Journal of Monetary Economics 23 (1989) 335–344. North-Holland

THE LONG-RUN BEHAVIOR OF THE VELOCITY OF CIRCULATION
A Review Essay

James D. HAMILTON*

University of Virginia, Charlottesville, VA 22901, USA

U.S. inflation has remained low since 1982 despite a dramatic surge in money growth. Considerable attention has deservedly been devoted to accounting for this dramatic decline in the velocity of U.S. M1.[1]

Macroeconomists immersed in the study of this important issue will find a refreshingly broader perspective in Michael Bordo and Lars Jonung's monograph, *The Long-Run Behavior of the Velocity of Circulation* (Cambridge University Press, 1987). This book is not about this recent episode in particular; indeed, data used in most of the empirical analysis end prior to the recent plunge in velocity. Instead, Bordo and Jonung invite us to consider up to a century's worth of data from four-score countries. Their book inquires into the nature and causes of the secular trends in velocity on this global, century-long scale. An understanding of these trends seems vital for designing a monetary policy that promotes stability of prices and output and for interpreting recent U.S. developments with some measure of historical perspective.

Bordo and Jonung's thesis is that technical progress in the financial sector introduces two competing influences on the long-run trend in velocity, each dominant at separate stages in the secular process of industrialization of a particular country. At a phase in development represented by the United States in the late nineteenth century, cash and demand deposits increasingly come to be used for settling transactions, replacing earlier reliance on barter and payments in kind. As a result of this increasing monetization of the economy, demand for transactions balances grows more rapidly than income, and velocity would be expected to be declining during this stage of development. Later in development, as represented by the postwar U.S., improving

*I am indebted to Michael Bordo, Charles Engel, Marjorie Flavin, and Ronald Michener for comments on an earlier draft of this essay.

[1]See for example Stone and Thornton (1987) and the articles in the January 1987 issue of *Contemporary Policy Issues*.

0304-3932/89/$3.50

economic security and stability permit a range of widely traded, highly liquid securities (such as stocks, bonds, and certificates of deposit) that come to dominate money as a store of value. Furthermore, technological innovation in the form of credit cards, rapid transfer of funds, and modern cash management techniques facilitates economizing on money balances. This in Bordo and Jonung's view accounts for the secular rise in U.S. M1 velocity between 1946 and 1980. Thus, the plot of velocity for any given developed country is seen as exhibiting a typical U-shaped pattern.

Traditional theories of money demand identify income, inflation, and interest rates as the principal determinants of velocity. If money demand exhibits greater than unit elasticity with respect to income, then economic growth would induce a secular downward trend in velocity. Friedman and Schwartz (1963, p. 639) defended this interpretation, seeing money as a luxury good whose demand increases more than proportionately with income. One is tempted to view the first downward jag of the 'U' in these terms. Arguing against this position are familiar models such as that due to Baumol (1952) and Tobin (1956) which imply less than unit income elasticity of money demand, making the early downward trend in velocity harder to understand. On the other hand, rising inflation introduces a motive for economizing on cash balances, and whatever one makes of the earlier declining trend of velocity, the fact that the reversal of this trend coincides with the postwar rising rates of inflation common to most developed countries might seem considerably less mysterious.

A focus on inflation as the key variable driving postwar velocity has the further advantage of offering a simpler account of the decline in velocity witnessed over the past five years.

One of the very nice features of Bordo and Jonung's research is that it employs three quite different methodologies and a large number of different data sets to explore the various implications of their hypothesis. They thus offer the reader a number of independent bases on which to accept their interpretation of the data. Chapter 4 employs regression estimates of conventional money demand specifications for 100 years of data on five major industrial countries. Chapter 5 presents a detailed historical case study of institutional developments in Sweden, while chapter 6 seeks to correlate the postwar trend in velocity with development status for a cross-section of 84 countries. I hope in this essay to review some methodological issues common to the broader literature on money demand, and towards that end I shall focus exclusively on the first of Bordo and Jonung's three approaches.

In chapter 4, Bordo and Jonung consider a benchmark regression representing traditional theories of money demand. The log of velocity (V_t) is expressed as a function of a nominal interest rate (i_t), expected inflation (π_t^e) which was proxied by the fitted values of a univariate autoregression for actual inflation, the log of real GNP per capita (Y_t), and a smoothed version of the last series (Y_t^{P}) interpreted as permanent real GNP per capita:[2]

$$\log V_t = \alpha_0 + \alpha_1 i_t + \alpha_2 \pi_t^{\mathrm{e}} + \alpha_3 \log Y_t + \alpha_4 \log Y_t^{\mathrm{P}} + \varepsilon_t. \tag{1}$$

Eq. (1), which Bordo and Jonung think of as a money demand equation transformed by the accounting identity $MV \equiv PY$, is estimated by least squares with Cochrane–Orcutt adjustment for autocorrelated residuals.

Bordo and Jonung then add to regression (1) a battery of variables thought to represent the degree of sophistication and stability of the financial sector. These include the ratio of currency to a broader monetary aggregate, the share of the labor force in agriculture, the ratio of total nonbank financial assets to total financial assets, and a six-year moving average of the standard deviation of Y_t. Bordo and Jonung find that the coefficients on this last set of variables tend to be statistically significant and have the predicted sign in most of the countries analyzed.

For a framework for discussing the issues raised by Bordo and Jonung, I turn to a model similar to that in Lucas and Stokey (1987) and Lucas (1987, ch. VI). Most of us would agree that the institution of money arises from frictions that prohibit trade in the full range of Arrow–Debreu contingent claims. Lucas and Stokey model this friction by postulating a category of goods (C_{1t}) that must be paid for with cash:

$$P_t C_{1t} \leq M_t. \tag{2}$$

The representative consumer framework serves nicely here, allowing all variables to be expressed in per capita terms. A second category of goods (C_{2t}) is not so restricted, and trade in these is permitted on the basis of credit (payable in cash acquired by the consumer at the end of the trading day). The goods are produced by entrepreneurs according to the linear technology

$$C_{1t} + \theta_t C_{2t} = Y_t, \tag{3}$$

where per capita real income (Y_t) is nonstorable and exogenous to the household's decision problem. I've introduced the coefficient θ_t into the

[2] Bordo and Jonung use $\log(Y_t/Y_t^{\mathrm{P}})$ in place of $\log(Y_t)$ in expression (1) to represent cyclical factors. Such a regression of course yields numerically identical parameter estimates as eq. (1), with my parameter α_4 equalling the difference between Bordo and Jonung's permanent and cyclical coefficients.

Lucas–Stokey framework to try to represent Bordo and Jonung's concept of an exogenous change in the payments technology (Lucas and Stokey take $\theta_t \equiv 1$). A decrease in θ_t causes credit goods to become cheaper to produce relative to cash goods and so encourages substitution away from cash as a payments mechanism. Household utility is taken to be

$$\sum_{\tau=0}^{\infty} \beta^{\tau} U(C_{1,t+\tau}, C_{2,t+\tau}).$$

Payments to entrepreneurs on cash transactions are received at the beginning of the day, paid for with the household's start-of-period money holdings M_t. Credit obligations are honored at the end of the day, paid for with sale of the household's period t endowment Y_t. Since entrepreneurs are unable to spend either of these sources of funds until the day following, they are indifferent between (a) receiving a dollar on a cash transaction and (b) receiving the promise of a dollar on a credit transaction. Hence from eq. (3), if P_t denotes the money price of good 1, then in equilibrium the money price of the endowment (Y_t) will also be P_t while the price of good 2 must be $P_t\theta_t$.

The household decision problem is thus to maximize

$$\mathrm{E}_t \sum_{\tau=0}^{\infty} \beta^{\tau} U(C_{1,t+\tau}, C_{2,t+\tau}),$$

subject to the cash-in-advance constraint (2) as well as the budget constraint

$$M_{t+1} = M_t + P_t[Y_t - C_{1t} - \theta_t C_{2t}] + X_t,$$

with X_t denoting any sources of funds (such as transfer payments) that are exogenous or predetermined from the point of view of the household's decision problem.

One can solve the above problem with dynamic programming techniques. I offer an intuitive derivation here. The consumer is free to contemplate two alternative uses of his current nominal noncash endowment (P_tY_t). On the one hand, a marginal dollar of this endowment might be pledged towards credit purchases this period, from which the gain in utility would be $(1/P_t\theta_t)\cdot U_2(C_{1t}, C_{2t})$. Alternatively, the dollar might be devoted to incrementing the stock of cash to be carried into next period and then used for purchases of the cash good $C_{1,t+1}$. The present value of the utility benefit associated with the latter strategy is $\beta\,\mathrm{E}_t(1/P_{t+1}) \cdot U_1(C_{1,t+1}, C_{2,t+1})$.[3] The consumer will choose to equate these gains at the margin, or[4]

$$\frac{1}{\theta_t P_t} U_2(C_{1t}, C_{2t}) = \beta\,\mathrm{E}_t\left[\frac{1}{P_{t+1}} U_1(C_{1,t+1}, C_{2,t+1})\right].$$

Using the equilibrium condition (3), this equation becomes

$$\frac{1}{\theta_t P_t} U_2(C_{1t}, [Y_t - C_{1t}]/\theta_t)$$

$$= \beta\,\mathrm{E}_t\left[\frac{1}{P_{t+1}} U_1(C_{1,t+1}, [Y_{t+1} - C_{1,t+1}]/\theta_{t+1})\right]. \tag{4}$$

If we define velocity V_t in the usual sense,

$$M_t \cdot V_t = P_t \cdot Y_t,$$

we see that, if the cash-in-advance constraint (2) is binding, then

$$V_t = [Y_t/C_{1t}]. \tag{5}$$

Changes in inflation, output, or the payments technology thus effect changes in velocity in this framework through their influence on the portion of income that consumers choose to devote to cash goods (C_{1t}/Y_t).

To understand the theory of velocity embodied in eq. (4), consider first a world in which income, the payments technology, and inflation are all constant $(C_{1t} = C_1,\ Y_t = Y,\ \theta_t = \theta,\ P_{t+1}/P_t = 1 + \pi$ for all $t)$. Then (4) becomes

$$\frac{U_2(C_1, [Y - C_1]/\theta)}{\theta \cdot U_1(C_1, [Y - C_1]/\theta)} = \frac{\beta}{1 + \pi}. \tag{6}$$

Eq. (6) implicitly characterizes cash purchases C_1 as a function of income Y, payments technology θ, and inflation π.

[3]This formula exploits an envelope condition that characterizes the optimal $(C_{1,t+1}, C_{2,t+1})$. If the cash-in-advance constraint is not binding at date $t+1$, then the household is indifferent between (a) devoting an additional dollar to $C_{1,t+1}$ and (b) devoting an additional dollar to $C_{2,t+1}$. In this case, the value of a marginal dollar is adequately measured by its contribution to utility if devoted entirely to good 1. Alternatively, if the cash-in-advance constraint is binding, then the utility from a dollar spent on $C_{1,t+1}$ exceeds that from a dollar spent on $C_{2,t+1}$, and the formula is equally valid.

[4]See eq. (12) in Lucas (1987, p. 77).

Suppose C_1 and C_2 are normal goods.[5] If we compare two steady-state economies with the same output but different inflation rates, differentiation of (6) reveals that the one with higher inflation will have a lower portion of purchases going to cash goods C_1; thus from (5), higher inflation increases the velocity V.

Next consider two steady-state economies with the same inflation rate but different levels of output. If $U(.)$ is homothetic, $U_2(\lambda C_1, \lambda[Y - C_1]/\theta)/U_1(\lambda C_1, \lambda[Y - C_1]/\theta)$ is a constant for all λ, meaning the two economies would have the same portion of income devoted to cash versus noncash purchases and the same velocity. If one is to attribute the initial secular decline in velocity to rising income as in the Friedman–Schwartz luxury hypothesis, this marginal rate of substitution must be a decreasing function of λ.

Finally, one calculates from (6) that an increase in the relative cost of credit goods (θ) leads to a reduction in equilibrium velocity V.

Once we get away from analysis of the time-invariant case (6), the velocity V_t implicit in (4) will be a function of Y_t, P_t, θ_t, and all information that is useful in forecasting these variables. That money demand is inherently a forward-looking concept is acknowledged in Bordo and Jonung's use of proxies for expected inflation and income. However, these proxy forecasts are based on restrictive use of past information. In principle, *any* variable Z_t that is useful in forecasting future inflation or income belongs in a money demand function, and we could be justified in interpreting the least-squares coefficient on that variable Z_t as telling us about the sensitivity of money demand to income and inflation as distinct from the *causal* effect of Z_t on velocity independent of its role as a forecaster of future Y_t or π_t.

This observation raises what would appear to be a fundamental identification problem for this line of inquiry. One might view the objective of Bordo and Jonung to be to identify the separate contributions to velocity made by the variables y_t, P_t, and θ_t. However, any shock to the payments technology θ_t will by construction affect the log of velocity at date t ($\log V_t = \log P_t + \log Y_t - \log M_t$). Since velocity is highly autocorrelated, the shock to θ_t is also likely to be useful in forecasting next period's velocity ($\log V_{t+1} = \log P_{t+1} + \log Y_{t+1} - \log M_{t+1}$). Thus, if θ_t exerts a material impact on current money demand, it probably also will prove useful in forecasting inflation (π_{t+1}) or income (Y_{t+1}). If something matters for current money demand, it's unlikely not to matter for future inflation or output.

Our dilemma is thus as follows: if a seemingly extraneous factor Z_t is found to be statistically significant in accounting for velocity, it might prove difficult to determine whether it belongs in the regression because it is useful in

[5] That is, $U_1U_{21} - U_2U_{11} > 0$ and $U_2U_{12} - U_1U_{22} > 0$.

forecasting inflation and output, in which case we might say its contribution is entirely conventional, or whether it appears because it proxies an exogenous shock to the money demand function through a variable such as θ_t.

The issue of expectations aside, identification and simultaneous equations bias are surely a concern in studying these questions. As a theoretical matter, simultaneity could account for the empirical success of apparently unconventional arguments in the money demand function. To see this point, consider the following stripped-down example, in which variables are expressed as deviations from means. In logarithmic terms, the money demand equation is

$$m_t - p_t = \beta y_t - k i_t + \varepsilon_t, \tag{7}$$

with m_t, p_t, and y_t denoting the logs of nominal M2, prices, and real output, respectively, and i_t is the nominal interest rate. Define the M2 multiplier as the ratio of M2 to the monetary base:

$$m_t \equiv b_t + \lambda_t, \tag{8}$$

where b_t denotes the log of the monetary base and λ_t the log of the M2 multiplier. The M2 multiplier depends in part on currency and reserve decisions by banks and the public, which decisions are influenced by interest rates:

$$\lambda_t = \alpha i_t + u_t. \tag{9}$$

For purposes of illustration, suppose that this dependence of the money supply process on interest rates in (9) is the sole potential source of simultaneous equations bias. Realistically, there are many other sources of concern and the issues discussed below become all the more serious. To analyze the simplest example of this potential problem, though, let us assume that p_t, y_t, and b_t are all econometrically exogenous, say mutually uncorrelated and uncorrelated with the independent disturbances ε_t and u_t.

In their chapter 4 regressions (though not in the postwar cross-sectional analysis), Bordo and Jonung add the ratio of currency to money $(-\lambda_t)$ to a conventional money demand equation such as (7) to proxy a shock to the payments technology θ_t. Suppose, however, that eq. (7) really characterizes the truth, and institutional factors play no role in money demand. What would happen in such a world if we estimated a regression of the form

$$m_t - p_t = \phi_1 y_t + \phi_2 i_t + \phi_3 \lambda_t + e_t, \tag{10}$$

and proceeded to test the hypothesis that $\phi_3 = 0$ as a test of the adequacy of (7) as a theory of money demand?

Note that the regressors $\boldsymbol{x}_t$ in eq. (10) can be written

$$\boldsymbol{x}_t = \begin{bmatrix} y_t \\ i_t \\ \lambda_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/(\alpha+k) & -1/(\alpha+k) \\ 0 & \alpha/(\alpha+k) & k/(\alpha+k) \end{bmatrix} \begin{bmatrix} y_t \\ \beta y_t + p_t + \varepsilon_t - b_t \\ u_t \end{bmatrix}.$$

Similarly, the dependent variable can be expressed as

$$m_t - p_t = \begin{bmatrix} (\beta y_t + \varepsilon_t) & (b_t + u_t - p_t) \end{bmatrix} \begin{bmatrix} \alpha/(\alpha+k) \\ k/(\alpha+k) \end{bmatrix}.$$

The OLS estimate of (10) is given by

$$\hat{\boldsymbol{\phi}} = \left[\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'/T \right]^{-1} \left[\sum_{t=1}^{T} \boldsymbol{x}_t (m_t - p_t)/T \right],$$

which converges in probability to

$$\hat{\boldsymbol{\phi}} \xrightarrow{\mathrm{p}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & k & -\alpha \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_y^2 & \beta\sigma_y^2 & 0 \\ \beta\sigma_y^2 & [\beta^2\sigma_y^2 + \sigma_p^2 + \sigma_\varepsilon^2 + \sigma_b^2] & 0 \\ 0 & 0 & \sigma_u^2 \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \beta\sigma_y^2 & 0 \\ [\beta^2\sigma_y^2 + \sigma_\varepsilon^2] & -[\sigma_p^2 + \sigma_b^2] \\ 0 & \sigma_u^2 \end{bmatrix} \begin{bmatrix} \alpha/(\alpha+k) \\ k/(\alpha+k) \end{bmatrix}$$

$$= \begin{bmatrix} \beta\gamma \\ -k\gamma \\ 1-\gamma \end{bmatrix},$$

where

$$\gamma \equiv (\sigma_p^2 + \sigma_b^2)/(\sigma_p^2 + \sigma_b^2 + \sigma_\varepsilon^2).$$

Notice that γ lies between 0 and 1.

In order for the OLS regression estimate $\hat{\phi}$ to be giving us sensible inference about the true money demand parameters $(\beta, -k, 0)$, it must be the case that γ is close to unity, that is, it must be the case that the variance of disturbances to the money demand equation (σ_ε^2) is inconsequential relative to that of disturbances to prices and the monetary base $(\sigma_p^2 + \sigma_b^2)$. If there are significant disturbances to the money demand equation, a variable such as $-\lambda_t$, the ratio of currency to M2, might appear statistically significant in a regression of the form of (10), not because it has anything to do with money demand but instead because it is important for the description of the money supply process.

Unfortunately, the real world is messier still. The Federal Reserve has not been setting the monetary base independently of economic conditions, but instead responds to interest rates, prices, and income, as does the M2 multiplier. Neither the price level nor the level of real output are econometrically exogenous; indeed, we are interested in studying money demand in order to measure the consequences of monetary policy for these variables. The problems only become more serious as the specification becomes more realistic. And attempting to estimate (4) by Hansen and Singleton's (1982) generalized method of moments is no panacea for this source of simultaneous equations bias [Garber and King (1984)].

Faced with this predicament, we basically have three options. The first is the road selected by Bordo and Jonung – repeat the analysis with as many different data sets and countries in different stages of development as possible. When one finds, as they document, evidence of a consistent, reproducible pattern that is robust across a large number of different specifications, one begins to establish a compelling scientific case that there is a predictable regularity in the correlations warranting a structural interpretation. The institutional details of the money supply process can't be the same in all these countries and over such a long period of time. If the OLS estimates are yielding stable coefficients, one is developing a compelling case that it is indeed money demand that has been picked up, an argument due to Hendry and Richard (1983). On the other hand, to the extent that the structural equation being estimated is a behavioral hypothesis in which people's expectations about the future are asserted to matter a great deal for their current behavior, the true money demand function should not be invariant across changes in the regime for inflation and income [Lucas (1976)] and stability of a specification such as (1) is somewhat puzzling.

A second option is to retreat into reduced forms. Since essentially all the variables are endogenous, in practice this means turning to the vector autoregressions of Sims (1980). We must bear in mind, however, that the questions one can legitimately address with VARs are restricted to those that can be framed in terms of forecasts. An interesting question is whether variables that seem quite specific to the payments institutions are useful in forecasting the individual components of velocity. Important and useful though this question is, we need more information in order to govern our monetary policy wisely. For as soon as we change the process for the money supply, the reduced-form coefficients will change, rendering any conditional forecasts we derived from the reduced form unreliable [Marschak (1953)].

Our third option is true structural analysis, thinking seriously about the real-world determinants of the money supply, prices, and output. It is clear enough that if we could carry it off, this approach would be the best. Unfortunately the execution is quite treacherous, and many of today's generation of macroeconometricians are spooked by the task.

True structural analysis of money demand may not be completely hopeless. If one thinks in a careful way about the determinants of weekly data on M1, there are many nontrivial details about which our knowledge is relatively good. The specific operation of lagged reserve requirements and target bands for the Federal funds rate, for example, may offer useful identifying restrictions that allow us to distinguish factors affecting money demand, the M1 multiplier, and the Federal Reserve.

It is a difficult exercise, but, as the recent numbers on inflation reveal, a very important one, to try to sort out empirically the determinants of money demand. Bordo and Jonung have pushed the frontiers forward considerably by documenting an extremely widespread phenomenon in the long-run behavior of velocity.

References

Baumol, William, 1952, The transactions demand for cash: An inventory theoretic approach, Quarterly Journal of Economics 56, 545–556.

Bordo, Michael D. and Lars Jonung, 1987, The long-run behavior of the velocity of circulation: The international evidence (Cambridge University Press, Cambridge).

Friedman, Milton and Anna J. Schwartz, 1963, A monetary history of the United States: 1867–1960 (Princeton University Press, Princeton, NJ).

Garber, Peter M. and Robert G. King, 1984, Deep structural excavation? Identification problems in Euler equation methods, Mimeo. (University of Rochester, Rochester, NY).

Hansen, Lars P. and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, Econometrica 50, 1029–1054.

Hendry, David F. and Jean-Francois Richard, 1983, The econometric analysis of economic time series, International Statistical Review 51, 111–163.

Lucas, Robert E., Jr., 1976, Econometric policy evaluation: A critique, in: K. Brunner and A. Meltzer, eds., The Phillips curve and labor markets, Carnegie–Rochester conference series on public policy, Vol. 1 (North-Holland, Amsterdam) 19–46.

Lucas, Robert E., Jr., 1987, Models of business cycles (Basil Blackwell, New York, NY).

Lucas, Robert E., Jr. and Nancy L. Stokey, 1987, Money and interest in a cash-in-advance economy, Econometrica 55, 491–513.

Marschak, Jacob, 1953, Economic measurements for policy and prediction, in: William C. Hood and Tjalling C. Koopmans, eds., Studies in econometric method (Wiley, New York, NY) 1–26.

Stone, Courtenay C. and Daniel L. Thornton, 1987, Solving the 1980's velocity puzzle: A progress report, Federal Reserve Bank of St. Louis Review 69, no. 7, 5–23.

Sims, Christopher A., 1980, Macroeconomics and reality, Econometrica 48, 1–48.

Tobin, James, 1956, The interest-elasticity of the transactions demand for cash, Review of Economics and Statistics 38, 241–247.