Perhaps the most consistent result in Table 5 is that Model 3, the unrestricted STVECM, is either the worst or tied-for worst forecasting model. Using the Mean Squared Prediction Error (MSPE) and Mean Absolute Error (MAE), Model 3 was ranked last for practically all forecast horizons. Using more robust measures, the Median Squared Prediction Error (MedSPE) and Median Absolute Error (MedAE), Model 3 and Model 4 were both ranked last among the five models. The evidence in Table 5, then, does not support the claim that money nonlinearly Granger causes output.

In light of Swanson's (1998) in-sample linear analysis, which suggests that money does Granger cause output, it is interesting to examine the relative out-of-sample forecasting performance of Model 1 and Model 2. Using the MSPE and MAE criteria, Model 2 dominates Model 1, suggesting that money does not Granger cause output. These results tend to be reversed, however, when the MedSPE and MedAE are used.

Table 6 presents pair-wise out-of-sample forecast comparisons. The first four panels present such comparisons with the same criteria used in Table 5. The fifth panel uses the Harvey *et al.* (1997) modification of the Diebold and Mariano (1995) statistic (MDM) to test, at the 5% level, whether the reduction in MSPE obtained with use of one model over another is statistically significant. The highest values appear in the third column and the lowest values appear in the third row, indicating that Model 3's forecasts tend to be rather strongly dominated by the forecasts of the other models. The sixth panel compares the model forecasts via the Diebold and Mariano (1995)-type test of forecast encompassing introduced by Harvey *et al.* (1998) (MDMFE). Once again, the forecasts of Model 3 are strongly dominated by the other models' forecasts. While the MDMFE results do not suggest that money nonlinearly Granger causes output, they do show that the forecasts generated by the restricted STVECMs Models 4 and 5 tend to either strongly dominate or equal, respectively, the forecasts of the unrestricted VECM, Model 1.

Returning once more to the comparison between Model 1 and Model 2, the last two panels of Table 6 show that the Model 1's improvement over Model 2 is never statistically significant. On the other hand, the MDMFE results show that for nearly 90% of the forecast horizons, Model 2's forecast encompasses the forecast of Model 1. Therefore, the MDM and MDMFE results do not support Swanson's (1998) finding that money Granger causes output.

Finally, it is interesting to examine the relative performance of the different models based on the various forecast criteria across forecast horizons. We have done so but are unable to uncover any consistent patterns of relative forecast performance across these forecast steps. Accordingly, we do not present any results here.

5 Conclusions

Our paper presents several important results. First, we strongly reject the null hypothesis of linearity for a 4-variable VECM of industrial production, money, prices, and interest rates, using quasi-velocity as the pre-specified cointegrating vector. This linearity testing indicates a certain form of nonlinear Granger causality from money to output, since the top ranked transition variable is a lag of the quarterly growth rate of M2.

Second, results from our rolling fixed window in-sample analysis vary quite strongly across use of the predictive accuracy tests given by comparison of the AIC and SIC values for the estimated models. Using the AIC the STVECMs strongly dominate the VECMs. Moreover, the AIC-based model comparisons suggest that allowing money to both linearly and nonlinearly Granger cause output generates considerable improvement in the STAR model's in-sample performance. However, using the SIC the unrestricted VECM dominates all other models across all windows of data.

Third, in our simulated out-of-sample forecasting exercise, the unrestricted STVECM is unambiguously the worst performing model, suggesting that money does not nonlinearly Granger cause output. But it is intriguing to see that the restricted STVECMs perform very favorably relative to the unrestricted VECM. Indeed, the STVECM restricted so that money does not appear as a regressor in the output equation generates forecasts which tend to strongly encompass the forecasts obtained from the unrestricted VECM. This suggests that, all else equal, allowing for nonlinear Granger causality from money to output in the STVECM greatly reduces the model's out-of-sample forecasting performance. It would be interesting to explore in more detail exactly why this is the case.

Our paper demonstrates the importance of considering both in-sample and out-of-sample analysis of causality. This is so not only for the question of whether money nonlinearly Granger causes output within the STVECM framework, but also for comparisons between linear VECMs. Focusing solely on the two linear VECMs studied, our out-of-sample forecasting exercise implies that the causal link from money to output may be a good deal less strong than Swanson's (1998) study suggests.

References

Bernanke, B.S. and A.S. Blinder, 1992, The federal funds rate and the channels of monetary transmission, *American Economic Review* **82**, 901-921.

Brown, B.Y. and R.S. Mariano, 1989, Predictors in dynamic nonlinear models: large sample behavior, *Econometric Theory* **5**, 430-452.

Christiano, L.J. and L. Ljungqvist, 1988, Money does Granger cause output in the bivariate money-output relation, *Journal of Monetary Economics* **22**, 217-235.

Clements, M.P. and J. Smith, 1997, The performance of alternative forecasting methods for SETAR models, *International Journal of Forecasting* **13**, 463-475.

Cover, J., 1992, Asymmetric effects of positive and negative money-supply shocks, *Quarterly Journal of Economics* **107**, 1261-1282.

Davidson, R. and J.G. MacKinnon, 1985, Heteroskedasticity-robust tests in regression directions, *Annales de l'INSEE* **59/60**, 183-218.

Diebold, F.X. and R.S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**, 253-263.

Dufour, J.-M. and E. Renault, 1998, Short run and long run causality in time series: theory, *Econometrica* **66**, 1127-1162.

Dufour, J.-M. and D. Tessier, 1993, On the relationship between impulse response analysis, innovation accounting, and Granger causality, *Economics Letters* **42**, 327-333.

Eitrheim, Ø. and T. Teräsvirta, 1996, Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* **74**, 59-75.

Friedman, B.M and K.N. Kuttner, 1993, Another look at the evidence on money-income causality, *Journal of Econometrics* **57**, 189-203.

Granger, C.W.J. and T. Teräsvirta, 1993, *Modeling Nonlinear Economic Relationships*, Oxford: Oxford University Press.

Granger, C.W.J., 1999, An overview of nonlinear macroeconometric empirical models, unpublished manuscript, Department of Economics, University of California, San Diego.

Hamilton, J.D., 1983, Oil and the macroeconomy since World War II, *Journal of Political Economy* **91**, 228-248.

Hamilton, J.D., 1996, This is what happened to the oil price-macroeconomy relationship, *Journal of Monetary Economics* **38**, 215-220.

Harvey, D.I., S.J. Leybourne, and P. Newbold, 1997, Testing the equality of prediction mean squared errors, *International Journal of Forecasting* **13**, 281-291.

Harvey, D.I., S.J. Leybourne, and P. Newbold, 1998, Tests for forecast encompassing, *Journal of Business and Economic Statistics* **16**, 254-259.

Hendry, D.F. and G.E. Mizon, 1993, Evaluating dynamic econometric models by encompassing the VAR, in P.C.B. Phillips (ed.), *Models, Methods, and Applications: Essays in Honor of A.R. Bergstrom*, Cambridge, MA.: Basil Blackwell, 272-300.

Hiemstra, C. and J.D. Jones, 1994, Testing for linear and nonlinear Granger causality in the stock price-volume relation, *Journal of Finance* **49**, 1639-1664.

Hooker, M.A., 1996, What happened to the oil price-macroeconomy relationship?, *Journal of Monetary Economics* **38**, 195-213.

Koop, G.M., M.H. Pesaran, and S.M. Potter, 1996, Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* **74**, 119-147.

Lin, J-L. and C.W.J. Granger, 1994, Forecasting from nonlinear models in practice, *Journal of Forecasting* **13**, 1-9.

Phillips, P.C.B., 1997, The ET interview: Professor Clive Granger, *Econometric Theory* **13**, 252-303.

Pötscher, B.M. and I.V. Prucha, 1997, *Dynamic Nonlinear Econometric Models - Asymptotic Theory*, Berlin: Springer-Verlag.

Ravn, M.O. and M. Sola, 1999, Business cycle dynamics: predicting transitions with macrovariables, in P. Rothman (ed.), *Nonlinear Time Series Analysis of Economic and Financial Data*, Boston: Kluwer Academic Press, 231-265.

Skalin, J. and T. Teräsvirta, 1999, Another look at Swedish business cycles: 1861-1988, *Journal of Applied Econometrics* **14**, 359-378.

Stock, J.H. and M.W. Watson, 1989, Interpreting the evidence on money-income causality, *Journal of Econometrics* **40**, 161-181.

Swanson, N.R., 1998, Money and output viewed through a rolling window, *Journal of Monetary Economics* **41**, 455-473.

Swanson, N.R., 1999, Finite sample properties of a simple LM test for neglected nonlinearity in error-correcting regression equations, *Statistica Neerlandica* **53**, 76-95.

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**, 265-275.

Teräsvirta, T., 1994, Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208-218.

Thoma, M.A., 1994, Subsample instability and asymmetries in money-income causality, *Journal of Econometrics* **64**, 279-306.

Weise, C.L., 1999, The asymmetric effects of monetary policy: a nonlinear vector autoregression approach, *Journal of Money, Credit, and Banking* **31**, 85-108.

White, H. and I. Domowitz, 1984, Nonlinear regression with dependent observations, *Econometrica* **52**, 143-161.

Wooldridge, J.M., 1990, A unified approach to robust, regression-based specification tests, *Econometric Theory* **6**, 17-43.

Wooldridge, J.M., 1991, On the application of robust, regression-based diagnostics to models of conditional means and conditional variance, *Journal of Econometrics* **47**, 5-46.

Table 1: Linearity Testing: Top Ten Ranked Candidate Transition Variables

	System-wide tests			Output equation tests		
Rank	S_1	S_2	S_3	S_1	S_2	S_3
1	$\Delta_3 m_{t-4}$	$\Delta_3 m_{t-1}$	$\Delta_3 m_{t-4}$	$\Delta_3 m_{t-4}$	$\Delta_3 m_{t-1}$	$\Delta_3 m_{t-4}$
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
2	$\Delta_3 m_{t-3}$	$\Delta_3 m_{t-3}$	Δm_{t-3}	Δm_{t-5}	$\Delta_3 m_{t-5}$	Δm_{t-5}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
3	Δm_{t-3}	$\Delta_3 m_{t-2}$	$\Delta_3 m_{t-3}$	$\Delta_3 m_{t-5}$	$\Delta_3 m_{t-4}$	$\Delta_3 m_{t-5}$
	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
4	$\Delta_3 m_{t-2}$	Δm_{t-3}	$\Delta_3 m_{t-2}$	$\Delta_3 m_{t-3}$	$\Delta_3 m_{t-3}$	$\Delta_3 p_{t-1}$
	(0.00)	(0.01)	(0.00)	(0.00)	(0.03)	(0.01)
5	$\Delta_3 m_{t-5}$	Δm_{t-1}	Δm_{t-5}	$\Delta_3 m_{t-1}$	$\Delta_3 m_{t-6}$	$\Delta_3 m_{t-3}$
	(0.00)	(0.01)	(0.00)	(0.01)	(0.03)	(0.01)
6	$\Delta_3 m_{t-6}$	$\Delta_3 p_{t-4}$	$\Delta_3 p_{t-3}$	Δm_{t-6}	$\Delta_3 y_{t-1}$	$\Delta_3 m_{t-2}$
	(0.00)	(0.01)	(0.00)	(0.01)	(0.03)	(0.02)
7	Δm_{t-4}	$\Delta_3 p_{t-3}$	Δm_{t-4}	$\Delta_3 m_{t-2}$	$\Delta_3 m_{t-2}$	$\Delta_3 m_{t-1}$
	(0.00)	(0.01)	(0.01)	(0.01)	(0.04)	(0.01)
8	$\Delta_3 m_{t-1}$	$\Delta_3 m_{t-6}$	$\Delta_3 m_{t-1}$	Δm_{t-4}	$\Delta_3 y_{t-4}$	Δm_{t-6}
	(0.00)	(0.02)	(0.01)	(0.01)	(0.06)	(0.02)
9	Δm_{t-5}	$\Delta_3 m_{t-5}$	$\Delta_3 m_{t-5}$	$\Delta_3 m_{t-6}$	$\Delta_3 y_{t-3}$	Δm_{t-4}
	(0.01)	(0.03)	(0.01)	(0.04)	(0.06)	(0.02)

The first three columns report the top ten ranked candidate transition variables, ranked by p-value, for robust multivariate versions of the Luukkonen *et al.* (1988) S_1, S_2, and S_3 LM-type tests for the linear VECM. The last three columns report the same for robust versions of these tests computed for the output growth rate equation only. p-values for these tests appear in parentheses. All tests are based on the 1960:01-1994:12 sample period. F versions of the tests were used.

Table 2: Model Definitions

Model name	Definition
Model 1	Unrestricted VECM: money linearly Granger causes output
Model 2	VECM restricted so that money and error-correction term do not appear in the output growth rate equation: money does not linearly Granger cause output
Model 3	Unrestricted STVECM: money nonlinearly Granger causes output
Model 4	STVECM restricted so that money and error-correction term do not appear in output growth rate equation: money does not Granger cause output
Model 5	STVECM restricted so that money and error-correction parameters in output growth rate equation do not vary across regimes: money only linearly Granger causes output

In Models 3, 4, and 5, the 4th lagged quarterly (log) change in M2 is used as the transition variable.

Table 3: Model Selection by AIC and SIC

	Rank j					Average
Model i	1	2	3	4	5	Rank
				AIC		
1	1.2	3.7	23.2	71.8	0.0	3.7
2	0.0	0.0	0.0	9.1	90.9	4.9
3	41.1	53.9	5.0	0.0	0.0	1.6
4	0.0	2.1	70.1	18.7	9.1	3.4
5	57.7	40.3	1.7	0.4	0.0	1.5
				SIC		
1	100.0	0.0	0.0	0.0	0.0	1.0
2	0.0	93.4	6.6	0.0	0.0	2.1
3	0.0	0.0	2.1	54.4	43.6	4.4
4	0.0	0.0	20.8	30.7	48.6	4.3
5	0.0	6.6	70.5	14.9	7.9	3.2

The first five columns report the percentage of windows for which Model i had rank j, $i, j = 1, \ldots, 5$, using the AIC or SIC, and the sixth column reports the average rank of Model i across all windows. The models were estimated for 241 windows of length 15 years, with the initial observation for the windows running from 1960:01 to 1980:01. See Table 2 for model definitions.

Table 4: Pair-wise Model Comparisons by AIC and SIC

Model i	Model j 1	2	3	4	5
			AIC		
1		100.0	3.7	28.2	2.5
2	0.0		0.0	9.1	0.0
3	96.3	100.0		97.9	41.9
4	71.8	90.9	2.1		0.4
5	97.5	100.0	58.1	99.6	
			SIC		
1		100.0	100.0	100.0	100.0
2	0.0		100.0	100.0	93.4
3	0.0	0.0		49.8	8.7
4	0.0	0.0	50.2		22.0
5	0.0	6.6	91.3	78.0	

The table reports the percentage of windows for which Model i was ranked higher than Model j, using the AIC and SIC. The models were estimated for 241 windows of length 15 years, with the initial observation for the windows running from 1960:01 to 1980:01. See Table 2 for model definitions.

Table 5: Out-of-sample Forecasting Ranks

Model i	Rank j 1	2	3	4	5	Average Rank
				MSPE		
1	0.0	3.3	80.0	13.3	3.3	3.2
2	43.3	53.3	0.0	3.3	0.0	1.6
3	0.0	3.3	0.0	0.0	96.7	4.9
4	50.0	40.0	6.7	3.3	0.0	1.6
5	6.7	0.0	13.3	80.0	0.0	3.7
				MedSPE		
1	43.3	33.3	23.3	0.0	0.0	1.8
2	26.7	13.3	16.7	10.0	33.3	3.1
3	3.3	13.3	23.3	33.3	26.7	3.7
4	6.7	6.7	23.3	36.7	26.7	3.7
5	20.0	33.3	13.3	20.0	13.3	2.7
				MAE		
1	0.0	40.0	43.3	16.7	0.0	2.8
2	46.7	33.3	20.0	0.0	0.0	1.7
3	0.0	0.0	0.0	3.3	96.7	5.0
4	50.0	23.3	13.3	10.0	3.3	1.9
5	3.3	3.3	23.3	70.0	0.0	3.6
				MedAE		
1	43.3	33.3	23.3	0.0	0.0	1.8
2	26.7	13.3	16.7	10.0	33.3	3.1
3	3.3	13.3	23.3	33.3	26.7	3.7
4	6.7	6.7	23.3	36.7	26.7	3.7
5	20.0	33.3	13.3	20.0	13.3	2.7

The table summarizes results for out-of-sample forecasting, for the 5 models and 241 estimation windows, across forecasting horizons $k = 1, \ldots, 30$. The four panels in the table show the percentage of forecast horizons Model i had Rank j as determined by the following forecast criteria: Mean Squared Prediction Error (MSPE); Median Squared Prediction Error (MedSPE); Mean Absolute Error (MAE); and Median Absolute Error (MedAE). See Table 2 for model definitions.

Table 6: Pair-wise Model Comparisons by Out-of-Sample Forecast Performance

	Model j				
Model i	1	2	3	4	5
			MSPE		
1		0.0	96.7	6.7	80.0
2	100.0		96.7	46.7	93.3
3	3.3	3.3		3.3	0.0
4	93.3	53.3	96.7		93.3
5	20.0	6.7	100.0	6.7	
			MedSPE		
1		70.0	86.7	90.0	73.3
2	30.0		56.7	53.3	50.0
3	13.3	43.3		56.7	0.0
4	10.0	46.7	43.3		30.0
5	26.7	50.0	80.0	70.0	
			MAE		
1		20.0	100.0	26.7	76.7
2	80.0		100.0	50.0	96.7
3	0.0	0.0		3.3	0.0
4	73.3	50.0	96.7		86.7
5	23.3	3.3	100.0	13.3	
			MedAE		
1		70.0	86.7	90.0	73.3
2	30.0		56.7	53.3	50.0
3	13.3	43.3		56.7	0.0
4	10.0	46.7	43.3		30.0
5	26.7	50.0	80.0	70.0	

continued on next page

continued from previous page

	Model j				
Model i	1	2	3	4	5
			MDM		
1		0.0	43.3	0.0	6.7
2	20.0		66.7	0.0	23.3
3	0.0	0.0		0.0	0.0
4	23.3	0.0	70.0		40.0
5	0.0	0.0	53.3	0.0	
			MDMFE		
1		0.0	86.7	0.0	73.3
2	86.7		76.7	80.0	76.7
3	10.0	3.3		3.3	23.3
4	70.0	73.3	96.3		90.0
5	60.0	0.0	100.0	3.3	

The table presents pair-wise model comparisons based on out-of-sample forecasting results for the 5 models and 241 estimation windows, across forecasting horizons $k = 1, \ldots, 30$. The six panels in the table compare the models' forecasts according to the following forecast criteria: Mean Squared Prediction Error (MSPE); Median Squared Prediction Error (MedSPE); Mean Absolute Error (MAE); and Median Absolute Error (MedAE), whether the modified Diebold-Mariano statistic of Harvey *et al.* (1997) rejects the null hypothesis that Model i's forecast performance as measured by MSPE is not superior to that of Model j at the 5% significance level (MDM), and whether the modified Diebold-Mariano forecasting encompassing statistic of Harvey *et al.* (1998) does not reject the null hypothesis that Model i's forecast encompasses Model j's forecast at the 5% significance level (MDMFE). See Table 2 for model definitions.

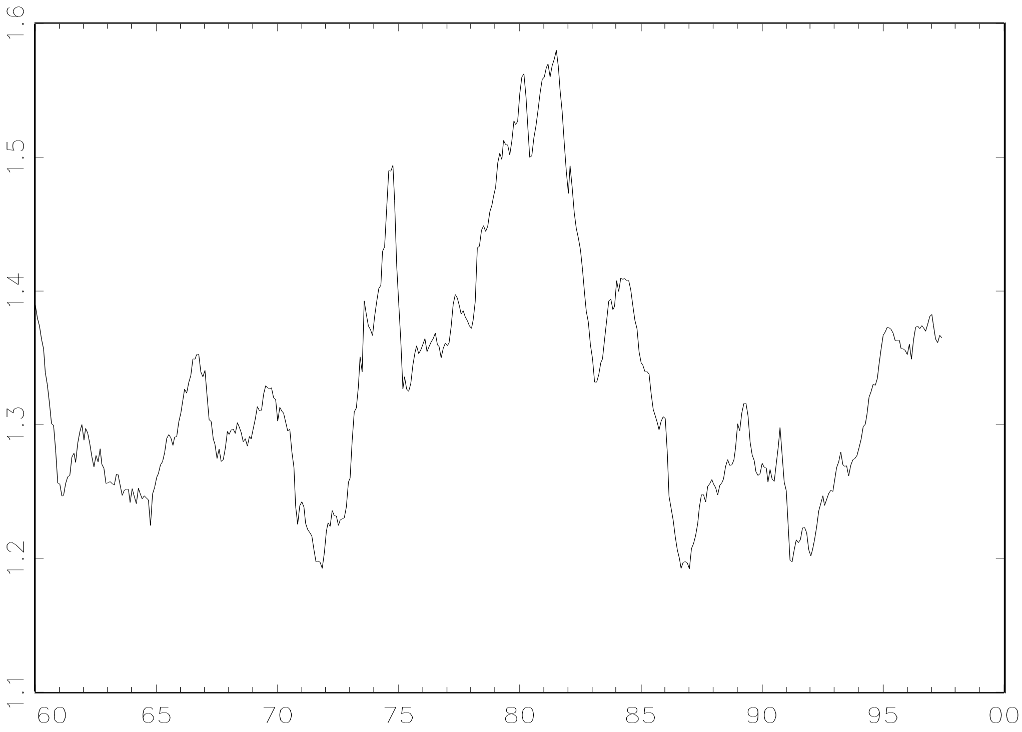


Figure 1: The graph shows the pre-specified error-correction term $z_t = p_t + y_t - m_t$, with p_t the log of the producer price index, y_t the log of industrial production, and m_t the log of nominal M2. We call this 'quasi'-velocity since we use industrial production, which is not a measure of income.

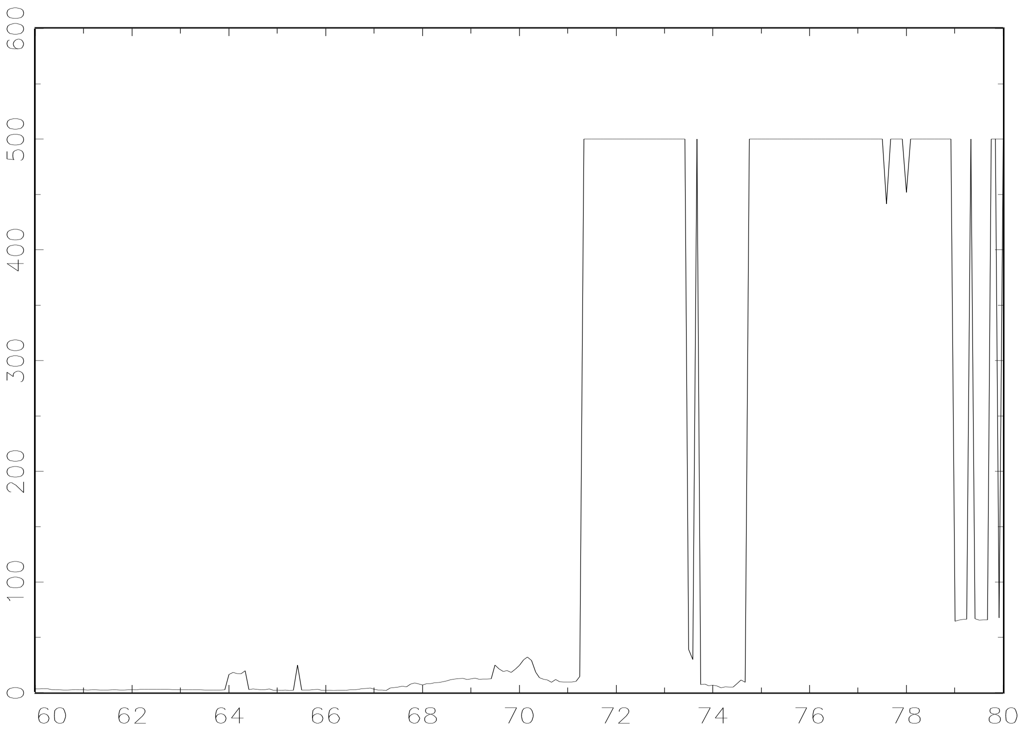
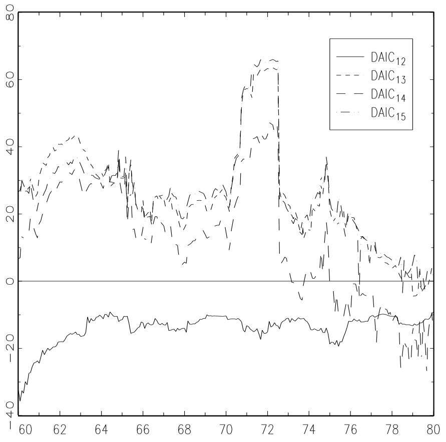
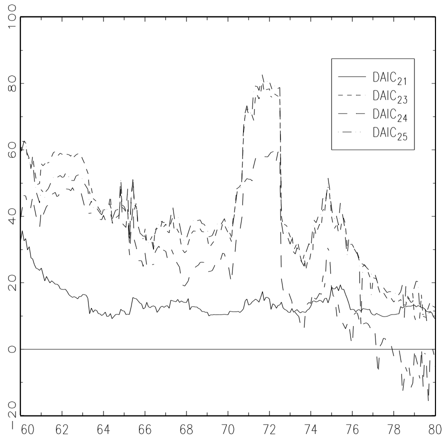


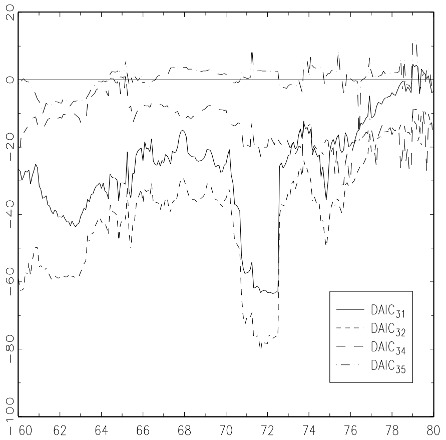
Figure 2: Time series plot of the estimated values of the smoothness parameter γ across the 241 15-year windows for the unrestricted STVECM, that is, Model 3 described in Table 2. Our estimation program imposes an upper bound of 500 on the point estimate of γ. At this value of the smoothness parameter, the switching in the STVECM is indistinguishable from discrete threshold-type switching behavior. The dates listed on the horizontal axis represent the initial observation for each 15-year window.



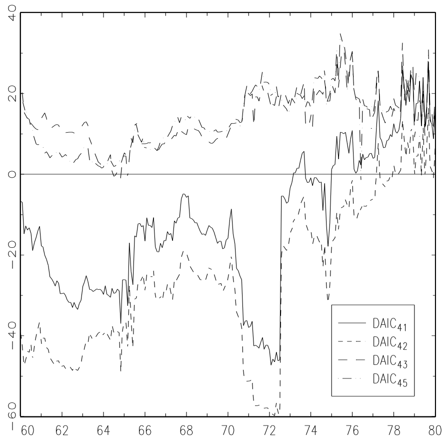
(a) Model 1



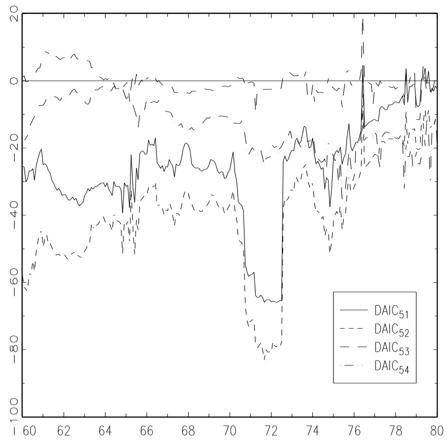
(b) Model 2



(c) Model 3

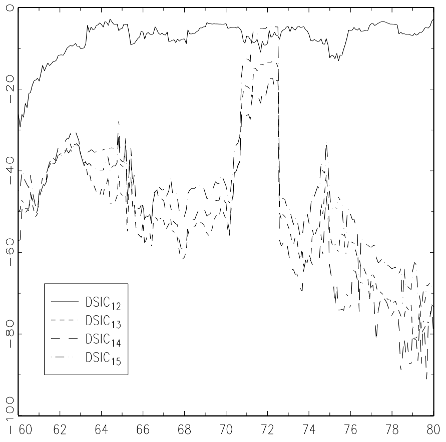


(d) Model 4

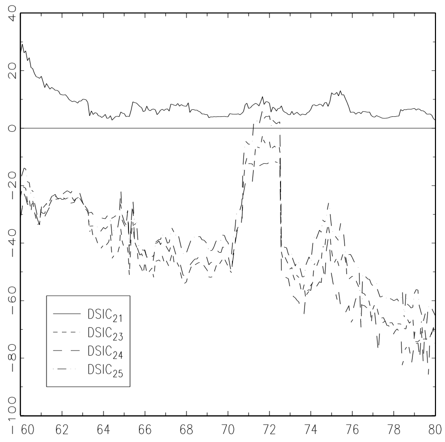


(e) Model 5

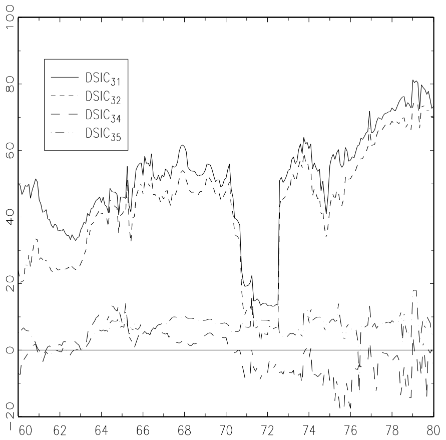
Figure 3: The variable DAIC_{ij} measures the difference between the AIC values for Model i and Model j. A negative value of DAIC_{ij} implies that Model i is preferred over Model j. The dates listed represent the initial observation for each 15-year window.



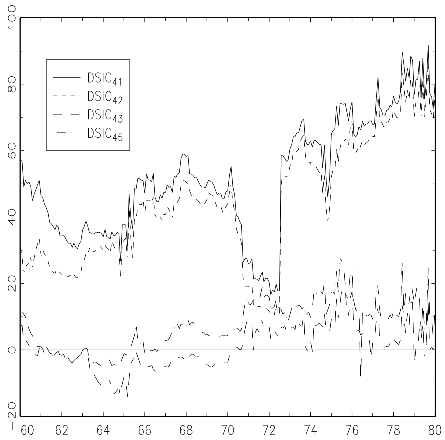
(a) Model 1



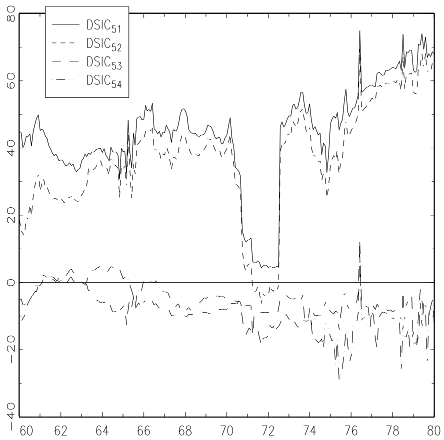
(b) Model 2



(c) Model 3



(d) Model 4



(e) Model 5

Figure 4: The variable DSIC_{ij} measures the difference between the SIC values for Model i and Model j. A negative value of DSIC_{ij} implies that Model i is preferred over Model j. The dates listed represent the initial observation for each 15-year window.