The Demand for Money during Hyperinflations under Rational Expectations: I

Author(s): Thomas J. Sargent

Source: *International Economic Review*, Feb., 1977, Vol. 18, No. 1 (Feb., 1977), pp. 59-82

Published by: Wiley for the Economics Department of the University of Pennsylvania and Institute of Social and Economic Research, Osaka University

Stable URL: https://www.jstor.org/stable/2525769

INTERNATIONAL ECONOMIC REVIEW
Vol. 18, No. 1, February, 1977

THE DEMAND FOR MONEY DURING HYPERINFLATIONS UNDER RATIONAL EXPECTATIONS: I*

BY THOMAS J. SARGENT[1]

1. INTRODUCTION

This paper proposes methods for estimating the demand schedule for money that Cagan used in his famous study of hyperinflation [**3**]. Wallace and I [**8**] pointed out that under assumptions that make Cagan's adaptive expectations scheme equivalent with assuming rational expectations, Cagan's estimator of α, which is the slope of the log of the demand for real balances with respect to expected inflation, is not statistically consistent. This is interesting in light of a paradox that emerged when Cagan used his estimates of α to calculate the sustained rates of inflation associated with the maximum flow of real resources that the creators of money could command by printing money. This "optimal" rate of inflation turns out to be $-1/\alpha$. For each of the seven hyperinflations, the reciprocal of Cagan's estimate of $-\alpha$ turned out to be less, and often very much less, than the actual average rate of inflation. The data are shown in Table 1, which reproduces a table of Cagan's. Cagan's estimates imply that the creators of money expanded the money supply at rates that far exceeded the sustained rates which maximized the real revenues they could obtain. A natural

TABLE 1

	(1)	(2)	(3)
Austria	.117	12	47
Germany	.183	20	322
Greece	.244	28	365
Hungary I	.115	12	46
Hungary II	.236	32	19,800
Poland	.435	54	81
Russia	.327	39	57

Column (1) $= -1/\hat{\alpha}$ (continuous compounding), rate per month that maximizes revenue of money creator.
Column (2) $= (e^{-1/\alpha} - 1)$ 100 (neglects compounding).
Column (3) = average actual rate of inflation per month
Source: Cagan's Figure 9, [**3**, (81)].

* Manuscript received August 4, 1975; revised March 17, 1976.
[1] Research on this paper was supported by the Federal Reserve Bank of Minneapolis, which doesn't necessarily endorse the conclusions. Helpful comments on an earlier draft were received from Christopher Sims, Thomas Turner, John Gewcke, Milton Friedman, Jacob Frankel, Robert E. Lucas, Jr., and Rusdu Saracoglu. Rusdu Saracoglu performed the calculations using a computer program that he wrote for estimating bivariate mixed moving average, autoregressive models.

first thing to consider in explaining this apparently irrational behavior by the creators of money is the possibility that it is a statistical artifact, namely, a consequence of using bad estimates of α.

This paper aims to complete a task begun by Wallace and me [8], namely, the analysis of Cagan's model of hyperinflation under circumstances in which Cagan's "adaptive" scheme for forming expectations of inflation is equivalent with expectations that are "rational" in Muth's sense [6]. The model is a very simple simultaneous-equations model of the inflation-money creation process, one equation of which turns out to be identical with Cagan's simple portfolio equilibrium equation. As Wallace and I have argued, Cagan's use of single equations methods exposed him to the possibility of severe simultaneous-equations bias. The present paper uses the full information maximum likelihood estimator to obtain a consistent estimator of Cagan's model. It also obtains an expression for the statistical inconsistency of Cagan's estimator for his model under the circumstances in which adaptive expectations coincide with rational expectations.

One way of justifying imposing rational expectations on Cagan's model is that it enables one to specify a complete model of the inflation-money creation process in a very economical way. This is a virtue, since the time series from the hyperinflations are too short to permit estimating complicated parameterizations. But a more important reason for using the hypothesis of rational expectations to complete Cagan's model is that doing so delivers an econometric model that is seemingly consistent with the exogeneity (or "causal") structure exhibited by the money creation-inflation process during the seven hyperinflations studied by Cagan. Empirical tests by Wallace and me typically indicated substantial evidence of feedback from inflation to money creation, with markedly less feedback from money creation to inflation. Cagan's model under rational expectations predicts a particular extreme version of such a pattern: it predicts that inflation "causes" (in Granger's sense) money creation with no reverse feedback (or "causality") from money creation to inflation. Cagan's model with rational expectations thus seems to provide one way of explaining the Granger-causal structure exhibited in the data.

Cagan's paper is rightly regarded as one of the best pieces of empirical work ever done in economics. His model and his estimation method have been applied with apparent success to a number of additional countries experiencing high inflation rates, but rates falling short of those characterizing hyperinflations.[2] The key substantive conclusion that has been drawn from Cagan's study, and those subsequent studies as well, is that even in the apparently chaotic conditions of rampant inflation it is possible to isolate a stable demand schedule for money having real balances varying inversely with the expected rate of inflation. In the light of the results of this paper, that conclusion must be severely modified. First, it is shown below that under conditions that make Cagan's model equivalent with assuming "rational" expectations, the slope parameter α is not econometrical-

[2] Among such studies are some of those in Meiselman [5].

ly identifiable. To identify α requires imposing a restriction on the covariance of the disturbances to the demand for money and to the supply of money. Neither economic theorizing nor intuition seems to provide a ready restriction on that covariance. Proceeding on the "neutral" assumption that that covariance is zero, one can extract estimates of α. But even then, the estimates of α are characterized by large standard errors.

From a technical point of view, this paper is an exercise in applying vector time series models. The key references are Granger [**4**], Sims [**9**], Wilson [**10**], Porter [**7**], and Zellner and Palm [**11**]. The model studied here is an interesting one from the point of view of the vector time series model, since it is one in which inflation "causes" money creation in Granger's sense, although these two series are supposed to be perfectly in phase, so that neither one "leads" the other. The model thus provides an example that illustrates the difference between Granger's causality and simple notions of the lead of one series over another. The model is also interesting because it illustrates the very important difference between Granger causality and a separate notion of causality often used by economists, namely, that of invariance with respect to an intervention. The present model predicts that money "causes" inflation in the sense that a given change in the stochastic process or "feedback rule" governing the money supply will produce a determinate change in the stochastic process for inflation. The stochastic process for inflation is an invariant function of the stochastic process governing money creation. In Cagan's model with rational expectations imposed, inflation Granger-causes money creation with no reverse Granger causality from money to inflation because the system is operating under a particular money supply rule that in effect prevents the money supply from being of any use in predicting subsequent rates of inflation. If there is a change in monetary regime, that is, a switch in the money supply rule, the economic model predicts that the Granger-causality structure of the money-inflation process will change.

2. THE MODEL

Cagan's model of hyperinflation builds on a demand schedule for real balances of the form

$$m_t - p_t = \alpha\pi_t + u_t, \qquad \alpha < 0 \tag{1}$$

where m is the log of the money supply (which is always equal to the log of the money demand); p is the log of the price level; π_t is the expected rate of inflation, i.e., the public's psychological expectation of $p_{t+1} - p_t$; and u_t is a random variable with mean zero. I have omitted a constant term from (1), though one would be included in empirical work. Cagan assumed that π_t was formed via the adaptive expectations scheme

$$\pi_t = \frac{1-\lambda}{1-\lambda L}(p_t - p_{t-1})$$

or

$$\pi_t = \frac{1-\lambda}{1-\lambda L} x_t \tag{2}$$

where $x_t = p_t - p_{t-1}$, the rate of inflation, and where L is the lag operator defined by $L^n x_t = x_{t-n}$.

Under rational expectations we require that

$$\pi_t = E_t x_{t+1}, \tag{3}$$

where $E_t x_{t+1}$ is the mathematical expectation of x_{t+1} conditional on information available as of time t.[3] Using (3) and recursions on (1), it is straightforward to show that under rational expectations we must have[4]

$$\pi_t = E_t x_{t+1} = \frac{1}{1-\alpha} \sum_{j=1}^{\infty} \left(\frac{-\alpha}{1-\alpha}\right)^{j-1} E_t \mu_{t+j} - \frac{1}{1-\alpha} \sum_{j=1}^{\infty} \left(\frac{-\alpha}{1-\alpha}\right)^{j-1} (E_t u_{t+j} - E_t u_{t+j-1}) \tag{4}$$

where $\mu_t = m_t - m_{t-1}$, the percentage rate of increase of the money supply. Equation (4) characterizes the (systematic part of the) stochastic process for inflation as a function of the (systematic part of the) stochastic process for money creation. The model asserts that (4) is invariant with respect to interventions in the form of changes in the stochastic process governing money creation. In this sense, since changes in the stochastic process for money creation are supposed to produce predictable changes in the stochastic process for inflation, money "causes" inflation.

For Cagan's adaptive expectation scheme (2) to be equivalent to rational expectations we require:

$$\frac{1-\lambda}{1-\lambda L} x_t = \frac{1}{1-\alpha} \sum_{j=1}^{\infty} \left(\frac{-\alpha}{1-\alpha}\right)^{j-1} E_t \mu_{t+j} - \frac{1}{1-\alpha} \sum_{j=1}^{\infty} \left(\frac{-\alpha}{1-\alpha}\right)^{j-1} (E_t u_{t+j} - E_t u_{t+j-1}). \tag{5}$$

The necessary and sufficient condition for (5) to hold for all α and all t is

[3] I assume that the information avilable consists (at least) of observations of current and past μ's and current and past x's. Thus $E_t x_{t+1} \equiv E[x_{t+1} \mid \mu_t, \mu_t, \ldots, x_t, x_{t-1}, \ldots]$. Similarly, where z_t is any arbitrary random variable, I will write $E_t z_{t+1}$ for $E[z_{t+1} \mid \mu_t, \mu_{t-1}, \ldots, x_t, x_{t-1}, \ldots]$.

[4] Substituting (3) into (1), first differencing, and shifting the time subscripts forward one period gives

$$\mu_{t+1} - x_{t+1} = \alpha E_{t+1} x_{t+2} - \alpha x_{t+1} + (u_{t+1} - u_t).$$

Taking expectations conditional on information available at time t gives

$$E_t x_{t+1} = \frac{1}{1-\alpha} E_t \mu_{t+1} - \frac{\alpha}{1-\alpha} E_t x_{t+2} - (E_t u_{t+1} - E_t u_t).$$

Recursion on the above difference equation shows that equation (4) is indeed a solution to that equation.

$$E_t\mu_{t+j} - E_t(u_{t+j} - u_{t+j-1}) = \frac{1-\lambda}{1-\lambda L} x_t .$$

For an arbitrary μ process, there exists a disturbance process u_t satisfying the above restriction, one in which $E_t(u_{t+j} - u_{t+j-1})$ is a complicated function of lagged x's and lagged μ's. From my point of view, however, the most fruitful conditions to impose are the following two that are sufficient (though clearly not necessary) to satisfy (5). The first condition is

(6) $$u_t = u_{t-1} + \eta_t$$

where η_t is a serially uncorrelated random term with mean zero and variance σ_η^2; I assume that $E[\eta_t | \mu_{t-1}, \mu_{t-2}, \ldots, x_{t-1}, x_{t-2} \ldots] = 0$. According to (6), u takes a random walk. Equation (6) implies that

$$E_t u_{t+j} = u_t, \qquad j \geq 0$$

which implies that

$$E_t u_{t+j} - E_t u_{t+j-1} = 0 \qquad \text{for all} \quad j \geq 1.$$

The second of my pair of sufficient conditions for (5) is

(7) $$E_t\mu_{t+j} = E_t\mu_{t+1} \qquad \text{for} \quad j > 1,$$

so that a constant rate of money creation is expected to occur over the entire future. Assuming (6) and (7) then implies that the appropriate version of (5) is[5]

$$\left(\frac{1-\lambda}{1-\lambda L}\right) x_t = E_t\mu_{t+1} \frac{1}{1-\alpha} \sum_{j=1}^{\infty} \left(\frac{-\alpha}{1-\alpha}\right)^{j-1}$$

or

(8) $$\frac{1-\lambda}{1-\lambda L} x_t = E_t\mu_{t+1} .$$

A process that satisfies (8) is

[5] To see that process (9) satisfies (8), write (9) as

$$\mu_{t+1} = (1-\lambda)x_{t+1} + \frac{(1-\lambda)\lambda}{1-\lambda L} x_t + \varepsilon_{t+1} .$$

Taking expectations conditional on information available at t, we have

$$E_t\mu_{t+1} = (1-\lambda)E_t x_{t+1} + \frac{(1-\lambda)\lambda}{1-\lambda L} x_t .$$

But $E_t x_{t+1} = \frac{1-\lambda}{1-\lambda L} x_t$, so that we have

$$E_t\mu_{t+1} = ((1-\lambda) + \lambda)\left(\frac{1-\lambda}{1-\lambda L}\right) x_t$$

$$E_t\mu_{t+1} = \frac{1-\lambda}{1-\lambda L} x_t ,$$

as required.

$$(9)\qquad \mu_t = \left(\frac{1-\lambda}{1-\lambda L}\right)x_t + \varepsilon_t (= E_t x_{t+1} + \varepsilon_t)$$

where ε_t is a serially uncorrelated random term with mean zero and variance σ_ε^2, and that satisfies

$$E(\varepsilon_t | x_{t-1}, x_{t-2}, \ldots, \mu_{t-1}, \mu_{t-2}, \ldots) = 0\,.$$

According to (9), the rate of money creation equals the expected rate of inflation plus a random term. Equation (9), which has been arrived at in a purely mechanical fashion by pursuing the implications of the assumption that Cagan's adaptive expectations scheme is rational, is nevertheless of interest as an hypothesis about the government's behavior. For example, if the government is creating money to finance a large part of a roughly fixed rate of real government purchases, then there is a presumption that inflation and expected inflation will feed back into money creation, an implication with which (9) is consistent. Thus, when π_t increases, causing $m_t - p_t$ to fall and thereby causing p_t to rise with a fixed m_t, money depreciates in value, prompting the creators of money to increase the rate of printing money in order to maintain their command over the flow of real resources (see Sargent and Wallace [**8**]). Alternatively, equation (9) is compatible with a "real bills" regime in which the monetary authority sets out to supply whatever money the public demands at some fixed nominal interest rate or some fixed real money supply. Equation (9) looks like a rule in which the monetary authority is attempting to peg the (rate of growth of the) real money supply. During the German hyperinflation, German monetary officials in effect repeatedly acknowledged that they were operating under a real-bills regime, acknowledgments made in efforts to argue that their actions were not causing the inflation but were merely responses to it.

The foregoing establishes that if equations (6) and (9) obtain, Cagan's adaptive expectations scheme is compatible with rational expectations and with the portfolio balance condition that he assumed. Under these assumptions, inflation and money creation form a bivariate stochastic process given by

$$(10)\qquad \mu_t - x_t = \alpha(1-L)\left(\frac{1-\lambda}{1-\lambda L}\right)x_t + \eta_t$$

$$(9)\qquad \mu_t = \left(\frac{1-\lambda}{1-\lambda L}\right)x_t + \varepsilon_t\,.$$

Equation (10) was obtained by first differencing (1) and then substituting for π_t from (2) and for $u_t - u_{t-1}$ from (6). The process (10)–(9) can be rewritten as

$$(11)\qquad (1-L)x_t = (\lambda + \alpha(1-\lambda))^{-1}(1-\lambda L)(\varepsilon_t - \eta_t)$$

$$(12)\qquad (1-L)\mu_t = [(\lambda + \alpha(1-\lambda))^{-1}](1-\lambda)(\varepsilon_t - \eta_t) - \varepsilon_{t-1} + \varepsilon_t.$$

Equations (11) and (12) can be derived directly from (10) and (9); alternatively, see Sargent and Wallace for a somewhat different but equivalent way of deriving (11) and (12).

The statistical model (11)–(12) was constructed in a fashion to guarantee the condition

$$E_t x_{t+1} = \frac{1-\lambda}{1-\lambda L} x_t$$

a condition that implies that μ does not Granger cause x. For the above equation states that once lagged x's are taken into account, lagged μ's don't help predict current x, which is Granger's definition of μ's not causing x. It bears mentioning that the statistical model inherits its Granger-causal structure in large part from the particular conditions (6) and (7). The statistical model (11)–(12) is *not* invariant with respect to an intervention in the form of a change in the money supply rule. Rather, it is equation (4) that is supposed to be invariant with respect to interventions in the form of changes in monetary regime. According to (4), changes in the μ_t process — which show up in changes in the (functions) $E_t\mu_{t+j}$ — result in changes in the systematic part of the inflation process, $E_t x_{t+1}$. Thus, one cannot expect the Granger-causal structure of the present model to survive interruptions in monetary regimes.

3. THE BIAS IN CAGAN'S ESTIMATOR

A convenient way to evaluate the (asymptotic) bias in Cagan's estimator is first to obtain a bivariate Wold representation[6] for $(\Delta x_t, \Delta\mu_t)$. Write (11) and (12) as

$$(13) \qquad (1-L)x_t = \phi(1-\lambda L)(\varepsilon_t - \eta_t)$$

$$(14) \qquad (1-L)\mu_t = \phi(1-\lambda)(\varepsilon_t - \eta_t) + (1-L)\varepsilon_t$$

where $\phi = (\lambda + \alpha(1-\lambda))^{-1}$. Next decompose ε_t according to

$$\varepsilon_t = E[\varepsilon_t | \varepsilon_t - \eta_t] + v_t$$

or

$$(15) \qquad \varepsilon_t = \rho(\varepsilon_t - \eta_t) + v_t$$

where $E[v_t | \varepsilon_t - \eta_t] = 0$ and ρ is the regression coefficient of ε_t on $(\varepsilon_t - \eta_t)$. Substituting (15) into (14) gives

$$(16) \qquad (1-L)\mu_t = [\phi(1-\lambda) + \rho(1-L)](\varepsilon_t - \eta_t) + (1-L)v_t$$

Since v_t is orthogonal to $(\varepsilon_t - \eta_t)$ and is serially uncorrelated by construction (recall that $v_t = \varepsilon_t - \rho(\varepsilon_t - \eta_t)$, where ε_t and η_t are serially uncorrelated), it follows that (13) and (16) form a (triangular) bivariate Wold representation for $(\Delta x_t, \Delta\mu_t)$ with fundamental noises $(\varepsilon_t - \eta_t)$ and v_t. The existence of a triangular bivariate Wold representation verifies that Δx is econometrically exogenous with

[6] The most readily accessible reference in economics on the multivariate Wold representation is Sims [9], especially the appendix.

respect to $\Delta\mu$ and that $\Delta\mu$ does not cause Δx in Granger's sense (see Sims [9]). It also makes it very easy to determine the population projection of $\Delta\mu$ on current and past Δx's, from which the asymptotic bias in Cagan's estimator is calculable.

From (13) notice that

$$\varepsilon_t - \eta_t = \phi^{-1}\frac{(1-L)}{1-\lambda L}x_t. \tag{17}$$

To obtain the projection of $\Delta\mu_t$ against current and past (and future) Δx's, substitute (17) into (16) to get

$$(1-L)\mu_t = (\phi(1-\lambda) + (1-L)\rho)\phi^{-1}\frac{(1-L)}{1-\lambda L}x_t + (1-L)v_t.$$

Dividing through by $(1-L)$ gives

$$\mu_t = \left(\frac{1-\lambda}{1-\lambda L} + \frac{(1-L)\rho\phi^{-1}}{1-\lambda L}\right)x_t + v_t$$

$$\mu_t = \left(\frac{1-\lambda+\rho(\lambda+\alpha(1-\lambda))(1-L)}{1-\lambda L}\right)x_t + v_t. \tag{18}$$

Recall from (13) that the v_t process is orthogonal to the x process. Therefore, equation (18) gives the projection of μ_t on x. Substracting x_t from both sides gives the projection of $\mu_t - x_t$ against x_t:

$$\mu_t - x_t = \left(\frac{1-\lambda+\rho(\lambda+\alpha(1-\lambda))(1-L)-(1-\lambda L)}{1-\lambda L}\right)x_t + v_t$$

or

$$\mu_t - x_t = \frac{[-\lambda+\rho(\lambda+\alpha(1-\lambda))](1-L)}{(1-\lambda L)}x_t + v_t. \tag{19}$$

Operating on (19) with the "summation" operator $(1-L)^{-1}$ gives

$$m_t - p_t = \frac{[-\lambda+\rho(\lambda+\alpha(1-\lambda))]}{1-\lambda L}x_t + \xi_t \tag{20}$$

where $\xi_t = \xi_{t-1} + v_t$. Equation (20) is the projection that Cagan estimated by (nonlinear) least squares regression. Notice that the residuals in (20) follow a random walk. It is noteworthy in this regard that the residuals in Cagan's and Barro's estimates of (20) were highly serially correlated, Barro reporting very low values for Durbin-Watson statistics.

Now Cagan regarded the projection (20) as giving estimates of the equation

$$m_t - p_t = \frac{\alpha(1-\lambda)}{1-\lambda L}x_t + u_t. \tag{21}$$

Least squares regression consistently estimates the parameters of the population projection (20) — only those parameters are not in general the same ones Cagan took them to be. Comparison of (20) with (21) shows that Cagan's estimator of

λ is consistent but that his estimator of α is not in general consistent, and will obey

$$\text{plim}\,\hat{\alpha}(1 - \hat{\lambda}) = [-\lambda + \rho(\lambda + \alpha(1 - \lambda))]$$

which implies that

$$\text{(22)} \qquad \text{plim}\ \hat{\alpha} = \rho\alpha + (1 - \rho)\left(\frac{-\lambda}{1 - \lambda}\right).$$

Notice that if $\rho = 0$, which will be true if $\varepsilon_t = 0$ for all t, (22) implies

$$\text{plim}\ \hat{\alpha} = \frac{-\lambda}{1 - \lambda},$$

which is an expression that Wallace and I derived and used. On the other hand, if $\eta_t = 0$ for all t, so that there is no noise in the portfolio balance schedule, from (15) $\rho = 1$, which with (22) implies

$$\text{plim}\,\hat{\alpha} = \alpha,$$

so that in this special case Cagan's estimator of α is consistent (and furthermore unbiased as it turns out, since $v_t = 0$ for all t).

On the special assumption $\sigma_{\varepsilon\eta} = 0$, we have

$$\rho = \frac{E(\varepsilon_t \cdot (\varepsilon_t - \eta_t))}{E(\varepsilon_t - \eta_t)^2} = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\eta^2}.$$

Alternatively, multiplying (15) by η_t, taking expectations, and rearranging gives

$$\rho = \frac{\sigma_{v\eta}}{\sigma_\eta^2},$$

so that ρ is the regression coefficient of the residual v in the regression of $(\mu - x)$ against current and past $(1 - L)x$ on the disturbance in the demand for money. If $v = \eta$, then $\rho = 1$.

An estimate of ρ could be obtained in the following way, again on the special assumption $\sigma_{\varepsilon\eta} = 0$. Multiplying (15) by ε_t, taking expectations and rearranging gives

$$\rho = 1 - \frac{\sigma_{v\varepsilon}}{\sigma_\varepsilon^2}.$$

The magnitude $\sigma_{v\varepsilon}/\sigma_\varepsilon^2$ is the regression coefficient of v on ε. The residual v can be estimated by the residual in (the first difference of) Cagan's equation. The variable ε_t can be extracted using the methods described below in Section 5. Then an estimate of ρ could be prepared using the above equation. It would be possible to use that estimate of ρ to correct Cagan's estimate of α by applying the formula

$$\text{plim}\ \hat{\alpha} = \rho\alpha + (1 - \rho)\left(\frac{-\lambda}{1 - \lambda}\right).$$

The calculations in this section provide a useful exercise in interpreting systems in which one variable (x) is econometrically exogenous with respect to (is not Granger-caused by) another variable (μ). In such a system, as Sims's Theorem 2 assures us and as the preceding calculations verify, the regression of μ on past, present, and future x's is one-sided on the present and past. Thus, there exist representations (models) of the (μ, x) process in which μ and $\mu-x$ are each one-sided linear functions of past and present x's with disturbances that are orthogonal to past, present, and future x's — so that in these relations x is strictly exogenous with respect to μ and $\mu-x$, respectively. But the representation in which x is econometrically exogenous with respect to $(\mu-x)$ — which is the relation that can be consistently estimated by least squares or generalized least squares — is *not* the demand function for money, which is the structural relation we are interested in estimating. The reason is that in the structural relation (21), u_t is not in general orthogonal to the x process. The upshot is that finding that x is exogenous with respect to $\mu-x$ does not guarantee that the one-sided $\mu-x$ on x distributed lag regression which is estimable by single equation methods corresponds to the structural relation that we're interested in.

4. A CONSISTENT ESTIMATOR

Equations (11) and (12) form a bivariate first-order moving average process in $(1-L)\mu_t$ and $(1-L)x_t$. Assuming that the white noises ε_t and η_t are jointly normally distributed, the likelihood function of a sample of length T observations, $t=1,\ldots, T$, generated by (11)–(12) can be written down. To apply the method of maximum likelihood, it is most convenient to write the model in its vector autoregressive form. First note that from (9) we can write

$$\varepsilon_t = \mu_t - \frac{1-\lambda}{1-\lambda L} x_t . \tag{23}$$

Next from (11) we have

$$\varepsilon_t - \eta_t = \frac{(\lambda + \alpha(1-\lambda))(1-L)}{1-\lambda L} x_t \tag{24}$$

Substituting (24) into (23) and rearranging gives

$$\eta_t = \mu_t - \left(\frac{1-\lambda + (\lambda + \alpha(1-\lambda))(1-L)}{1-\lambda L}\right) x_t . \tag{25}$$

In vector notation equations (23) and (25) can be written

$$\begin{bmatrix} \varepsilon_t \\ \\ \eta_t \end{bmatrix} = \begin{bmatrix} -\dfrac{(1-\lambda)}{1-\lambda L} & 1 \\ \\ -\dfrac{[1-\lambda + (\lambda + \alpha(1-\lambda))(1-L)]}{1-\lambda L} & 1 \end{bmatrix} \begin{bmatrix} x_t \\ \\ \mu_t \end{bmatrix} .$$

Multiplying both sides of the equation by $(1-\lambda L)\cdot I$ where I is the $2x2$ identity matrix, gives

$$\begin{bmatrix}(1-\lambda L)\varepsilon_t\\(1-\lambda L)\eta_t\end{bmatrix}=\begin{bmatrix}-(1-\lambda) & 1-\lambda L\\-[1-\lambda+(\lambda+\alpha(1-\lambda))(1-L)] & 1-\lambda L\end{bmatrix}\begin{bmatrix}x_t\\\mu_t\end{bmatrix}$$

or

$$\begin{bmatrix}\varepsilon_t\\\eta_t\end{bmatrix}-\lambda I\begin{bmatrix}\varepsilon_{t-1}\\\eta_{t-1}\end{bmatrix}=\begin{bmatrix}-(1-\lambda) & 1\\-(1+\alpha(1-\lambda)) & 1\end{bmatrix}\begin{bmatrix}x_t\\\mu_t\end{bmatrix}+\begin{bmatrix}0 & -\lambda\\\lambda+\alpha(1-\lambda) & -\lambda\end{bmatrix}\begin{bmatrix}x_{t-1}\\\mu_{t-1}\end{bmatrix}.$$

Let

$$G_0=\begin{bmatrix}-(1-\lambda) & 1\\-(1+\alpha(1-\lambda)) & 1\end{bmatrix}.$$

Premultiplying the preceding equation by

$$G_0^{-1}=\begin{bmatrix}1 & -1\\1+\alpha(1-\lambda) & -(1-\lambda)\end{bmatrix}/(\lambda+\alpha(1-\lambda))$$

gives

$$G_0^{-1}\begin{bmatrix}\varepsilon_t\\\eta_t\end{bmatrix}-\lambda IG_0^{-1}\begin{bmatrix}\varepsilon_{t-1}\\\eta_{t-1}\end{bmatrix}=\begin{bmatrix}x_t\\\mu_t\end{bmatrix}+G_0^{-1}\begin{bmatrix}0 & -\lambda\\\lambda+\alpha(1-\lambda) & -\lambda\end{bmatrix}\begin{bmatrix}x_{t-1}\\\mu_{t-1}\end{bmatrix}$$

or

(26) $$\begin{bmatrix}a_{1t}\\a_{2t}\end{bmatrix}-\lambda I\begin{bmatrix}a_{1t-1}\\a_{2t-1}\end{bmatrix}=\begin{bmatrix}x_t\\\mu_t\end{bmatrix}+G_0^{-1}\begin{bmatrix}0 & -\lambda\\\lambda+\alpha(1-\lambda) & -\lambda\end{bmatrix}\begin{bmatrix}x_{t-1}\\\mu_{t-1}\end{bmatrix}$$

where

$$\begin{bmatrix}a_{1t}\\a_{2t}\end{bmatrix}\equiv G_0^{-1}\begin{bmatrix}\varepsilon_t\\\eta_t\end{bmatrix}.$$

Computing

$$G_0^{-1}\begin{bmatrix}0 & -\lambda\\\lambda+\alpha(1-\lambda) & -\lambda\end{bmatrix}$$

explicitly and rearranging the above equation gives

(27) $$\begin{bmatrix}x_t\\\mu_t\end{bmatrix}=\begin{pmatrix}1 & 0\\(1-\lambda) & \lambda\end{pmatrix}\begin{pmatrix}x_{t-1}\\\mu_{t-1}\end{pmatrix}+\begin{bmatrix}a_{1t}\\a_{2t}\end{bmatrix}-\lambda I\begin{bmatrix}a_{1t-1}\\a_{2t-1}\end{bmatrix}.$$

Equation (27) is a vector first order autoregression, first-order moving average process. The random variables a_{1t}, a_{2t} are the innovations in the x and μ processes, respectively. They are the one period-ahead forecasting errors for x_t and μ_t, respectively. The a's are related to the ε's and η's appearing in the structural equations of the model by

$$\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} = G_0^{-1} \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}$$

(28)
$$\begin{bmatrix} a_{1t} \\ \\ a_{2t} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\lambda + \alpha(1-\lambda)}(\varepsilon_t - \eta_t) \\ \\ \dfrac{1-\lambda}{\lambda + \alpha(1-\lambda)}(\varepsilon_t - \eta_t) + \varepsilon_t \end{bmatrix}.$$

Notice that the first equation of (27) can be written as

$$(1 - L)x_t = (1 - \lambda L)a_{1t}.$$

It is straightforward to write this in the autoregressive form

(29)
$$x_t = \left(\frac{1-\lambda}{1-\lambda L}\right)x_{t-1} + a_{1t}.$$

Since $E_{t-1}a_{1t}=0$, we have

$$E_{t-1}x_t = \left(\frac{1-\lambda}{1-\lambda L}\right)x_{t-1}.$$

The second equation of (27) can be written as

$$(1 - \lambda L)\mu_t = (1 - \lambda)x_{t-1} + (1 - \lambda L)a_{2t}$$

But from (29) we have $(1-\lambda)x_{t-1}=(1-\lambda L)x_t-(1-\lambda L)a_{1t}$, which when substituted into the above equation gives

$$(1 - \lambda L)\mu_t = (1 - \lambda L)x_t - (1 - \lambda L)a_{1t} + (1 - \lambda L)a_{2t}$$

or

(30)
$$\mu_t = x_t + a_{2t} - a_{1t}.$$

From (30), it follows that

(31)
$$E_{t-1}\mu_t = E_{t-1}x_t.$$

The triangular character of representation (27) demonstrates that μ does not "cause" in Granger's sense (i.e., help predict, once lagged own values are taken into account) the variable x. That is, x is econometrically exogenous with respect to μ.[7] On the other hand, x_t does cause the variable μ_t. Even stronger, the model implies that $E_{t-1}\mu_t=E_{t-1}x_t=\dfrac{(1-\lambda)}{(1-\lambda L)}x_{t-1}$ so that lagged μ's don't

[7] Sims [9] proved the equivalence of Granger causality with econometric exogeneity.

help predict μ once lagged x's are taken into account.[8] That x causes μ in Granger's sense is not to be confused with x's "leading" μ in any National Bureau sense. On the contrary, according to (30), x_t and μ_t are "in phase" with one another, neither one leading the other. (The phase of their cross-spectrum equals zero at all frequencies.) Evidence that x leads μ would not be consistent with the model being studied here.

The vector autoregressive, moving average process (27) is in a form that can be estimated by the maximum likelihood estimator described by Wilson **[10]**. It is essential that the matrices multiplying current $\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$ and current $\begin{bmatrix} x_t \\ \mu_t \end{bmatrix}$ both be identity matrices in order to apply the method, so that each a_i process can be interpreted as the residual from a vector autoregression either for μ_t or x_t. This is by way of getting things in a form in which the likelihood function of $\begin{bmatrix} x_t \\ \mu_t \end{bmatrix}$ equals the likelihood function of $\begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$.

Let

$$a_t = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix},$$

and let D_a be the covariance matrix of a_t,

$$D_a = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = Ea_t a_t'.$$

The likelihood function of the sample $t = 1, \ldots, T$ can now be written as

$$\text{(32)} \quad L(\lambda, \sigma_{11}, \sigma_{12}, \sigma_{22} \mid \mu_t, x_t) = (2\pi^{-T}) \mid D_a \mid^{-T/2} \exp\left(-\frac{1}{2} \sum_{t=1}^{T} a_t' D_a^{-1} a_t \right)$$

Given initial values for (a_{10}, a_{20}) or equivalently for (ε_0, η_0), and given a value of λ, equation (26) or (27) can be used to solve for a_t, $t = 1, \ldots, T$. (I will take $a_{10} = a_{20} = 0$.)

[8] Wallace and I were mistaken when we asserted that "the system is one in which expectations of money creation could equally well be formed as a distributed lag of past rates of money creation," **[8**, (337)]. It is true that

$$E[\mu_t \mid \mu_{t-1}, \ldots] = \frac{1-\gamma}{1-\gamma L} \mu_{t-1},$$

where γ is a parameter that depends on the ratio of the variance of ε_t to the variance of η_t. However, in the model $E[\mu_t \mid \mu_{t-1}, \ldots] \neq E[\mu_t \mid \mu_{t-1}, \ldots, x_{t-1}, \ldots]$. Instead, $E[\mu_t \mid \mu_{t-1}, \ldots, x_{t-1}, \ldots] = E[\mu_t \mid x_{t-1}, \ldots]$, which, of course, has a smaller prediction error variance than $E[\mu_t \mid \mu_{t-1}, \ldots]$. The erroneous statements on page 337 of Sargent and Wallace **[8]** amount to an assertion that the Wold representation of the $x_t - \mu_t$ process contains only one noise, so that lagged values of *either* x *or* μ exhaust all information in the past values of x *and* μ useful for predicting *either* x *or* μ. That is wrong, as the triangular Wold representation derived in Section 3 of this paper verifies. The upshot of all this is that it was not necessary for Sargent and Wallace to posit measurement errors in the money supply to rationalize the empirical observation that x causes μ. That is already an implication of the system free of measurement errors.

Wilson notes that maximizing (32) is equivalent with minimizing with respect to λ the determinant of the estimated covariance matrix of the a_t's,

$$(33)\qquad |\hat{D}_a| \equiv |T^{-1}\sum_{t=1}^{T} \hat{a}_t\hat{a}_t'|$$

where the $\hat{a}_t's$ are determined by solving (27) recursively and so depend on λ. The covariance matrix of the $a_t's$ is estimated as

$$\hat{D}_a = T^{-1}\sum_{t=1}^{T} \hat{a}_t\hat{a}_t'$$

evaluated at the value of λ that minimizes (33). The resulting estimates are known to be statistically consistent (see Wilson [**10**]).

Notice that α does not appear explicitly in the likelihood function, but only indirectly by way of the elements of D_a, namely, σ_{11}, σ_{12}, and σ_{22}. That this must be so can be seen by inspecting representation (27), in which λ appears explicitly but α does not. On the basis of the *four* parameters λ, σ_{11}, σ_{12}, and σ_{22} that are identified by (27), i.e., that characterize the likelihood function (32), we can think of attempting to estimate the *five* parameters of the model α, λ, σ_ε^2, σ_η^2, and $\sigma_{\varepsilon\eta}$. Not surprisingly, some of the parameters are underidentified. In particular, while λ and σ_ε^2 are identified, α, σ_η^2, and $\sigma_{\varepsilon\eta}$ are not separately identified. To see that α and $\sigma_{\varepsilon\eta}$ are not identified consider the following argument. From equation (28), we know that the identifiable parameters σ_{11}, σ_{12}, and σ_{22} are related to the structural parameters σ_ε^2, σ_η^2, $\sigma_{\varepsilon\eta}$, α, and λ by

$$(34)\qquad \sigma_{11} = \left(\frac{1}{\lambda + \alpha(1-\lambda)}\right)^2(\sigma_\varepsilon^2 + \sigma_\eta^2 - 2\sigma_{\varepsilon\eta})$$

$$(35)\qquad \sigma_{12} = \frac{(1-\lambda)}{(\lambda + \alpha(1-\lambda))^2}(\sigma_\varepsilon^2 + \sigma_\eta^2 - 2\sigma_{\varepsilon\eta}) + \frac{1}{\lambda + \alpha(1-\lambda)}(\sigma_\varepsilon^2 - \sigma_{\varepsilon\eta})$$

$$(36)\qquad \sigma_{22} = \left(\frac{(1-\lambda)}{(\lambda + \alpha(1-\lambda))}\right)^2(\sigma_\varepsilon^2 + \sigma_\eta^2 - 2\sigma_{\varepsilon\eta}) + \sigma_\varepsilon^2 + 2\left(\frac{1-\lambda}{\lambda + \alpha(1-\lambda)}\right)(\sigma_\varepsilon^2 - \sigma_{\varepsilon\eta}).$$

These equations imply

$$(37)\qquad \sigma_{12} = (1-\lambda)\sigma_{11} + \frac{1}{\lambda + \alpha(1-\lambda)}(\sigma_\varepsilon^2 - \sigma_{\varepsilon\eta})$$

$$(38)\qquad \sigma_{22} = (1-\lambda)^2\sigma_{11} + \sigma_\varepsilon^2 + \frac{2(1-\lambda)}{\lambda + \alpha(1-\lambda)}(\sigma_\varepsilon^2 - \sigma_{\varepsilon\eta}).$$

Do there exist offsetting changes in α and $\sigma_{\varepsilon\eta}$ that leave both of these equations satisfied with σ_{11}, σ_{22}, and σ_{12} unchanged? That is, holding λ and σ_ε^2 constant, can we change α and $\sigma_{\varepsilon\eta}$ in offsetting ways that leave σ_{11}, σ_{12}, and σ_{22} constant? The answer is yes, as can be seen by differentiating (37) and (38) and setting $d\sigma_{12} = d\sigma_{11} = d\sigma_{22} = d\lambda = d\sigma_\varepsilon^2 = 0$:

$$0 = (1-\lambda)(\lambda+\alpha(1-\lambda))^{-2}(\sigma_\varepsilon^2 - \sigma_{\varepsilon\eta})d\alpha + (\lambda+\alpha(1-\lambda))^{-1}d\sigma_{\varepsilon\eta} = 0 \tag{39}$$

$$0 = 2(1-\lambda)^2(\lambda+\alpha(1-\lambda))^{-2}(\sigma_\varepsilon^2 - \sigma_{\varepsilon\eta})d\alpha + 2(1-\lambda)(\lambda+\alpha(1-\lambda))^{-1}d\sigma_{\varepsilon\eta} = 0. \tag{40}$$

Dividing the second equation by $2(1-\lambda)$ gives the first equation, which proves that if $d\alpha$ and $d\sigma_{\varepsilon\eta}$ obey equation (39), both equations (37) and (38) will remain satisfied. Thus, there exist offsetting changes in α and $\sigma_{\varepsilon\eta}$ that leave the identifiable parameters λ, σ_{11}, σ_{12}, and σ_{22} unaltered. It follows that $\sigma_{\varepsilon\eta}$ and α are not separately identifiable. It is evident from (27) or (32) that λ is identified. To see that σ_ε^2 is identifiable, simply recall that ε_t obeys the feedback rule

$$\mu_t = \frac{1-\lambda}{1-\lambda L}x_t + \varepsilon_t, \tag{9}$$

so that given λ, and samples of μ_t and x_t, σ_ε^2 is identifiable as the variance of the residual in the above equation.

To proceed to extract estimates of α it is necessary to impose a value of $\sigma_{\varepsilon\eta}$. I propose to impose the condition $\sigma_{\varepsilon\eta}=0$, so that shocks to the money supply rule and shocks to portfolio balance are uncorrelated. It is straightforward to calculate

$$\begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & \sigma_\eta^2 \end{bmatrix} = E\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}[\varepsilon_t\ \eta_t] = G_0 D_a G_0'$$

$$= \begin{bmatrix} (1-\lambda)^2\sigma_{11} + 2(1-\lambda)\sigma_{12} + \sigma_{22}, & (1-\lambda)(1+\alpha(1-\lambda))\sigma_{11} - (2-\lambda+\alpha(1-\lambda))\sigma_{12} + \sigma_{22} \\ (1-\lambda)(1+\alpha(1-\lambda))\sigma_{11} - (2-\lambda+\alpha(1-\lambda))\sigma_{11} + \sigma_{22}, & (1+\alpha(1-\lambda))^2\sigma_{11} - 2(1+\alpha(1-\lambda))\sigma_{12} + \sigma_{22} \end{bmatrix}.$$

Imposing $\sigma_{\varepsilon\eta}=0$, we have

$$0 = \sigma_{\varepsilon\eta} = (1-\lambda)(1+\alpha(1-\lambda))\sigma_{11} - (2-\lambda+\alpha(1-\lambda))\sigma_{12} + \sigma_{22},$$

which implies that α is to be estimated by

$$\alpha = \frac{-\sigma_{11}}{(1-\lambda)\sigma_{11} - \sigma_{12}} + \frac{(2-\lambda)\sigma_{12}}{(1-\lambda)((1-\lambda)\sigma_{11} - \sigma_{12})} - \frac{\sigma_{22}}{(1-\lambda)((1-\lambda)\sigma_{11} - \sigma_{12})}. \tag{41}$$

Let this estimator of α be

$$\hat{\alpha} = g(\lambda, \sigma_{11}, \sigma_{12}, \sigma_{22}) = g(\theta)$$

where $\theta=(\lambda, \sigma_{11}, \sigma_{12}, \sigma_{22})$. Let Σ_θ be the estimated asymptotic covariance matrix of θ. Then the asymptotic variance of $\hat{\alpha}$ will be estimated as

$$\operatorname{var} \hat{\alpha} = \left(\frac{\partial g}{\partial \theta}\right)_{\hat{\theta}} \Sigma_\theta \left(\frac{\partial g}{\partial \theta}\right)'_{\hat{\theta}}$$

where $(\partial g/\partial \theta)_{\hat{\theta}}$ is the (1×4) vector of partial derivatives of g with respect to θ evaluated at the maximum likelihood estimates $\hat{\theta}$. The asymptotic covariance matrix of $(\lambda, \sigma_{11}, \sigma_{12}, \sigma_{22})$ is given by

$$\Sigma_\theta = \frac{1}{T}\begin{bmatrix} T\sigma_\lambda^2 & 0 & 0 & 0 \\ 0 & 2\sigma_{11}^2 & 2\sigma_{11}\sigma_{12} & 2\sigma_{12}^2 \\ 0 & 2\sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22}+\sigma_{12}^2 & 2\sigma_{12}\sigma_{22} \\ 0 & 2\sigma_{12}^2 & 2\sigma_{12}\sigma_{22} & 2\sigma_{22}^2 \end{bmatrix}$$

where $T\sigma_\lambda^2$ is estimated by

$$T\sigma_\lambda^2 = \left[-\frac{\partial^2 \log L}{\partial \lambda^2}\right]_{\hat{\theta}}^{-1}$$

and where $\log L$ is the natural logarithm of the likelihood function (32). Notice that the maximum likelihood estimate of λ is asymptotically orthogonal to the estimates $\sigma_{11}, \sigma_{12}, \sigma_{22}$. The preceding formula for Σ_θ can be derived by applying results of Wilson [**10**] and Anderson [**1**, (159–161)]. In the computations summarized below, the components σ_{11}, σ_{12}, and σ_{22} were estimated by

$$\hat{D}_a = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_{22} \end{pmatrix} = T^{-1} \sum_{t=1}^{T} \hat{a}_t \hat{a}_t',$$

the maximum likelihood estimator. The term

$$\left(-\frac{\partial^2 \log L}{\partial \lambda^2}\right)_{\hat{\theta}}$$

was estimated numerically in the course of minimizing (33) to obtain the maximum likelihood estimates.

It bears emphasizing that α is identifiable at all only on the basis of a restriction on $\sigma_{\varepsilon\eta}$, and that the estimator of α obtained by imposing $\sigma_{\varepsilon\eta}=0$ depends sensitively on the covariance matrix of the errors in forecasting x_t and μ_t from the past. The estimates of α thereby obtained ought to be regarded as very delicate.

5. AN ALTERNATIVE ESTIMATOR

If it is assumed that $\sigma_{\varepsilon\eta}=0$, so that shocks to the demand for money and to the supply of money are uncorrelated, an instrumental variable estimator is available. From equations (30) and (29) we have that

$$E_{t-1}x_t = E_{t-1}\mu_t = \frac{1-\lambda}{1-\lambda L}x_{t-1},$$

and that

(42) $$x_t - E_{t-1}x_t = a_{1t} = \frac{1}{\lambda + \alpha(1-\lambda)}(\varepsilon_t - \eta_t)$$

(43) $$\mu_t - E_{t-1}x_t = a_{2t} = \frac{1-\lambda}{\lambda + \alpha(1-\lambda)}(\varepsilon_t - \eta_t) + \varepsilon_t.$$

Notice that

(44) $$a_{2t} - (1-\lambda)a_{1t} = \varepsilon_t.$$

This suggests the following procedure. Estimate by maximum likelihood the univariate first-order moving average process for Δx_t, i.e.,

$$(1-L)x_t = (1-\lambda L)a_{1t}$$

where $a_{1t} = (\lambda + \alpha(1-\lambda))^{-1}(\varepsilon_t - \eta_t)$ is "white." This will yield consistent estimates of λ and permit estimating the forecast errors. The forecasts $E_{t-1}x_t$ can be estimated from the above equation as

$$E_{t-1}x_t = x_{t-1} - \lambda a_{1t-1}.$$

Use of (44) shows that estimates of ε_t can be extracted according to

(45) $$\varepsilon_t = (\mu_t - E_{t-1}x_t) - (1-\lambda)(x_t - E_{t-1}x_t).$$

On the assumption that ε_t is uncorrelated with η_t, ε_t is a valid instrument for estimating equation (1): it is correlated with the regressors but orthogonal to the disturbance. Letting $\tilde{\varepsilon}_t$ be the estimates of ε_t obtained by applying (45), I propose fitting the first-stage regression

$$\hat{x}_t = \sum_{i=0}^{n} \hat{w}_i \tilde{\varepsilon}_{t-1}$$

where the hatted values denote least squares estimates. Then Cagan's equation (1) would be estimated by applying (nonlinear) least squares to the second-stage regression

(46) $$m_t - p_t = \alpha\left(\frac{1-\lambda}{1-\lambda L}\right)\hat{x}_t + u_t - \frac{\alpha(1-\lambda)}{1-\lambda L}(\hat{x}_t - x_t).$$

This procedure yields consistent estimates of α and λ on the assumption that

$$\operatorname*{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} \tilde{\varepsilon}_t\eta_t = 0,$$

a condition that the orthogonality of ε_t and η_t goes a long way toward delivering.

6. TESTING THE MODEL

Representation (27) shows that the model is a special case of the general vector first-order moving average, first order autoregressive process

$$(27') \quad \begin{bmatrix} x_t \\ \mu_t \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{pmatrix} x_{t-1} \\ \mu_{t-1} \end{pmatrix} + \begin{pmatrix} a_{1t} \\ a_{2t} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} a_{1t-1} \\ a_{2t-1} \end{pmatrix}$$

where in (27) *seven* linear restrictions have been placed on the eight parameters ($c_{11}, c_{12}, c_{21}, c_{22}, b_{11}, b_{12}, b_{21}, b_{22}$) of (27′) so that the systematic part of (27) only involves the single parameter λ. The model (27) can be tested by relaxing some subset of the seven restrictions that were imposed on (27′) to get (27), maximizing the likelihood function under the less restrictive parameterization, and calculating the pertinent χ^2 statistic. Let $L(x_t, \mu_t; \theta_0)$ be the maximum of the likelihood function under parameterization (27), which is Cagan's model under rational expectations. Let $L(x_t, \mu_t; \theta, q)$ be the maximum of the likelihood function under (27′) with q of the seven restrictions in (27) being relaxed. Then

$$-2\log\left(\frac{L(x_t, \mu_t; \theta_0)}{L(x_t, \mu_t; \theta, q)}\right)$$

is asymptotically distributed as $\chi^2(q)$. High values of the test statistic lead to rejection of representation (27). Below this test is implemented under several alternative relaxations of the restrictions on (27).

7. EMPIRICAL RESULTS

For Cagan's and Barro's data, respectively, Tables 2 and 3 report the estimates obtained using the maximum likelihood estimator and the assumption that $\sigma_{\varepsilon\eta}$

TABLE 2
ESTIMATES FOR CAGAN'S DATA (STANDARD ERRORS IN PARENTHESES)
(x AND μ ARE DEVIATIONS FROM RESPECTIVE MEANS)

Country	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\sigma}_{11}$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$
GERMANY	.6774	−5.973	.0625	.0158	.0091
Oct '20–July '23	(.0533)	(4.615)	(.0147)	(.0048)	(.0022)
AUSTRIA	.7537	−.3113	.0385	.0148	.0085
Feb '21–Aug '22	(.0589)	(1.5695)	(.0119)	(.0051)	(.0026)
GREECE	.4587	−4.086	.0675	.0245	.0279
Feb '43–Aug '44	(.0884)	(2.970)	(.0208)	(.0109)	(.0086)
HUNGARY I	.4183	−1.841	.0362	.0089	.0060
Aug '22–Feb '24	(.0668)	(.3978)	(.0112)	(.0038)	(.0019)
RUSSIA	.6259	−9.745	.0524	.0138	.0205
Feb '22–Jan '24	(.0728)	(10.742)	(.0145)	(.0070)	(.0057)
POLAND	.5364	−2.529	.0566	.0149	.0089
May '22–Nov '23	(.0722)	(.8562)	(.0175)	(.0059)	(.0027)

TABLE 3

ESTIMATES FOR BARRO'S DATA (STANDARD ERRORS IN PARENTHESES)
(x AND μ ARE DEVIATIONS FROM RESPECTIVE MEANS)

County	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\sigma}_{11}$	$\hat{\sigma}_{12}$	$\hat{\sigma}_{22}$
AUSTRIA	.6373	−3.979	.0584	.0161	.0081
Apr '21–Dec '22	(.0739)	(2.805)	(.0172)	(.0056)	(.0024)
GERMANY	.5921	−2.344	.1806	.0653	.0263
Feb '21–Aug '23	(.0510)	(1.223)	(.0445)	(.0165)	(.0065)
HUNGARY	.4323	−1.705	.0280	.0071	.0038
Nov '21–Feb '24	(.0559)	(.2782)	(.0072)	(.0023)	(.0010)
POLAND	.4790	−2.043	.0319	.0063	.0040
Feb '22–Jan '24	(.0533)	(.3537)	(.0089)	(.0025)	(.0011)

TABLE 4

CAGAN'S ESTIMATES OF α, λ TOGETHER WITH CONFIDENCE BAND FOR α.

Country	$\hat{\lambda}$	$\hat{\alpha}$	(α_e, α_u)
AUSTRIA	.95	−8.55	−(4.43, 30.0)
Jan '21–Aug '22			
GERMANY	.82	−5.46	−(5.05, 6.13)
Sept '20–July '23			
GREECE	.86	−4.09	−(2.83, 32.5+)*
Jan '43–Aug '44			
HUNGARY	.90	−8.70	−(6.36, 42.2+)*
July '22–Feb '24			
HUNGARY	.86	−3.63	−(2.55, 4.73)
July '45–Feb '46			
POLAND	.74	−2.30	−(1.74, 3.94)
Apr '22–Nov '23			
RUSSIA	.70	−3.06	−(2.66, 3.76)
Dec '21–Jan '24			

(α_e, α_u) = 90 percent confidence band calculated by Cagan using likelihood ratio method.

*α_u actually exceeds right-hand figure in parentheses.

Source: Cagan's Table 3, [**3**, (43)].

TABLE 5

BARRO'S ESTIMATES OF λ AND α

Country	$\hat{\lambda}$	$\hat{\alpha}$*
AUSTRIA	.829	−4.09
Jan '21–Dec '22		(−3.6, −4.5)
GERMANY	.824	−3.79
Jan '21–Aug '23		(−3.3, −4.3)
HUNGARY	.861	−5.53
Oct '21–Feb '24		(−4.6, −6.9)
POLAND	.709	−2.56
Jan '22–Jan '24		(−2.1, −3.3)

* 95 percent confidence intervals in parentheses beneath each estimate.

Source: Barro, Table 3.

$=0$. Asymptotic standard errors are in parentheses beneath each estimator. Cagan's and Barro's estimates are reported in Tables 4 and 5 for convenience. For Cagan's data, the maximum likelihood estimator recovers estimates of α that are in most cases characterized by large standard errors. In particular, for the important German case, a case in which Cagan had apparently estimated α with a tight confidence band, my estimate of α has a big standard error, one nearly as big as the point estimate itself. Evidently, the estimate of α is not statistically significantly different from zero even at modest confidence levels, at least if we are willing to use the asymptotic (normal) distribution of the estimates.[9] For the Austrian and Russian cases, my estimate of α is smaller than its standard error. Only in the case of Hungary I, and to a lesser extent in the case of Poland, is the standard error of α small relative to the point estimate of α. Interestingly enough, for Hungary I my estimate of α of -1.84 is much smaller in absolute value than Cagan's estimate of -8.70. The reciprocal of $+1.84$ is .54, while the average monthly rate of inflation in the Hungary I case was .46. In the case of Hungary I, my estimate of α suggests that the paradox with which I began this paper, the apparent tendency of creators of money to print money "too fast", was not present. For what it is worth, then, my estimate of α for Hungary I tends to explain away the paradox. For the other countries, the point estimates do not explain away the paradox. However, in each case, values of α that would cause the paradox to disappear do exist within confidence intervals of two standard errors on each side of the point estimate of α. This suggests that perhaps the paradox ought not to be taken as having been seriously confirmed since the estimates of α on which it is based seem so shaky.

Notice that my estimates of λ are always lower than Cagan's. That is an unexpected result, since according to the model, Cagan's estimate of λ and my maximum likelihood estimator are each consistent. The systematic difference in estimates as between the two estimators may reflect the inadequacy of the model.

For Barro's data, the maximum likelihood estimates are reported in Table 3. For Austria and Germany, the estimated asymptotic standard errors of α are fairly large relative to the point estimates, while for Hungary I and Poland they are much smaller. As with Cagan's data, my estimate of α is much smaller than is Barro's for Hungary I. My estimate is somewhat smaller than Barro's for Poland. As with Cagan's data, my estimate of λ is always smaller than Cagan's.

The main conclusion that I draw from these estimates is that even under the restriction $\sigma_{\varepsilon\eta}=0$, the slope parameter α is usually poorly estimated. When to this is added the observations that α is not even identifiable unless $\sigma_{\varepsilon\eta}$ is restricted and that economics does not seem to restrict $\sigma_{\varepsilon\eta}$, the uncertainty about α only increases. It seems correct to conclude that, with the possible exception of Hungary I, I have not been able to estimate very well the slope of the portfolio balance schedule.

This is not to say, however, that the model is necessarily defective. It is cer-

[9] Actually, the normality of the asymptotic distribution is conjectural. See Porter, [7].

tainly conceivable that the model approximated reality quite well even though α cannot be estimated well or isn't even identifiable. As pointed out in Section 6, the proper way to test the model is to "overfit" the vector moving average, autoregressive representation (27), and to test whether the restrictions imposed by (27) are violated. For overfitting, I have estimated each of the six parameterizations reported in Table 6. For each parameterization, the chi-square statistic described in Section 6 was computed, and is reported in Table 7 for Cagan's data and in Table 8 for Barro's data. High values of the χ^2 statistic lead to rejection of the null hypothesis that model (27) is adequate.

For Cagan's data, at the .95 confidence level, the model is rejected relative to parameterization 5 for Russia, relative to parameterizations 1, 2, 4, and 5 for Hungary I, and relative to parameterizations 2, 5, and 6 for Austria. For Germany, Greece, and Poland, the model is not rejected relative to any of the six parameterizations at the .95 confidence level. For three of the hyperinflations, then, overfitting representation (27) does turn up evidence that would prompt rejection of the model. However, it surprised me just how adequately the model does seem to perform relative to the six parameterizations in Table 6. Representation (27) is a very stark, highly restricted parameterization; indeed, the systematic part of the vector autoregression has only one parameter. I had expected the model to be rather decisively rejected by these overfitting tests. It is remarkable that the model seems to survive those tests for even three of the hyperinflations.

TABLE 6

PARAMETERIZATIONS FOR OVERFITTING

$$\begin{bmatrix} x_t \\ \mu_t \end{bmatrix} = C \begin{bmatrix} x_{t-1} \\ \mu_{t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} + B \begin{bmatrix} a_{1t-1} \\ a_{2t-1} \end{bmatrix}$$

1. $$C = \begin{bmatrix} c & 0 \\ (1-\lambda) & \lambda \end{bmatrix}, \quad B = \begin{bmatrix} -\lambda & 0 \\ 0 & -\lambda \end{bmatrix}$$

2. $$C = \begin{bmatrix} 1 & c \\ 1-\lambda & \lambda \end{bmatrix}, \quad B = \begin{bmatrix} -\lambda & 0 \\ 0 & -\lambda \end{bmatrix}$$

3. $$C = \begin{bmatrix} 1 & 0 \\ 1-\lambda & \lambda \end{bmatrix}, \quad B = \begin{bmatrix} -\lambda & b \\ 0 & -\lambda \end{bmatrix}$$

4. $$C = \begin{bmatrix} c_1 & c_1 \\ 1-\lambda & \lambda \end{bmatrix}, \quad B = \begin{bmatrix} -\lambda & 0 \\ 0 & -\lambda \end{bmatrix}$$

5. $$C = \begin{bmatrix} 1 & c \\ (1-\lambda) & \lambda \end{bmatrix}, \quad B = \begin{bmatrix} -\lambda & b \\ 0 & -\lambda \end{bmatrix}$$

6. $$C = \begin{bmatrix} 1 & 0 \\ c_1 & c_2 \end{bmatrix}, \quad B = \begin{bmatrix} -\lambda & 0 \\ 0 & -\lambda \end{bmatrix}$$

TABLE 7

CAGAN'S DATA RESULTS OF OVERFITTING — CHI SQUARE STATISTICS

Country	Parameterization Number					
	1	2	3	4	5	6
	$\chi^2(1)$	$\chi^2(1)$	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(2)$	$\chi^2(2)$
GERMANY Oct '20–July '23	.52	1.12	2.06	.95	3.37	2.14
RUSSIA Feb '22–Jan '24	.21	3.05	2.84	3.90	7.79	.97
GREECE Feb '43–Aug '44	1.04	1.53	.25	4.14	1.87	.40
HUNGARY I Aug '22–Feb '24	4.13	7.57	3.13	7.57	7.62	.24
POLAND May '22–Nov '23	.19	.04	.22	.31	.56	.53
AUSTRIA Feb '21–Aug '22	2.77	4.97	.63	4.97	10.05	7.13

Significance Levels: $\chi^2(1)_{.05}=3.84$ $\chi^2(2)_{.05}=5.99$
$\chi^2(1)_{.01}=6.63$ $\chi^2(2)_{.01}=9.21$

TABLE 8

BARRO'S DATA RESULTS OF OVERFITTING — CHI SQUARE STATISTICS

Country	Parameterization Number					
	1	2	3	4	5	6
	$\chi^2(1)$	$\chi^2(1)$	$\chi^2(1)$	$\chi^2(2)$	$\chi^2(2)$	$\chi^2(2)$
GERMANY Feb '21–Aug '23	1.272	.382	.3	3.5	.33	$\simeq 0.$
HUNGARY I Nov '21–Feb '24	5.424	7.6	1.232	7.63	8.49	.39
POLAND Feb '22–Jan '24	1.58	.528	.184	.528	.66	8.8
AUSTRIA Apr '21–Dec '22	.502	3.11	.006	3.97	3.13	$\simeq 0.$

Significance Levels: $\chi^2(1)_{.05}=3.84$ $\chi^2(2)_{.05}=5.99$
$\chi^2(1)_{.025}=5.02$ $\chi^2(2)_{.025}=7.37$
$\chi^2(1)_{.01}=6.63$ $\chi^2(2)_{.01}=9.21$

For Barro's data, at the .95 confidence level the chi-square statistics call for rejecting representation (27) relative to parameterizations (1), (2), (4), and (5) for Hungary I. The statistics do not call for rejection of (27) for Germany, Poland, or Austria.

Notice that for both Cagan's and Barro's data, the overfitting tests reject representation (27) for the case of Hungary I, a case for which my estimator of α obtained the tightest confidence band.

8. CONCLUSIONS

This paper has applied maximum likelihood techniques to derive a consistent estimator of a bivariate, rational expectations version of Cagan's model of hyperinflation. The estimator, in principle, eliminates the simultaneous-equations, asymptotic bias that characterizes Cagan's estimator. Application of the maximum likelihood estimator typically yields "loose" estimates of the slope parameter of the demand schedule for money. The estimates are so loose that confidence bands of two standard errors on each side of them include values that would imply that the creaters of money were inflating at rates that maximized their command over real resources, thus maybe resolving the "paradox" with which I began this paper. While perhaps this resolves the paradox, it does so in a destructive way, by suggesting that the demand for money in hyperinflation has not been isolated as well as might have been thought. This is not a very satisfactory state of affairs in which to leave the subject. In a subsequent paper, I intend to describe further efforts to isolate the demand schedule for money, using a technique which for special reasons cannot be applied to Cagan's model. Use of that technique will be shown to require abandoning the assumption of adaptive (geometric lag, unit-sum) expectations. The technique will be shown to break down under the singular circumstance that the model in the present paper is the correct one. However, the results of my "overfitting" tests, to the extent that they do not always emphatically reject the model in the present paper, suggest that the prospects for success are not great for using such a technique. It could just be true that the model in this paper is the "correct" one, so that even though the portfolio balance schedule was exactly the one Cagan assumed, the nature of the money supply regimes in effect during the hyperinflations makes difficult or impossible estimating the slope of that portfolio balance schedule.

University of Minnesota, U.S.A.

REFERENCES

[1] ANDERSON, T. W., *An Introduction to Multivariate Statistical Analysis* (New York: John Wiley and Sons, 1958).

[2] BARRO, ROBERT J., "Inflation, The Payments Period, and the Demand for Money," *Journal of Political Economy,* LXXVIII, (November/December 1970), 1228–1263.

[3] CAGAN, PHILLIP, "The Monetary Dynamics of Hyperinflation," in M. Friedman, ed., *Studies in the Quantity Theory of Money* (Chicago: University of Chicago Press, 1956).

[4] GRANGER, C. W. J., "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica,* XXXVII, (July, 1969), 424–438.

[5] MEISELMAN, DAVID, *The Varieties of Monetary Experience* (Chicago: University of Chicago Press, 1970).

[6] MUTH, JOHN F., "Rational Expectations and the Theory of Price Movements," *Econometrica,* XXIX (July, 1961), 315–335.

[7] PORTER, RICHARD D., "Multiple Time Series Containing Unobserved Components,"

unpublished manuscript, Board of Governors, Federal Reserve System, (September, 1974).

[8] SARGENT, THOMAS J. AND NEIL WALLACE, "Rational Expectations and the Dynamics of Hyperinflation," *International Economic Review,* XIV, (June, 1973), 328–350.

[9] SIMS, CHRISTOPHER A., "Money, Income, and Causality," *American Economic Review,* LXII (September, 1972), 540–552.

[10] WILSON, G. T., "The Estimation of Parameters in Multivariate Time Series Models," *Journal of the Royal Statistical Societies B,* 1973 (1), 76–85.

[11] ZELLNER, A. AND F. PALM, "Time Series Analysis and Simultaneous Equation Econometric Models," *Journal of Econometrics,* II (May, 1974), 17–54.