*Macroeconomic Dynamics*, **5**, 2001, 506–532. Printed in the United States of America.
DOI: 10.1017.S1365100500000444

# MULTIVARIATE STAR ANALYSIS OF MONEY–OUTPUT RELATIONSHIP

**PHILIP ROTHMAN**
*East Carolina University*

**DICK VAN DIJK AND PHILIP HANS FRANSES**
*Econometric Institute*
*and*
*Erasmus University Rotterdam*

This paper investigates the potential for nonlinear Granger causality from money to output. Using a standard four-variable linear (subset) vector error-correction model (VECM), we first show that the null hypothesis of linearity can be rejected against the alternative of smooth-transition autoregressive nonlinearity. An interesting result from this stage of the analysis is that the yearly growth rate of money is identified as one of the variables that may govern the switching between regimes. Smooth-transition VECM's (STVECM's) are then used to examine whether there is nonlinear Granger causality in the money–output relationship 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 analyses. For the in-sample analysis, we compare alternative models using the Akaike information criteria, which can be interpreted as a predictive accuracy test. The results show that allowing for both nonlinearity and for money–output causality leads to considerable improvement in model's in-sample performance. By contrast, the out-of-sample forecasting results do not suggest that money is nonlinearly Granger causal for output. They also show that, according to several criteria, the linear VECM's dominate the STVECM's. However, these forecast improvements seldomly are statistically significant at conventional levels.

**Keywords:** Nonlinear Granger Causality, Akaike Information Criterion, Prediction, Multivariate Nonlinear Time-Series Model

We thank participants of the Tinbergen Institute/Macroeconomic Dynamics Conference on Nonlinear Modeling of Multivariate Macroeconomic Relations, the Econometric Society European Winter Meeting, and the 8th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics, seminar participants at the Sveriges Riksbank, and, in particular, the Special Issue editor Timo Teräsvirta and three Anonymous referees 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. Address correspondence to: Philip Hans Franses, Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands; e-mail: franses@few.eur.nl.

 1365-1005/01 $9.50 

## 1. INTRODUCTION

Macroeconomists have long been interested in whether "money matters." Over the past two decades, this question has frequently been investigated by testing whether 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 (VAR's); 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 VAR's that 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 Akaike and Schwarz Bayesian information criteria (AIC and BIC, respectively). 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 that 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 Granger causality in the money–output relationship using linear VAR's and vector error-correction models (VECM's), 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 structures for macroeconomic data has been reported in the literature, though most of these findings are based on univariate analysis; see Granger (2001) 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 that 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 that governs the 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 from theirs. Their analysis is single-equation and they cast the question of nonlinear Granger causality within the framework of additive nonlinearity as developed by Eitrheim and Teräsvirta (1996); see also Péguin-Feissolle and Teräsvirta (1999). 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. In contrast to Swanson (1998) and earlier studies, in our linear analysis, we use subset VECM's, that is, VECM's for which zero restrictions are placed on some of the coefficients. We specify our baseline linear model by sequentially eliminating the variable with the lowest $t$-ratio until all remaining coefficients have $t$-ratios greater than some threshold value. We vary the threshold value, as recommended by Brüggemann and Lütkepohl (2000), such that this procedure is equivalent to sequentially removing the variables, the elimination of which yields the largest improvement in the value of a prespecified model selection criterion. This subset linear VECM is specified using practically the full sample, and the resulting specification is imposed on the linear models estimated for each of the rolling windows; we do the same in specifying the STAR models used. Finally, 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 prespecification 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.

Besides the use of nonlinear models, another contribution of our paper is that we consider out-of-sample forecasting-based tests of Granger causality. Heretofore, most studies of the Granger-causal relationship between money and output have been based on in-sample fits. Accordingly, we are among the first in the literature to focus on comparison of out-of-sample forecasting performance as a test of whether money Granger-causes output; Hess and Porter (1993), Thoma and Gray (1998), Black et al. (2000), Chao et al. (2001), and Amato and Swanson (in press) also use out-of-sample forecasting to study this question, but they consider only linear models. Although in-sample comparisons of models with and without money indeed are consistent with what has become known as a test of Granger causality, 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 postsample forecasting exercise, we compare the performance of the various models across a relatively long range of forecast horizons. Use of multiple forecast horizons in this exercise is important since Dufour and Renault (1998) show that, for linear projections, noncausality at one forecast horizon does not imply noncausality at all horizons in the presence of auxiliary variables.

Several of the time series we use are subject to continual revision, due to, for example, incomplete data collection and seasonal adjustments. In addition, the monetary aggregates are occasionally subjected to redefinitions. For these and related reasons, researchers sometimes make use of real-time data, so as to simulate construction of forecasts made in real time; see, for example, Boschen and Grossman (1982), Mankiw et al. (1984), and Diebold and Rudebusch (1991). Since analysis of such real-time forecasts is not the goal of our paper, we make use of revised data. It is important nonetheless to mention that our analysis certainly could be carried out with real-time data, and such an exercise could possibly document behavior relevant for an improved understanding of the effects of monetary policy. Indeed, Amato and Swanson (in press) and Chao et al. (2001) use real-time data in their out-of-sample forecasting studies of money–output causality.

The paper closest to ours is that of 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. However, the primary contrast between the two papers is that our main focus is on causality testing, whereas Weise's (1999) chief concern is with monetary shock asymmetry, which he examines through use of generalized impulse response analysis of the type introduced by Koop et al. (1996). On this point, we note that, at least for linear models, it is known that a zero impulse response is not necessarily equivalent to noncausality; see Dufour and Tessier (1993).

Another paper that is similar in spirit to ours is that of Thoma (1994), who reports the important result that the $p$-values of conventional money–income causality tests across expanding 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. Although Thoma (1994) does include an out-of-sample forecasting exercise in his study, his 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. Note further that some of the state-dependent models estimated by Thoma (1994) can be interpreted as first-order approximations to STAR models.

Our paper proceeds as follows. In Section 2, we introduce the multivariate STAR model and outline a possible specification procedure for such models. In particular, 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, in Section 5, we conclude the paper.

## 2. MULTIVARIATE STAR MODELS AND LINEARITY TESTING

Let $\boldsymbol{x}_t = (x_{1t}, \ldots, x_{kt})'$ be a $(k \times 1)$ vector time series. In our case, we have $\boldsymbol{x}_t = (y_t, m_t, p_t, i_t)'$, 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 $i_t$ the 90-day Treasury bill rate. The data are for the U.S. economy and cover the sample period 1959:01–1999:12. The industrial production and M2 series are seasonally adjusted, whereas the producer price index and Treasury bill rate are not. Because seasonal adjustment involves application of a two-sided moving-average filter to the time series, this may have consequences for causality testing. Further treatment of this issue is, however, beyond the scope of our paper; see Lee and Siklos (1997) for discussion. We also note that industrial production is a volume index and, as such, is a real variable as opposed to a nominal variable. The data are taken from databases made publicly available by the Federal Reserve Bank of St. Louis and the Board of Governors of the Federal Reserve System.

A $k$-dimensional smooth-transition vector error-correction model (STVECM) can be specified as

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

where $\Delta_j$ denotes the $j$th difference operator, defined as $\Delta_j x_t = x_t - x_{t-j}$ for integers $j \neq 0$ and $\Delta_1 \equiv \Delta$, $\boldsymbol{\mu}_i$, $i = 1, 2$, are $(k \times 1)$ vectors, $\boldsymbol{\alpha}_i$, $i = 1, 2$, are $(k \times r)$ matrices, $\mathbf{z}_t = \beta' \boldsymbol{x}_t$ for some $(k \times r)$ matrix $\beta$ denoting the error-correction terms, $\boldsymbol{\Phi}_{i,j}$, $i = 1, 2$, $j = 1, \ldots, p-1$, are $(k \times k)$ matrices, and $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{kt})$ 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 0 and 1. In this paper, we allow the transition variable $s_t$ to be either a function of lagged components of $\boldsymbol{x}_t$ or a lagged exogenous variable.

The STVECM can be thought of as a regime-switching model that allows for two regimes associated with the extreme values of the transition function, $G(s_t; \gamma, c) = 0$ and $G(s_t; \gamma, c) = 1$, where the transition from one regime to the other is smooth. In this paper, we restrict attention to the logistic transition function

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

where $\hat{\sigma}_s$ is the sample standard deviation of $s_t$. The parameter $c$ in (2) can be interpreted as the threshold or border between the two regimes, in the sense that the logistic function changes monotonically from 0 to 1 as $s_t$ increases, and $G(c; \gamma, c) = 0.5$. The parameter $\gamma$ determines the smoothness of the change in the value of the logistic function and thus the smoothness of the transition from one regime to the other. As $\gamma$ becomes very large, the change of $G(s_t; \gamma, c)$ from 0 to 1 becomes almost instantaneous at $s_t = c$ and, consequently, the logistic function $G(s_t; \gamma, c)$ approaches the indicator function $I[s_t > c]$, defined as $I[A] = 1$ if $A$ is

https://doi.org/10.1017/S1365100501023045 Published online by Cambridge University Press

true and $I[A] = 0$ otherwise. Hence, the STVECM (1) with (2) nests a two-regime threshold vector error-correction model (TVECM) as a special case; see Balke and Fomby (1997) and Tsay (1998) for discussion. Finally, note that when $\gamma \to 0$, the logistic function becomes equal to a constant (equal to 0.5) and when $\gamma = 0$, the STVECM model reduces to a linear VECM.

The procedure we follow for specifying STVECM's 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 $\boldsymbol{x}_t$; that is,

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

where the lag order $p$ should be such that the residuals $\hat{\boldsymbol{\epsilon}}_t$ are approximately white noise and have zero autocorrelations at all lags. The AIC applied in an unrestricted linear VAR model with a deterministic linear trend for our four-dimensional vector $\boldsymbol{x}_t$ selects $p = 7$ as the appropriate lag order. However, this choice of lag length leaves a considerable amount of serial correlation in the residuals. We find that the lag order needs to be increased to $p = 16$ to eliminate this. Because the resulting VECM contains a large number of parameters $[4 + (4 \times r) + (4 \times 4 \times 15)]$, we decided to use a subset VECM by imposing zero restrictions on coefficients in the $\boldsymbol{\Phi}_j$, $j = 1, \ldots, p-1$, matrices in (3).

Following the recommendations of Brüggemann and Lütkepohl (2000), we use a single-equation procedure by treating the individual equations in the VECM separately. We estimate the parameters in the $i$th equation of (3) by ordinary least squares (OLS) and sequentially delete the regressor with the smallest absolute value of the corresponding $t$-ratios, until all $t$-ratios of the remaining coefficients are greater than some threshold value $\tau$ in absolute value. We emphasize that in each iteration only a single regressor is eliminated, after which the reduced model equation is reestimated and new $t$-ratios are computed. We choose the threshold $\tau$ as a function of the iteration $l$ as

$$\tau = \tau_l = \sqrt{[\exp(\lambda_T/T) - 1](T - L + l - 1)}, \tag{4}$$

where $T$ denotes the effective sample size, $L = 1 + r + 4 \times (p-1)$ is the number of parameters in the unrestricted equation, and $\lambda_T$ is a sequence indexed by the sample size. As shown by Brüggemann and Lütkepohl (2000), by setting $\lambda_T$ equal to the penalty term involved in an information criterion of choice, this procedure leads to the same final model as sequentially removing those regressors whose elimination yields the largest improvement in the value of this particular information criterion. In general, a single-equation information criterion for the $l$th iteration is given by $\mathrm{IC} = \log(\hat{\sigma}_l^2) + \lambda_T(L - l + 1)/T$, where $\hat{\sigma}_l^2$ is the estimate of the residual variance in the $l$th iteration and $L - l + 1$ is the number of parameters estimated in the $l$th iteration. The AIC and BIC are obtained by setting $\lambda_T = 2$ and $\lambda_T = \log T$, respectively. Here we use the AIC and hence set $\lambda_T$ equal to 2. It should be remarked that this procedure leads to the model that would be selected

https://doi.org/10.1017/S1365100501023045 Published online by Cambridge University Press

by applying the AIC to each equation individually. It is not guaranteed that this model also will minimize the AIC for the system as a whole. The simulation evidence of Brüggemann and Lütkepohl (2000), however, shows that the difference between the models selected by this single-equation approach and a comparable system approach is small in general. Also note that we only eliminate lagged first differences from the VECM, and always retain the intercept and error-correction terms.

We set the cointegrating rank $r = 2$ and prespecify the two cointegrating vectors as $(1, -1, 1, 0)'$ and $(0, 0, 0, 1)'$; that is, the first row of $\boldsymbol{z}_t$ is the (log) velocity of M2 and the second row of $\boldsymbol{z}_t$ is the 90-day Treasury bill rate. Since our measure of output is industrial production, which is not another measure of income, in contrast to GDP, it is perhaps more accurate to refer to our velocity measure as "quasi-velocity." Our choice of $r$ and the prespecification 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), Garratt et al. (2000), and Söderlind and Vredin (1996). The latter authors show that the Cooley and Hansen (1995) monetary equilibrium business-cycle model implies that both velocity and the nominal interest rate are stationary. To assess whether the the prespecified cointegrating vectors are acceptable, we have estimated a VECM (3) with lag length $p = 16$ and cointegrating rank $r = 2$, but without prespecifying the parameters in $\beta$. The normalized cointegrating vectors are equal to $(1, -1.17, 1.07, 0)'$ and $(0, -1.53, -0.72, 1)'$. The first vector thus is close to (log) velocity of money, but the second is substantially different from the 3-month T-bill interest rate. Nevertheless, because the standard errors of the second and third elements of the second cointegrating vector are very large, a likelihood ratio test does not reject the restrictions we impose at the 10% significance level.

The next step in the specification procedure consists of testing linearity against the alternative of a STVECM. In our application, we test linearity of the subset VECM obtained with the variable selection procedure discussed earlier but, for ease of presentation, we discuss below the linearity tests based on the unrestricted VECM (3). Testing linearity is hampered by the fact that the STVECM as given in (1) 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 $H_0' : \gamma = 0$. 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

$$\begin{aligned}\Delta \boldsymbol{x}_t &= \boldsymbol{M}_0 + \boldsymbol{A}_0 \boldsymbol{z}_{t-1} + \sum_{j=1}^{p-1} \boldsymbol{B}_{0,j} \Delta \boldsymbol{x}_{t-j} + \boldsymbol{M}_1 s_t + \boldsymbol{A}_1 \boldsymbol{z}_{t-1} s_t \\ &\quad + \sum_{j=1}^{p-1} \boldsymbol{B}_{1,j} \Delta \boldsymbol{x}_{t-j} s_t + \boldsymbol{e}_t, \qquad (5)\end{aligned}$$

https://doi.org/10.1017/S1365100501023045 Published online by Cambridge University Press

where $\boldsymbol{e}_t$ consists of the original shocks $\boldsymbol{\epsilon}_t$ and the error arising from the Taylor approximation. Note that, in (5), it is assumed that $s_t$ is not one of the variables in $\boldsymbol{x}_{t-j}$, $j = 1, \ldots, p-1$ or a linear combination thereof. If this is not the case, the term $\boldsymbol{M}_1 s_t$ should be excluded from (5). 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, $s_t$, $\boldsymbol{z}_{t-1}s_t$, and $\Delta \boldsymbol{x}_{t-j}s_t$, $j = 1, \ldots, p-1$ 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 $\chi^2$ distribution with $k(r+1) + (p-1)k^2$ degrees of freedom under the null hypothesis. The statistic, which is denoted as $S_1$, can be computed easily from an auxiliary regression of the residuals from the linear VECM under the null hypothesis on a constant, $\boldsymbol{z}_{t-1}$, $\Delta\boldsymbol{x}_{t-j}$, $s_t$, $\boldsymbol{z}_{t-1}s_t$, and $\Delta\boldsymbol{x}_{t-j}s_t$, $j = 1, \ldots, p-1$, 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 (5) with $s_t^i$, $\boldsymbol{z}_{t-1}s_t^i$, and $\Delta\boldsymbol{x}_{t-j}s_t^i$, $i = 2, 3$, $j = 1, \ldots, p-1$ 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 (5) 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). Because our time series, especially the inflation-rate and interest-rate series, appear to be quite heteroskedastic, we use robust versions of the LM statistics according to the procedures outlined by Wooldridge (1991), particularly "Procedure 3.1," to guard against spurious rejection of linearity.

Lundbergh and Teräsvirta (1998) present some simulation evidence suggesting that single-equation heteroskedasticity-robust linearity tests may not be very powerful, and therefore are of limited use in practice. We have conducted some Monte Carlo experiments to examine the size and power properties of both the single-equation and system linearity tests in our specific context. These simulations show that the tests appear to be conservative at conventional nominal significance levels and confirm the finding of Lundbergh and Teräsvirta (1998) that the tests are not extremely powerful. To conserve space, the results from our simulations are not shown here, but they are available upon request. Note, though, that we also have simulated the size properties of the standard single-equation and system tests in the presence of heteroskedasticity. These simulations suggest that robust tests should be preferred to nonrobust ones in such a case, since the estimated sizes at conventional significance levels were severely distorted upward. Thus, we feel that use of the robust tests is warranted in our case, given that our VECM residuals are highly heteroskedastic. Further, despite the low estimated power of these tests revealed by simulations, we believe that the ranking across a set of prospective transition variables revealed by heteroskedasticity-robust tests provides useful information in the STVECM modeling algorithm.

To identify an appropriate transition variable $s_t$, the LM statistics can be computed for several candidates—for example $s_{1t}, \ldots, s_{mt}$—and the one for which the $p$-value of the test statistic is smallest can be selected. Here, we consider the following different candidate transition variables: lagged yearly growth rates in output ($\Delta_{12} y_{t-d}$), lagged yearly growth rates in M2 ($\Delta_{12} m_{t-d}$), lagged annual inflation rates ($\Delta_{12} p_{t-d}$), lagged yearly changes in the 90-day Treasury bill rate ($\Delta_{12} i_{t-d}$), lagged yearly changes in the annual growth rates in M2 ($\Delta_{12}^2 m_{t-d}$), lagged yearly changes in the annual inflation rate ($\Delta_{12}^2 p_{t-d}$), lagged yearly changes in the federal funds rate ($\Delta_{12}\, ff_{t-d}$), and lagged yearly changes in the relative price of oil ($\Delta_{12} o_{t-d}$, with $o_t = p_t^{\mathrm{OIL}}/p_t$ and $p_t^{\mathrm{OIL}}$ the crude petroleum producer price index). The empirical and theoretical literature upon which we base our focus on these particular candidate transition variables is large. Much research has been done which suggests that these variables are reasonable measures of the "state of the economy" and/or the "state of policy." As such, our use of these variables is also motivated by much of the past decade's macroeconomic research on "state-dependent" dynamics; see, for example, Caplin and Leahy (1991) and Caballero and Hammour (1994). We do wish to mention, though, that we follow Weise (1999) in testing linearity with lagged changes in the inflation rate. Also, regarding our use of the growth rate of money as a candidate transition variable, we feel it is important to note the finding of Ravn and Sola (1999) in a Markov-switching framework that growth rates in the U.S. monetary aggregates tend to have significant effects on the probability of remaining in a recession but negligible effects on the probability of remaining in an expansion.

The reason why we use 12-month differences as transition variables is that we expect the regimes in the money–output relationship to be quite persistent because, for example, they are related to the business cycle or to monetary policy. Since all of the monthly time series contain a substantial amount of short-run fluctuations that do not necessarily represent changes in regime, our view is that monthly changes are not suitable as transition variables. Using 12-month differences effectively eliminates these short-run fluctuations. We test linearity with the above-mentioned variables for delays $d = 1, \ldots, d_{\max}$, where we set the maximum

https://doi.org/10.1017/S1365100501023045 Published online by Cambridge University Press

value of the delay parameter $d_{max}$ equal to 4, leading to a total of 32 different candidate transition variables.

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 1961:08–1997:06; 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 1961:08–1997:06 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 windows. This decision assumes that the indicator for the switching between regimes is stable over time. We feel that this is reasonable if the regimes are to be given a sensible economic interpretation.

The results of our linearity testing appear in Table 1, which reports the top-ranked candidate transition variables according to $F$ versions of the heteroskedasticity-robust versions of the $S_1$, $S_2$, and $S_3$ linearity tests. The table contains only candidate

**TABLE 1.** Linearity testing: Top 10 candidate transition variables[a]

| | Systemwide tests | | | Output-equation tests | | |
|---|---|---|---|---|---|---|
| Rank | $S_1$ | $S_2$ | $S_3$ | $S_1$ | $S_2$ | $S_3$ |
| 1 | $\Delta_{12}o_{t-4}$ (0.029) | | $\Delta_{12}p_{t-3}$ (0.025) | $\Delta_{12}m_{t-4}$ (0.033) | $\Delta_{12}y_{t-4}$ (0.036) | $\Delta_{12}m_{t-4}$ (0.052) |
| 2 | $\Delta_{12}y_{t-1}$ (0.032) | | $\Delta_{12}p_{t-1}$ (0.037) | $\Delta_{12}m_{t-3}$ (0.079) | $\Delta_{12}m_{t-4}$ (0.082) | $\Delta_{12}m_{t-3}$ (0.093) |
| 3 | $\Delta_{12}p_{t-2}$ (0.033) | | $\Delta_{12}p_{t-2}$ (0.038) | $\Delta_{12}y_{t-2}$ (0.079) | $\Delta_{12}m_{t-3}$ (0.100) | |
| 4 | $\Delta_{12}m_{t-2}$ (0.034) | | $\Delta_{12}m_{t-2}$ (0.038) | $\Delta_{12}^2 p_{t-2}$ (0.097) | | |
| 5 | $\Delta_{12}p_{t-3}$ (0.038) | | $\Delta_{12}m_{t-3}$ (0.040) | | | |
| 6 | $\Delta_{12}m_{t-3}$ (0.039) | | $\Delta_{12}^2 p_{t-1}$ (0.040) | | | |
| 7 | $\Delta_{12}^2 p_{t-1}$ (0.040) | | $\Delta_{12}^2 p_{t-2}$ (0.044) | | | |
| 8 | $\Delta_{12}^2 p_{t-2}$ (0.042) | | $\Delta_{12}y_{t-1}$ (0.047) | | | |
| 9 | $\Delta_{12}y_{t-2}$ (0.044) | | $\Delta_{12}o_{t-4}$ (0.051) | | | |
| 10 | $\Delta_{12}o_{t-3}$ (0.045) | | $\Delta_{12}m_{t-1}$ (0.051) | | | |

[a] Columns 2–4 report the top-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 subset VECM. Only candidate transition variables for which the $p$-value is less than 0.10 are shown.

[b] The last three columns report the same for robust versions of these tests computed for the output growth-rate equation only. The $p$-values for these tests appear in parentheses. All tests are based on the 1961:08–1997:06 sample period. $F$ versions of the tests were used.

https://doi.org/10.1017/S1365100501023045 Published online by Cambridge University Press

transition variables for which the $p$-value is less than or equal to 0.10, with a maximum of 10 transition variables. Results for both multivariate (systemwide) tests and single-equation tests (based upon the output equation) are presented. Several features stand out in the table. First, the evidence for nonlinearity varies strongly across the different tests. On the basis of the $S_1$ and $S_3$ tests, the null hypothesis can be rejected at conventional significance levels for a large number of different transition variables, especially for the multivariate tests. By contrast, the $S_2$ test does not reject linearity, except for $s_t = \Delta_{12} y_{t-4}$ in the univariate test. Obviously, this reflects the large number of degrees of freedom used by the $S_2$ test, which leads to a reduction in power. Another possible explanation for the weak rejection of linearity is the finding from Monte Carlo simulations that the heteroskedasticity-robust linearity tests are quite conservative and not very powerful.

Second, different tests favor different candidate transition variables, in the sense that the top-ranked ones differ considerably. Hence it is not possible to select a single variable as the transition variable in the STVECMs based on the "minimum $p$-value rule" mentioned above. By combining the results for the different linearity tests, we decided to focus on the following three transition variables: the yearly growth rate in industrial production lagged 1 month ($\Delta_{12} y_{t-1}$), the yearly growth rate in money lagged 3 months ($\Delta_{12} m_{t-3}$), and the yearly change in the annual inflation rate lagged 1 month ($\Delta_{12}^2 p_{t-1}$). The first of these—$\Delta_{12} y_{t-1}$—is ranked 2 and 8 by the multivariate $S_1$ and $S_3$ tests, whereas for the latter test the difference between the $p$-values of higher-ranked transition variables is quite small. Furthermore, using $\Delta_{12} y_{t-1}$ leads to near rejection of linearity based on the multivariate $S_2$ and the univariate $S_1$ and $S_3$ tests, with $p$-values equal to 0.121, 0.122, and 0.149, respectively. For the multivariate $S_2$ test, this is in fact the smallest $p$-value among the 32 candidate transition variables we consider. Another reason for selecting $\Delta_{12} y_{t-1}$ as a transition variable is that this will effectively allow the model parameters to vary across the business cycle, as noted later in our discussion of the estimated transition function obtained through use of this transition variable; see Figure 2a.

The transition variable $\Delta_{12} m_{t-3}$ is ranked 2, 3, and 2 by the univariate $S_1$, $S_2$, and $S_3$ tests, respectively, and appears among the top 10 candidate transition variables for the multivariate $S_1$ and $S_3$ test statistics as well. The latter is the reason for preferring the 3-month lagged yearly growth rate of money over the 4-month lag, which is ranked higher by the univariate tests but does not appear in the top 10 multivariate test outcomes. Use of a lagged growth rate of money as a transition variable allows for smooth switching between regimes of slow and fast growth in M2.

Regarding the third transition variable that we use, $\Delta_{12}^2 p_{t-1}$ appears among the top 10 variables for the multivariate $S_1$ and $S_3$ tests, and leads to near rejections for the univariate $S_1$ and $S_3$ tests, with $p$-values equal to 0.111 and 0.132, respectively. Even though lags of the annual inflation rate itself lead to relatively strong rejections of linearity for the systemwide $S_1$ and $S_2$ tests, we prefer to use $\Delta_{12}^2 p_{t-1}$ because lags of $\Delta_{12} p_t$ do not lead to even borderline rejections in the single-equation tests.

We note that use of the change in the inflation rate as a transition variable generates switching between periods of decreasing and accelerating inflation.

Given that the lagged money growth rate appears to be a potentially useful transition variable and 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 and 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.

Last, it is interesting to note the relative rarity with which use of lags in the relative price of oil leads to rejections of linearity in Table 1, even though $\Delta_{12}o_{t-4}$ is the top-ranked transition variable according to the systemwide $S_1$ test. Likewise, lags of the federal funds rate never lead to rejection of the linearity null hypothesis at the 10% significance level. 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) measure of "net increases in oil prices." 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.

## 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 192 windows of data, each with a fixed length of 20 years. The first window covers the 1961:08–1981:07 sample, and the last is for the 1977:07–1997:06 period. Note that our window size is larger than Swanson's (1998), who considers 10-year and 15-year windows. The extra degrees of freedom yielded by the longer window are particularly important for estimation of the highly parameterized STVECM's we use.

To provide two important benchmarks, for each window we estimate two linear models. Model 1 is an unrestricted linear subset VECM and Model 2 is a linear subset VECM restricted so that money does not Granger-cause output. That is, in the equation for $\Delta y_t$ in Model 2, neither lags of $\Delta m_t$ nor lags of $\mathbf{z}_{1,t-1}$ (the first row of $\mathbf{z}_t$, i.e., the log velocity of money) appear, whereas these two variables are allowed to enter as regressors in the output growth-rate equation for Model 1. In Model 2, the regressors in the money, inflation-, and interest-rate equations are taken to be the same as in Model 1. The regressors in the output equation are determined by applying the subset strategy discussed earlier, starting with 15 lagged first differences of output growth, the inflation rate and the interest rate as

https://doi.org/10.1017/S1365100501023045 Published online by Cambridge University Press

regressors, in addition to an intercept and the lagged level of the 90-day Treasury bill rate. Given our prespecification of (the log of) M2 quasi-velocity and the 90-day Treasury bill rate as the error-correction terms, both Models 1 and 2 are estimated by seemingly unrelated regressions estimation. We use the same subset VECM's (and STVECM's discussed later) across all windows, where the variables to be included in the different equations are determined by the selection procedure applied to the full sample 1961:08–1997:06. Obviously, this is not the optimal procedure because the significance of variables may change over time. However, applying the subset selection procedure for each window separately would be too time-consuming for the nonlinear models.

For each window we estimate three classes of STVECM's. Model 3 is the unrestricted STVECM and Model 4 is a STVECM restricted so that money does not Granger-cause output. In Model 4, then, neither lags of $\Delta m_t$ nor lags of $\boldsymbol{z}_{1,t-1}$ appear in the equation for $\Delta y_t$. Model 5 imposes a weaker restriction; that is, the parameters associated with these two variables in $\boldsymbol{\Phi}_{1,j}$ and $\boldsymbol{\Phi}_{2,j}$, $j = 1, \ldots, p-1$, and $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are identical across regimes in the output growth-rate equation, so that, in Model 5, money only linearly Granger-causes output in a nonlinear multivariate model. We estimate these STVECM's with $\Delta_{12} y_{t-1}$, $\Delta_{12} m_{t-3}$, and $\Delta_{12}^2 p_{t-1}$ as the transition variable $s_t$. The model numbers are indexed by $y$, $m$, and $p$ to reflect the transition variable that is used.

To reduce the dimensionality of the STVECM's, we apply a subset strategy similar to the one used to specify the subset VECM's. Note that the STVECM's contain parameter restrictions across equations because the transition function $G(s_t; \gamma, c)$ is identical in all equations in the system. Hence, we use a system approach to select the relevant regressors in the subset STVECM. For example, for the case of Model 3, we start with the full STVECM given in (1) with $p = 15$ and sequentially delete the regressor with the smallest absolute value of the corresponding $t$-ratios, until all $t$-ratios of the remaining coefficients are greater (in absolute value) than the threshold value $\tau_l$, which is set according to (4) with $T$ replaced by $4T$, $L$ set equal to the number of the parameters in the unrestricted STVECM, and $\lambda_T = 2$. In contrast to the comparable single-equation approach, in this case there is no direct relation between this approach and the procedure that uses sequential elimination of those regressors that lead to the largest improvement in the AIC for the complete system. In Table 2, we provide a list of all model names, definitions, and number of parameters in the models. Note that for each window we estimate a total of 11 different models—2 linear and 9 nonlinear.

Estimation of the parameters in the subset STVECM is a relatively straightforward application of nonlinear generalized least squares (NGLS), 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, $\gamma$ and $c$, estimates of $\boldsymbol{\mu}_i$, $\boldsymbol{\alpha}_i$, $\boldsymbol{\Phi}_{i,j}$, $i = 1, 2$, $j = 1, \ldots, p-1$, can be obtained by OLS. A convenient method to obtain

https://doi.org/10.1017/S1365100501023045 Published online by Cambridge University Press

**TABLE 2.** Model definitions

| | | Number of parameters | | | |
|---|---|---|---|---|---|
| | | | STVECM's[a] ($s_t =$) | | |
| Model | Definition | VECM's | $\Delta_{12}y_{t-1}$ | $\Delta_{12}m_{t-3}$ | $\Delta_{12}^2 p_{t-1}$ |
| 1 | Unrestricted VECM: Money linearly Granger-causes output. | 92 | | | |
| 2 | VECM restricted so that lagged growth rates and velocity of money do not appear in the output growth-rate equation: Money does not linearly Granger-cause output. | 91 | | | |
| 3 | Unrestricted STVECM: Money nonlinearly Granger-causes output. | | 263 | 234 | 253 |
| 4 | STVECM restricted so that lagged growth rates and velocity of money do not appear in output growth-rate equation: Money does not Granger cause output. | | 242 | 206 | 249 |
| 5 | STVECM restricted so that lagged growth rates and velocity of money in output growth-rate equation do not vary across regimes: Money only linearly Granger-causes output | | 264 | 220 | 238 |

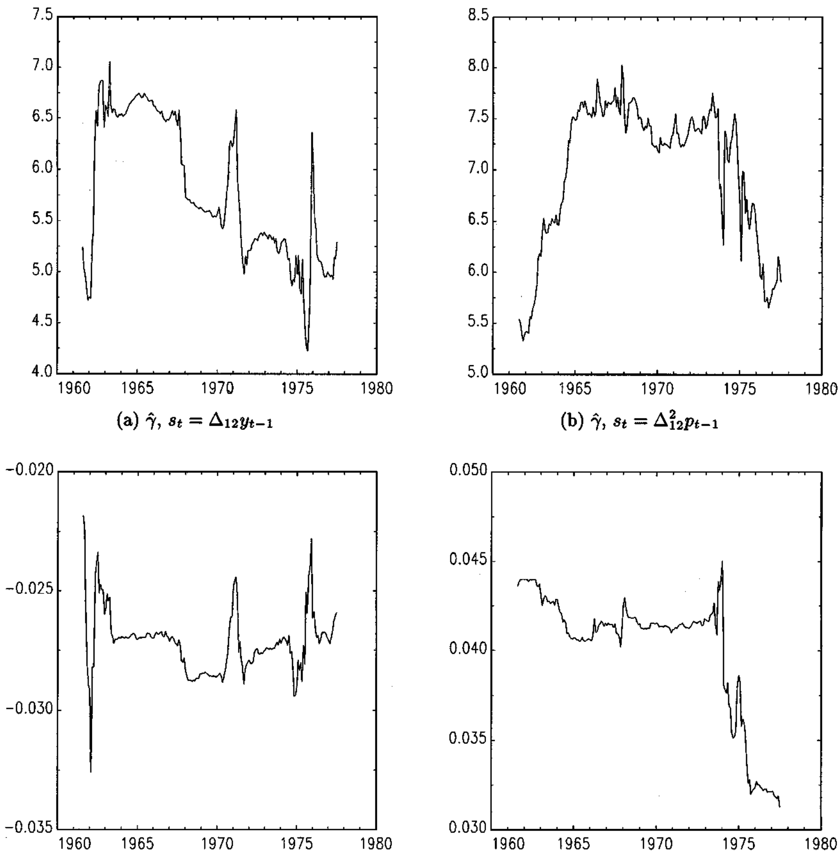
[a]The model numbers of the STVECM's are indexed by $y$, $m$, and $p$ to reflect the transition variable that is used.

sensible starting values for the nonlinear optimization algorithm then is to perform a two-dimensional grid search over $\gamma$ and $c$. Furthermore, the objective function (the log of the determinant of the residual covariance matrix) can be concentrated with respect to $\boldsymbol{\mu}_i, \boldsymbol{\alpha}_i, \boldsymbol{\Phi}_{i,j}, i = 1, 2, j = 1, \ldots, p-1$. This reduces the dimensionality of the NGLS estimation problem considerably because the objective function needs to be minimized with respect to the two parameters $\gamma$ and $c$ only.

Figure 1 presents time-series plots of the NGLS point estimates of $\gamma$ and $c$ across all 192 windows for Models 3$y$ and 3$p$; results for Models 4$y$, 4$p$, 5$y$, and 5$p$ are very similar. These graphs reveal that in both models the estimates of $\gamma$ and $c$ are reasonably constant. The estimates of $\gamma$ are such that the transition between the two regimes is reasonably smooth. In the STVECM's that use $\Delta_{12}m_{t-3}$ as the transition variable, $\hat{\gamma}$ consistently hits the upper bound allowed by the estimation program ($\hat{\gamma} = 500$) across all windows; therefore, these are not shown here. At this value of the smoothness parameter, the switching in the STVECM is of the far more discrete threshold type. The estimate of the location parameter $c$ in Model 3$m$ fluctuates around 0.085, such that the upper regime associated with $G(s_t; \gamma, c)$ only becomes active for large values of the growth rate of money.

Figure 2 shows the value of the transition function $G(s_t; \gamma, c)$ over time in Models 3$y$, 3$m$, and 3$p$ estimated for the full sample 1961:08–1997:06. Note that the different models imply a different nonlinear structure, in the sense that

https://doi.org/10.1017/S1365100501023045 Published online by Cambridge University Press

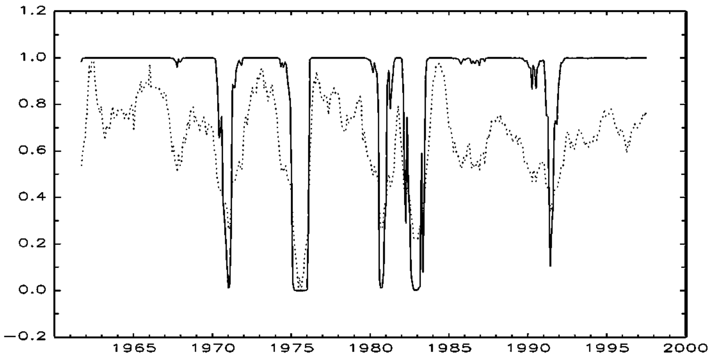



(a) $\hat{\gamma}$, $s_t = \Delta_{12} y_{t-1}$ (b) $\hat{\gamma}$, $s_t = \Delta_{12}^2 p_{t-1}$

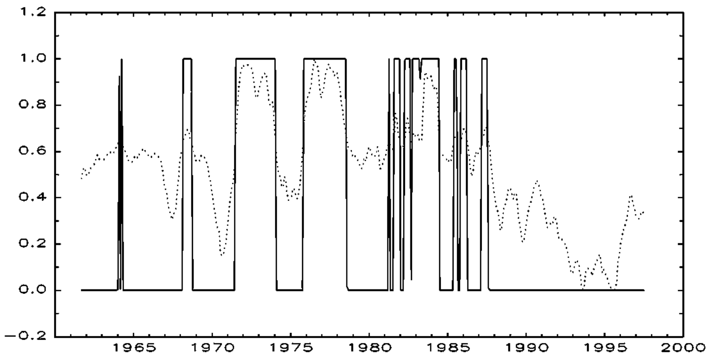
(c) $\hat{c}$, $s_t = \Delta_{12} y_{t-1}$ (d) $\hat{c}$, $s_t = \Delta_{12}^2 p_{t-1}$

**Figure 1.** Time-series plots of the estimated values of the smoothness parameter $\gamma$ and the location parameter $c$ in (2) across the 192 twenty-year windows for the unrestricted STVECM's (Model 3 in Table 2) which use the yearly growth rate in industrial production ($\Delta_{12} y_{t-1}$) or the yearly change in the annual inflation rate ($\Delta_{12}^2 p_{t-1}$) as transition variable. Dates listed on the horizontal axis represent the initial observation for each 20-year window.

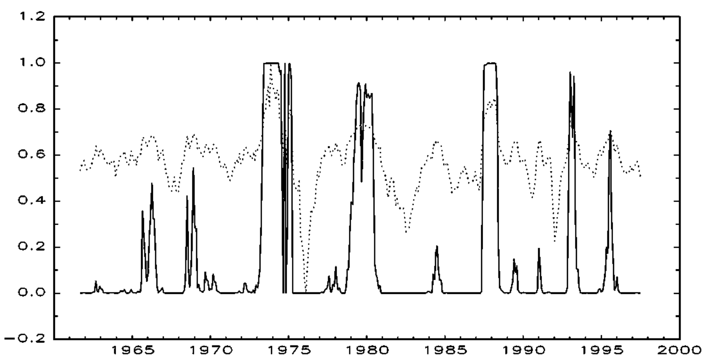
the regime switches do not occur simultaneously. It is difficult to associate the time-series variation in these values for Models $3m$ and $3p$ with significant events in the U.S. macroeconomy over this sample period. The behavior of Model $3y$'s estimated transition function captures the five recessions that the U.S. economy has experienced since the end of 1961. This is not surprising, of course, because the transition variable $s_t$ is the yearly change in industrial production and the logistic function $G(s_t; \gamma, c)$ in (2) is a nonlinear monotonic transformation of that variable. Since the transition variable is the yearly, not monthly, change in




(a) Model 3y - $s_t = \Delta_{12} y_{t-1}$




(b) Model 3m - $s_t = \Delta_{12} m_{t-3}$




(c) Model 3p - $s_t = \Delta_{12}^2 p_{t-1}$

**Figure 2.** Time-series plot of the value of the transition function $G(s_t; \gamma, c)$ in the unrestricted STVECM's (Models $3y$, $3m$, and $3p$) estimated over the full sample 1961:08–1997:06 (solid line) together with the (scaled) transition variable $s_t$ (dotted line).

https://doi.org/10.1017/S1365100501023045 Published online by Cambridge University Press

industrial production over the previous year, the regime switches do not coincide exactly with the business-cycle peaks and troughs identified by the National Bureau of Economic Research, but usually occur a few months later.

Noting several advantages in doing so, Swanson (1998) compares AIC and BIC across models with and without money to test Granger causality, calling such comparisons "predictive accuracy" tests. In that 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.

We followed this suggestion of Swanson (1998) and tested for linear (possibly nonlinear) Granger causality from money to output by comparing the values of AIC and BIC for the estimated linear and nonlinear VECM's. Note that it is not quite clear how to calculate the number of parameters when applying model selection criteria to rank models that involve unidentified nuisance parameters, such as the STVECM. For example, the STVECM can be reduced to a linear VECM by just setting the smoothness parameter $\gamma$ in the logistic transition function equal to zero. Thus, one might argue that it is not fair to count the sum of all parameters in the STVECM when comparing the nonlinear model with a nested linear VECM using the AIC or BIC. To our knowledge, alternatives to the standard model selection criteria have not been developed yet and therefore we have chosen to use the traditional criteria. Finally, because the penalty for additional parameters is considerably higher in the BIC than in the AIC, this problem is more serious for the BIC, especially given the relatively small number of observations in the 20-year moving windows. Therefore, we report only results obtained using the AIC, but results for the BIC are available upon request.

Table 3 reports the ranking of the 11 models across all 192 windows as determined by the AIC and Table 4 presents pairwise comparisons. With this criterion, six of the STVECM's—those using $\Delta_{12}^2 p_{t-1}$ and $\Delta_{12} m_{t-3}$ as transition variables—are, on average, ranked higher than the two linear models, implying that use of the STAR approach can provide a substantial improvement in modeling this vector of time-series data. The STVECM's estimated using the lagged growth rate in output as the transition variable, however, are the three lowest ranked models, so that use of estimated transition functions that are rather closely correlated with U.S. business-cycle movements leads to relatively poor in-sample model performance.

It is interesting to examine in more detail the relative performance of the estimated STVECM's. For example, via the AIC, Models $5p$, $3p$, and $4p$ are the top three, suggesting that the yearly change in the annual inflation rate consistently dominates the two other selected transition varibles. While Model $3p$ strongly dominates Model $4p$ in pairwise AIC comparisons, such comparisons also show that Model $5p$ strongly dominates both Models $3p$ and $4p$, suggesting that money only linearly Granger-causes output in the STVECM framework. Nonetheless, for approximately 43% of the windows, an unrestricted STVECM (either Model $3p$ or Model $3m$) is the top-ranked model via the AIC, providing nontrivial evidence that money does indeed nonlinearly Granger-cause output.

Finally, according to the AIC, Model 1 is ranked higher than Model 2 and, in pairwise comparisons, Model 1 dominates Model 2 about 95% of the time. These results strongly suggest that money Granger-causes output and they are consistent with Swanson's (1998) linear analysis.

**TABLE 3.** Model selection and comparison by AIC[a]

| | Rank $j^b$ | | | | | | | | | | | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | rank[c] |
| 1 | 0.0 | 15.1 | 2.6 | 0.0 | 5.7 | 12.0 | 64.6 | 0.0 | 0.0 | 0.0 | 0.0 | 5.9 |
| 2 | 0.0 | 2.6 | 15.1 | 2.1 | 1.0 | 2.6 | 7.8 | 68.8 | 0.0 | 0.0 | 0.0 | 6.8 |
| 3*y* | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 100.0 | 0.0 | 0.0 | 9.0 |
| 4*y* | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 100.0 | 11.0 |
| 5*y* | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 100.0 | 0.0 | 10.0 |
| 3*m* | 18.8 | 0.0 | 0.0 | 15.6 | 4.2 | 55.7 | 1.0 | 4.7 | 0.0 | 0.0 | 0.0 | 4.8 |
| 4*m* | 0.0 | 0.0 | 0.0 | 6.3 | 65.1 | 4.7 | 7.3 | 16.7 | 0.0 | 0.0 | 0.0 | 5.6 |
| 5*m* | 0.0 | 0.0 | 0.0 | 57.3 | 6.3 | 7.3 | 19.3 | 9.9 | 0.0 | 0.0 | 0.0 | 5.2 |
| 3*p* | 24.0 | 47.4 | 10.9 | 0.0 | 17.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.4 |
| 4*p* | 1.6 | 13.5 | 66.1 | 1.0 | 0.0 | 17.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.4 |
| 5*p* | 55.7 | 21.4 | 5.2 | 17.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.8 |

[a] Models were estimated for 192 windows of length 20 years, with the initial observation for the windows running from 1961:08 to 1977:07. See Table 2 for model definitions.
[b] Percentage of windows for which the various models had rank $j$, $j = 1, \ldots, 11$ using the AIC.
[c] Average rank of models across all windows.

**TABLE 4.** Pairwise model comparisons by AIC[a]

| | Model $j$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Model $i$ | 1 | 2 | 3*y* | 4*y* | 5*y* | 3*m* | 4*m* | 5*m* | 3*p* | 4*p* | 5*p* |
| 1 | | 94.3 | 100.0 | 100.0 | 100.0 | 7.8 | 24.5 | 29.7 | 17.7 | 17.7 | 17.7 |
| 2 | 5.7 | | 100.0 | 100.0 | 100.0 | 6.8 | 23.4 | 26.6 | 17.7 | 17.7 | 17.7 |
| 3*y* | 0.0 | 0.0 | | 100.0 | 100.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 4*y* | 0.0 | 0.0 | 0.0 | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 5*y* | 0.0 | 0.0 | 0.0 | 100.0 | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3*m* | 92.2 | 93.2 | 100.0 | 100.0 | 100.0 | | 37.0 | 40.1 | 18.8 | 18.8 | 18.8 |
| 4*m* | 75.5 | 76.6 | 100.0 | 100.0 | 100.0 | 63.0 | | 21.9 | 0.0 | 0.0 | 0.0 |
| 5*m* | 70.3 | 73.4 | 100.0 | 100.0 | 100.0 | 59.9 | 78.1 | | 0.0 | 0.0 | 0.0 |
| 3*p* | 82.3 | 82.3 | 100.0 | 100.0 | 100.0 | 81.3 | 100.0 | 100.0 | | 90.1 | 24.0 |
| 4*p* | 82.3 | 82.3 | 100.0 | 100.0 | 100.0 | 81.3 | 100.0 | 100.0 | 9.9 | | 6.8 |
| 5*p* | 82.3 | 82.3 | 100.0 | 100.0 | 100.0 | 81.3 | 100.0 | 100.0 | 76.0 | 93.2 | |

[a] Percentage of windows for which Model $i$ was ranked higher than Model $j$ using the AIC. Models were estimated for 192 windows of length 20 years, with the initial observation for the windows running from 1961:08 to 1977:07. See Table 2 for model definitions.

## 4. OUT-OF-SAMPLE FORECASTING

For each of the 11 models discussed earlier, we generate a sequence of 192 out-of-sample forecast profiles as follows: Starting with the estimated model for the 1961:08–1981:07 window, we compute a set of out-of-sample forecasts for forecast horizons $h = 1, \ldots, 30$, so that the out-of-sample forecast period is 1981:08–1984:03. Then we roll the fixed 20-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 1997:07–1999:12.

It is important to recall that we use the same subset VECM's and STVECM's and the same transition variables in all windows. This is based on the variable selection procedure and linearity testing done in Section 2 using the sample period 1961:08–1997:06, so that in this sense we do make use of some out-of-sample information. 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 straightforward, since they are linear. It is well known that computing multistep-ahead forecasts for nonlinear models is more difficult because 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 multistep-ahead forecasts for the nonlinear Models 3, 4, and 5. The bootstrap approach is favored over the Monte Carlo method because no assumptions are required about the distribution of the underlying stochastic error terms.

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 those ranks are computed across all 30 forecast horizons. Arguably the most consistent result in Table 5 is that, across all forecast accuracy criteria, which include the mean square prediction error (MSPE), median square prediction error (MedSPE), and mean absolute error (MAE), Models 1 and 2— the VECM's—are ranked either first or second. Thus, the linear models appear to strongly dominate the nonlinear models in out-of-sample forecasting. The first three panels in Table 6's pairwise comparisons confirm the relative average rankings of the VECM's given in Table 5. Using the MSPE, Model 1 ranks higher than Model 2, but using the other measures, Model 2 is ranked higher than Model 1.

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, Model 1 ranks higher than Model 2, suggesting that money does linearly Granger-cause output. However, this result is uniformly reversed via the MedSPE and MAE. Therefore, use of more robust forecast comparison measures implies that money does not linearly Granger-cause output.

**TABLE 5.** Out-of-sample forecasting ranks[a]

| Model $i$ | Rank $j$ 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | Average rank |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **MSPE** | | | | | | | | | | | | |
| 1 | 33.3 | 26.7 | 16.7 | 10.0 | 13.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.4 |
| 2 | 3.3 | 36.7 | 40.0 | 16.7 | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.8 |
| 3$y$ | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.3 | 96.7 | 11.0 |
| 4$y$ | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 6.7 | 93.3 | 0.0 | 0.0 | 8.9 |
| 5$y$ | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.3 | 0.0 | 0.0 | 93.3 | 3.3 | 9.9 |
| 3$m$ | 0.0 | 0.0 | 0.0 | 0.0 | 3.3 | 20.0 | 23.3 | 46.7 | 3.3 | 3.3 | 0.0 | 7.4 |
| 4$m$ | 26.7 | 16.7 | 23.3 | 13.3 | 6.7 | 13.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.0 |
| 5$m$ | 0.0 | 0.0 | 3.3 | 20.0 | 10.0 | 20.0 | 30.0 | 13.3 | 3.3 | 0.0 | 0.0 | 6.6 |
| 3$p$ | 0.0 | 0.0 | 3.3 | 0.0 | 10.0 | 30.0 | 33.3 | 23.3 | 0.0 | 0.0 | 0.0 | 6.6 |
| 4$p$ | 16.7 | 13.3 | 3.3 | 23.3 | 10.0 | 13.3 | 10.0 | 10.0 | 0.0 | 0.0 | 0.0 | 4.3 |
| 5$p$ | 20.0 | 6.7 | 10.0 | 16.7 | 43.3 | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.7 |
| **MedSPE** | | | | | | | | | | | | |
| 1 | 23.3 | 26.7 | 20.0 | 16.7 | 10.0 | 0.0 | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 2.8 |
| 2 | 46.7 | 30.0 | 10.0 | 13.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.9 |
| 3$y$ | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.3 | 10.0 | 16.7 | 70.0 | 10.5 |
| 4$y$ | 0.0 | 3.3 | 0.0 | 0.0 | 0.0 | 3.3 | 0.0 | 10.0 | 50.0 | 26.7 | 6.7 | 9.0 |
| 5$y$ | 0.0 | 0.0 | 3.3 | 0.0 | 3.3 | 0.0 | 0.0 | 0.0 | 23.3 | 53.3 | 16.7 | 9.5 |
| 3$m$ | 0.0 | 3.3 | 0.0 | 0.0 | 0.0 | 13.3 | 23.3 | 46.7 | 6.7 | 0.0 | 6.7 | 7.6 |
| 4$m$ | 0.0 | 10.0 | 16.7 | 13.3 | 13.3 | 23.3 | 16.7 | 6.7 | 0.0 | 0.0 | 0.0 | 5.0 |
| 5$m$ | 0.0 | 0.0 | 10.0 | 3.3 | 26.7 | 20.0 | 16.7 | 16.7 | 6.7 | 0.0 | 0.0 | 6.1 |
| 3$p$ | 0.0 | 3.3 | 10.0 | 20.0 | 10.0 | 6.7 | 33.3 | 16.7 | 0.0 | 0.0 | 0.0 | 5.7 |
| 4$p$ | 20.0 | 16.7 | 13.3 | 10.0 | 16.7 | 20.0 | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 3.6 |
| 5$p$ | 10.0 | 6.7 | 16.7 | 23.3 | 20.0 | 13.3 | 3.3 | 0.0 | 3.3 | 3.3 | 0.0 | 4.3 |
| **MAE** | | | | | | | | | | | | |
| 1 | 23.3 | 43.3 | 3.3 | 10.0 | 13.3 | 3.3 | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 2.7 |
| 2 | 36.7 | 43.3 | 16.7 | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.9 |
| 3$y$ | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 100.0 | 11.0 |
| 4$y$ | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.3 | 0.0 | 3.3 | 93.3 | 0.0 | 0.0 | 8.9 |
| 5$y$ | 0.0 | 0.0 | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 96.7 | 0.0 | 9.8 |
| 3$m$ | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 16.7 | 30.0 | 46.7 | 3.3 | 3.3 | 0.0 | 7.5 |
| 4$m$ | 13.3 | 3.3 | 46.7 | 10.0 | 10.0 | 16.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.5 |
| 5$m$ | 0.0 | 0.0 | 0.0 | 20.0 | 13.3 | 10.0 | 36.7 | 16.7 | 3.3 | 0.0 | 0.0 | 6.3 |
| 3$p$ | 0.0 | 0.0 | 3.3 | 3.3 | 13.3 | 26.7 | 20.0 | 33.3 | 0.0 | 0.0 | 0.0 | 6.6 |
| 4$p$ | 23.3 | 3.3 | 10.0 | 23.3 | 16.7 | 13.3 | 10.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.9 |
| 5$p$ | 3.3 | 6.7 | 16.7 | 30.0 | 33.3 | 10.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.1 |

[a]The table summarizes results for out-of-sample forecasting for the 11 models and 192 estimation windows, across forecasting horizons $k = 1, \ldots, 30$. The three 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); and mean absolute error (MAE). See Table 2 for model definitions.

https://doi.org/10.1017/S1365100501023045 Published online by Cambridge University Press

**Table 6.** Pairwise model comparisons by out-of-sample forecast[a]

| | Model $j$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Model $i$ | 1 | 2 | $3y$ | $4y$ | $5y$ | $3m$ | $4m$ | $5m$ | $3p$ | $4p$ | $5p$ |
| | | | | | MSPE | | | | | | |
| 1 | 0.0 | 66.7 | 100.0 | 100.0 | 100.0 | 100.0 | 56.7 | 100.0 | 100.0 | 63.3 | 70.0 |
| 2 | 33.3 | 0.0 | 100.0 | 100.0 | 100.0 | 100.0 | 50.0 | 100.0 | 96.7 | 70.0 | 70.0 |
| $3y$ | 0.0 | 0.0 | 0.0 | 0.0 | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| $4y$ | 0.0 | 0.0 | 100.0 | 0.0 | 96.7 | 6.7 | 0.0 | 3.3 | 0.0 | 0.0 | 0.0 |
| $5y$ | 0.0 | 0.0 | 96.7 | 3.3 | 0.0 | 3.3 | 0.0 | 3.3 | 0.0 | 0.0 | 0.0 |
| $3m$ | 0.0 | 0.0 | 100.0 | 93.3 | 96.7 | 0.0 | 0.0 | 16.7 | 36.7 | 20.0 | 0.0 |
| $4m$ | 43.3 | 50.0 | 100.0 | 100.0 | 100.0 | 100.0 | 0.0 | 96.7 | 90.0 | 60.0 | 63.3 |
| $5m$ | 0.0 | 0.0 | 100.0 | 96.7 | 96.7 | 83.3 | 3.3 | 0.0 | 53.3 | 36.7 | 23.3 |
| $3p$ | 0.0 | 3.3 | 100.0 | 100.0 | 100.0 | 63.3 | 10.0 | 46.7 | 0.0 | 13.3 | 3.3 |
| $4p$ | 36.7 | 30.0 | 100.0 | 100.0 | 100.0 | 80.0 | 40.0 | 63.3 | 86.7 | 0.0 | 36.7 |
| $5p$ | 30.0 | 30.0 | 100.0 | 100.0 | 100.0 | 100.0 | 36.7 | 76.7 | 96.7 | 63.3 | 0.0 |
| | | | | | MedSPE | | | | | | |
| 1 | 0.0 | 30.0 | 100.0 | 96.7 | 96.7 | 96.7 | 83.3 | 93.3 | 83.3 | 60.0 | 83.3 |
| 2 | 70.0 | 0.0 | 100.0 | 96.7 | 96.7 | 96.7 | 100.0 | 100.0 | 93.3 | 70.0 | 86.7 |
| $3y$ | 0.0 | 0.0 | 0.0 | 13.3 | 20.0 | 6.7 | 0.0 | 3.3 | 0.0 | 0.0 | 3.3 |
| $4y$ | 3.3 | 3.3 | 86.7 | 0.0 | 70.0 | 13.3 | 3.3 | 10.0 | 6.7 | 0.0 | 6.7 |
| $5y$ | 3.3 | 3.3 | 80.0 | 30.0 | 0.0 | 6.7 | 3.3 | 6.7 | 6.7 | 0.0 | 6.7 |
| $3m$ | 3.3 | 3.3 | 93.3 | 86.7 | 93.3 | 0.0 | 16.7 | 20.0 | 20.0 | 3.3 | 3.3 |
| $4m$ | 16.7 | 0.0 | 100.0 | 96.7 | 96.7 | 83.3 | 0.0 | 73.3 | 53.3 | 40.0 | 40.0 |
| $5m$ | 6.7 | 0.0 | 96.7 | 90.0 | 93.3 | 80.0 | 26.7 | 0.0 | 50.0 | 23.3 | 26.7 |
| $3p$ | 16.7 | 6.7 | 100.0 | 93.3 | 93.3 | 80.0 | 46.7 | 50.0 | 0.0 | 13.3 | 26.7 |
| $4p$ | 40.0 | 30.0 | 100.0 | 100.0 | 100.0 | 96.7 | 60.0 | 76.7 | 86.7 | 0.0 | 50.0 |
| $5p$ | 16.7 | 13.3 | 96.7 | 93.3 | 93.3 | 96.7 | 60.0 | 73.3 | 73.3 | 50.0 | 0.0 |
| | | | | | MAE | | | | | | |
| 1 | 0.0 | 30.0 | 100.0 | 96.7 | 96.7 | 100.0 | 76.7 | 100.0 | 93.3 | 66.7 | 70.0 |
| 2 | 70.0 | 0.0 | 100.0 | 100.0 | 96.7 | 100.0 | 83.3 | 100.0 | 100.0 | 73.3 | 90.0 |
| $3y$ | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| $4y$ | 3.3 | 0.0 | 100.0 | 0.0 | 96.7 | 6.7 | 0.0 | 3.3 | 3.3 | 0.0 | 0.0 |
| $5y$ | 3.3 | 3.3 | 100.0 | 3.3 | 0.0 | 3.3 | 0.0 | 3.3 | 3.3 | 0.0 | 3.3 |
| $3m$ | 0.0 | 0.0 | 100.0 | 93.3 | 96.7 | 0.0 | 0.0 | 16.7 | 36.7 | 10.0 | 0.0 |
| $4m$ | 23.3 | 16.7 | 100.0 | 100.0 | 100.0 | 100.0 | 0.0 | 96.7 | 83.3 | 60.0 | 70.0 |
| $5m$ | 0.0 | 0.0 | 100.0 | 96.7 | 96.7 | 83.3 | 3.3 | 0.0 | 43.3 | 30.0 | 20.0 |
| $3p$ | 6.7 | 0.0 | 100.0 | 96.7 | 96.7 | 63.3 | 16.7 | 56.7 | 0.0 | 3.3 | 3.3 |
| $4p$ | 33.3 | 26.7 | 100.0 | 100.0 | 100.0 | 90.0 | 40.0 | 70.0 | 96.7 | 0.0 | 56.7 |
| $5p$ | 30.0 | 10.0 | 100.0 | 100.0 | 96.7 | 100.0 | 30.0 | 80.0 | 96.7 | 43.3 | 0.0 |

**TABLE 6.** (Continued.)

| | Model $j$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Model $i$ | 1 | 2 | $3y$ | $4y$ | $5y$ | $3m$ | $4m$ | $5m$ | $3p$ | $4p$ | $5p$ |
| | | | | | | MDM | | | | | |
| 1 | | 0.0 | 6.7 | 16.7 | 3.3 | 76.7 | 0.0 | 16.7 | 10.0 | 3.3 | 10.0 |
| 2 | 0.0 | | 6.7 | 16.7 | 3.3 | 83.3 | 0.0 | 6.7 | 13.3 | 6.7 | 6.7 |
| $3y$ | 0.0 | 0.0 | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| $4y$ | 0.0 | 0.0 | 6.7 | | 3.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| $5y$ | 0.0 | 0.0 | 3.3 | 0.0 | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| $3m$ | 0.0 | 0.0 | 3.3 | 3.3 | 3.3 | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| $4m$ | 0.0 | 0.0 | 6.7 | 16.7 | 3.3 | 73.3 | | 20.0 | 3.3 | 10.0 | 10.0 |
| $5m$ | 0.0 | 0.0 | 6.7 | 10.0 | 3.3 | 46.7 | 0.0 | | 6.7 | 0.0 | 0.0 |
| $3p$ | 0.0 | 0.0 | 6.7 | 13.3 | 3.3 | 16.7 | 0.0 | 0.0 | | 0.0 | 0.0 |
| $4p$ | 0.0 | 0.0 | 6.7 | 16.7 | 3.3 | 33.3 | 0.0 | 20.0 | 6.7 | | 0.0 |
| $5p$ | 3.3 | 0.0 | 6.7 | 16.7 | 3.3 | 43.3 | 3.3 | 20.0 | 46.7 | 3.3 | |
| | | | | | | MDMFE | | | | | |
| 1 | | 53.3 | 96.7 | 96.7 | 96.7 | 86.7 | 53.3 | 70.0 | 73.3 | 60.0 | 60.0 |
| 2 | 53.3 | | 100.0 | 96.7 | 96.7 | 76.7 | 36.7 | 46.7 | 66.7 | 70.0 | 56.7 |
| $3y$ | 93.3 | 93.3 | | 93.3 | 93.3 | 93.3 | 93.3 | 93.3 | 93.3 | 93.3 | 93.3 |
| $4y$ | 80.0 | 80.0 | 100.0 | | 96.7 | 80.0 | 83.3 | 80.0 | 80.0 | 80.0 | 80.0 |
| $5y$ | 93.3 | 93.3 | 100.0 | 93.3 | | 93.3 | 93.3 | 93.3 | 93.3 | 93.3 | 93.3 |
| $3m$ | 0.0 | 0.0 | 96.7 | 80.0 | 96.7 | | 0.0 | 0.0 | 33.3 | 20.0 | 6.7 |
| $4m$ | 56.7 | 33.3 | 100.0 | 96.7 | 96.7 | 90.0 | | 96.7 | 80.0 | 56.7 | 46.7 |
| $5m$ | 43.3 | 36.7 | 96.7 | 86.7 | 93.3 | 73.3 | 46.7 | | 56.7 | 40.0 | 40.0 |
| $3p$ | 40.0 | 23.3 | 100.0 | 90.0 | 93.3 | 46.7 | 26.7 | 33.3 | | 33.3 | 33.3 |
| $4p$ | 30.0 | 50.0 | 96.7 | 96.7 | 96.7 | 60.0 | 40.0 | 40.0 | 63.3 | | 46.7 |
| $5p$ | 33.3 | 13.3 | 100.0 | 93.3 | 93.3 | 56.7 | 20.0 | 40.0 | 96.7 | 30.0 | |

[a] The table presents pairwise model comparisons based on out-of-sample forecasting results for the 11 models and 192 estimation windows, across forecasting horizons $k = 1, \ldots, 30$. The five 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); 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 forecast-encompassing the 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.

While no STVECM ever appears among the top two average-ranked forecasting models, it is useful to note that the MSPE rankings show that one of the STVECM's (either Model $4m$, $4p$, or $5p$) is the top-ranked model for approximately 65% of the windows. According to the MedSPE and MAE rankings, one of these STVECM's is the top-ranked model for roughly 30% to 35% of the time. Thus, out-of-sample forecast improvement apparently is achievable with the STVECM approach. Given these results, it might be informative to conduct a forecasting experiment in which, for example, various forecast combinations of the VECM's and STVECM's are also studied.

Although some of the STVECM's appear to forecast relatively well at least occasionally, this generally is not the case for the unrestricted STVECM's, Models $3y$, $3m$, and $3p$. Indeed, for each transition variable and each forecast accuracy measure, the unrestricted STVECM has the lowest average rank across the three STVECM's considered. The evidence in Table 5, then, does not support the claim that money nonlinearly Granger-causes output. This finding is corroborated by the pairwise results in the first three panels of Table 6.

The fourth panel of Table 6 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 one model over another is statistically significant. On the whole the values in this table are relatively low, showing that the MSPE reductions are statistically significant in only a relatively few cases. The highest values appear in the seventh column, indicating that Model $3m$'s forecasts tend to be rather strongly dominated by the forecasts of the other models. The fifth panel compares the model forecasts via the Diebold–Mariano type of forecast-encompassing test introduced by Harvey et al. (1998) (MDMFE). In many cases the MDMFE comparisons imply that two models' forecast encompass one another, a strong exception being the tests between Model $3m$'s forecasts and those of Models 1, 2, $4m$, and $5m$. Although the MDMFE results do not suggest that money nonlinearly Granger-causes output, it is puzzling that Model $3y$'s forecasts encompass the forecasts of Model 1 and 2 in more than 90% of the windows, given that Model $3y$ is ranked last across all criteria in Table 5.

Returning once more to the comparison between Model 1 and Model 2, when the MDM criterion is used Model 1's MSPE reductions over those of Model 2 are never statistically significant. The MDMFE comparisons show that these two models' forecast encompass the forecasts of the other model with the same frequency. Therefore, the MDM and MDMFE results do not support Swanson's (1998) finding that money Granger-causes output.

Considering our use of the MDM and MDMFE tests, it is important to note some recent work in the out-of-sample forecast comparison literature. For example, the tests we use ignore any modification for error in estimation of model parameters, a problem considered by West (1996), West and McCracken (1998), Clark and McCracken (1999), and McCracken (1999). Further, Clark and McCracken's (1999) and McCracken's (1999) analyses identify problems associated with use of the MDM and MDMFE tests when the models being compared are nested, as are some of the models that we consider in our forecasting exercise. Clark and McCracken (1999) tabulate critical values for the distribution of the Diebold–Mariano test for the case in which the competing models are nested, while McCracken (1999) tabulates critical values for Granger causality-type Diebold–Mariano tests. However, given the dependence of these critical values on nuisance parameters, and since they are often close to the standard normal critical values commonly used, we do not use these alternative critical values.

It is also 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 such results here.

Finally, we feel it is useful to recall Granger and Teräsvirta's (1993) warning that superior in-sample performance obtained by nonlinear models will be matched out-of-sample only to the extent that this latter period shares similar nonlinear features with the earlier period. To examine whether the sequence of out-of-sample periods employed in this paper is indeed insufficiently "nonlinear" would require additional analysis not yet carried out. But such an analysis might be quite informative and could help explain the relatively poor out-of-sample forecasting performance of the STVECM's.

## 5. CONCLUSIONS

Our paper contains several important results. First, using a relatively conservatively sized heteroskedasticity-robust test, we reject at conventional significance levels the null hypothesis of linearity for a four-variable VECM of industrial production, money, prices, and interest rates, employing a prespecified cointegrating vector, with use of many candidate transition variables. This test result indicates a certain form of nonlinear Granger causality from money to output, since lags of the growth rate of M2 are often among the top-ranked transition variables.

Second, results of the in-sample predictive accuracy test, given by comparison of the AIC values for the estimated models on rolling fixed windows, clearly demonstrate that many of STVECM's dominate the linear VECM's. 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.

Third, in our simulated out-of-sample forecasting exercise, the linear VECM's dominate rather unambiguously via four measures of forecast accuracy. However, further testing using one of these measures, the mean square prediction error, shows that these forecast improvements are rarely significant at conventional significance levels. In addition, in most cases, pairwise forecast-encompassing tests show that both models' forecasts tend to encompass one another. Although none of our forecasting results suggest that money nonlinearly Granger-causes output, it is intriguing to observe that one unrestricted STVECM generates forecasts that encompass the VECM forecasts more than 90% of the time. This is especially interesting given that this particular unrestricted STVECM performs so relatively poorly based upon both our in-sample analysis and evaluation of out-of-sample forecasting using alternative indicators of forecast accuracy.

Our paper demonstrates the importance of considering both in-sample and out-of-sample analyses of causality. This is true not only for the question of whether money nonlinearly Granger-causes output within the STVECM framework, but also for comparisons between linear VECM's. Focusing solely on the two linear VECM's studied, our out-of-sample forecasting exercise implies that the causal link from money to output may be a good deal weaker than Swanson's (1998) study suggests. Similarly, our in-sample and out-of-sample results are consistent with the findings of both Amato and Swanson (in press) and Chao et al. (2001).

## REFERENCES

Amato, J. & N.R. Swanson (in press) The real-time predictive content of money for output, *Journal of Monetary Economics*.

Balke, N.S. & T.B. Fomby (1997) Threshold cointegration. *International Economic Review* 38, 627–646.

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

Black, D.C., P.R. Corrigan, & M.R. Dowd (2000) New dogs and old tricks: Do money and interest rates still provide information content for forecasts of output and prices? *International Journal of Forecasting* 16, 191–205.

Boschen, J.F. & H.I. Grossman (1982) Tests of equilibrium macroeconomics using contemporaneous monetary data. *Journal of Monetary Economics* 10, 309–333.

Brown, B.Y. & R.S. Mariano (1989) Predictors in dynamic nonlinear models: Large sample behavior. *Econometric Theory* 5, 430–452.

Brüggemann, R. & H. Lütkepohl (2000) Lag Selection in Subset VAR Models with an Application to a U.S. Monetary System. SFB 373 discussion paper 2000-37, Humboldt University.

Caballero, R. & M. Hammour (1994) The cleansing effects of recession. *American Economic Review* 84, 1350–1368.

Caplin, A.S. & J. Leahy (1991) State-contingent pricing and the dynamics of money and output. *Quarterly Journal of Economics* 106, 683–708.

Chao, J., V. Corradi, & N.R. Swanson (2001) Out of sample test for Granger causality. *Macroeconomic Dynamics* 5, 598–620.

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

Clark, T.E. & M.W. McCracken (1999) Granger Causality and Tests of Equal Forecast Accuracy and Encompassing. Research working paper 99-11, Federal Reserve Bank of Kansas City.

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

Cooley, T.F. & G.D. Hansen (1995) Money and the business cycle. In T.F. Cooley (ed.), *Frontiers of Business Cycle Research*. Princeton, NJ: Princeton University Press.

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

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

Diebold, F.X. & G.D. Rudebusch (1991) Forecasting output with the composite leading index: A real time analysis. *Journal of the American Statistical Association* 86, 603–610.

Dufour, J.-M. & E. Renault (1998) Short run and long run causality in time series: Theory. *Econometrica* 66, 1127–1162.

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

Eitrheim, Ø. & T. Teräsvirta (1996) Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics* 74, 59–75.

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

Garratt, A., K. Lee, M.H. Pesaran, & Y. Shin (2000) A structural cointegrating VAR approach to macroeconometric modelling. In S. Holly & M. Weale (eds.), *Econometric Modelling: Techniques and Applications*, pp. 94–131. Cambridge, UK: Cambridge University Press.

Granger, C.W.J. (2001) Overview of nonlinear macroeconometric empirical models. *Macroeconomic Dynamics* 5, 466–481.

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

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, & P. Newbold (1997) Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13, 281–291.

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

Hendry, D.F. & 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*, pp. 272–300. Cambridge, MA: Basil Blackwell.

Hess, G.D. & R.D. Porter (1993) Comparing interest-rate spreads and money growth as predictors of output growth: Granger causality in the sense Granger intended. *Journal of Economics and Business* 45, 247–268.

Hiemstra, C. & 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, & S.M. Potter (1996) Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* 74, 119–147.

Lee, H.S. & P. Siklos (1997) The role of seasonality in economic time series—reinterpreting money-output causality in U.S. data. *International Journal of Forecasting* 13, 381–391.

Lundbergh, S. & T. Teräsvirta (1998) Modelling Economic High-Frequency Time Series with STAR-GARCH Models. Working Paper Series in Economics and Finance 291, Stockholm School of Economics.

Luukkonen, R., P. Saikkonen, & T. Teräsvirta (1988) Testing linearity against smooth transition autoregressive models. *Biometrika* 75, 491–499.

Mankiw, N.G., D.E. Runkle, & M.D. Shapiro (1984) Are preliminary announcements of the money stock rational forecasts? *Journal of Monetary Economics* 14, 15–27.

McCracken, M.W. (1999) Asymptotics for Out of Sample Tests of Causality. Working paper 99-1, Louisiana State University.

Péguin-Feissolle, A. & T. Teräsvirta (1999) A General Framework for Testing the Granger Noncausality Hypothesis. Working Paper Series in Economics and Finance 343, Stockholm School of Economics.

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

Pötscher, B.M. & I.V. Prucha (1997) *Dynamic Nonlinear Econometric Models—Asymptotic Theory*. Berlin: Springer–Verlag.

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

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

Söderlind, P. & A. Vredin (1996) Applied cointegration analysis in the mirror of macroeconomic theory. *Journal of Applied Econometrics* 11, 363–381.

Stock, J.H. & 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.

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.

Thoma, M.A. & J.A. Gray (1998) Financial market variables do not predict real activity. *Economic Inquiry* 36, 522–539.

Tsay, R.S. (1998) Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* 93, 1188–1202.

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

West, K.D. (1996) Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.

West, K.D. & M.W. McCracken (1998) Regression based tests of predictive ability. *International Economic Review* 39, 817–840.

White, H. & 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.