August 1976

no. 86

THEORY AND IMPLICATIONS OF CURRENCY SUBSTITUTION

by

Lance Girton and Don Roper

May 1976

THEORY AND IMPLICATIONS OF CURRENCY SUBSTITUTION*

Lance Girton and Don Roper **

One of the interesting features in the study of international monetary theory is that some problems arise in a multi-currency context that are not so obvious in "domestic" monetary theory. One such problem which has received little attention, concerns a phenomenon that might be called "currency substitution". By currency substitution, abbreviated CS, we mean the degree that currencies are substitutes in the portfolios of ultimate wealth holders. The word "currency" refers specifically to money (to be defined more precisely below) as opposed to interest bearing assets recorded in capital account entries of balance of payments accounts.

The meaning of CS can be conveyed with some examples. Substitution was possible between the old Czarist currency and the new Bolshevik currency as they both circulated with a variable exchange rate for a period after the Russian Revolution. With the introduction in the early part of October, 1974, of the Turkish lira in Cyprus

*The ideas developed in this paper originated in conversations between the authors and Russell Boyer during his consultantships with the Federal Reserve Board in the summers of 1972 and 1973. This paper extends the analysis found in Boyer (1972) and (1973) which, will be partly summarized in the present paper. Specific statements of the relation between his work and ours will be given in the text. Boyer has also made helpful comments on the present work, but he is not responsible for the particular development of the ideas in the paper. The authors are also grateful for comments received in a international finance seminar at Princeton University.

**The views expressed herein are solely those of the authors and do not necessarily represent the views of the Federal Reserve System.

(or at least in the part of Cyprus controlled by Turkey), the residents could freely substitute between the lira and the Cyprus pound at a market determined rate of exchange. The fact that transactions can be effected in one country with currency from another country — especially in border areas where tourism, travel, and/or smuggling is prevalent — is another example of CS. In the last quarter of the 19th century and until World War I, there was a monetary union among Denmark, Sweden, and Norway. All three currencies circulated in each of the member countries and residents had a choice over the currency used to effect many transactions.

It is not clear that the phrase "currency substitution" is the most appropriate name; there are a couple of reasons to think that "currency mobility" is more appropriate. In the first place, the international immobility of money plays a role in international monetary theory similar to the role played by the immobility of factors of production in international trade theory. Despite terminology suggesting that "money leaves the country" when there is a net capital outflow, we know that money *per se, usually* doesn't leave the country — it stops, so to speak, at the foreign exchange market where it is exchanged for another currency. Just as it is interesting to explore the implications for real trade theory of allowing for some degree of factor mobility, it is interesting to explore the implications for multi-currency monetary theory of allowing for currency mobility.

In the second place, currency substitution is very similar to what is frequently called "capital mobility." Just as the mobility of capital tends to equalize interest rates, the mobility of money tends to equalize real rates of return on monies.

But recent developments in the theory of financial capital movements suggest that "mobility" more appropriately refers to the efficiency with which markets operate whereas "substitution" more appropriately refers to a characteristic of portfolio preferences.[1] Since the term "substitution" more accurately describes the characteristic of interest it will be adopted here despite the precedents for the use of the term "mobility" and the intuitive appeal of this term.

It is important to separate substitution in demand from supply considerations. This point can be illustrated with a seemingly trivial example of nickels and dimes. The fact that nickels and dimes exchange at a fixed rate of 2 for 1 reflects the fact that the U.S. mint stands ready to purchase or sell nickels against dimes in whatever quantities necessary to maintain the fixed rate of exchange. Moreover, even if the authorities let the nickel-dime rate "float,"[2] and the coins,

---

[1] See, for instance, William Branson (1968) and Girton and Henderson (1976).

[2] Some people are tempted to argue that since both coins are issued by the same authority and since they are both measured in the same units, cents, such that their exchange rate (10 cents/5 cents = 2) is a dimensionless number, then the government sets the so-called "exchange rate" by fiat and it cannot "float." But this intuitive response has causation running in the wrong direction. The fact that both coins are referred to with the word "cents" is a linguistic convention *made possible by* the fact that the rate of exchange has been and is expected to remain fixed by government pegging. An example of a "floating" exchange rate between two liabilities of the same monetary authority is found in the 1960's when U.S. mints were unable to obtain enough silver to continue pegging the exchange rate between Federal Reserve notes and some U.S. coins with silver content. The result (as reported by Friedman and Schwartz (1970) who, in turn, cite the *Wall Street Journal* of April 23 and June 8, 1964) was that some businesses paid a 5% premium to obtain sufficient quantities of coins. In 1974 some banks paid up to a 10% premium for pennies in bulk. Friedman and Schwartz also stress the difference between substitutability in supply and in demand although they do not specify how one would determine the degree of demand substitution.

continued to trade at the rate of 2 for 1, that alone still not imply that the two coins are perfect substitutes. Nickels and dimes, like other financial assets, are stocks whose real demands are determined by (own and cross) rates of return. The degree of substitutability in demand between two financial assets is reflected in the degree to which demand considerations force their rates of return into equality. Two assets like nickels and dimes would be considered perfect substitutes if demand considerations forced their rates of return into equality. In short, a stable rate of exchange between two assets, fixed by a pegging operation implies nothing about the degree of substitution on the demand side.

The paper is organized as follows: In section I a model of two currencies and one other asset, capital, is developed in which the rates of growth of the two monies are taken as exogenous. The implicaitons of different degrees of currency substitution for the stability and determinancy of the exchange rate and the composition of currencies within portfolios is examined. The necessity of imperfect CS for balance of payments disequilibria is demonstrated. In section II the rates of growth of the two monies are endogenized by assuming profit maximizing behavior on the part of the money issuers. The implications of currency substitution for the profit-maximizing rates of monetary expansion and inflation are examined. It is demonstrated that the competition between money issuers in the presence of perfect substitution between currencies leads to the (socially) optimal quantities of monies.

The paper abstracts from considerations of risk, uncertainty, and the forward market.

I. Implications of CS with Exogenous Money Growth Rates

A portfolio balance model of three assets, two of which are currencies, will be developed in this section. The model will be dynamic and monetary growth rates will be taken as exogenous parameters. The implications of currency substitution for both the differential inflation rate and the exchange rate will be examined.

The following notation will be used:

F_k = real demand for the kth asset ($k = 1,2,3$)

r_k = real interest rate on kth asset (r_k^* = anticipated real rate)

P_k = price of goods in terms of currency k ($k = 1,2$)

$\rho = \ln(P_1/P_2)$ = logrithmic ratio of prices

$\pi_k = d\ln P_k/dt$ = inflation rate in terms of currency k (= 1,2)

$\delta = \pi_1 - \pi_2 = \dot{\rho}$ = differential inflation rate

W = real wealth = $F_1 + F_2 + F_3$

The asset demand functions are assumed to depend on the anticipated real returns and real wealth,

$$F_k = F_k(r_1^*, r_2^*, r_3^*, W), \qquad (k = 1,2,3).$$

It is assumed that an equal change in all anticipated real returns leaves the three asset demand functions unchanged. Therefore, the functions can be expressed as depending on the differential returns and can be written as

$$F_1 = L_1(r_1^*-r_2^*,\ r_1^*-r_3^*,\ W) \qquad F_2 = L_2(r_2^*-r_3^*,\ r_2^*-r_3^*,\ W)$$

$$F_3 = f_3(r_1^*-r_3^*,\ r_2^*-r_3^*,\ W)$$

The anticipated real interest rate on money is found by taking the nominal or market rate, i_k, and substracting the anticipated rate of inflation. In symbols, we have $r^*_k = i_k - \pi^*_k$ $(k = 1,2)$. The two currencies or monies are assumed to be the liabilities of two separate institutions -- whether two central banks or commercial banks -- that do not pay interest on their liabilities. Thus, i_k will be taken as zero (or at least constant) for the analysis.[1] For expositional purposes, the (anticipated) real return on the third asset, r^*_3, will be taken as constant.[2] Suppressing the constants, i_1, i_2, and r^*_3, the sum of the asset-demand functions can be expressed as

$$L_1(\delta^*, \pi^*_1, W) + L_2(\delta^*, \pi^*_2, W) + f_3(\pi^*_1, \pi^*_2, W) = W. \tag{1}$$

At a point in time real wealth, $W(t)$, is given so that the partial derivatives of the three functions are constrained in the following manner:[3]

(i) $\partial L_1/\partial\delta^* + \partial L_2/\partial\delta^* = 0$ (ii) $\partial L_1/\partial W + \partial L_2/\partial W + \partial f_3/\partial W = 1$

(iii) $\partial L_1/\partial\pi^*_1 + \partial f_3/\partial\pi^*_1 = 0$ (iv) $\partial L_2/\partial\pi^*_2 + \partial f_3/\partial\pi^*_2 = 0$

Constraint (i) reflects substitution between the two currencies due to an anticipated differential inflation rate. The third (iii) and fourth (iv) constraints reflect substitution between the two currencies and the third asset, capital. The derivative, $\partial L_k/\partial\pi^*_k$ measures the usual "own" effect of anticipated inflation whereas $\partial L_k/\partial\delta^*$ reflects substitution between the two currencies.

---

[1]The role of this assumption in distinguishing money from other assets has been emphasized by James Tobin (1969).

[2]This is only for expositional convenience since the impact of r^*_3 will be subsequently shown to cancel out in our analysis of the differential inflation rate.

[3]It is impossible, of course, for δ^* to change with π^*_1 and/or π^*_2 changing. But this fact does not alter the validity of the constraints (i)-(iv) as written.

The first constraint (i) implies that the responses in real demands for currencies one and two, induced by a change in δ^*, will be equal and opposite in sign. This assumes that, at a point in time, real wealth will be unaffected by the reallocation of portfolios between the two currencies. But if one money is converted into another money at an exchange rate of E (price of currency two in terms of currency one), then the realization of this reallocation of monies will leave real wealth W, unaffected only if $E = P_1/P_2$. The necessity of this condition becomes more apparent by considering the consequence of its not holding. Suppose, for instance, that an increase in δ^* induced wealth holders to switch from currency one to currency two when $P_2 > P_1/E$ and $E = 1$. Since each unit of currency one would be worth more in terms of goods than a unit of currency two, the reallocation would cause real wealth to fall. Consequently, the constant real wealth assumption behind constraint (i) is predicated on the conditions that $E = P_1/P_2$.[1]

It is important to distinguish the equality $E = P_1/P_2$ from the usual notion of purchasing power parity (PPP). PPP is typically regarded as a statement equating price levels in different <u>countries</u>. But if two currencies circulate within each country, the ratio of their values (in terms of the same set of goods) can equal the exchange rate without implying that consumption or production price indices between the two countries are equal. Our analysis is concerned with currencies rather than countries and with the values of monies denominated in the same numeraire good rather than the money values of baskets of goods in different countries.

---

[1]This condition implies that $M_1/P_1 = M_1/EP_1$. More generally, the "real" quantity of M_k is independent of the weights used in the price deflator.

Values of the two currencies are assumed to not change instantaneously. A partial adjustment model will be used to explain the rate of inflation in currency K as depending on the percent, excess of currency k. The inflation rate in currency one is given by[1]

(2) $\pi_1 = \gamma_1[\ln M_1^s(t) - \ln M_1^d(t)]$

where $M_1^d = P_1 L_1(\delta*, \pi_1^*, W)$

(3) $M_1^s(t) = M_1(0)\exp(u_1 t)$

u_1 = exogenous exponential growth rate of currency one.

The natural logarithm of the nominal money demand for currency one, M_1^d, can be expressed in linear form as

(4) $\ln M_1^d = \ln P_1 + \ln L_1(0) + (\partial \ln L_1/\partial \pi_1^*)\pi_1^* + (\partial \ln L_1/\partial \delta^*)\ \delta^* + (\partial \ln L_1/\partial w)dw$

where equation (4) is the linear part of a Taylor's series expanded around $\ln L_1(0) = \ln L_1(0, 0, w(0))$ and $\ln P_1(0)$.[2] It will be convenient to label the two partial derivatives as

(5) $\alpha_1 = -\partial \ln L_1/d\pi_1^* > 0$ and $\sigma_1 = -\partial \ln L_1/\partial \delta^* > 0$.

Substituting (3)-(5) into (2) yields

(6) $\pi_1 = \gamma_1[c_1 + u_1 - \ln P_1 + \alpha_1\pi_1^* + \sigma_1\delta^* \quad -(\partial \ln L_1/\partial w)dw]$

where $c_1 = \ln[M_1(0)/L_1(0)]$.

The identical analysis for the inflation rate in currency two yields

(7) $\pi_2 = \gamma_2[c_2 + u_2 - \ln P_2 + \alpha_2\pi_2^* - \sigma_2\delta^* -(\partial \ln L_2/\partial w)dw]$

---

[1] A similar adjustment equation is suggested by Friedman (1966).

[2] In equation (4) the initial price level, $P_1(0)$ has been taken as unity so its log would vanish ($\ln P_1(0) = 0$). The expansion has been around the initial value of the logarithm of wealth such that $\partial \ln L/\partial w$ is an elasticity where $w = \ln W$.

Since we want to examine relative currency values and differential inflation rates (as opposed to absolute currency values and inflation rates), the two currency demand functions are assumed to have the same responsiveness to their own inflation rates[1] ($\alpha_1 = \alpha_2 = \alpha$), the same wealth elasticities ($\partial \ln L_1/\partial w = \partial \ln L_2/\partial w$), and possible differences in adjustment speeds will be neglected ($\gamma_1 = \gamma_2 = \gamma$). Using these assumptions the differential inflation rate can be found by substracting equation (7) from (6) to obtain[1]

$$\dot{\rho} = \pi_1 - \pi_2 = \gamma[c + \upsilon t - \ln P_1 + \ln P_2 + \alpha(\pi_1^* - \pi_2^*) + (\sigma_1 + \sigma_2)\ \delta^*]$$

$$(8) \qquad = \gamma[c + \upsilon t - \rho + \eta\ \delta^*]$$

where
$$c = c_1 - c_2 \gtrless 0$$
$$\upsilon = u_1 - u_2 \gtrless 0$$
$$\eta = \alpha + \sigma_1 + \sigma_2$$

The parameter η embodies both the own-effect α of the two currency demands plus the currency substitution parameter $\sigma (= \sigma_1 + \sigma_2)$. Since σ can range from zero to infinity, $\bar{\eta}$ can range from α to infinity.

Equation (8) is the basic differential equation of the model. To complete the model the determination of δ^* must be specified and assumptions concerning the formation of anticipations are required. In order to separate the fundamental implications of currency substitution from those that might depend on particular behavioral assumptions regarding anticipations formation, two different specifications of δ^* will be considered. In the first case (Figures I-III) it will be assumed that anticipations are realized at all times such that $\delta^* = \dot{\rho}$. In the second case it will be assumed that δ^* is determined by an adaptive expectations model.

---

[1] The anticipated real rate of return on the third asset, r_3^*, would have canceled out in the $\alpha\delta^*$ term in equation (8) had it been carried forward to this point. Consequently, the earlier assumption of a constant r_3^* was for expositional convenience only.

When the actual differential rate, $\delta(=\dot{\rho})$ is equated with δ^*, equation (8) reduces to a single, linear, non-homogeneous, differential equation expressible in the form

(9) $$\dot{\rho} = \gamma(c + \upsilon t - \rho + \eta\dot{\rho})$$

or $$\dot{\rho} = \frac{-\gamma}{1-\eta\gamma}(\rho - (c+\upsilon t)).$$

Equation (9) is plotted in Figure I for $\upsilon = u_1 - u_2 = 0$. One line is drawn for the stable case in which $\eta\gamma = (\alpha + 2\sigma)\gamma > 1$.

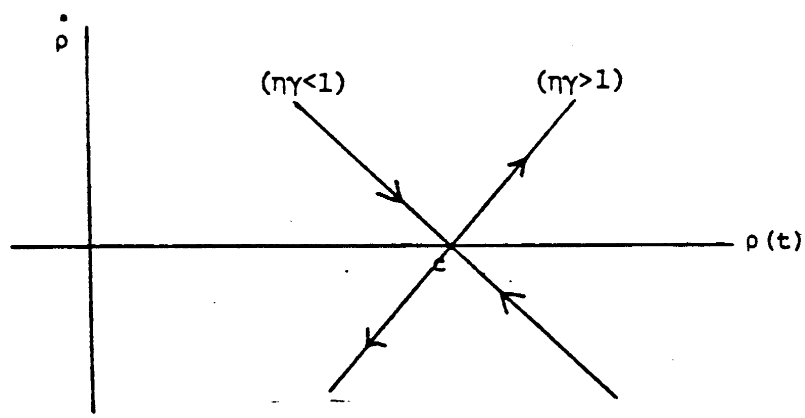

Figure I: Dynamics with Realized Expectations and $\upsilon = 0$

The condition $\upsilon = 0$ means that the secular supplies of monies are growing at the same rates -- that $u_1 = u_2$. Thus, relative prices have a stationary equilibrium at the point $c = \ln[M_1(0)/L_1(0)] - \ln[M_2(0)/L_2(0)]$.

In the more general case, there is a secular differential inflation caused by differential monetary growth rates. Figure II is drawn for the case in which $\upsilon = u_1 - u_2 > 0$ such that currency one depreciates relative to currency two. Figure II also assumes that currency substitution is sufficiently low (σ is sufficiently small) that $\eta < 1/\gamma$ and the system is stable.

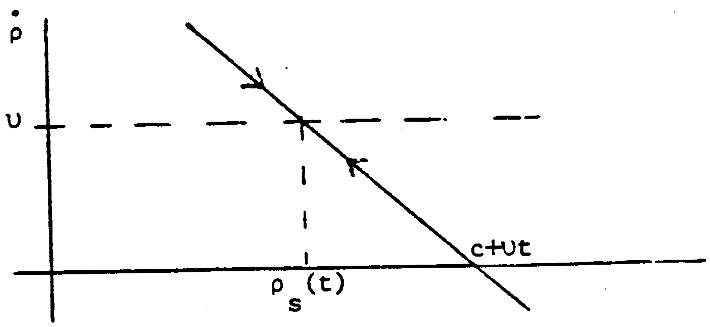

Figure II: Stable System with $\upsilon > 0$

The equilibrium position in Figure II is stationary in $\dot{\rho}$ (=υ) but ρ, of course, moves over time. At any point in time the ρ-intercept is $c + \upsilon t > \rho_s(t)$. When there is no currency substitution, the slope of the curve is $-\gamma/(1-\eta\gamma) = -\gamma/(1-\alpha\gamma)$. With an increase in CS, η becomes larger and the curve rotates around its ρ-intercept in a clockwise direction. When $\eta > 1/\gamma$, the curve assumes a positive slope and the system again becomes unstable as shown in Figure III. The equilibrium value, $\rho_s(t)$ (found by the intersection of the horizontal $\dot{\rho} = \upsilon$ line and the positively sloped curve) moves further to the right with greater degrees of CS.

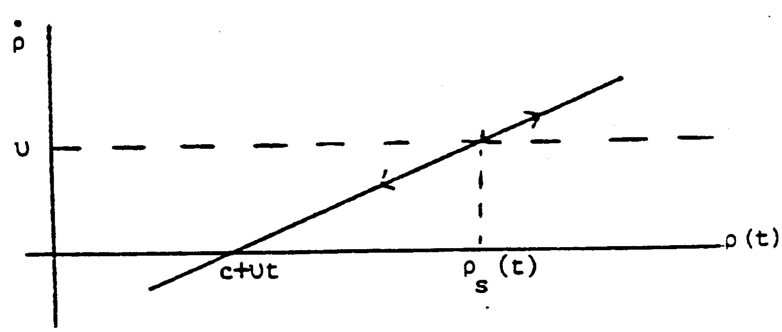

Figure III: Moving Equilibrium with High CS

Figure III can be used to illustrate two points. First, as Boyer (1973) demonstrated with a somewhat different model, CS causes instability. This analysis assumes that the rate of exchange between the two currencies is sufficiently flexible to allow for secular differential inflation rates and that there is no interest paid on the monies to compensate for the differential inflation rate. More attention will be given to the underlying reasons for this instability in section II.

The second point to draw from Figure III is that in the limit, as η goes to infinity, the phase line for equation (9) continues to rotate clockwise until it coincides with the abscissa. In that case there is no equilibrium point or intersection between $\dot{\rho} = \upsilon$ and the phase line unless $\upsilon = 0$. In other words there is only one differential inflation rate (viz., $\dot{\rho} = \upsilon = 0$) that is consistent with perfect currency substitution. This result, however, is not surprising. It is analogous to the usual result concerning perfect substitution between other financial assets. With perfect substitution between bonds (or with perfect "capital mobility") the real rates of return on bonds will be identical. Except for the fixed money interest rate money is not different than other financial assets such that perfect substitution equalizes real rates of return which, in the case of money, implies equal inflation rates.

Thus far, the analysis has been based upon the assumption that the actual differential inflation rate is perfectly anticipated at all points in time. It is useful to determine whether the results are robust to alternative assumptions regarding the formation of inflationary expectations. The alternative considered here is the adaptive expectations mechanism. Inflationary expectations for each country separately can be expressed as

(10) $\dot{\pi}_1^* = \beta_1(\pi_1 - \pi_1^*)$ and $\dot{\pi}_2^* = \beta_2(\pi_2 - \pi_2^*)$.

Continuing to abstract from possible differences between currencies such that $\beta_1 = \beta_2 = \beta$, the anticipated differential inflation rate is revised according to

$$\dot{\delta}^* = \dot{\pi}_1^* - \dot{\pi}_2^* = \beta(\pi_1 - \pi_2^* - (\pi_1 - \pi_2^*)) \quad \text{or}$$

(11) $\dot{\delta}^* = \beta(\delta - \delta^*) = \beta(\dot{\rho} - \delta^*)$.

Equations (9) and (6) constitute a system of linear differential equations in ρ and δ^*. They can be written in matrix form as

$$\begin{bmatrix} 1 & 0 \\ -\beta & 1 \end{bmatrix} \begin{bmatrix} \dot{\rho} \\ \dot{\delta}^* \end{bmatrix} = - \begin{bmatrix} \gamma & -\gamma\eta \\ 0 & \beta \end{bmatrix} \begin{bmatrix} \rho - c \\ \delta^* \end{bmatrix} + \begin{bmatrix} \upsilon \\ 0 \end{bmatrix} \cdot t$$

and in more succinct notation as —

(12) $A\dot{x} = - Bx + Tt$.

The homogeneous system is stable if and only if $\eta < \frac{1}{\gamma} + \frac{1}{\beta}$. The system becomes cyclical before it becomes unstable and, if $\gamma = \beta$, it is cyclical for all positive values of η. The particular integral of this system, which we can denote with the subscript "s" for "steady state" is

(13) $x_s = (-B^{-1}A + It)B^{-1}T$.

The particular integral gives the secular equilibrium values of ρ and δ^*. When (11) is written out, it reduces to the very simple expressions

$$\delta_s^* = \upsilon \quad \text{and} \quad \rho_s(t) = c + \upsilon t + (\eta - 1/\gamma)\upsilon.$$

These conditions are depicted graphically in Figure IV.

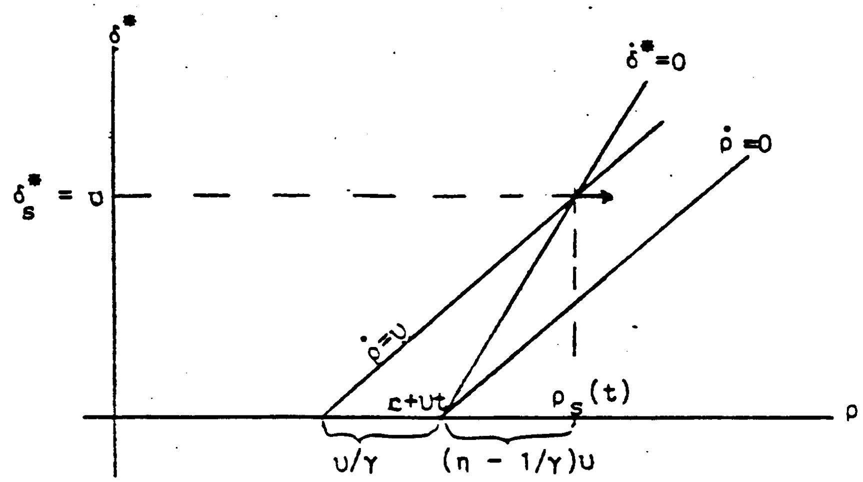

Figure IV: Interpretation of the Steady State ($\eta\gamma > 1$)

Steady state equilibrium is located in Figure IV by the intersection of the $\delta_p^* = \upsilon$ and the $\dot{\delta}^* = 0$ lines. The intercept on the ρ-axis of both the $\dot{\delta}^* = 0$ and the $\dot{\rho} = 0$ lines is found at $c + \upsilon t$. As before, the equilibrium values move over time if υ is non-zero. In the steady state the value of currency one continually falls relative to the value of currency two at the rate of υ. The intersection of $\delta_s^* = \upsilon$ and $\dot{\delta}^* = 0$ must also be consistent with $\dot{\rho} = \upsilon$ as shown in Figure IV.

With an increase in CS the curves again rotate clockwise around their ρ-intercepts and the equilibrium value of ρ, at a point in time, is further to the right. The value of $\rho_s(t)$ is greater than $c + \upsilon t$ by the amount $(\eta - 1/\gamma)\upsilon$ and, as CS becomes complete, $(\eta - 1/\gamma)$ approaches infinity. This means that the real value of the outstanding stock of currency one approaches zero (relative to the real value of currency two).

For a given non-zero value of υ, high currency substitution drives one currency out of circulation[1] in the sense that it reduces its (relative) real value to zero. Thus, there will be one inflation rate with complete CS regardless of the number of currencies in circulation. But only if υ is zero can more than one currency circulate.

The determination of equilibrium positions must be distinguished from the problem of stability. Although the condition for stability, $\eta < 1/\gamma + 1/\beta$, is not as stringent as before, it can be easily violated because $\eta = \alpha + 2\sigma$ can take on values approaching infinity. Since there are now two differential equations and, therefore, two characteristic roots of the system, cycles are possible and, in fact, rather likely. The cyclical behavior of the system is depicted in Figure V where the dynamics of the homogenous system has been combined with the secular rightward movement of the curves.

---

[1]The phraseology sounds like Gresham's Law but, in fact, it is quite different. The currency that is "driven out of circulation" in Gresham's Law is exported or melted down for bullion. In this model the demand falls towards zero for the currency whose value is expected to depreciate most rapidly.

Since gold and greenbacks were circulating at a market determined exchange rate after the Civil War and before the resumption of specie payments in 1879, the appreciating currency (initially gold) might have driven greenbacks out of circulation had it not been for the expectations of resumption at the pre-war price.

The Belgium franc and German mark circulated together in 1914 in the occupied part of Belgium. The franc had been allowed to float against gold since the currency plates had been sent to London just prior to the German invasion. Not unexpectedly, the francs appreciated, were hoarded, and disappeared from circulation until after the war. We are grateful to Lothar Guder for this example.

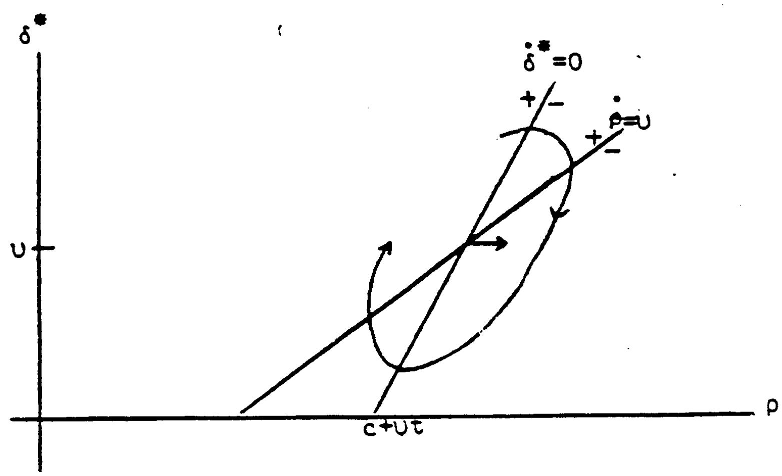

Figure V: Cyclical Adjustment Combined with Secular Movement

The system adjusts in a clockwise direction and, immediately after the variables cross the $\dot{\rho} = \upsilon$ curve from the left, the curves are then moving faster than $\rho(t)$ such that $\rho(t)$ appears to be decreasing. In other words, the elipse is drawn relative to the curves which are themselves moving at the rate of υ vis-a-vis the ρ-axis.[1]

---

[1]If the adaptive expectations and partial adjustment models were equivalent to one another, their combined use would not produce this cyclical behavior. Writers who have demonstrated their equivalence (e.g., Fiege (1967) and Waud (1968) have allowed real cash balances to adjust to the decrepancy in the money market rather than just the price level. Their adjustment models are certainly appropriate for individual behavior but, for the economy as a whole, we think that the adjustment equations (4) and (5) seem to be more appropriate.

Since the portfolio balance constraint between the two currencies required that $E = P_1/P_2$, the preceding analysis has implications for the rate of exchange as well as relative prices. Each of the preceding graphs would be unaffected if lnE $(= \rho = \ln P_1/P_2)$ were plotted on the horizontal axes. The cyclical adjustment in Figure V and the instability associated with high values of CS are as applicable to lnE as to ρ. The possibility of an unstable exchange rate has been emphasized by Boyer (1973) and used as an argument for fixed exchange rates between currencies that are good substitutes.[1]

Most of the preceding implications of currency substitution can be summarized under two major points, one concerning instability and the other concerning indeterminancy. The instability of the system has already been stressed. Although the exact stability conditions differed depending on the model of expectations formation, the fact that σ and, therefore, $\eta = \alpha + 2\sigma$ can go to infinity implied that instability is assured for high degrees of currency substitution. Whether adjustment paths are cyclical was found to depend on the particular assumption regarding inflationary expectations -- cyclical adjustment is not an inherent feature of currency substitution.

---

[1]For instance, the substitutability between U.S. dollars and Panamanian notes in the portofolios of Panama residents could be so great that the exchange rate would be unstable. In public comment on the policies of the new central bank of New Guinea, it has been proposed that the new currency should be allowed to float. But substitution between the new currency and the Australian dollar would produce instability. A discussion of New Guinea's monetary system is found in Arndt (1970).

The second major point to be drawn from the analysis is that complete currency substitution implies indeterminancy. When two currencies are perfect substitutes there is, in effect, only one currency and one rate of return or inflation rate.[1] Given this inflation rate and real wealth the real value of total currency outstanding will be determinant and can be labeled

$$(14) \qquad m = M_1/P_1 + M_2/P_2 = L_1(0, \pi_1^*, w) + L_2(0, \pi_2^*, w) = L(\pi^*, w)$$

When the two currencies are perfect substitutes, wealth holders are indifferent to the allocation of their portfolio between currencies one and two such that the composition is indeterminant. If the condition $E = P_1/P_2$ is substituted into equation (14) the result is

$$(15) \qquad m = M_1/(P_2E) + M_2/P_2 = (1/P_2)(M_1/E + M_2) = L(\pi^*, w)$$

The public will be indifferent to any values of M_1 and M_2 that leave m, P_1, P_2, and E unchanged.

An alternative way the indeterminancy manifests itself is that any set of values of P_1, P_2, and E are equilibrium values if they leave m unchanged and do not violate the condition $E = P_1/P_2$. This indeterminancy can be demonstrated from the preceding analysis for both assumptions regarding inflationary expectations. Rewriting equation (9) for convenience,

$$(9) \qquad \dot{\rho} = \frac{-\gamma}{1 - \eta\gamma} \quad (\rho - (c+\upsilon t)),$$

it is clear that the value of $\dot{\rho}$ is zero for all values of ρ if η is infinite. In Figures I through III, the phase lines coincide with the horizontal axis when σ and, therefore, η are infinite. Any values of ρ and E are equilibrium values provided only that m is unaffected and that $\ln E = \rho$. There is no pressure in the system for ρ or E to change since they are in meta-stable equilibrium.

[1] Perfect substitution might characterize, for instance, Vatican lira and Italian lira or Federal Reserve notes of different district banks. As these examples illustrate, perfect substitution does not imply indistinguishability. The analysis of this paper assumes that the two currencies, regardless of their degree as substitution, are distinguishable. The role of distinguishability has been emphasized by B. Klein (1974).

This indeterminancy can also be seem in the system depicted by Figures IV and V. If η is infinite, all the phase lines concide with the horizontal axis. This means that $\dot{\rho}$ and $\dot{\delta}^*$ are no longer influenced by the value of ρ or lnE. If the system is off the horizontal axis, it will still be unstable[1] -- both δ^* and ρ will go to plus or minus infinity. But the speed with which they go to infinity is not influenced by ρ. If the system is initially anywhere on the horizontal axis, there is no reason for it to move away.

The indeterminancy of the composition of currencies and of the exchange rate (subject to the conditions specified above) implies that no intervention is necessary to keep the exchange rate fixed. With perfect substitution between two currencies, transactions that can be effected with one currency can just as easily be accomplished with the second currency. Consequently, there need be no reason for persons to exchange currencies before purchasing certain goods. There might still be a balance of payments or redistribution of money between two regions, but no official intervention is necessary to peg the exchange rate. The measure of balance-of-payments disequilibrium defined with regard to intervention (as opposed to, say, residency in geographical regions) presupposes imperfect substitution between the currencies. Just a real trade theory presupposes imperfect factor mobility[2], balance-of-payments theory presupposes imperfect currency substitution.

---

[1]Here again is a difference between the system of Figures I-III and the system of Figures IV-V. In the first system, instability ceases with an infinite value of η and the system must be on the horizontal axis. In the second system, the system remains unstable with infinite currency substitution if the initial condition is not on the horizontal axis.

[2]This point is associated with R. Mundell (1957).

II. Competition Between Money Issuers and the Optimum Quantity of Money

In section I the rates of monetary expansion, u_1 and u_2, were taken as exogenous parameters. In this section the money growth rates are endogenized by assuming that the issuance of each currency is determined by profit maximizing behavior. The purpose is to determine the impact of currency substitution on the profit maximizing money growth rates and inflation rates.

The introduction of competition between money issuers forms a bridge between two separate strands of literature. On the one hand, there has been the literature concerned with the profit-maximizing inflation and money growth rates for a single issuer of money.[1] On the other hand, a newer literature has emerged concerned with the inflation and money growth rates necessary to induce people to hold the optimum quantity of money.[2] With the introduction of competition we are able to join the two sets of literature and determine conditions under which profit maximizing on the part of money issuers will produce the optimum quantity of money.

---

[1]Some of the early articles that have developed the idea of inflation as a tax have been Simon Newcomb (1865), J. Maynard Keynes (1922), and Milton Friedman (1953).

[2]A partial list of some of the main articles would include George Tolley (1957), James Tobin (1968), Alvin Marty (1968), and Milton Friedman (1969).

The steady state flow of real revenue obtained by the issuer of money one can be expressed as [1]

(16) $R_1 = \dot{M}_1/P_1 = u_1 m_1 = (\pi_1 + \lambda)m_1$

where $m_1 = M_1/P_1$

and λ = exponential rate of growth of real wealth.[2]

Leondardo Auernheimer (1974) has pointed out that the rate of inflation which maximizes the steady state flow of revenue will be different (lower) than the rate of inflation that maximizes the capitalized value of future money creation. Maximizing R_1 with respect to π_1 fails to take into account the height of the growth path of m_1 as determined by different choices of the steady state inflation rate. To the integral of the discounted value of R_1, therefore, must be added the one-and-for-all shift of m_1 (as it changes from an arbitrary inial value, m_1^o) when the bank of issue selects the optimum rate of monetary expansion. The present value (V) of future money creation beginning from t = 0 can be expressed as the sum of two components,

(17) $V_1 = \int_0^\infty (\pi_1 + \lambda)m_1(t)dt + m_1(0) - m_1^o$

where $m_1(0)$ is that value of m_1 immediately following the introduction (at time t = 0) of the steady state and profit-maximizing inflation rate. Using the linearized money demand functions from section I (equation (4)), the equations for m_1 and $\overset{o}{m}_1(t)$ can be written as

$$m_1^o = C_1\exp[-\alpha(\pi_1^o + r) - \sigma(\pi_1^o - \pi_2)]$$

and $m_1(t) = C_1 \exp[\lambda t - \alpha(\pi_1(t) + r) - \sigma(\pi_1(t) - \pi_2)]$

where π_1^o is the arbitrary inflation rage existing just prior to t = 0

and $\pi_1(t)$ is the inflation rate existing after t = 0.

---

[1]It is assumed in this section that π_i always equals π_i^*.

[2]The assumption of a constant growth rate for real wealth is appropriate for steady state analysis. The assumption was not employed in the previous section dealing with cyclical problems.

When the above expressions for $m_1(t)$ and m_1^o are substituted in (17), V_1 becomes[1]

(18) $$V_1 = C_1(\pi_1+\lambda)\exp[(-\alpha-\sigma)\ \pi_1-\alpha r+\pi_2] \int_0^\infty \exp[(\lambda-r)\tau]d\tau$$

$$+ C_1\exp[-\alpha r+\sigma\pi_2](\exp[(-\alpha-\sigma)\pi_1] - \exp[(-\alpha-\sigma)\pi_1^o]).$$

When integrated this yields[2]

(19) $$V_1 = \frac{(\pi_1+\lambda)}{(r-\lambda)} C_1\exp[(-\alpha-\sigma)\pi_1 -\alpha r-\sigma\pi_2] + \ldots$$

where the second term (represented by dots) is the same as in equation (18). When $\partial V_1/\partial\pi_1$ (calculated assuming π_2 is constant) is set to zero, the profit maximizing rate of inflation is found to be[3]

(20) $$\hat{\pi}_1 = (\alpha+\sigma)^{-1} -r.$$

Continuing with the assumption of section I that $\alpha_1 = \alpha_2 = \alpha$, the condition for a maximum V_2 with respect to π_2 (treating π_1 as a constant) yields the identical result, viz., $\hat{\pi}_2 = (\alpha+\sigma)^{-1} -r.$[4] In subsequent discussion, therefore, we can refer to the optimum rate of inflation without having to denote currency one or two in particular.

---

[1]Our derivation of the profit-maximizing inflation rate follows Auernheimer except for the presence of the π_2 and σ terms.

[2]The integration assumes that $\lambda < r$; otherwise, the integral does not converge at a finite value. A discussion of this assumption is found in Auernheimer.

[3]One can also show that $V_1''(\hat{\pi}_1) < 0$ to assure that $V_1(\hat{\pi}_1)$ is a maximum.

[4]One must rememberthat the exponential demand functions for the two monies were found as (semi-log) linear approximations in section I about a zero value of δ. For larger values of the differential inflation rate, the approximations will violate the original wealth constraint. By retaining the assumption that $\alpha_1 = \alpha_2 = \alpha$, however, $\delta = \pi_1 - \pi_2 = 0$ such that the exponential approximations will not violate the wealth constraint. If the assumption $\alpha_1 = \alpha_2$ were relaxed, it would be necessary to use non-exponential demand functions in order to preserve the wealth constraint. We have worked through such an exercise with linear demand functions but the results are not reported here since the analysis is extremely cumbersome and the results are the same.

Note also that the $\delta = 0$ condition implies that the system, although unstable as shown in section I, will remain in equilibrium.

Several implications of currency substitution follow immediately from equation (20). In the first place, larger values of σ will lower the profit-maximizing rates of inflation. Beginning from a point of zero currency substitution ($\sigma = 0$), equation (20) gives Auerheimer's result that $\hat{\pi} = \alpha^{-1} - r$. As σ increases, the profit-maximizing rate of inflation, $\hat{\pi}$, decreases monotonically as inspection of (20) shows. The greater the value of σ, the more easily a money issuer can capture a larger share of the market by lowering his inflation rate. In short, profit-maximizing competition between money issuers implies that greater substitution between their products will induce them to offer higher rates of return or (given zero interest payments on monies) lower rates of inflation.

The other implications of equation (20) concern the impact of perfect substitution between monies one and two. When σ approaches infinity, $\hat{\pi}_1$ and $\hat{\pi}_2$ approach r, the real yield on the alternative asset, capital. One can substitute $-\hat{\pi}_1 = r = -\hat{\pi}_2$ into equation (19) to show that $V(\hat{\pi}_1, \hat{\pi}_2) = 0$ when money issuers are competitive and currencies are perfect substitutes. All profits from the issuance of money are competed away when currencies are perfect substitutes. And, as one might expect from the usual theory of competition, the optimal production of monies is assured. The condition $-\hat{\pi} = r$ implies that ultimate wealth holders will hold sufficient cash balances for their marginal non-pecuniary return on money to equal the social cost of producing money, viz., zero.[1] Thus, the profit-maximizing and welfare-maximizing rates of inflation converge in the presence of competition and perfect substitution between currencies.

---

[1] The interpretation of the $-\hat{\pi} = r$ condition in terms of the marginal non-pecuniary services from holding money is developed in Friedman (1969).

When the profit-maximizing inflation rate induced by competition under perfect CS is substituted in equation (17), a simplification arises that permits further interpretation. The substitution of $\hat{\pi} = -r$ into (17) yields

$$(21) \qquad V = \int_0^\infty (\lambda - r)m(0)\exp[(\lambda-r)\tau]d\tau \quad + \quad m(0) - m^o$$

$$= -m(0) \qquad\qquad + m(0) - m^o$$

where the integrand was found from $um(t)\exp(-rt) = (\lambda+\pi)m(t)\exp(-rt) = (\lambda-r)m(0)\exp[(\lambda-r)t]$. The fact that the integral yields the negative value, $-m(0)$, is explained as follows: From the point $t = 0$ forward, the bank expands its liabilities at the rate λ generating a flow of real revenue, $\lambda m(t)$. But netted against this $\lambda m(t)$ term is the negative term $\pi m(t) = -rm(t)$ which represents a contraction of the bank's liabilities outstanding. Along institutional lines, one can think of the bank receiving interest payments of $rm(t)$. If the interest were paid in money, the bank could destroy the money and its equity holdings would, in effect, yield no return. If the interest on its equity holdings were paid in kind, the bank would have to sell its equity-received-as-interest to retire money outstanding to produce the deflation, $-\pi = r$. The value of the flow of equity purchased ($\lambda m(t)$) minus the flow of equity sold ($rm(t)$), when discounted by $\exp(\lambda-r)$, yields the net liability, $-m(0)$.

Consider now the remaining expression in (21), $m(0) - m^o$. Assuming that $\pi^o > r$ and $\sigma = \infty$ before $t = 0$, then $m^o = 0$. Then $m(0) - m^o = m(0)$ represents the share of the market that a bank can obtain by issuing currency bearing the competitive rate of return, r. The present value, V, in equation (21) can be interpreted as the net value of entering the banking business at $t = 0$. The entering bank immediately issues currency and purchases equities worth $m(0)$. When matched against the present value of providing more money for growth and servicing old and new balances at a return r, $V = -m(0) + m(0) = 0$.

The profit-maximizing condition, $-\hat{\pi} = r$, can be interpreted more generally as a condition for setting the real return on money equal to the real return on the alternative asset.[1] Using the notation from section I where i represented an explicit, pecuniary interest payment on money, the profit-maximizing condition can be re-expressed as $i - \hat{\pi} = r$. This condition leaves the actual inflation rate indeterminant as illustrated in Figure VI.[2]

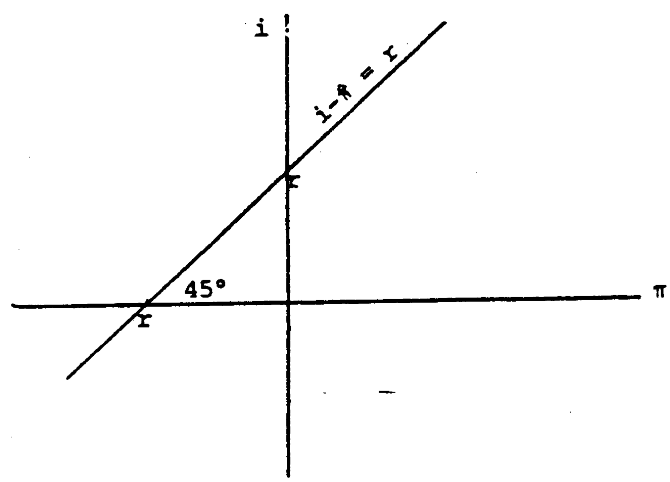

Figure VI: Profit-Maximizing Values of $\hat{i}$ and π

---

[1]This point can be justified by returning to the development of the money demand functions in section I. There it was assumed that the demand for money (say) one could be converted from $L_1(r_1-r_2,\ r_1-r_3,\ w) = L_1[(i_1-\pi_1)-(i_2-\pi_2),\ (i_1-\pi_2)-r_3,\ w]$ to $L_1[\pi_1-\pi_2,\ \pi_1+r_3,\ w]$ if the interest payments for money were constant. Without that assumption, the maximization procedure of this section could have been carried out with respect to r_1 holding r_2 constant producing the profit maximizing rate of real return $\hat{r}_1 = (\alpha + \sigma)^{-1} - r_3$. With perfect CS ($\sigma = \infty$), this reduces to $\hat{r}_1 = r_3$. The $\underline{r}$ used in the text, of course, is the real return on the third asset, r_3.

[2]The line drawn in Figure VI holds for $\sigma = \infty$, and there is a family of parallel lines for different values of σ.

The interpretation of equation (21) applies only to the horizontal intercept of the line in Figure VI. At that point the bank is creating money at the rate λ while simultaneously destroying interest proceeds from its equity holdings and thereby causing the deflation $-\pi = r$. But the integral of equation (21) equals the negative number, $-m(0)$, for any point along the curve in Figure VI. At the vertical intercept, for instance, the bank is returning the interest it receives to money holders as explicit interest on cash balances and, thereby, eliminating the deflation. At points further to the right on the curve, the bank is creating money in excess of growth but giving the "excess" away to money holders in proportion to their holdings -- i.e., as explicit interest that exactly compensates for the inflation. Both the bank and the public are indifferent to all points along the 45° line in Figure VI.

The profit-maximizing values of the inflation rate and interest rate imply that the real return on money is invariant to the rate of growth of money. In the money-and-growth literature, a change in the rate of growth of money is usually non-neutral because it alters the real return on money and, therefore, other real variables. But with competition between banks and perfect substitution between their products, money, in effect, carries an escalator clause since the higher inflation rates are automatically compensated by higher explicit interest payments. Neutrality in this dynamic sense reflects the fact that money issuers' profits are invariant to u and π -- there is no redistribution of resources from the private sector to money issuers as a result of larger or smaller rates of monetary creation.

A final implication of competition under perfect currency substitution is that the rate of change of the exchange rate will equal the differential interest payments on monies.[1] If both money issuers pay zero nominal interest (and generate deflations of $\hat{\pi} = -r$), then the exchange rate will be stable. And even if the banks choose different nominal interest rate payments on their respective monies, the rate of change of the exchange rate will be constant.

Although the constancy of the exchange rate is implied by competitive behavior and zero interest payments on monies, it can also be produced by collusion. In the face of increased currency substitution, banks of money issue could respond by pursuing the same inflation rates in order to eliminate the incentives of money holders to reallocate their portfolios between currencies. This would enable money issuers to stabilize their market shares and retain monopoly profits. Since the collusion would induce them to pursue the same inflation rates, the exchange rate would be stable. Thus, while a stable exchange rate can be generated by competition under perfect CS, it is also consistent with collusion for any degree of currency substitution.

---

[1] This is predicated on our previous assumption that the exchange rate must equal the ratio of the values of the two monies in terms of the same bundle of goods. The difference between this assumption and the usual interpretation of purchasing power parity was discussed in section I.

References

Arndt, H.W. (1971), "The Future of New Guinea's Monetary System", Eonomic Record, March.

Auernheimer, L. (1974), "The Honest Government's Guide to the Revenue from the Creation of Money", Journal of Political Economy, 82 (May/June).

Boyer, R. (1972), "Nickels and Dimes", Federal Reserve Board, unpublished manuscript.

__________(1973), "Substitutability between Currencies and between Bonds: A Theoretical Analysis of Gresham's Law", University of Western Ontario, unpublished manuscript.

Branson, W. (1968), Financial Capital Flows in the U.S. Balance of Payments, North-Holland.

Cagan, P. (1956), "The Monetary Dynamics of Hyperinflations", in Studies in the Quantity Theory of Money, ed. by Milton Friedman, University of Chicago Press, Chicago.

Feige, E. (1967), "Expectations and Adjustments in the Monetary Sector", American Economic Review, LVII (May), pp. 462-473.

Friedman, M. (1953), "Discussion of the Inflationary Gap", in Essays in Positive Economics, Chicago, pp. 253-57.

__________(1966), "Interest Rates and the Demand for Money," J. of Law and Economics.

__________(1969), "The Optimum Quantity of Money", in The Optimum Quantity of Money and Other Essays, Aldine Publishing Co., Chicago.

__________and A. J. Schwartz (1970), Monetary Statistics of the United States, National Bureau of Economic Research, Columbia University Press.

Girton, L. and D. Henderson (1976), "Financial Capital Movements and Central Bank Behavior in a Two-Country, Short-run Portfolio Balance Model", Journal of Monetary Economics, February.

Keynes, J. M. (1922), "Inflation as a Method of Taxation", Manchester Guardian Commercial, Julu 27.

Klein, B. (1974), "The Competitive Supply of Money," Journal of Money Credit and Banking (Nov.)

Marty, A. L. (1968) "The Optimal Rate of Growth of Money," Journal of Political Economy, 76 (July/Aug.).

Mundell, R. (1957), "International Trade and Factor Mobility." American Economics Review, XLVII (June).

Newcomb, S. (1865), Financial Policy during the Southern Rebellion, J.D. Appleton Co., New York. Reprinted, 1969, Greenwood Press, New York.

Tobin, J. (1969), "A General Equilibrium Approach to Monetary Theory", Journal of Money Credit and Banking, I, February.

__________(1968), "Notes on Optimal Monetary Growth," Journal of Political Economy 76 (July/Aug.).

Tolley, G. S. (1957), "Providing for Growth of the Money Supply." Journal of Political Economy 65 (Dec.).

Waud, R. (1968), "Misspecification in the 'Partial Adjustment' and "Adaptive Expectations' Models", International Economic Review, 9, June, pp. 204-217.