Transaction Services, Inflation, and Welfare

S. Rao Aiyagari, Toni Braun,
and Zvi Eckstein*

ABSTRACT

This paper is motivated by a variety of empirical observations on the comovements of currency velocity, inflation, and the relative size of the "credit services" sector. By the credit services sector we mean the part of banking and credit sector which provides alternative means of transactions to using currency as well as other services which help people economize on currency. We incorporate the credit services sector into a monetary growth model. Our model makes two specific and new contributions. The first is to show that direct quantitative evidence on the welfare cost of low inflation using measures of the relative size of an appropriately defined credit services sector for the U.S.—essentially the cost incurred by banks and credit unions in providing demand deposit and credit card services—is consistent with the welfare cost measured using an estimated money demand curve following the classic analysis of Bailey (1956) and the more recent analysis of Lucas (1993). Both of these measures amount to about 0.5 percent of GNP. The second contribution is in providing welfare cost of inflation estimates over a range of inflation rates which have some new features. We find that the total welfare cost of inflation remains bounded at about 5 percent of consumption.

1. Introduction

This paper is motivated by a variety of empirical observations, to be described later, on the comovements of currency velocity, inflation and the relative size of the "credit services" sector. By the credit services sector we mean the part of banking and credit sector which provides alternative means of transactions to using currency as well as other services which help people economize on currency. These observations motivate our development of a monetary growth model in which a costly credit services sector provides alternative means of transactions to using currency - as in the model of Gillman [1993].

Aside from being consistent with the empirical observations which motivate our modeling, our model makes two specific and new contributions. The first is to show that direct quantitative evidence on the welfare cost of low inflation is consistent with the welfare cost measured by using an estimated money demand curve following the classic analysis of Bailey [1956] and the more recent analysis of Lucas [1993]. This is done by showing that the model establishes a link between money demand and the welfare cost of inflation on the one hand and the welfare cost of inflation and the relative size of the credit services sector on the other hand.[1] We then show that

---

[1]Regarding the welfare implications of resources used for providing transaction services Lucas [1993, p. 33] writes, "In a monetary economy, it is in everyone's private interest to try to get someone else to hold non-interest bearing cash and reserves. But someone has to hold it all, so all of these efforts must simply cancel out. All of us spend several hours per year in this effort, and we employ thousands of talented and highly-trained people full-time in the financial industry to help us. These person-hours--hundreds of billions of dollars worth--are simply thrown away, wasted on a task that should not have to be performed at all."

when the model is calibrated from a semi-log money demand function estimated from U.S. data the resulting welfare cost estimate is remarkably consistent with direct measures of the relative size of an appropriately defined credit services sector for the U.S. - essentially the cost incurred by banks and credit unions in providing demand deposit and .credit card services. Both of these measures amount to about 0.5 percent of GNP.

The second contribution of this paper is that it provides welfare cost of inflation estimates which have some new features. We find that the total welfare cost of inflation is bounded at a fairly low level. The total welfare cost of inflation in our model reflects two distinct effects of inflation. The first is the effect noted above that inflation affects the share of total output that is devoted to transaction services provided by the credit services sector. The second is that inflation distorts labor supply and investment decisions and thereby affects total output as in Cooley and Hansen [1989,1991] and Stockman [1985]. It turns out that the nature of the inflation induced distortions in labor supply and capital accumulation depend on the functional form of the money demand function. For our specification which fits U.S. data quite well, these costs are bounded no matter how high inflation is. Further, the relative size of the credit services sector also remains bounded with inflation. Consequently, the *total* welfare cost of inflation remains bounded with inflation. For our parameterization this bound occurs at about 5 percent of consumption.

The rest of this paper is organised as follows. In section 2 we describe the empirical observations concerning comovements in inflation, M-0 velocity, and the share of the banking and credit sector, which motivate our

2

study and modeling approach. In section 3 we describe our model and show that it is consistent with these observations. In section 4 we provide direct quantitative evidence on the welfare cost of low inflation and show that this evidence is consistent with the welfare cost measured using an estimated money demand curve. In Section 5 we describe our model's quantitative implications for the welfare cost of inflation and in Section 6 we offer some concluding remarks concerning the robustness of our results to alternative specifications.

2. Co-movements in Inflation, Velocity, and the Relative Size of the Banking and Credit Sector

In this section we provide some evidence of comovements among M-0 velocity, the relative size of the banking and credit sector, and inflation. This evidence is of two types. First we show that in high inflation countries movements in velocity and the relative size of the banking and credit sector tend to parallel movements in inflation. Our evidence comes from Israel, Argentina and Brazil which have experienced episodes of very high inflation during the eighties, Austria, Hungary, Poland and Germany which experienced hyperinflations after WWI. Second we show that for the U.S. even after accounting for the comovements of velocity with inflation there is residual comovement between velocity and the relative size of the banking and credit sector in post-WWII data.

We should acknowledge that the relative size of the banking and credit sector is an imperfect proxy for the relative size of the credit services sector. Banks perform many functions other than those that help people economize on currency and many functions that help people economize on

currency are performed outside of banks. Unfortunately, data limitations compel us to use these crude proxies. However, we feel it is plausible that for countries which have experienced episodes of very high inflation rates over a relatively short period of time much of the change in the relative size of the banking and credit sector is due to changes in services which help people economize on currency. In our discussion of velocity we use the monetary base as the measure of money. We feel this is appropriate for our purposes since the relevant distinction in our analysis is between costlessly provided outside money and costly inside monies.

We first describe observations from some high inflation countries.

High Inflation Countries

The strong association between the relative size of the banking and credit sector and inflation is most transparent in economies that experienced an accelerating inflation and then an end to the inflation period. In Figure 1a we display data on inflation, the share of employment in banks, and the total number of bank accounts for Israel which experienced a period of high and accelerating inflation from 1970-1985. This Figure shows a significant upward trend in the share of employment in banks and in the number of bank accounts from 1968 to 1985. In July 1985 the Israeli government implemented a stabilization program which resulted in an abrupt drop of the annual inflation rate from a high of close to 500 percent to a low of 16-20 percent. It can be seen that during the period 1986 to 1989 the share of employment in banks and the total number of bank accounts dropped.[2]

---

[2]Data on the number of teller machines per 100 people and the area of banks per 1000 people from Aiyagari and Eckstein [1994, Table 1] also show a

4

Argentina has also experienced a protracted period of accelerating inflation as shown in Figure 1b. In April 1991 the Argentinian government implemented a stabilization program that abruptly reduced the annual inflation rate from a high of close to 350 percent to 10 percent over a period of several months. Figure 1b shows a strong positive relationship between the employment share in the banking sector, and the CPI inflation rate.[3] The banking employment share increases rapidly during the years 1975-1976, a period in which the annual inflation rate exceeded 150 percent. The banking employment share peaks in 1980 and then gradually falls thereafter. The number of bank branches shows a similar pattern of comovement with inflation.

Brazil is a third country that has experienced a long period of accelerating inflation during the eighties. The latest effort to reduce the inflation rate in Brazil began in December of 1994 and it is still too soon to get data for the post inflation period. However, recent reports (see e.g. the *Economist* January 21, 1995) suggest that the story of Israel and Argentina is now being repeated in Brazil. Figure 1c displays data on the value added share of the financial sector in GDP, and inflation, in Brazil. Once again we see that the long period of rising inflation is accompanied by an increase in the relative size of the banking sector. Evidence on check clearing is also consistent with this view. During the late eighties Brazil cleared nearly twice as many checks a year (relative to GNP) as the U.S. and

---

similar pattern of comovement with inflation.

[3]The measure of banking sector employment summarized in this Figure includes employees in private banks, public banks and other financial entities. Employees of the central bank of Argentina are excluded.

they were cleared faster.[4] Evidence from Brazil and Argentina also indicates that banks largely abandon traditional banking activities during high inflations and focus instead on activities that help individuals economize on their holdings of cash.[5]

Wicker [1986] documents sharp increases in unemployment in the banking sector in Austria, Hungary and Poland during the post-stabilization period in the 1920's. With reference to the Hungarian case he says (1986, p. 358), *"The most striking thing about these figures (incidence of trade union unemployment) is the extraordinary increase in the number of unemployed in the financial sector - 31.5 percent of the total net increase of 13,000. All of this increase can be attributed to the ending of hyperinflation which had increased substantially the money market as well as other operations of commercial banking."* For Austria, Wicker [1986] describes a similar pattern where 10,000 workers in the banking sector lost their jobs immediately after stabilization. With regard to the German hyperinflation of the early 1920's

---

[4]In Brazil 7 billion checks a year were cleared during the late eighties. For purposes of comparison the Federal reserve system cleared about 19 billion checks in 1990. If we scale down the U.S. figure using the fact that GDP in Brazil is about 20 percent of GDP in the U.S. we get a figure of 3.8 billion checks a year. With this scale adjustment Brazil cleared nearly twice as many checks a year as the U.S. Moreover, in Brazil 95 percent of all checks cleared within twenty four hours whereas in the U. S. only 85 percent of all checks cleared within twenty four hours.

[5]In Brazil, the maximum term on loans in 1990 was 180 days and most loans had a maturity of less than 60 days. In Argentina (which also experienced very high inflation during the eighties) seventy percent of bank deposits in January of 1991 were held in accounts that had a maturity of less than 13 days. Moreover, during periods of high inflation banks make huge returns on the "float" they hold. The *Economist* (January 21, 1995) notes that Brazilian banks made as much from managing float as from traditional banking activities. In December of 1994 when inflation fell to about 1.25 percent per month from over 30 percent per month six months earlier, revenues from float dropped to 5 percent from 25 percent of total revenue.

and its stabilization, Garber [1982], Graham. [1930] .and Bresciani-Turroni [1937] indicate that there was a substantial increase in employment in the banking sector during the period of accelerating inflation and a decrease in employment in that sector after stabilization.

A Low Inflation Country: U.S.

Figure 2a displays the time series of actual inverse velocity and its fitted values using interest rates and a semi-log money demand equation with a unit income elasticity over the period 1930-89. In Figure 2b we display the time series of actual inverse velocity as well as its fitted values using interest rates and the employment share of the banking and the credit sector.[6] A comparison of these two Figures clearly demonstrates that the banking and credit sector employment share helps explain movements in velocity even after accounting for movements in inflation.[7]

In the next section we describe our monetary growth model and show that this model is consistent with the observations described in this section.

---

[6]Our definition of the banking and credit sectors consists of the "banking" and "credit agencies other than banks" sub-sectors of the Finance, Insurance and Real Estate (FIRE) of the NIPA. Our measure of GNP is the usual GNP less net exports and government purchases.

[7]Melnick [1995] reports similar evidence for Israel. In addition to the contemporaneous effects displayed in Figure 2, vector autoregression (VAR) analysis reveals significant dynamic interactions. For the vector of observables consisting of inverse velocity, the nominal interest rate, and the employment share of the banking and credit sector, we find that the shocks corresponding to inverse velocity and the employment share explain most of the variance in these variables. Specifically, the employment share variable explains up to half the variance in inverse velocity and inverse velocity explains up to twenty percent of the variance in the employment share upto horizons of ten quarters.

3. A Monetary Growth Model With Credit Goods Production

Our model is of the cash-in-advance (CIA) variety; however, the key distinctive feature of our model is that cash goods and credit goods are perfect substitutes in consumption and investment but differ in their production technologies; specifically, production of credit goods is a resource using activity.[8] The model is a competitive equilibrium model and is described in terms of the behavior of its three decision units, namely, the households, the producers, and the government, and equilibrium conditions. We start by describing the technology and the behavior of the representative producer.

The Technology and Producer Optimization

Total output, denoted Y_t^* is produced using capital (K_t) and labor (N_t) via a constant returns to scale production function F. Total output can be used to produce goods (Y_t) for investment and consumption on a one-to-one basis or can be used to produce credit services (S_t) where one unit of credit services requires q_{st} units of total output. Therefore, we have

---

[8]This approach is unlike the cash goods/credit goods model of Lucas and Stokey [1987] in which cash goods and credit goods are perfect substitutes in production but not in preferences and is similar to the approach in Gillman [1993]. We think that our approach is sensible since gasoline is gasoline regardless of whether one pays for it with cash or check or credit card. The distinction ought to be traced to the additional resources in the form of the services of the banking sector or credit card companies that are required for the purchase of "credit gasoline" as opposed to "cash gasoline." This implies that differences in the technology of producing cash versus credit gasoline will be reflected in different prices for credit gasoline and cash gasoline. This is indeed the case for gasoline at many gas stations around the U. S. at which different prices are listed depending on whether one pays for gas with cash or with credit. Aside from descriptive realism one advantage of adopting our modeling strategy is that data can be brought to bear in calibrating the model and developing the model's quantitative implications.

$$(3.1) \qquad Y_t^* = F(K_t, \theta_t N_t) = Y_t + q_{st} S_t.$$

The labor augmenting technology shock θ_t is assumed to follow a trend stationary stochastic process with average growth rate of g. The cost of producing credit services q_{st} is assumed to follow a stationary stochastic process.[9]

The output of goods Y_t is assumed to be uniformly distributed across a continuum of types indexed by $z \in [0,1]$. A unit of type z good can be used in two ways. It can either be used to produce a unit of the cash good on a one-to-one basis or it can be combined with $\mathcal{R}(z, \varepsilon_t)$ units of credit services and used to produce a unit of the credit good. We assume that $\mathcal{R}$ is strictly increasing in z with $\mathcal{R}(0, \varepsilon_t) = 0$. That is, a good that is indexed with a higher value of z requires more services to be transformed into a credit good.[10] ε_t is a random variable assumed to follow an exogenous stationary stochastic process independent of θ_t and q_{st}.

---

[9]Our specification is equivalent to one where there are two different sectors producing goods and credit services with production functions that are identical except for a scale factor which is q_{st}. The value added share of credit services in total output is then also equal to the shares of labor and capital in the credit services sector. This specification is adopted for convenience and is reasonably consistent with the data. Later we will assume that the production function F is Cobb-Douglas so that data on the labor share can be used to estimate its parameter. Empirically, labor share in the banking and credit sector (as defined earlier, see note (?) is not much different than the labor share in GNP. The goods producing sector of the model is identified with GNP less net exports, government purchases and value added in the banking and credit sector. Using U.S. annual data from 1947 to 1989, we find labor shares of 0.62 and 0.59 in the banking and credit sector and the goods producing sector, respectively.

[10]Our specification of the technology for producing credit goods is similar to that of Gillman [1993].

A unit of type z good (either cash good or credit good) can either be consumed or used for gross investment, i.e., to produce new capital goods.[11] Letting i_t and $i_t(z)$ be gross investment and the amount of type z good used for gross investment, the technology for gross investment is of the following Leontief fixed coefficients type.

$$(3.2) \qquad i_t = \inf_z \{i_t(z)\}.$$

We now describe some implications of producer optimization. It's obvious that $p_{st}/p_{1t} = q_{st}$ where p_{1t} and p_{st} denote the prices of cash goods and credit services, respectively. Further, $w_t = \theta_t F_2(K_t, \theta_t N_t)$ and $r_t = F_1(K_t, \theta_t N_t)$ where w_t and r_t are the wage and the rental on capital in units of the cash good, respectively. Letting $p_{2t}(z)$ denote the price of type z credit goods we have

$$(3.3) \qquad p_{2t}(z) = p_{1t} + p_{st}\mathcal{R}(z, \varepsilon_t) = p_{1t}[1 + q_{st}\mathcal{R}(z, \varepsilon_t)].$$

which follows from the fixed coefficients technology for the production of credit goods. Notice that our assumptions on $\mathcal{R}(.)$ imply that the credit price of a good is increasing in the type index z.

The Representative Household

There is a representative infinitely lived household which has one unit

---

[11]Thus, we do not arbitrarily designate consumption goods as cash goods and investment goods as credit goods as in, for e.g., Cooley and Hansen [1989]. In our model both cash goods as well as credit goods may be used for consumption and/or investment.

of labor endowment available each period. The household consumes the amount $c_t(z)$ of type z goods and supplies the amount n_t of labor input in each period t. The household's preferences are given by the following expected discounted sum of utilities of consumption and leisure

$$(3.4a) \qquad E_0\{\Sigma_{t=0}^{\infty} \beta^t U(c_t, 1-n_t)\}, \quad 0 < \beta < 1, \text{ where}$$

$$(3.4b) \qquad c_t = \inf_z\{c_t(z)\}$$

is (composite) consumption.[12]

The household purchases $\chi_t(z)$ units of type z good and uses these goods for consumption and gross investment. In view of (3.2) and (3.4b) we have that

$$(3.5a) \qquad i_t(z) = i_t, \quad c_t(z) = c_t, \quad \chi_t(z) = \chi_t, \text{ for all } z, \text{ where}$$

$$(3.5b) \qquad \chi_t = c_t + i_t = c_t + k_{t+1} - (1-\delta)k_t,$$

where k_t is the stock of capital that the household has at the beginning of period t and δ is the depreciation rate of capital.

We now describe the household's optimization problem. We start with the household's CIA and budget constraints. The total purchases χ_t will be partly in the form of cash goods and partly in the form of credit goods. Since the cash price of goods is constant across types whereas the credit price of goods is increasing in the type index z there will be a particular

---

[12]Our specification is the same as that used by Schreft [1992] and turns out to be much more convenient since it enables us to retain the one-good structure of the standard monetary growth model.

cut-off index z_t^* such that the household will purchase goods with indices below z_t^* as credit goods and goods with indices above z_t^* as cash goods. Thus, credit goods purchases equal $\chi_t z_t^*$ and cash goods purchases equal $\chi_t(1-z_t^*)$.

In addition to capital, the household begins period t with m_t units of money and b_t units (in terms of face value) of nominal bonds. The household also receives nominal lump sum transfer payments from the government in the amount X_t. As is usual in CIA models of money there is a financial market in which the household can rearrange its portfolio of money and bonds. Once this is done the financial market closes and the goods markets (for purchasing cash goods and credit goods) open. In the cash goods market the household can purchase cash goods subject to the following CIA constraint.

(3.6) $(m_t+X_t)/p_{1t} + b_t/p_{1t} - b_{t+1}/[(1+R_t)p_{1t}] \geq \chi_t(1-z_t^*).$

In (3.6) R_t is the nominal interest rate from t to t+1. Note that the left side of (3.6) is the amount of cash the household has available after the close of the financial market.

The household's budget constraint is as follows where $\int_0^{z_t^*} \chi_t p_{2t}(z)dz$ is the total nominal cost of purchasing credit goods.

(3.7) $(m_t+X_t)/p_{1t} + b_t/p_{1t} + w_t n_t + r_t k_t \geq \chi_t(1-z_t^*) + \int_0^{z_t^*} \chi_t p_{2t}(z)dz/p_{1t} +$
$$m_{t+1}/p_{1t} + b_{t+1}/[(1+R_t)p_{1t}].$$

Note that wage and rental income is received after the close of the financial market and cannot be used for cash goods purchases. Further, labor

and capital services are treated as costless credit goods, i.e., they do not require credit services for exchange.[13]

The household maximizes the expected discounted sum of utilities in (3.4a) subject to the following constraints: total purchases equal purchases for consumption and investment (3.5b), the CIA constraint (3.6) and the budget constraint (3.7). The solution to the consumer's optimization problem is characterized by the following first order necessary conditions (FONCs).

(3.8a) $U_{l,t}/U_{c,t} = w_t/(1+\tau_t),$

(3.8b) $U_{c,t}/(1+\tau_t) = \beta E_t[U_{c,t+1}(p_{1t}/p_{1,t+1})(1+R_{t+1})/(1+\tau_{t+1})],$

(3.8c) $U_{c,t} = \beta E_t[(1-\delta+r_{t+1}/(1+\tau_{t+1}))U_{c,t+1}],$

(3.8d) $1+R_t = p_{2t}(z_t^*)/p_{1t},$

where

(3.9) $1+\tau_t = (1-z_t^*)(1+R_t) + \int_0^{z_t^*} p_{2t}(z)dz/p_{1t},$

and $U_{l,t}$ and $U_{c,t}$ denote the marginal utilities of leisure and consumption, respectively, at date t.

Government

The government sets the money growth rate $x_t \equiv (M_{t+1}-M_t)/M_t$ in such a way that x_t follows a stationary stochastic process which is independent of $\{\theta_t, q_{st}, \varepsilon_t\}$.

---

[13]A brief discussion of these assumptions is given in the concluding section.

Equilibrium

The following conditions (which are pretty self-explanatory) need to hold in equilibrium.

(3.10a) $F(K_t, \theta_t N_t) = Y_t + q_{st} S_t,$

(3.10b) $\chi_t = Y_t = C_t + K_{t+1} - (1-\delta)K_t,$

(3.10c) $Y_t \int_0^{z_t^*} \mathcal{R}(z, \varepsilon_t) dz = S_t,$

(3.10d) $(k_t, n_t, c_t, m_t) = (K_t, N_t, C_t, M_t),$ K_0, M_0 given.

In (3.10c) the expression on the left is the total amount of credit services used in producing credit goods.

This completes the description of the model.

We now show that the qualitative predictions of the model are consistent with the empirical observations in section 2. These observations are that inflation, the relative size of the banking and credit sector, and velocity tend to comove and that there is residual comovement between the relative size of the banking and credit sector and velocity even after accounting for the comovement due to inflation. The key element of our model on which its implications rest is the link between money demand and the value added share of the credit services sector, which we now proceed to derive.

Money Demand

The money demand function in our model is derived from the household's

FONC (3.8d) which is a Baumol [1952] type condition that sets the opportunity cost of cash equal to the cost of credit services for the marginal good. A household can either purchase an extra unit of the cash good at price p_{1t} by borrowing in the financial market at the interest rate R_t, thereby reducing its cash holding at t+1, or it can purchase a unit of the marginal credit good (with index z_t^*) at the price $p_{2t}(z_t^*)$ and reduce its cash holding at date t+1. Since cash goods and credit goods are perfect substitutes in consumption and investment the marginal condition (3.8d) must hold.

We can use (3.3) to rewrite (3.8d) in the following way.

$$(3.11) \qquad R_t = q_{st} \mathcal{R}(z_t^*, \varepsilon_t).$$

This condition determines the cut-off index z_t^* as illustrated in Figure 3. Goods with indices lower than z_t^* are purchased as credit goods and goods with indices higher than z_t^* are purchased as cash goods. Holding the relative price of credit services fixed an increase in the nominal interest rate leads to an increase in the fraction of goods bought on credit (since $\mathcal{R}$ is increasing) and, hence, a decrease in the fraction of goods bought with cash.

Letting $m_t = M_{t+1}/p_{1t}$ denote real balances and noting that (3.10b) and the CIA constraint imply that $z_t^* = 1 - m_t/Y_t$, we can rewrite (3.11) to obtain the following inverse money demand function for our model.

$$(3.12) \qquad R_t = q_{st} \mathcal{R}(1 - m_t/Y_t, \varepsilon_t).$$

From (3.12) we see that the money demand function for our model reflects the technology for producing credit goods.

Credit Services

Let $\phi_t = q_{st}S_t/Y_t{}^*$ denote the value added share of credit services in GNP. We now derive the relation between ϕ_t and money demand. To exhibit this it is convenient to let a_t denote q_{st} times the integral in (3.10c). Thus a_t measures the amount of credit services used (in units of goods) per unit of goods produced (see area A in Figure 3 for an illustration). Note that a_t is also the area under the inverse money demand curve normalized by the highest possible value of real balances. To see this let A_t denote the area under the inverse money demand curve and note that

$$(3.13) \qquad a_t \equiv \int_0^{z_t^*} q_{st}\mathcal{R}(z,\varepsilon_t)dz = \int_{m_t}^{Y_t} q_{st}\mathcal{R}(1-m/Y_t,\varepsilon_t)dm/Y_t = A_t/Y_t.$$

The relation between a_t and ϕ_t is as follows.

$$(3.14) \qquad a_t \equiv \int_0^{z_t^*} q_{st}\mathcal{R}(z,\varepsilon_t)dz = q_{st}S_t/Y_t = \phi_t/(1-\phi_t).$$

This equation exhibits the key link in our model between the value added share of the credit services sector and money demand. This link arises because in our model the area under the inverse money demand curve measures the amount of credit services used which implies that anything that affects velocity will affect the value added share.

16

Comovements with inflation

Equation (3.14) implies a positive comovement between inflation and the relative size of the credit services sector. As inflation and the nominal interest rate rise the opportunity cost of cash goods rises and pushes consumers to purchase less cash goods and more credit goods (see 3.12 and Figure 3). Thus, real balances fall, velocity rises, and the share of credit goods in total goods rises. The fall in real balances raises the normalised area under the inverse money demand curve a_t. Equivalently, the greater quantity of credit goods requires a greater quantity of credit services (see 3.14 and Figure 3). Hence, capital and labor move into the credit services sector and its value added share (ϕ_t) rises. Thus our model can account for the comovement among inflation, velocity and the relative size of the banking and credit sector.

Our model is also consistent with the strong linkage between inverse velocity and the employment share of the banking and credit sector that remains after accounting for their comovement due to inflation (see Figures 2 and the discussion in section 2). This is because, in addition to inflation, improvements in the technologies for producing credit services and credit goods can also generate comovements between velocity and the value added share of credit services. Such technological improvements can be interpreted as decreases in q_{st} and in ε_t. Both of these lower the cost of buying goods on credit and thereby increase the share of credit goods, reduce money demand and, hence, raise currency velocity. Noting that $z_t^* = 1-m_t/Y_t$ and using (3.12) and (3.14), we can express m_t/Y_t and ϕ_t as functions of $(R_t, q_{st}, \varepsilon_t)$.

(3.15a) $\quad m_t/Y_t = \mathcal{M}(R_t, q_{st}, \varepsilon_t),$

(3.15b) $\quad \phi_t = \Phi(R_t, q_{st}, \varepsilon_t).$

Therefore, movements in velocity unexplained by interest rates ought to be correlated with movements in the value added share unexplained by interest rates. Alternatively, the value added share ϕ_t (equivalently, the employment share) ought to improve the fit of the money demand equation. Thus our model is consistent with Figures 2 discussed in section 2.[14]

Having shown that our model is consistent with the qualitative evidence linking inflation, currency velocity and the banking and credit sector, we now turn to examine direct evidence on the welfare cost of inflation which is consistent with an empirically estimated money demand function. This is one of the two main contributions of this paper and is described in the next section.

4. Direct measure of the welfare cost of inflation

As we saw before a key feature of our model is that a rise in inflation leads to an expansion of the credit services sector. It is easy to see that resources allocated to this sector represent a social waste. In order to focus on this misallocation of resources let's assume temporarily that labor supply is inelastic and fixed at n and that gross investment is exogenously fixed at the constant fraction $(g+\delta)$ of the capital stock so that the

---

[14]It is also obvious that if the variables $(R_t, q_{st}, \varepsilon_t)$ are serially correlated then the there will be dynamic linkages among the vector of observables $(R_t, m_t/Y_t, \phi_t)$. Thus, our model is also qualitatively consistent with the VAR based results on dynamic linkages among these variables mentioned in section 2.

capital stock grows exogenously at the same rate as θ. We can rewrite (3.10b) using (3.10a) as follows.

$$(4.1) \qquad c_t + K_{t+1} - (1-\delta)K_t = Y_t = (1-\phi_t)F(K_t, \theta_t n_t).$$

Now note that when the nominal interest rate is zero all goods are sold as cash goods and the share of credit goods z_t^* is zero (see 3.11 and Figure 3). Hence, the normalised area under the inverse money demand curve a_t is zero and, thereby, the value added share ϕ_t is zero (see 3.13 and 3.14). It follows that consumption is at its highest. When the nominal interest is positive, ϕ is positive and acts as a tax on GNP reducing consumption and, hence, welfare. It's obvious that ϕ is an exact measure of the loss in consumption relative to GNP when total resources are held fixed. *Thus, we have shown that resources diverted to the credit services sector from the goods production sector are misallocated and that the value added share in the credit services sector is one component of the welfare cost of inflation.*

Therefore, one way to measure the welfare cost of inflation is by directly measuring the relative size of an appropriately defined credit services sector. A second way to measure the welfare cost of inflation is by calculating the area under the inverse money demand curve (holding Y_t fixed). To see that this is equivalent note that the welfare cost (denoted ΔW) can be written as

$$(4.2) \quad \Delta W_t = -\Delta C_t = \phi F(K_t, \theta_t n_t) = q_s S_t = q_s Y_t \int_0^{z_t^*} \mathcal{R}(z, \varepsilon) dz$$

$$= \int_{m_t}^{Y_t} q_s \mathcal{R}(1-m/Y_t, \varepsilon)dm.$$

Thus, from the last equality in (4.2) we see that the welfare cost corresponds exactly to the area under the inverse money demand curve.[15]

We now provide empirical evidence from U.S. data which shows that the area under an estimated money demand curve does indeed match well with direct measures of the relative size of an appropriately defined credit services sector for the U.S.

We first describe how the model is parameterized.

Parameterization

We start with the specification of the money demand function. The only restrictions on the functional form of money demand are that it have a unit income elasticity, be decreasing in the nominal interest rate, and be bounded even as the nominal interest rate goes to zero (see 3.12). The last restriction is due to the CIA nature of our model so that money demand remains bounded by Y_t. Thus our model accommodates a wide variety of functional forms for money demand within a one-sector monetary growth model with standard specifications of preferences, technology, and growth, and can be calibrated from empirically estimated money demand functions. We use the empirically popular semi-log form of money demand which fits U.S. data quite well (see Lucas 1993). This is obtained from the following specification of

---

[15]This is the component of the welfare cost of inflation captured in the models of Lucas [1993]. Interestingly, Lucas finds that his numbers are approximately the same as the area under the inverse money demand curve; in our model they are exactly the same.

the $\mathcal{R}(.)$ function

(4.3) $\mathcal{R}(z,\varepsilon_t) = -\eta[\ln(1-z)-\varepsilon_t]$.

This leads to the following semi-log specification of money demand.

(4.4) $\ln(m_t) = -R_t/(q_{st}\eta) + \ln(Y_t) + \varepsilon_t$.

We use the monetary base as our measure of money. As noted earlier this is appropriate for our purposes since the relevant distinction in our model is between costlessly provided outside money and costly inside monies. We assume that q_{st} is constant, and use a value of 10 for the semi-elasticity of money demand (with respect to annual interest rates), i.e., we set $q_s\eta = 0.1$ for annual interest rates. This is the value we estimated by OLS using annual data from 1900-91 on the monetary base, net national product and the commercial paper rate. This value is close to the value estimated by Lucas [1993] who uses M1 as his definition of money. We set the mean of ε in (4.3) and (4.4) to zero and use a model period of 2 months which is consistent with our definition of money. Since the model period is 1/6 of an year the value of $q_s\eta$ is adjusted accordingly to $0.1/6$.[16]

The rest of the model specification is as follows. The utility discount factor $\beta = 0.9516$ (annual) and the utility function is taken to be

---

[16]Note that with our specification in (4.3) and (4.4) and the choice of the period length as two months, an annual nominal interest rate of 10 percent and the estimated semi-elasticity of 0.1 (with respect to annual nominal interest rates) imply a currency velocity of about 2.7 over a two month period or about 16 over an annual period. This seems about right for recent data.

21

log-linear in consumption and leisure with the weight on consumption equal to 0.311. These values are chosen to match the real return to capital and the share of hours worked in total hours. The production function is taken to be Cobb-Douglas with a capital share parameter of 0.36. The annual depreciation rate of capital is set at 0.06 and the annual exogenous trend growth rate of the economy is set at 0.02.

In Figure 4 we display a time series of the costs (as a percentage of GNP) of U. S. commercial banks arising from the provision of transaction services – specifically demand deposits and credit cards. The numbers behind Figure 6 are constructed as follows. The Functional Cost and Profit Analysis (FCPA) report on commercial banks provides data on the volume of demand deposits and the costs associated with the demand deposit function for banks participating in the report. Using these we calculate costs as a perecentage of demand deposits for the reporting banks and then multiply this ratio by total demand deposits and then divide by GNP in order to get an estimate of costs as a percentage of GNP associated with the demand deposit function. The FCPA reports for commercial banks also provide numbers on total credit card balances and the costs associated with the credit card function for reporting banks. These numbers are used to calculate costs as a percentage of total credit card balances for reporting banks. This number is then multiplied by total credit card balances and divided by GNP to obtain an estimate of costs as a percentage of GNP associated with the credit card function. We then take the sum of costs as a percentage of GNP associated with the demand deposit function and the credit card function is a rough

estimate of costs associated with transaction services.[17] As can be seen these costs average to about 0.5 percent of GNP. In Figure 5 we display the values of ϕ for various values of the nominal interest rate. These are calculated using (3.14) where a_t is calculated from the estimated semi-log money demand function. For moderate inflation rates, say 3-8 percent annually, the corresponding values of ϕ are about 0.5 percent of GNP.[18] Thus, these values are remarkably consistent with independent measures of the costs of transaction services displayed in Figure 4. This is striking evidence in favor of our model and the estimated money demand function.

In the next section we focus on the second main contribution of our paper - our model's implications for the welfare cost of inflation with

---

[17]We implicitly assume that all of the costs associated with credit cards are arising from the transaction services provided and that none are arising from the borrowing feature. The *Survey of Consumer Finances 1989* reports that average monthly charges for convenience users (those who do not incur finance charges) was $524 in 1989 whereas average outstanding balances were $2090 in 1989. This suggests that convenience users had average balances of $262 (half of $524) and, therefore, may have accounted for only about 10 percent of the total costs associated with credit cards. If this is right then the costs asociated with credit card use for transactions may be considerably smaller than shown in Figure 4. Figure 4 also ignores credit unions which may be an important omission because credit unions have a substantial share of the credit card business and costs as a percentage of outstanding balances are significantly lower for credit unions than for commercial banks. For instance, in 1993 credit cards issued by credit unions had outstanding balances which were more than three times larger than on credit cards issued by commercial banks. Further, costs as a fraction of outstanding balances were 0.055 for credit unions compared to 0.213 for commercial banks. Averaging over commercial banks and credit unions yields a figure of 0.091 for costs as a fraction of outstanding balances which is slightly less than half the figure based on commercial banks alone. This factor also suggests that costs associated with credit cards may be lower than shown in Figure 4. Unfortunately, whereas the FCPA reports for commercial banks go back almost 30 years, the FCPA reports for credit unions are available only for the years 1992 and 1993.

[18]The real interest rate is assumed to be 7 percent annually so that annual nominal interest rates of 10-15 percent correspond to annual inflation rates of 3-8 percent.

particular emphasis on some new implications.

5. The Welfare Cost of Inflation

The Friedman rule, i.e., setting the nominal interest rate R to zero, is clearly optimal in this economy. At the optimum all goods are cash goods – no credit services and credit goods are produced and the allocation is the same as in the standard growth model without money (see Figure 3).[19] We now discuss the welfare costs of a positive nominal interest rate using the zero nominal interest rate allocation as the benchmark.

The welfare cost can be decomposed into two parts. As explained in section 4, one part arises from the misallocation of existing resources away from the goods sector and into the credit services sector. This part is captured by the rise in ϕ, the value added share of the credit services sector.[20]

The second component of the welfare cost of inflation arises due to changes in labor input and the capital stock (when labor supply is elastic

---

[19]This allocation is not the same as the allocation that would arise in our model without money since in that case all goods are credit goods which are costly to produce. For instance, if the $\mathcal{R}(.)$ schedule is bounded then there is some value of R, say R_{max} at which money will lose value and all goods will be credit goods. Even if the $\mathcal{R}(.)$ schedule is not bounded so long as the area under it is finite there exists a well-defined equilibrium without money in which all goods are credit goods. This is the allocation that corresponds to not having money in the model and involves a welfare loss relative to having money in the economy. Therefore, analogous to overlapping generations models, our model can have a well defined non-monetary equilibrium which is worse, in welfare terms, than the monetary equilibrium. This is because money permits the purchase of the most costly credit goods with cash, thereby, saving some resources. This saving in resources (per unit of goods produced) corresponds to the area B+C illustrated in Figure 3. Thus, our model captures the welfare improving role of money in the economy.

[20]As shown earlier, it also corresponds to the traditional area under the inverse money demand curve measure of the welfare cost of inflation.

and investment is endogenous) due to the usual inflation distortions which have further effects on consumption and leisure and thereby on welfare.[21] These distortions can be seen in the household's FONCs (3.10a,c) and (3.11) where the term $(1+\tau_t)$ distorts the household's labor-leisure and intertemporal consumption choices. For this reason we will refer to τ_t as the effective inflation tax rate.[22]

A key implication of our analysis is that not only the misallocation

---

[21]This component is due to the change in total income which shifts the money demand curve. Normally, this effect is very small since inflation induced changes in labor supply and hence, total income, are very small. Consequently, the area under the inverse money demand curve captures most of the welfare cost of inflation. However, in our model this is not the case because we treat goods used for consumption symmetrically with goods used for investment and do not arbitrarily designate consumption goods as cash goods and investment goods as credit goods not requiring any credit services. As we will see this leads to an inflation tax effect on investment in our model which has significant additional welfare consequences not captured by the usual area under the inverse money demand curve. As Gillman [1993] has noted the first component of the welfare cost arises because it is costly in terms of resources to substitute away from cash and make transactions with means other than cash. This component would be absent if such substitution were either not possible or costless. On the one hand, if either $q_s = \infty$ or if $\mathcal{R}(.) \equiv \infty$ then goods could not be purchased with credit at all. In this case the only component of the welfare cost of inflation is the second one arising from changes in the total amount of resources; there would be no cost arising from misallocation of resources into the credit services sector. On the other hand, if either $q_s = 0$, or if $\mathcal{R}(.) \equiv 0$ then all goods can be purchased as credit goods at the same price as cash goods. Therefore, money disappears from the model and the model reduces to the standard real one-sector growth model.

[22]Condition (3.8c) is slightly different from the standard model in that it exhibits an inflation tax effect on the return to capital as in Stockman [1985]. This effect is absent in the traditional model because investment is treated as a credit good which does not require any credit services. In our model increasing the level of investment requires the household to buy more cash goods as well as more credit goods at a higher marginal price. This feature leads to an inflation tax effect on the return to capital and, hence, on investment. However, for reasons explained subsequently the magnitude of this effect can remain bounded and small even at very high inflation rates.

component but also the inflation tax component of the welfare cost of inflation depend on the nature of the credit services technology and hence on the nature of money demand. The former connection was already established in section 4. To see the latter connection notice that the inflation distortion appears in the form $(1+\tau_t)$ rather than the more usual $(1+R_t)$. Since the the $p_{2t}(z)$ schedule depends on the $\mathcal{R}(z,\varepsilon)$ schedule which determines the money demand function it follows from (3.9) that the effective inflation tax rate τ_t depends on the nature of the money demand function.

The above feature has very important implications for the effects of the inflation tax in our model. Specifically, depending on the nature of the money demand function the effective inflation tax rate τ_t may well remain bounded with inflation. To understand this note that $(1+R_t)$ is the shadow price of cash goods purchases and $\left[\int_0^{z_t^*} p_{2t}(z)dz/p_{1t}\right]/z_t^*$ is the average price of credit goods purchases. From (3.9) we then see that the effective inflation tax rate is a weighted average of the shadow price of cash goods purchases and the average price of credit goods purchases, weighted by the respective shares of cash goods purchases and credit goods purchases in total purchases. Therefore, by changing the mix of purchases between cash goods and credit goods the consumer can potentially limit his exposure to the inflation tax. Consequently, depending on the nature of the money demand function, the welfare losses due to the inflation tax effect can remain bounded.

It turns out that, for our semi-log specification of money demand, the effective inflation tax rate τ indeed remains bounded even as R goes to infinity. To see this we can rewrite (3.9) as follows after substituting for

$p_{2t}(z)$ from (3.3).

$$(5.1) \qquad \tau = (1-z^*)R + \int_0^{z^*} q_s \mathcal{R}(z)dz = \int_0^1 Min\{R, q_s\mathcal{R}(z)\}dz.$$

Thus, τ equals the area under the $q_s\mathcal{R}(.)$ schedule up to z^* plus the area under the height R from z^* to unity (area A+B in Figure 3).

The following three implications can be drawn from (5.1). First, we can see that when R is zero τ is zero and that τ is increasing in R. Second, τ is always less than R. This is because τ is a weighted average of the shadow price of cash goods and the prices of credit goods. Since only goods with credit prices below R are purchased as credit goods (see Figure 3), it follows that τ is always less than R. Third, τ remains bounded as R goes to infinity. To see this, note that from (3.8d) and the specification of the $\mathcal{R}(.)$ schedule (see 4.3), the first term in (5.1) goes to zero as R goes to infinity. The second term in (5.1) is exactly equal to the area under the normalised inverse money demand curve (see 3.13 and Figure 3) which approaches $q_s\eta$, the inverse of the semi-elasticity of money demand, as R goes to infinity. Therefore, given our parameter values the effective inflation tax rate τ is bounded above by 1.67 percent.

In Figure 5 we show the welfare cost of inflation (expressed as a percentage of consumption) as a function of the (annual) nominal interest rate. We show the misallocation component, the value added share ϕ and the total welfare cost (including the inflation tax component) separately.[23]

---

[23]The misallocation component differs slightly from the value added share because the welfare cost is expressed as a percentage of consumption and not GNP.

There are two key results to be noted in Figure 5. The first, and perhaps the most important, is that both components of the welfare cost of inflation (and, hence the total welfare cost of inflation) are bounded at less than 5 percent of consumption. The misallocation component behaves like the value added share asymptoting to a fairly low value even as inflation reaches very high values. As explained before (see 3.14 and 4.2) both the misallocation component and ϕ are closely related to a_t, the normalised area under the inverse money demand curve, which is bounded above by the inverse of the money demand semi-elasticity. The component of the welfare cost of inflation due to the inflation tax effects also remains bounded because the effective inflation tax rate remains bounded even as the inflation rate goes to infinity.

The second result is that at low to moderate inflation rates the inflation tax component, which is the difference between the total welfare cost and the misallocation component, is about two to three times the misallocation component. At very high inflation rates it's still about the same magnitude as the misallocation component. Thus, the inflation tax component is quite significant even though the effective inflation tax rate τ remains very low at all inflation rates.[24]

---

[24]This result is entirely due to the inflation tax effect on investment. In our model, as the annual nominal interest rate rises from zero to 10 percent, 30 percent, 50 percent, and 500 percent, the effective inflation tax rate rises from zero to 1 percent, 1.6 percent, 1.66 percent, and 1.67 percent. As a result the share of investment in output falls from 22 percent to 21.7 percent, 21.6 percent, 21.55 percent, and 21.55 percent. One way to see that the welfare consequences of the inflation distortion are arising mostly through investment is to consider the following results for an *alternative* version of the model in which goods used for investment do not require any credit services; thus investment goods are always purchased as credit goods at the same price as cash goods. This alternative version is obtained by making the following changes in the model of section 3. Replace

Finally, Figure 5 also reports welfare cost calculations that take the transition period into account. The transitional dynamics were approximated by log-linearizing the equations characterizing the model's solution path around the steady state where the nominal interest rate is zero. As can be seen in Figure 5, taking the transition period into account can have a significant impact on the welfare cost of inflation. In Figure 5 the welfare cost of a 10 percent nominal interest rate falls from 2.0 percent to 0.5 percent of consumption when the transition is taken into account.

6. Concluding Remarks

We end the paper with a brief discussion of the robustness of our results to alternative specifications. Two features of our specification seem especially worth mentioning. The first is the Leontief specification of composite investment and consumption. This specification makes it possible for us to retain the one good structure of the standard monetary growth model. If some substitutability were allowed then this convenience would be lost since the relative price of goods with different z indices would change with inflation. Further, the empirical evidence presented in section 4 is

---

χ_t with c_t in (3.6) and (3.7), add investment goods purchases $k_{t+1} - (1-\delta)k_t$ to the right side of (3.7), and replace Y_t with c_t in (3.10c). It is then easy to verify that the effective inflation tax rate τ_{t+1} will no longer appear in (3.8c). In such a model, investment is no longer subject to an inflation tax. If we redo the welfare cost calculations for this version of the model, the misallocation component varies from 0.44 percent to 1.6 percent and 1.67 percent, and the total welfare cost varies from 0.52 percent to 1.61 percent and 1.67 percent, as the annual nominal interest rate varies from 10 percent to 50 percent and 100 percent, respectively. From these calculations it is obvious that the misallocation component and hence, the usual area under the inverse money demand curve, captures most of the welfare cost of inflation if investment requires no credit services.

quite consistent with our Leontief specification. The second feature of our specification is that all of labor is sold as "credit" labor and all of capital is rented as "credit" capital without using up any credit services. It is this feature which leads to the inflation distortion on labor supply and investment and the distinction between the misallocation component and the inflation distortion component of the welfare cost of inflation. If this feature of the specification were modified then all of the welfare cost of inflation would appear only as the misallocation of resources into the credit services sector. However, it's not clear that this would make much difference to our main conclusions regarding the welfare cost of inflation. With this modification there would be no inflation distortion component but the misallocation component would be larger. Hence, the overall welfare cost of inflation may not be affected. Further, the feature of the welfare cost that we emphasized, namely that the welfare cost remains bounded with inflation, is also likely to remain unaffected. This is because, given our specification of the money demand function which fits U.S. data quite well, the misallocation component is bounded at less than 5 percent of consumption. All in all we conclude that our main conlcusions are likely robust to alternative specifications of our model.

REFERENCES

Aiyagari, S. Rao and Zvi Eckstein, 1994. Interpreting monetary stabilization in a growth model with credit goods production, Research Department Working Paper 525, Federal Reserve Bank of Minneapolis. Forthcoming in *Financial Factors in Stabilization and Growth*, edited by Mario Blejer, Zvi Eckstein, Zvi Hercowitz and Leonardo Leiderman, Cambridge University Press, Cambridge, MA.

Bailey, Martin J. 1956. The welfare cost of inflationary finance, *Journal of Political Economy*, 64, 93-110.

Baumol, William J. 1952. The transactions demand for cash: An inventory-theoretic approach, *Quarterly Journal of Economics*, 66, 545-556.

Bresciani-Turroni, C. 1937. *The Economics of Inflation*, (George Allen and Unwin: London).

Cooley, T. and G. Hansen, 1989 The inflation tax in a real business cycle model, *American Economic Review*, 79, 733-48.

------ and ------. 1991. The welfare costs of moderate inflations, *Journal of Money, Credit, and Banking*, 23(3), Part 2, 483-503.

Garber, Peter M., 1982. Transition from inflation to price stability, in *Monetary Regimes and Protectionism*, Carnegie-Rochester Conference Series on Public Policy, Brunner K. and Meltzer, A. H. (editors), North-Holland: Amsterdam.

Gillman, Max. 1993. The welfare cost of inflation in a cash-in-advance economy with costly credit, *Journal of Monetary Economics*, 31, 97-115.

Graham, F. D. 1930. *Exchange, Prices and Production in Hyperinflation: Germany 1920-23*, (Russell and Russell: New York).

Lucas, Robert E. Jr. 1993. On the welfare cost of inflation, Department of Economics Working Paper, University of Chicago.

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

Melnick, Rafi, 1995. Financial services, cointegration, and the demand for money in Israel, *Journal of Money, Credit and Banking*, 27(1), 140-153.

Schreft, Stacy, 1992. Transaction costs and the use of cash and credit, *Economic Theory*, 2, 283-296.

Stockman, Alan, 1985. Anticipated inflation and the capital stock in a cash-in-advance economy, *Journal of Monetary Economics*, 8, 387-393.

Wicker, E. 1986. Terminating hyperinflation in dismembered Hapsburg monarchy, *American Economic Review*, 76(5), 350-64.
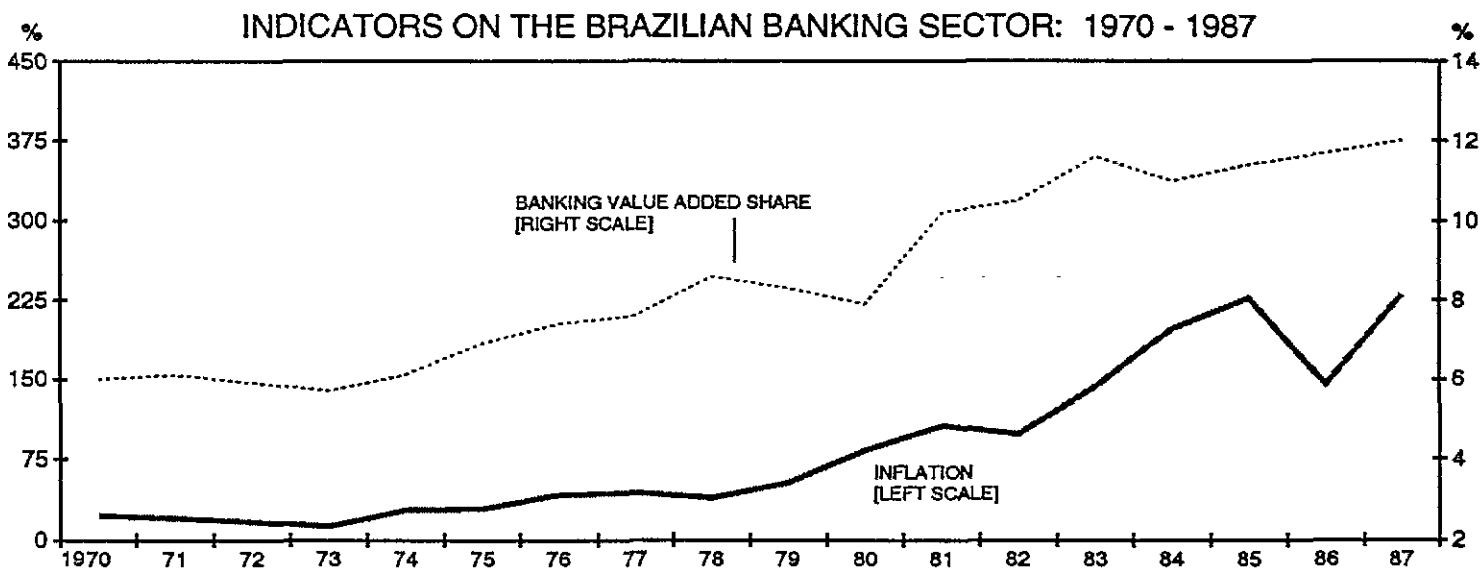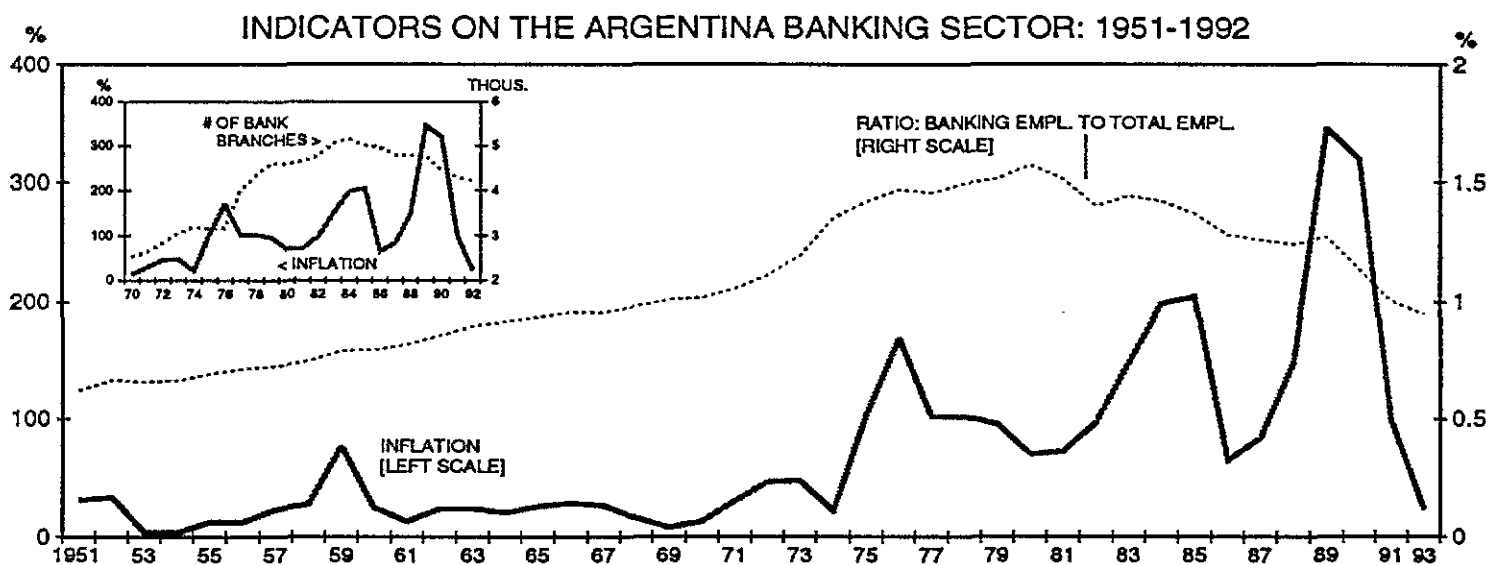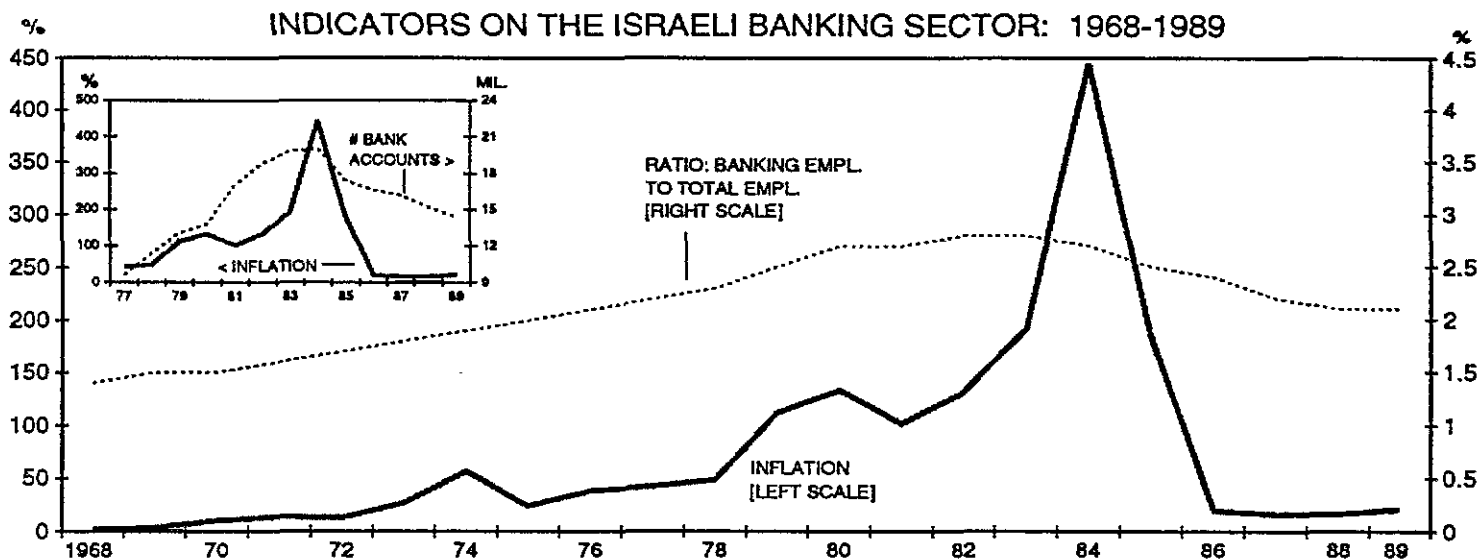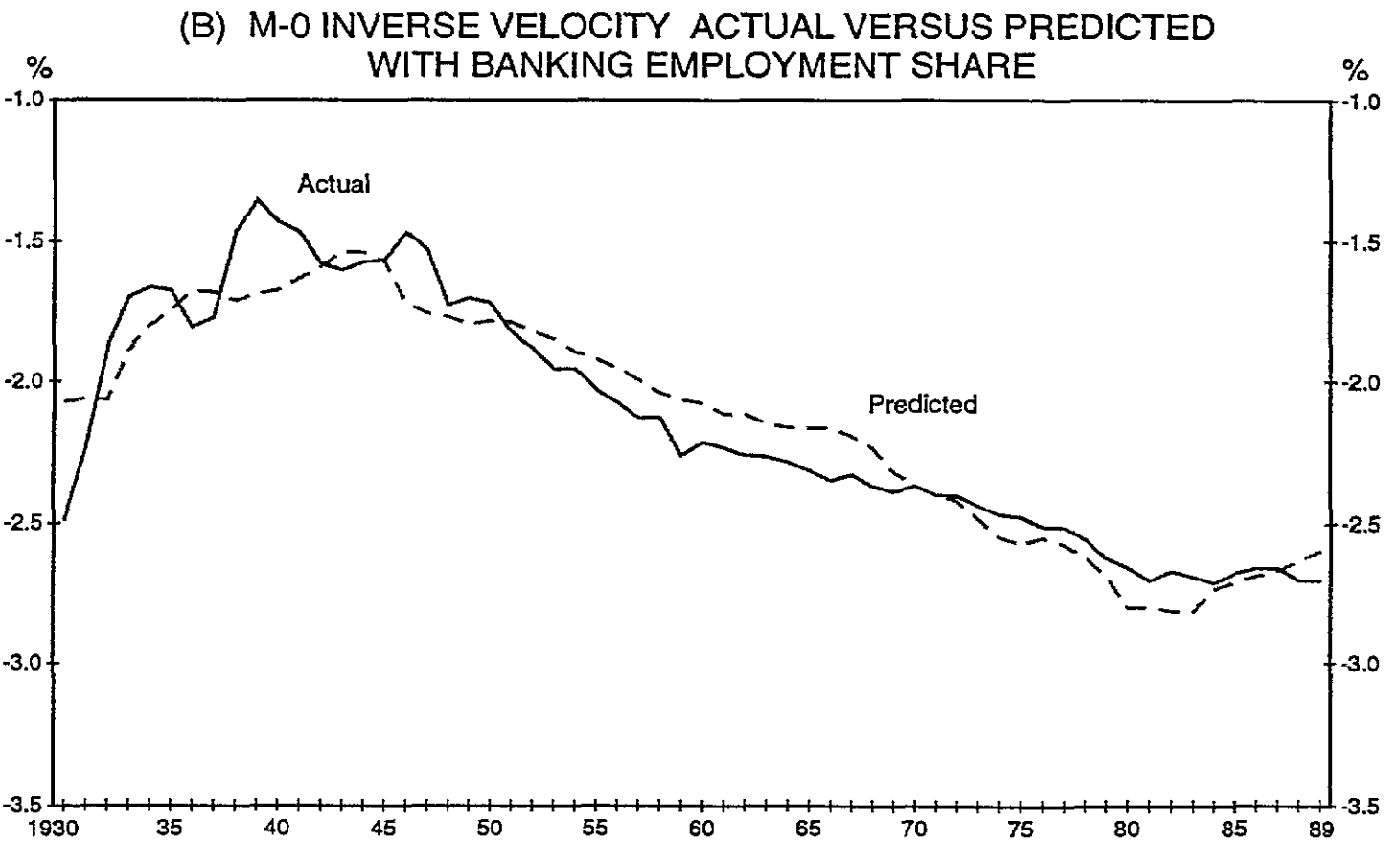
Figure 1

(A) M-0 INVERSE VELOCITY ACTUAL VERSUS PREDICTED

(B) M-0 INVERSE VELOCITY ACTUAL VERSUS PREDICTED WITH BANKING EMPLOYMENT SHARE

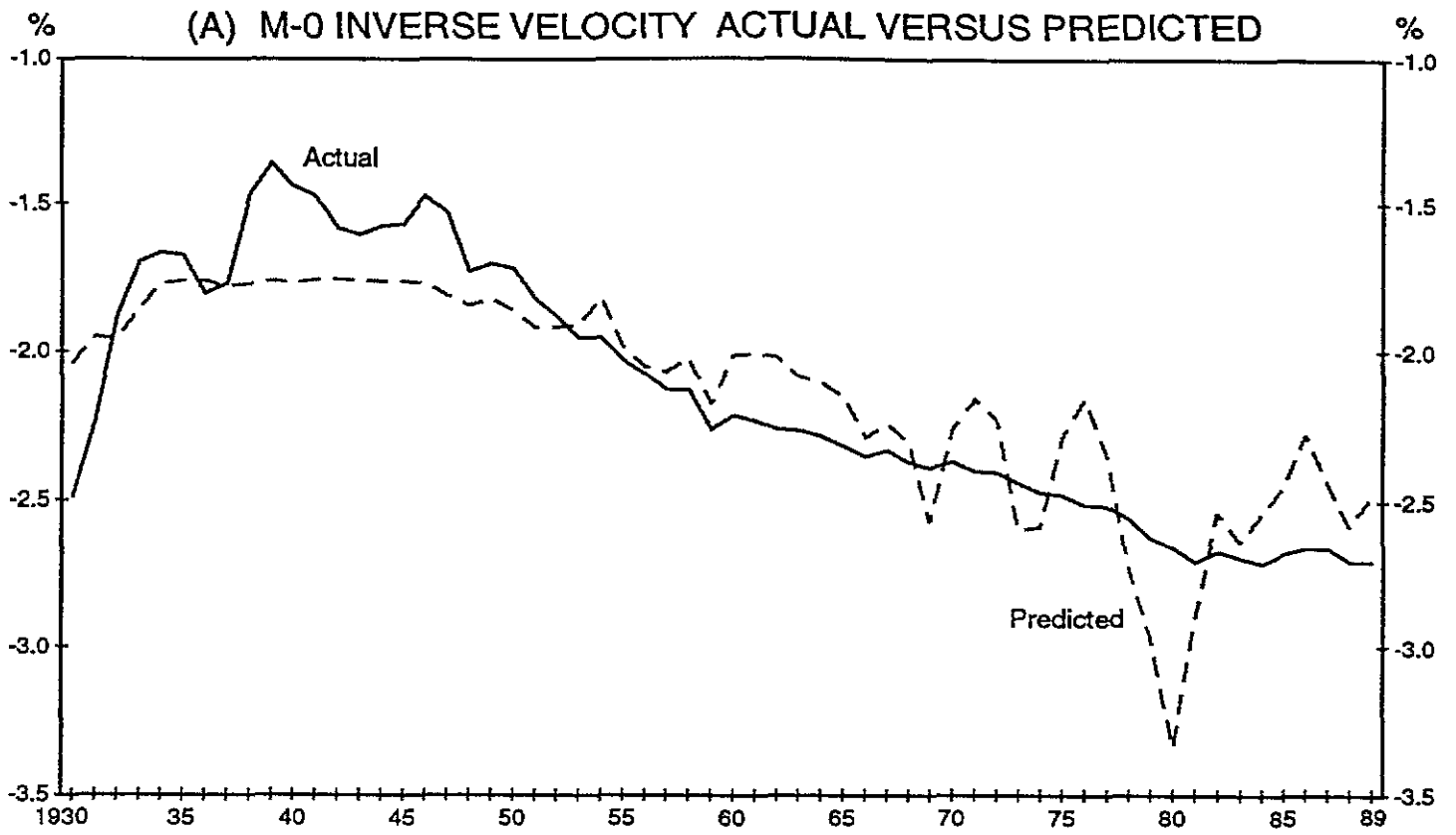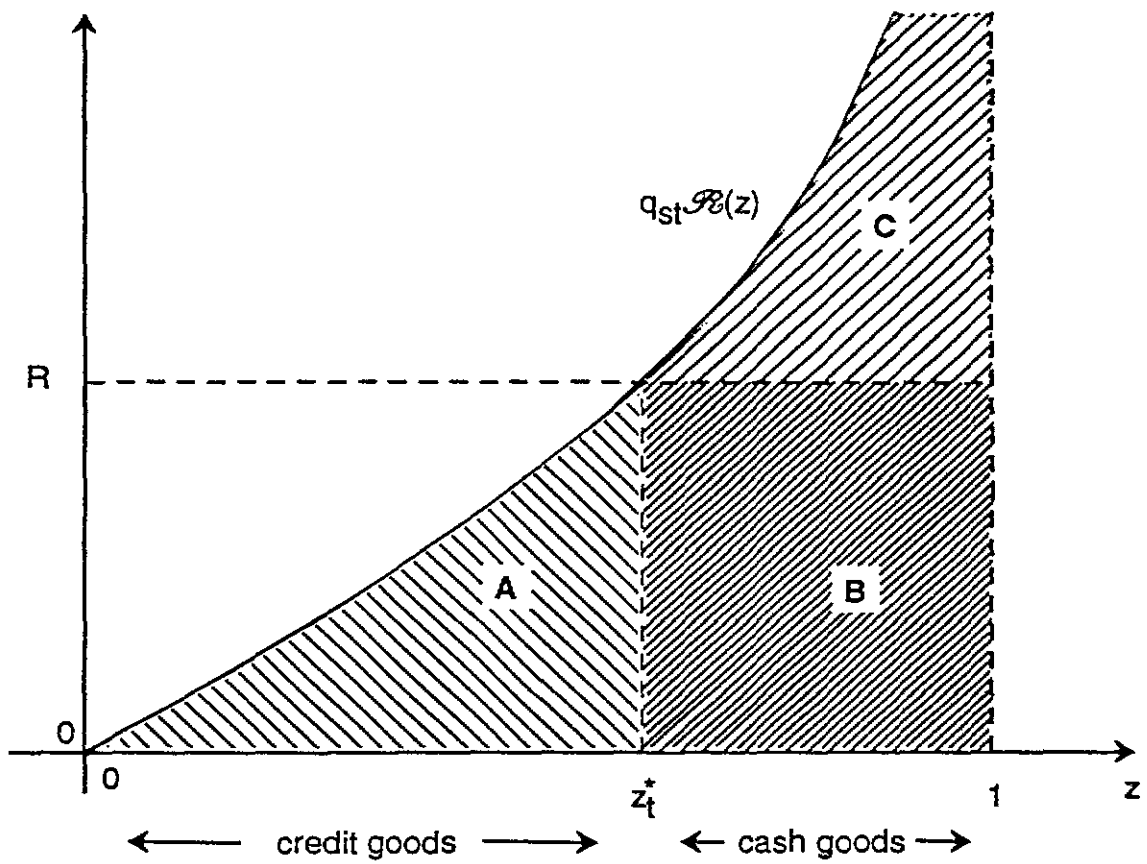**Figure 2**

A = Resources used for credit goods (a_t)

A+B = Effective inflation tax rate (τ_t)

B+C = Resource saving from money

Figure 3

COSTS ASSOCIATED WITH TRANSACTION SERVICES / GPD
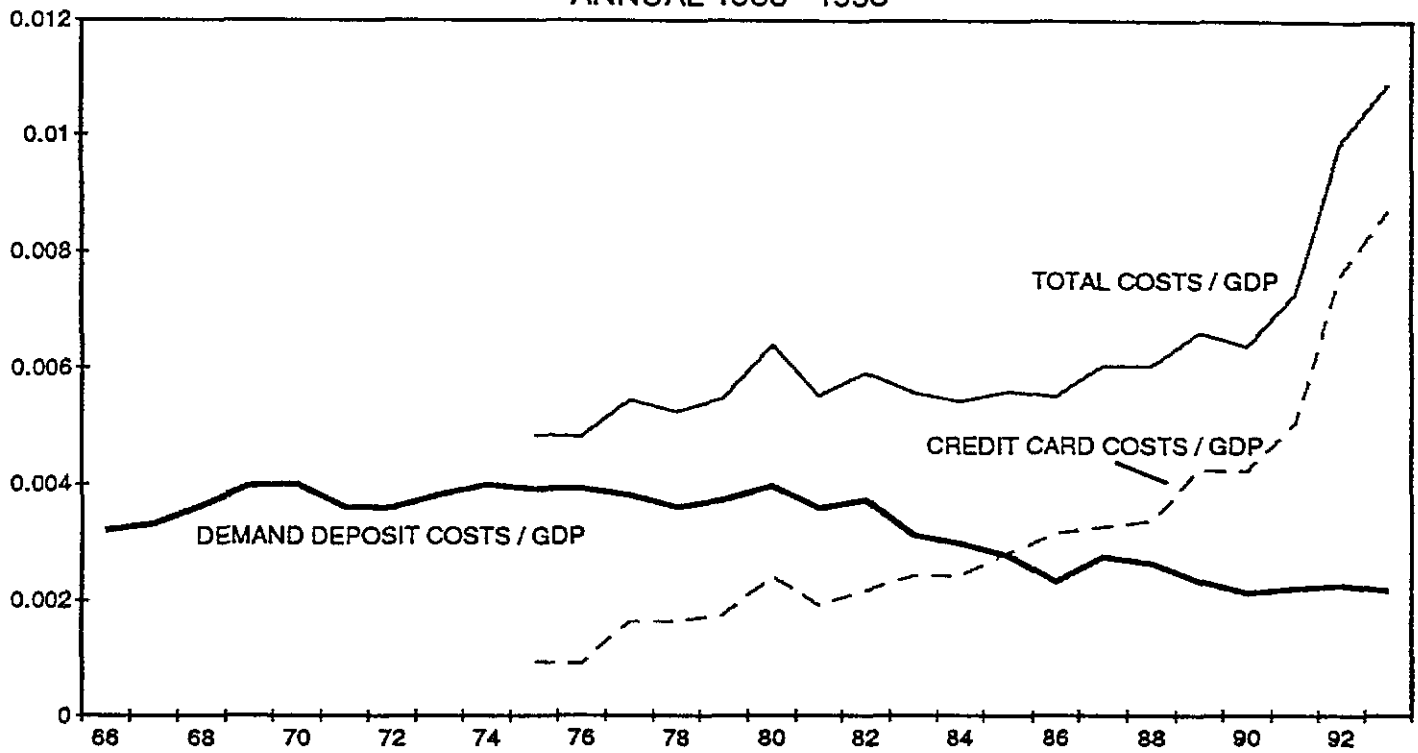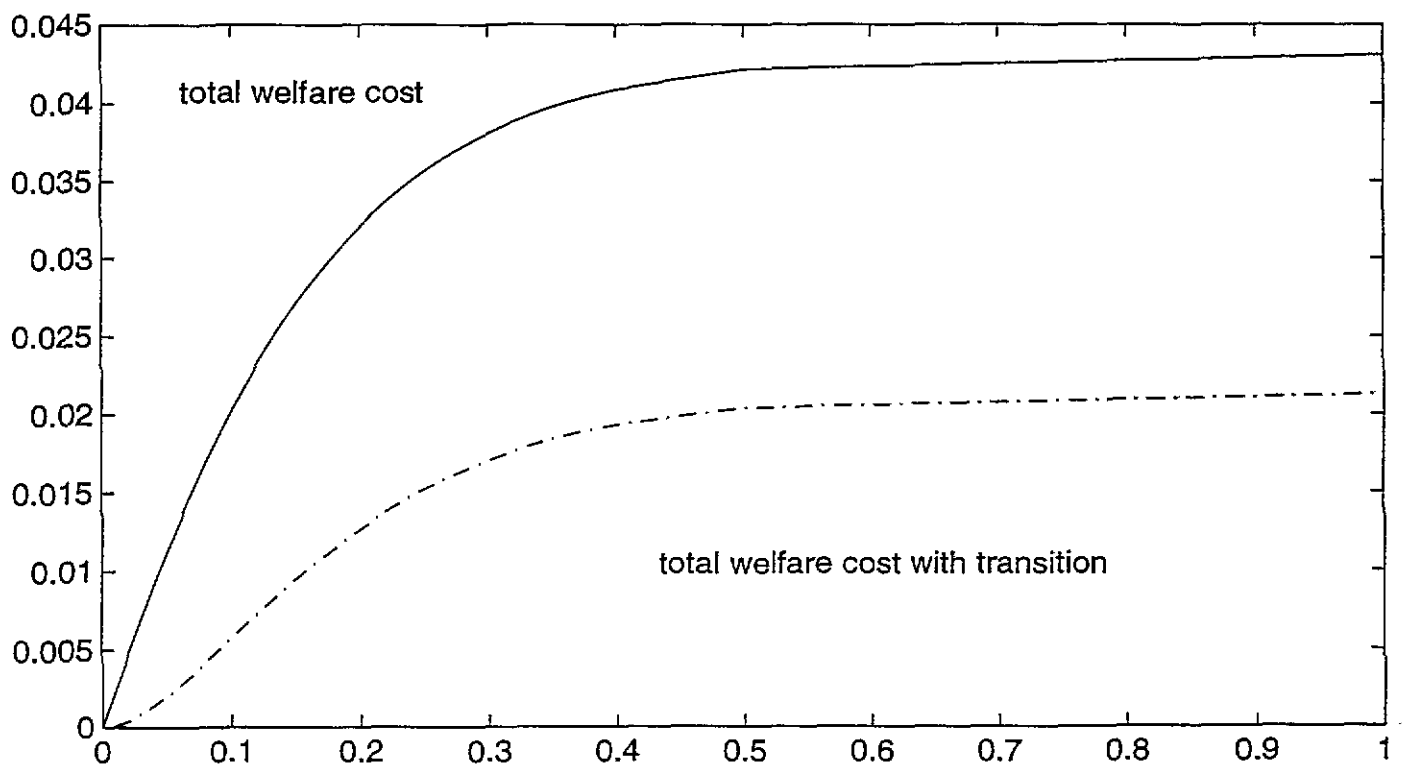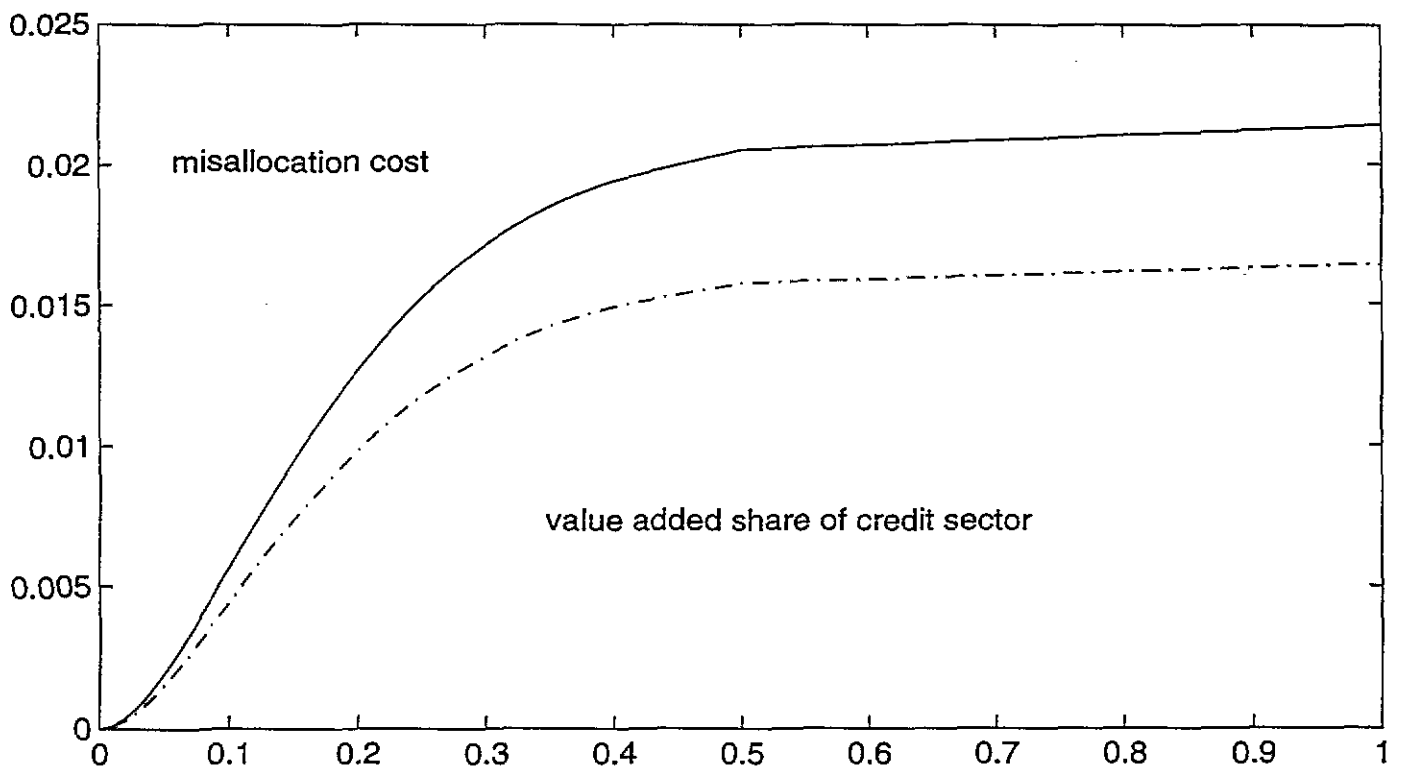ANNUAL 1966 - 1993

Figure 4

Figure 5