Richard G. Anderson is a vice president at the Federal Reserve Bank of St. Louis. Robert H. Rasche is a senior vice president and director of research at the Federal Reserve Bank of St. Louis. George Essig provided research assistance.

Eighty Years of Observations on the Adjusted Monetary Base: 1918–1997

Richard G. Anderson and Robert H. Rasche

A number of recent trends in empirical macroeconomic research—embedding long-run relationships in models via cointegration, modeling the correlation between seasonal cycles and business cycles, building endogenous growth models, and the renewed interest of policymakers in inflation targeting—have increased the importance of long time series of macroeconomic data. Many empirical phenomena suggested by this research are likely to occur only at relatively low frequencies, making long series of data essential. Among the more important of such time series are measures of the monetary base, which furnishes a monetary economy's nominal anchor.

Many empirical analyses of monetary policy assume that the stance of policy is adequately measured by the Federal Open Market Committee's target level for the federal funds rate. But the federal funds rate, or any other financial market interest rate or asset price, is not suitable as a measure of the stance of monetary policy when it is analyzed without corresponding quantity data. Prior to 1964, free reserves, not the federal funds rate, was the Federal Reserve's preferred indicator and frequent policy target (Brunner and Meltzer, 1964). More generally, models of stabilization policy with nonzero correlations among contemporaneous shocks suggest that both interest rates and monetary aggregates should appear in policy feedback rules (Poole, 1970; Friedman, 1990). Further, policies focused solely on interest rates may leave the economy's price level indeterminate, at least in the classes of models most useful for policy analysis (McCallum, 1986, 1997). Finally, as the outside money in a monetary economy, the monetary base furnishes the economy's nominal anchor (Patinkin, 1961; McCallum, 1997).

Previously published data on the adjusted monetary base begin in 1935, seasonally unadjusted, and in 1950, seasonally adjusted. In this analysis, we introduce new data for 1918 through 1935, and combine these data with the later-dated, published series.

THE MONETARY BASE

An economy's *monetary base* consists of those liabilities of the monetary authorities that are used as media of exchange by the nonbank public (households and firms other than depository institutions), and those liabilities that are used by depository institutions to settle interbank payments (Balbach and Burger, 1976). In the United States monetary system, the *monetary authorities* are the U.S. Treasury Department and the Federal Reserve System.

Previous studies of the St. Louis adjusted monetary base have examined data that begin in August 1935 (Andersen and Jordan, 1968; Tatom, 1980; Anderson and Rasche, 1996a). For these dates, the monetary base is equal to the sum of currency in circulation outside the Treasury and the Federal Reserve, plus the deposits of depository institutions at Federal Reserve Banks.

For dates before August 1935, the best-known measure of the monetary base is the high-powered money series of Friedman and Schwartz (1963), pp. 799-808. The measure introduced in this article differs from theirs in several respects. Our data are from *Banking and Monetary Statistics 1914-1941*, Table 101, pp. 369-71, and begin August, 1917.[1]

[1] These data, labeled "Money in Circulation" in *Banking and Monetary Statistics*, include gold coin in circulation prior to February 1934.

Currency in Circulation

The *Banking and Monetary Statistics* data measure currency in circulation as the monthly average of daily levels. Friedman and Schwartz's data measure currency as of the last day of the month. Our data are not seasonally adjusted, while Friedman and Schwartz's data are seasonally adjusted. By using such seasonally unadjusted data, we can create our adjusted monetary base as a chain index and, thereafter, examine separately its seasonality. We also exclude Friedman and Schwartz's adjustment for $287 million of missing gold coin; see Friedman and Schwartz (1963), pp. 463-64, footnote 45.

Deposits of Banks at the Federal Reserve

Similar to currency, the *Banking and Monetary Statistics* data measure the deposits held by banks at Federal Reserve Banks as the monthly average of daily balances. Friedman and Schwartz's data measure deposits as of the end of the month or the last Wednesday of the month. We also do not include an adjustment for Federal Reserve float. Friedman and Schwartz (1963, p. 748) include a measure of float that is constructed primarily from end-of-month figures, with some months interpolated from less frequent observations. In any case, the inclusion or exclusion of float makes little difference. Float usually is small. At the end of January 1936, for example, float was only about $2 million *(Annual Report of the Board of Governors of the Federal Reserve System for 1936*, June 1937, p. 74). We also omit clearing balances held by nonmember banks at the Federal Reserve Banks, due to the paucity of data. Friedman and Schwartz (1963, p. 748) create a time series for these balances by linear interpolation between the only published values: last day of the year figures in the Federal Reserve Board's *Annual Report*. These nonmember clearing balances are small, relative to total member bank deposits: $91 million and $123 million on December 31, 1935 and 1936, respectively *(Annual Report of the Board of Governors of the Federal Reserve System for 1935*, p. 83, and *Annual Report of the Board of Governors of the Federal Reserve System for 1936*, p. 73). Member bank deposits were $5.587 and $6.606 billion on these dates (*Banking and Monetary Statistics*, p. 332). These data also are not seasonally adjusted, while Friedman and Schwartz's published data are seasonally adjusted.

THE *RAM* ADJUSTMENT

Changes in statutory reserve requirements often cause depository institutions to change the amount of base money that they hold. These changes have no implication for the stance of monetary policy and should be removed from the adjusted monetary base.[2] One such adjustment—the *reserve adjustment magnitude*, or *RAM*—was introduced by Karl Brunner, and has been extended by Burger and Rasche (1977) and Anderson and Rasche (1996b). The RAM adjustment measures the amount by which changes in statutory reserve requirements—relative to those in effect during a specific base period—have changed the quantity of base money held by depository institutions.

Measuring RAM precisely requires a model of depository institutions' asset management that includes an explicit role for statutory requirements. Such a model is beyond the scope of this paper. The underlying concept, however, may be illustrated simply if we assume that all base money held by depository institutions is eligible to satisfy statutory reserve requirements. (In the United States, vault cash could not be used to satisfy reserve requirements between 1917 and 1959.) In this case, the amount of base money held by a depository institution, when statutory reserve requirements are relatively high, will be determined largely by the amount of its required reserves. Conversely, when statutory requirements are relatively low, the amount held will be largely determined by the depository's business needs, such as converting customer deposits into currency, making interbank wire transfers, and settling interbank check collection debits.

Let us suppose that a depository institution's demand function for base money may be written as $MB^D(d,rr)$, where d is the institution's deposit liabilities and rr is

[2] A broad cross-country study illustrating the importance of such adjustments is McCallum and Hargraves (1995).

RESERVE REQUIREMENTS IN THE UNITED STATES SINCE 1914

When the Federal Reserve System opened for business in May 1914, member banks became subject to statutory reserve requirements set by the Federal Reserve. (Banks that held federal, or national, charters were required by law to be members of the Federal Reserve System. Membership was optional for state-chartered banks.) In 1980, implementation of the Monetary Control Act made all depository institutions subject to the requirements set by the Federal Reserve.

The Federal Reserve's statutory reserve requirements specify three items: the type and amount of deposits subject to requirements, the reserve requirement ratio applicable to these deposits, and the bank's assets that are eligible to satisfy the requirements. For examples of how changes in these regulations have affected measurement of the adjusted monetary base, see Burger and Rasche (1977) and Anderson and Rasche (1996b).

Prior to 1972, reserve requirement ratios differed across three separate categories of member banks: central reserve city banks, reserve city banks, and country banks. Originally, there were three central reserve cities: New York, Chicago, and St. Louis. As discussed in the text of this article, St. Louis was reclassified as a reserve city in 1922. In 1972, these categories were eliminated and a new system of requirements initiated. In the new system, reserve requirement ratios vary with the amount of deposits held by a bank, not by its location.

Depository institutions satisfy their requirements today by holding cash-in-vault and deposits at Federal Reserve Banks. Between August 1917 and November 1959, however, only deposits at Federal Reserve Banks could be used to satisfy requirements; vault cash was not eligible. The eligibility of vault cash was phased in between December 1959 and December 1960. Details are available in *Banking and Monetary Statistics 1914-1941*, p. 401, and *Annual Report of the Board of Governors of the Federal Reserve System for 1972*, pp. 45-46.

Prior to the 1980, depository institutions that were not members of the Federal Reserve were subject to reserve requirements set by state regulators. Because most such requirements could be satisfied by holding interest-bearing liquid assets and/or deposits in banks, as well as vault cash, it seems unlikely that these requirements significantly affected the quantity of base money demanded by these institutions. Hence, these requirements do not enter into our measure of RAM.

the statutory reserve requirement ratio. If statutory reserve requirements are, at the margin, the binding constraint that determines the amount of base money held by the depository, then

$$\frac{\partial MB^D(d,rr)}{\partial rr} > 0 .$$

If the depository's business needs, rather than statutory requirements, are the binding constraint, then

$$\frac{\partial MB^D(d,rr)}{\partial rr} = 0 .$$

To measure RAM, we must be able to estimate (or infer) the sign of the derivative

$$\frac{\partial MB^D(d,rr)}{\partial rr}$$

at all dates in our sample, for each set of statutory reserve requirements.

To be more specific, let us denote a depository institution's level of required

WHY DOES RAM BEGIN IN 1917?

The RAM adjustment in this article begins in August 1917 for two reasons. First, the purpose of our analysis is to extend previously published *monthly* data to earlier dates, for use in subsequent econometric analysis. The earliest available month-average data on the daily level of the monetary base begin in August, 1917 (*Banking and Monetary Statistics 1914-1941*, pp. 369-71), although annual data are available for earlier dates.

Second, the data required to measure RAM are not available for dates prior to August 1917. The structure of the statutory reserve requirements that applied to Federal Reserve System member banks changed sharply on June 21, 1917. Prior to this date, member banks were required to satisfy a minimum of one-third of their required reserves with vault cash, and could at their discretion satisfy up to two-thirds with vault cash; separate minimum ratios were specified for satisfying requirements in vault cash and in Federal Reserve Bank deposits. On June 21, vault cash became ineligible to satisfy required reserves, and reserve requirement ratios for net demand deposits were reduced by 5 percentage points—to 13, 10 and 7 percent from 18, 15 and 12 percent for central reserve city, reserve city, and country banks, respectively. According to Federal Reserve Board staff, the changes were intended to "centralize" the holding of reserve balances (*Annual Report of the Board of Governors of the Federal Reserve System for 1935*, pp. 17-18). At the time, it was expected that the changes would have little effect on the overall demand for base money. If the decrease in a bank's required reserves was approximately equal to the amount of its vault cash, and if the bank's demand for vault cash was determined primarily by day-to-day operations rather than by statutory requirements, then the demand for base money might change little. Interestingly, this argument is similar to one made by the Bundesbank in 1995 when it also reduced reserve requirement ratios and made vault cash ineligible to satisfy requirements (*Monthly Report*, Deutsche Bundesbank, July 1995, pp. 25-26).

reserves during period t, as $RR(d_t,rr_t)$. Also, let us denote as $RR(d_t,rr_0)$ what the same institution's level of required reserves would have been during period t, if the statutory reserve requirements of a base period, denoted as period 0, had been in effect. (For all cases, we assume that sufficient data exist so as to permit calculation of the quantity $RR(d_t,rr_0)$.) Then, consider four cases:

Case 1: If $rr_0 = rr_t$, that is, the reserve requirement ratio has not changed, then RAM = 0.

Case 2: If:

$$\left.\frac{\partial MB^D(d,rr)}{\partial rr}\right|_{\substack{d=d_0\\ rr=rr_0}} = 0,$$

and

$$\left.\frac{\partial MB^D(d,rr)}{\partial rr}\right|_{\substack{d=d_t\\ rr=rr_t}} = 0,,$$

that is, if the business needs of the bank were the binding constraint in both the base period 0 and period t, then RAM = 0.[3]

Case 3: If both:

$$\left.\frac{\partial MB^D(d,rr)}{\partial rr}\right|_{\substack{d=d_0\\ rr=rr_0}} > 0,$$

[3] In the notation of the text, the vertical bar indicates that the derivative is evaluated at the values of *d* and *rr* shown at the bottom end of the bar.

and

$$\left.\frac{\partial MB^D(d,rr)}{\partial rr}\right|_{\substack{d=d_t\\ rr=rr_t}} > 0,$$

that is, if the statutory requirements were the binding constraint on the bank in both the base period *0* and period *t*, then the RAM adjustment for period *t* (conditional on the choice of period *0* as the base period) is:

$$RAM_t = RR(d_t, rr_0) - RR(d_t, rr_t).$$

Case 4: If:

$$\left.\frac{\partial MB^D(d,rr)}{\partial rr}\right|_{\substack{d=d_0\\ rr=rr_0}} > 0,$$

but

$$\left.\frac{\partial MB^D(d,rr)}{\partial rr}\right|_{\substack{d=d_t\\ rr=rr_t}} = 0,$$

that is, if the statutory requirements were binding in the base period but not in period *t*, then to measure RAM we must find the smallest reserve requirement ratio, say *rr**, for which

$$\left.\frac{\partial MB^D(d,rr)}{\partial rr}\right|_{\substack{d=d_t\\ rr=rr^*}} > 0.$$

Then, $RAM_t = RR(d_t, rr_0) - RR(d_t, rr^*)$. An empirical criterion for measuring RAM in this case was developed by Anderson and Rasche (1996b), based on statistical analysis of a large panel data set.

The above analysis assumes that the only change in statutory reserve requirements between periods *0* and *t* is a change in the reserve requirement ratio, *rr*. It also assumes that the data exist to calculate the counterfactual level of required reserves, $RR(d_t, rr_0)$. But these assumptions may not be satisfied if other aspects of the reserve requirement system—such as the categories of deposits subject to requirements—change. Then, a lack of data may make it impossible to calculate $RR(d_t, rr_0)$ and, hence, the value of RAM. One example of such a change is the shift in 1972 from a structure of reserve requirements based on location (with central reserve city, reserve city, and country categories), to a structure based on the amount of deposits held by a bank. A second example is the Monetary Control Act's extension in 1980 of statutory reserve requirements to nonmember depository institutions, which had not been subject to Federal Reserve requirements previously. In such cases, a new base period must be selected, and a new RAM series begun.

For long time series of data, multiple changes in statutory reserve requirements might require several separate segments of RAM. Nevertheless, within each such segment, the *adjusted* monetary base is equal to the sum of the monetary base and the appropriate RAM. As long as the resulting multiple segments overlap by at least one observation at each break point, they may be chained together to form a single measure of the adjusted monetary base. A procedure for doing so was suggested by Tatom (1980). His procedure is applied as follows. Beginning with the earliest-dated observation, move forward in time to the first overlapping break point in the data. At this date, calculate the ratio of the first observation in the next segment of data to the final observation in the prior segment, then multiply all earlier-dated data by this ratio. Repeat this procedure for all segments and all break points. An advantage of this procedure is that it preserves the growth rates of the series, and allows the final time series to be interpreted as a chained index-number measure of the Federal Reserve's policy actions.

Tatom (1980) calculated segments of RAM for three reserve-requirement base periods:

- RAM(1935) for August 1935 through December 1972 using the reserve requirement structure of August 1935 as the base period;
- RAM(1972) for December 1972 through January 1975, using December 1972 as the base period; and

- RAM(1975) for January 1975 through October 1980 using January 1975 as the base period.

An additional segment, RAM(1991) for dates beginning October 1980 was constructed by Anderson and Rasche (1996b), using the reserve requirement structure of January 1991 as the base period. These RAM adjustments permit us to construct a measure of the adjusted monetary base that begins in August 1935 and is chained in December 1972, December 1975, and October 1980. This measure has been published by the Federal Reserve Bank of St. Louis since early 1996 (Anderson and Rasche, 1996a).

To measure the adjusted monetary base during the earlier period between 1917 and 1935, we must measure the effects of three changes in statutory reserve requirements:

1) St. Louis was reclassified, as of July 1922, from a central reserve city to a reserve city.

2) The reserve requirement ratio applicable to U.S. government deposits was increased, as of August 1935, from zero to the same ratio as applicable to other demand deposits.

3) In August 1935, the method of calculating the amount of net demand deposits subject to reserve requirements was changed.

To adjust the monetary base for these events, we introduce an additional, fifth RAM, denoted RAM(1922), that spans the period from August 1917 to August 1935, and uses July 1922 as its base period. Its calculation is explained below.

In addition to these three changes, there were a number of changes in reserve requirements that we do not consider; see *Banking and Monetary Statistics 1914-1941*, p. 401. The Federal Reserve Act gave the Board the flexibility to reclassify banks in the outlying areas of larger cities and in annexed areas as country banks. Most changes consisted of reclassifying banks, branches, and cities among the categories of central reserve city, reserve city, and country. The data necessary to estimate the effect of these changes on the quantity of base money demanded are sparse. Further, because these changes affected relatively smaller banks, the aggregate effects likely are much smaller than those from the reclassification of St. Louis. We do not consider any of these changes in this analysis.

- Reclassification of St. Louis as a Reserve City: The structure of reserve requirements in effect as of 1914 designated three central reserve cities: New York City, Chicago, and St. Louis. Based on a petition by its bankers, St. Louis was reclassified by the Federal Reserve Board as of July 1, 1922, to be a reserve city, thereby reducing reserve requirements on the banks in St. Louis. Measuring RAM(1922) for dates prior to July 1922 requires deposit data for the affected banks. Unfortunately, as far as we have been able to determine, the Federal Reserve retains neither a list of these banks nor data on their deposits. Hence, our RAM for this period is based on published deposit data for the weekly reporting banks in St. Louis. (These data begin December 1917.)

We believe this proxy is satisfactory for several reasons. First, evidence suggests that in St. Louis the central reserve city and weekly reporting banks were the same banks. Data for weekly reporting banks as published in the *Federal Reserve Bulletin* from December 1917 through October 18, 1918, had separate classifications for central reserve city and reserve city banks. For the week of October 18, 1918, the *Bulletin* table shows 14 reporting central reserve city banks in St. Louis. Beginning with the week of October 25, 1918, the title on the table in the *Bulletin* was changed to "Reserve City Banks." There remained 14 reporting banks for that week in the city of St. Louis, however. (The number of reporting banks in the other central reserve cities, New York and Chicago, also did not change.) Our conclusion—that all central reserve city banks were weekly reporting banks—is further supported by call report data. The deposit and asset totals on published call reports for central reserve city banks in New York City and Chicago are approximately the same as the totals for the weekly reporting banks in these cities. Finally, bank directories of the period provide additional, corroborative evidence. From January 10, 1919, through July 3, 1919, the *Bulletin* tables show 15 weekly reporting banks in St. Louis. The Rand McNally *Bank Directory* for January 1919, identifies 16 member banks within the city of St. Louis. Of these banks, 15 were in or near the downtown core; one bank, the smallest, was on the outskirts of the city. We are confident that the largest 15 banks were both the weekly reporting and central reserve city banks in St. Louis, and hence, that it is accurate to use published data on net demand deposits at weekly reporting banks in St. Louis to measure RAM(1922).

• Reserve Requirements on U.S. Government Deposits: In August 1935, U.S. government deposits became subject to the same statutory reserve requirements as were applied to private demand deposits.[4] Government deposits previously had become exempt from reserve requirements in 1917 due to a provision in the Liberty Bond Act. Because demand deposits at central reserve city, reserve city, and country banks were subject to different reserve requirement ratios—13, 10, and 7 percent, respectively—to accurately measure the impact of this change on required reserves would require individual-bank data, which are not available. Below, we use an estimate published by the Federal Reserve Board in its *Annual Report* for 1935.

• Rules for the Calculation of Required Reserves: The change in the definition of net demand deposits subject to reserve requirements, as of August 1935, is described in *Banking and Monetary Statistics 1914–1941*, pp. 65–66:

> [Prior to the Banking Act of 1935] net demand deposits of a member bank … were made up of (1) the gross amount of all demand deposits except those due to other banks, and (2) the net excess (if any) of demand deposits due to other banks over demand balances due from other domestic banks and cash items in process of collection. From April 24, 1917, to August 23, 1935 … United States Government deposits were exempt by law from all reserve requirements and were, therefore, excluded from net demand deposits.
>
> The Banking Act of 1935 brought about a fundamental change in the definition of net demand deposits: it prescribed that reserves be carried against United States Government deposits, and permitted allowable deductions to be offset against total demand deposits instead of against demand deposits due to banks. Net demand deposits thus were defined as the excess of all demand deposits, including deposits due to banks and the United States Government, over demand balances due from other domestic banks (except Federal Reserve Banks, foreign banks or branches thereof, foreign branches of domestic banks, and private banks) and cash items in process of collection.

The amount by which this redefinition changed the required reserves of an individual member bank depended on that bank's size and mixture of deposits. The aggregate change in required reserves cannot be calculated from aggregate data.[5]

The measurement of RAM(1922) may now be described. First, from December 1917 through June 1922, RAM(1922) measures the difference between the required reserves that St. Louis central reserve city banks would have held if the statutory reserve requirements of July 1922 had been in effect, and their actual required reserves. St. Louis central reserve city banks, from December 1917 through June 1922, faced a 13 percent reserve requirement ratio on net demand deposits; a 10 percent ratio became effective as of July 1, 1922, the base period for RAM(1922). Hence, for t = December 1917 through June 1922, RAM(1922) is:

$$RAM(1922)_t = 0.10*(\text{Net Demand Deposits, St. Louis Weekly Reporting Banks})_t - (\text{Required Reserves, St. Louis Weekly Reporting Banks})_t.$$

[4] *Federal Reserve Bulletin*, September 1935, p. 618.

[5] See *Banking and Monetary Statistics 1914–1941*, p. 66, and especially footnote 13.

Table 1

RAM Adjustment Due to Reclassi cation of St. Louis from a Central Reserve City to a Reserve City, July 1, 1922

(Millions of Dollars)

	Jan.	Feb.	Mar.	Apr.	May	Jun.
1917	--	--	--	--	--	--
1918	–5.888	–5.828	–5.073	–5.612	–5.495	–5.466
1919	–6.100	–6.075	–6.092	–5.979	–6.023	–6.054
1920	–7.781	–7.545	–7.700	–7.265	–6.771	–6.782
1921	–6.877	–6.788	–6.570	–6.453	–6.180	–5.907
1922	–6.474	–6.636	-–6.523	–6.567	–6.564	–6.527

	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
1917	--	--	--	--	--	–5.687
1918	–5.364	–5.796	–5.715	–5.548	–5.483	–5.998
1919	–6.399	–6.888	–6.836	–6.812	–7.131	–7.350
1920	–6.773	–6.761	–6.619	–6.387	–6.344	–6.523
1921	–5.848	–5.881	–5.834	–5.961	–6.077	–6.203
1922	0	0	0	0	0	0

Its values, shown in Table 1, range between –$7.8 and –$5.0 million. Second, from July 1922 through July 1935, RAM(1922) equals zero. Finally, in August 1935, we set RAM (1922) equal to the reduction in the required reserves of member banks as reported by the Federal Reserve Board, $35 million.[6]

The five RAM adjustments used to construct the adjusted monetary base are plotted in Figure 1. In general, the size of a RAM adjustment during any specific period is smaller (larger) than during its base period if statutory reserve requirements are higher (lower) than during the base period. RAM (1935), for example, becomes large and negative during 1937-38 as a result of the Federal Reserve doubling the reserve requirement ratios on member banks. The small size of RAM(1922) relative to later adjustments suggests that minor errors in its measurement are unimportant and that further research to measure the effects of other changes in reserve requirements prior to 1935 would not be worthwhile.

SEASONAL ADJUSTMENT

Most empirical modeling is done with seasonally adjusted data, despite cautions that such filters may distort dynamic relationships; see for example Wallis (1974) and Harvey and Scott (1994). Seasonal adjustment of relatively long time series, such as the one in this analysis, is troublesome because of structural shifts in the data generating process. If the seasonal process is not separable from other parts of the data generating process—as suggested by Barsky and Miron (1989), Beaulieu, MacKie-Mason, and Miron (1992), and others—then seasonality might diminish or even vanish during periods of unusual economic activity. We identify such a shift during the Great Depression and World War II.

The growth rate and autocorrelation function of the adjusted monetary base, for our full sample and three subperiods, are shown

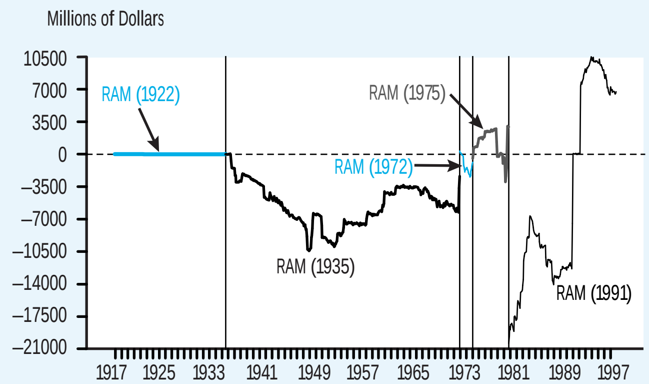
[6] *Twenty-Second Annual Report of the Board of Governors of the Federal Reserve System for 1935*, p.19.

in Figure 2. As expected, significant seasonal variation is apparent, except perhaps for the period stretching from early 1933 (the year of the third banking panic during the 1930s) through the late 1940s (approximately the end of the Federal Reserve's bond-pegging period). During the latter part of our sample, between 1950 and 1997, seasonal variation appears both more regular (as indicated by the autocorrelation function) and, generally, weaker (as indicated by the growth rates) than in earlier decades.

Our analysis of the seasonal variation in the adjusted monetary base proceeds in two steps. First, we examine the stability of seasonal variation decade-to-decade within the confines of deterministic seasonality. Next, we extend the analysis to allow stochastic seasonality, and summarize seasonal adjustment factors estimated by both the Bureau of the Census X11 and X12–regARIMA programs.[7] For X11, we use the program included in version 4.31 of the RATS econometrics package. For X12–regARIMA, we use the version for PC DOS dated June 1998, available from the Bureau of the Census at <ftp.census.gov>. The length of our time series, and the high degree of apparent noise during some periods, suggests that the extensive tests for outliers and sophisticated diagnostics contained in the X12-regARIMA package may be particularly valuable (Findley et. al., 1998).

Figure 1

Five Segments of RAM, December 1917-December 1997



Tests Based on Deterministic Seasonality

We begin our investigation of the seasonal properties of the adjusted monetary base by testing the hypothesis that the seasonal pattern differs decade-by-decade. For this test, we assume deterministic seasonality (represented by monthly dummy variables) and interact these variables with a set of decade-specific dummy variables.

The null hypothesis that there are no month/decade interactions over the full sample between 1918 and 1997 is rejected with an F-statistic of 1.78, 77 and 840 degrees of freedom, and a p-value less than 0.001. Next, we test the less restrictive hypothesis that there are no month/decade interactions between 1950 and 1997. The hypothesis is not rejected, with an F-statistic of 0.81, 44 and 840 degrees of freedom, and a p-value of 0.80. Third, we test the significance of the resemblance between the seasonal variation during the 1920s to that from 1950 through 1997. This null hypothesis—that the seasonal patterns are the same—is rejected at the 5 percent level; the F-statistic is 1.60, 66 and 840 degrees of freedom, and a p-value of 0.002. Finally, we test the null hypothesis that there is no month/decade interaction during the 1930s and 1940s. This hypothesis is rejected, with an F-statistic of 2.58, 22 and 840 degrees of freedom, and a p-value less than 0.001.

Hence, within the limits of assuming only deterministic seasonality, we conclude that seasonal fluctuations in the adjusted monetary base were:

- reasonably constant between 1950 and 1997;
- weakly similar during 1920-29 and 1950-97;
- different during 1930-49 than during the other decades in our sample.

The statistical conclusions of the previous paragraph are reinforced through the sequence of tier charts in Figures 3a and 3b. The eight panels of the figure display the level of the adjusted monetary base, by decade, nor-

[7] Deterministic and stochastic seasonality, and the X11 algorithms, are discussed by den Butter and Fase (1991).

Figure 2

Growth of the Adjusted Monetary Base

Various Periods, Log First Difference, Monthly, Percent Annual Rate

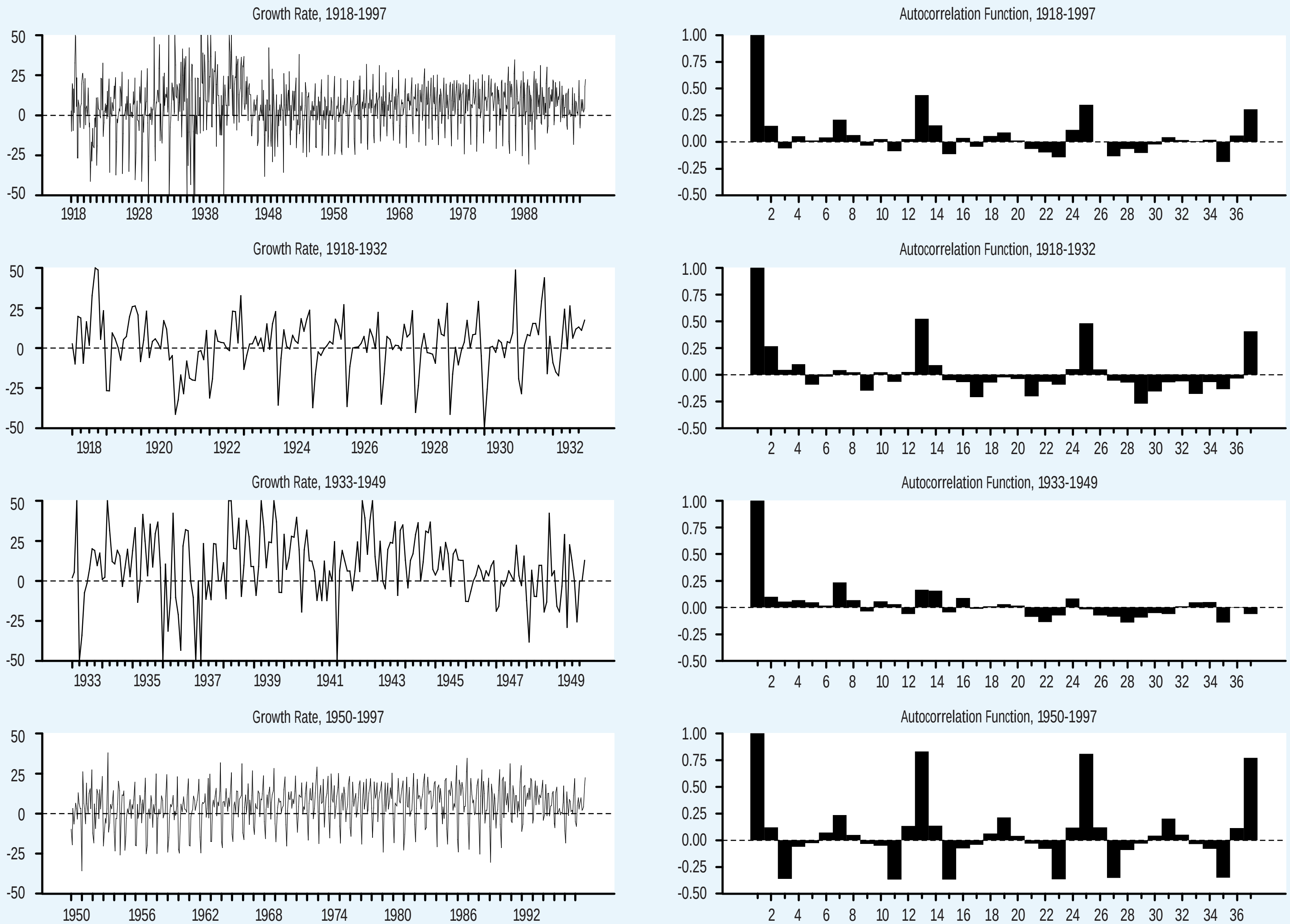
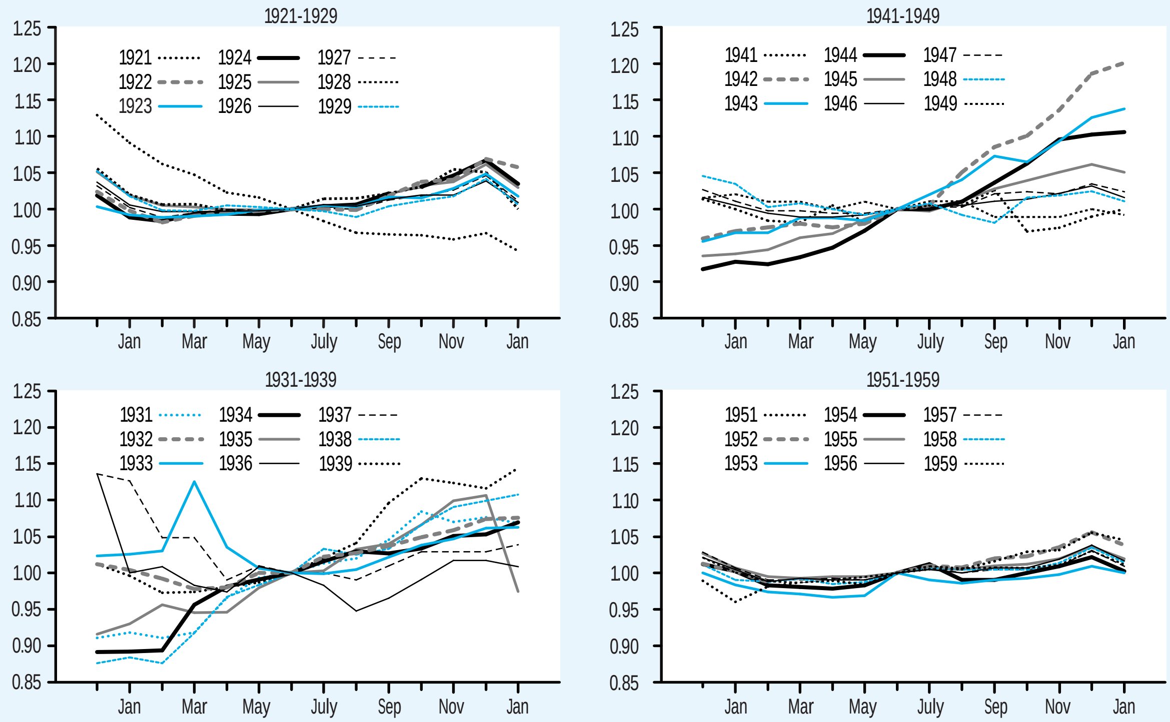


Figure 3a

Adjusted Monetary Base, NSA

Levels, Normalized to 1.0 in June



malized to the June level of each year. (Because of software limitations, only nine years of a decade are shown in each panel.) The historically unprecedented sharp drop in the level of the monetary base during 1920 and 1921 is apparent in the first panel. The other years of the decade display a clear, uniform seasonal pattern, with the exception, perhaps, of the second half of 1929. In contrast, little if any regular seasonal variation is apparent during the 1930s and 1940s, at least through 1945. The latter years of the 1940s hint at an emerging pattern similar to that observed during the 1920s. A more regular seasonal pattern does emerge during the 1950s, and the pattern tightens during the 1960s, 1970s, and 1980s. The 1990s display more variation, but the same underlying pattern is clear.

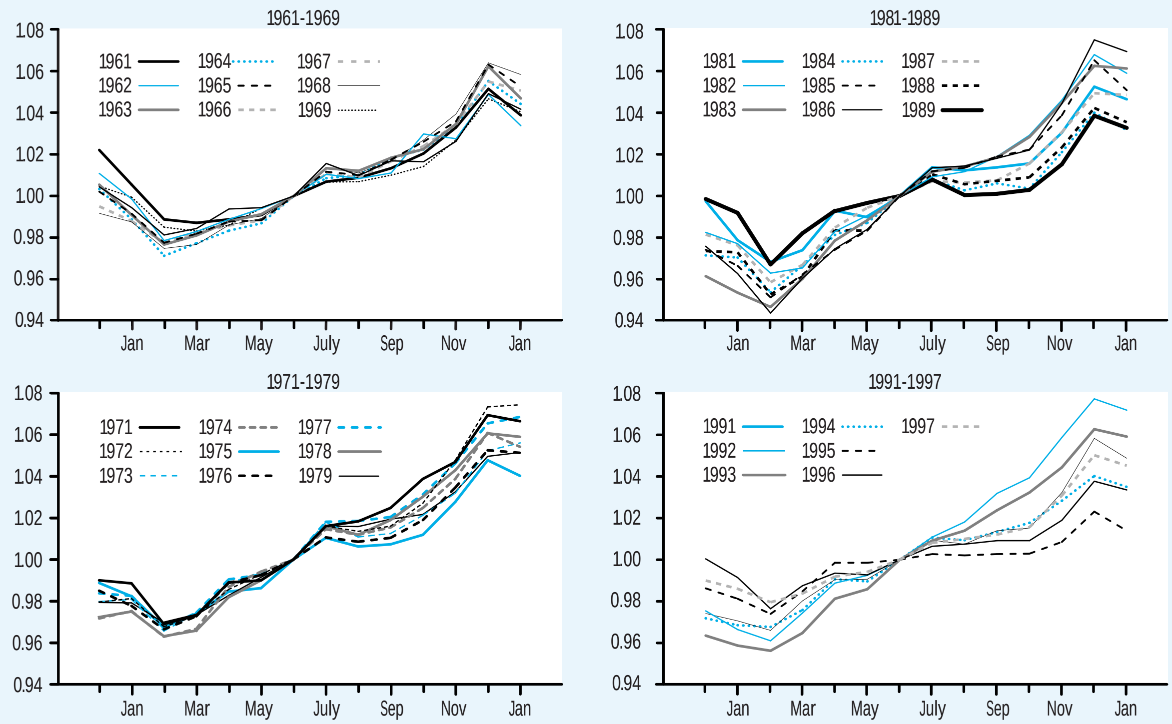
The stability of monthly seasonal variation across decades is further explored in Figure 4 through a series of month plots. In each panel, the growth rates for one month are shown for all years in the sample. Month plots are particularly useful for assessing when an uncertain seasonal pattern settles into a stable pattern, or vice versa. In Figure 4, the decades of the 1930s and 1940s are shaded, and 1950—the year when our earlier analysis suggests some emerging stability—is marked. The month plots suggest two conclusions. First, seasonal monthly growth rates tend to stabilize after 1950, and second, there may be significant time variation in the seasonal patterns (stochastic seasonality).

Overall, these results suggest that the adjusted monetary base displays a persistent seasonal pattern that becomes apparent during the 1920s, is interrupted by the Depression and World War II, and reestablishes itself after 1950 (as the Federal Reserve relaxes its pegging of government bond prices). *We emphasize that this seasonal data generating process is indeed very*

Figure 3b

Adjusted Monetary Base, NSA

Levels, Normalized to 1.0 in June



special. Because base money can neither be created nor destroyed by the private sector of the economy, this seasonal pattern necessarily reflects, in full extent, the Federal Reserve's actions to smooth seasonal fluctuations in the demand for base money and market interest rates.

Tests Based on Stochastic Seasonality

We next extend our analysis of seasonal variation to allow for time-varying, stochastic seasonality, by use of the Bureau of the Census X11 and X12–regARIMA seasonal adjustment programs. In this analysis, it is important to appreciate two aspects of the X11 algorithm: outlier replacement and the use of two-sided moving average filters. First, the X11 algorithm searches for outliers during estimation of seasonal adjustment factors, and replaces these data points with more moderate observations.[8] Outliers are detected by a sequential procedure. First, the X11 algorithm removes a stochastic (time-varying) trend from the data. Next, the algorithm searches, roughly in five-year intervals, for data points more than 1.5 standard deviations from the mean. By default, points between 1.5 and 2.5 standard deviations are replaced with a linear combination of values at 1.5 and 2.5 deviations; points beyond 2.5 standard deviations are replaced with values at 2.5 standard deviations. (These bounds may be changed by the user.)

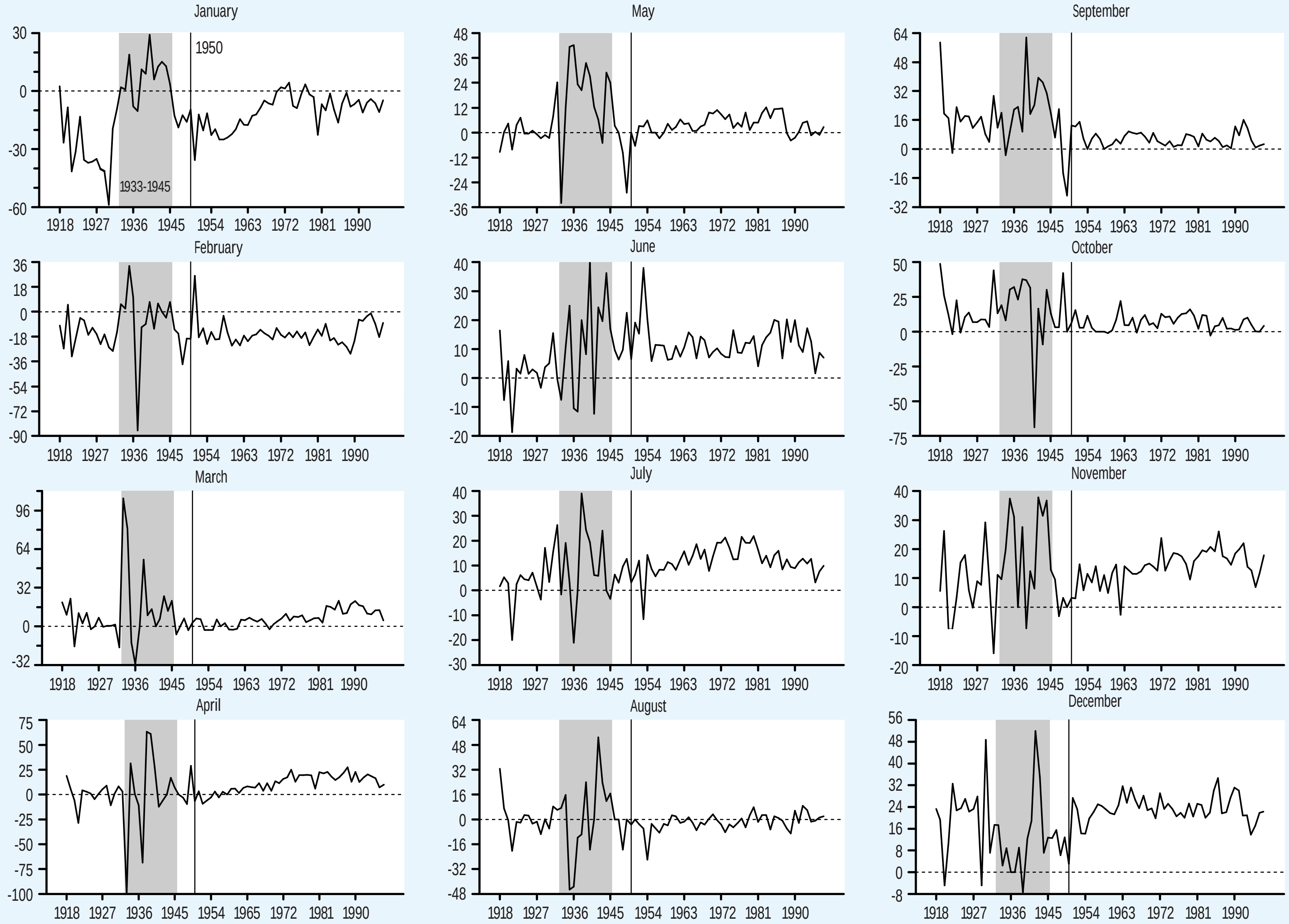
It is well known that, as a result of this replacement process, the X11 algorithm might (in some circumstances) produce apparently reasonable seasonal factors even for extremely noisy data. Some recent analysis suggests that this outlier replacement algorithm may be augmented by pre-filtering the series via the introduction of intervention terms in regression models with ARIMA disturbance processes (Findley et. al., 1998). Second, X11 and X12 use centered two-sided moving

[8] den Butter and Fase (1991) describe the algorithm.

Figure 4

Log of Adjusted Monetary Base, 1918-1997

Increase From Previous Month, by Month, Percent Annual Rate, NSA



average filters that eventually are—after all steps are completed and if series length permits—approximately 12 years wide. Nearer to the ends of the sample, the filters are truncated by folding the weights back onto the observed data. If there are structural breaks in the data generating process, such seasonal adjustment factors may be affected for a significant number of periods *before* the break.

The X12-regARIMA package consists of two parts. The first fits seasonal Box-Jenkins (*p,d,q*)x(*P,D,Q*) ARIMA models to the data series, with intervention terms as suggested by Box and Tiao (1975). These are regression models with ARIMA error processes, or regARIMA models. The X12–regARIMA program can generate several types of intervention variables: *additive outliers*, or single anomalous observations; *level shifts*, or permanent shifts in the series; *temporary changes*, where the level of the series is unusually high or low but decays for several periods back to its previous regime; and *ramps*, where the level may take several periods to move up or down permanently. After a sequence of general-to-specific tests, the accepted dummy variables are included as intervention terms in a dynamic regression model with a seasonal ARIMA disturbance. Forecasts from the regARIMA model also are appended to the series. Next, the X11 algorithm is used to obtain seasonal adjustment factors.

For messy data, there is some evidence that the X12 ARIMA-model pre-filtering may provide superior estimates of the seasonal adjustment factors. In our view, a recommendation by the X12 algorithm that a large number of intervention terms should be added to the ARIMA model also suggests that an unusual shock (or sequence of shocks) has disrupted the regular seasonal pattern of economic activity. If so, it perhaps is unwise to allow the X11 or X12 programs to replace a large number of observations because a subsequent X11 estimation might appear to find stable seasonal patterns when they are, in fact, not present.

Figure 5 shows growth rates and X11-estimated seasonal factors for various periods beginning with January 1918 and ending with December for the years 1929, 1932, 1933, and 1934. (Data shown in the figure include seasonal factors that are forecast by the X11 program for dates after the end of sample used for estimation.) Our previous analysis suggested a possible break in the data generating process circa 1932. This is confirmed by the estimates: Even with replacement of extreme observations, the seasonal adjustment factors for most months between 1928 and 1932 are strongly affected by the inclusion of data for 1933 and 1934. When estimating the period between 1918 and 1932, the X11 algorithm replaced most observations for the months of July through December between 1918 and 1920, as well January and February 1921; October 1921 through February 1922; December 1929 through February 1930; and October through December 1930. Historical events suggest that these replacements may be reasonable. The years between 1918 and 1921 were turbulent, with substantial gold flows associated with the war and its aftermath. The period between October and December 1930 includes the first banking crisis during the 1930s.

Results obtained from the X12 program are shown in Figure 6. Based on a seasonal (*1,1,1*)*x*(*0,1,1*) ARIMA model, the X12 algorithms suggested eight outliers, marked by vertical lines and labeled by type in Figure 6. Perhaps most prominent are the intervention variables in October 1931 when Britain left the gold standard and bank runs in the United States surged; in March 1933 the month of the U.S. "Bank Holiday;" and in early 1934, when the Federal Reserve sharply curtailed growth of Federal Reserve credit. We used the X11 algorithm within the X12–regARIMA package to calculate seasonal adjustment factors after accounting for these interventions. For four sample periods—each beginning in January 1918 and ending, respectively, in December of 1929, 1932, 1933, and 1934—the factors resembled those obtained from conventional X11 estimation, and hence, are not shown.

Seasonal variation between 1933 and 1949 is examined in detail in Figures 7 and 8. Our previous statistical results, based on deterministic seasonal effects, suggested that life simply was not the same during

Figure 5

Comparative Estimates of a Seasonally Adjusted Monetary Base

Estimated Over Period Indicated at Left

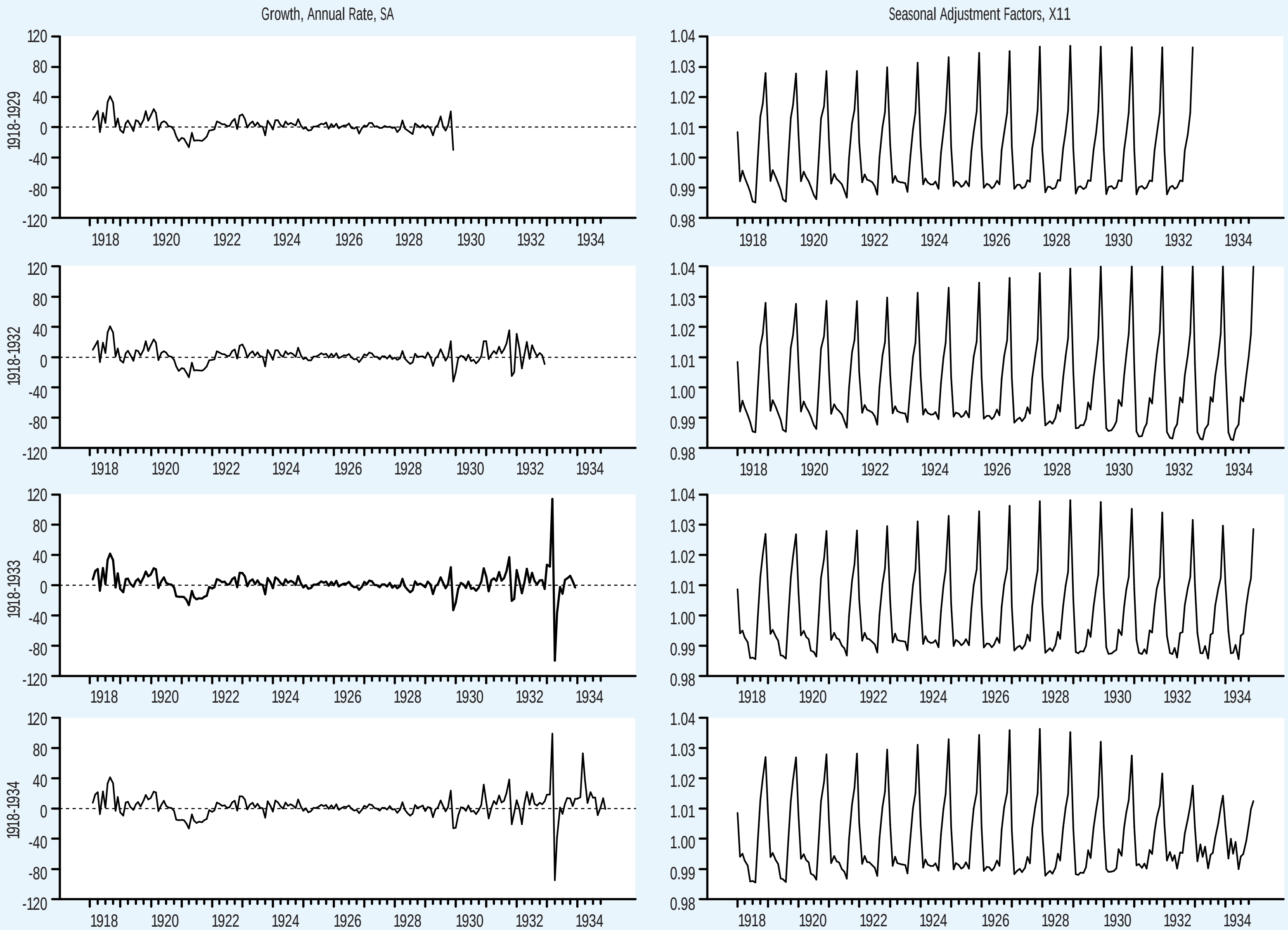
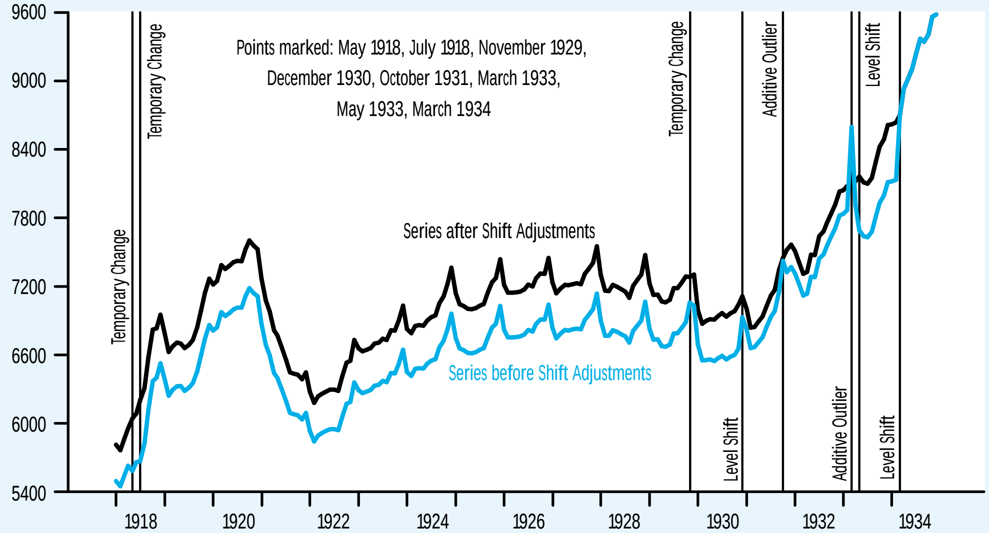


Figure 6

Outlier Analysis of Adjusted Monetary Base Using X12-regARIMA

January 1918-December 1934, Millions of Dollars, NSA



these decades. This, of course, is easy to rationalize: Extraordinary economic events such as the Great Depression and World War II likely interrupted previous seasonal patterns. During the Depression, many families had insufficient income to support their usual summer travel and holiday buying habits. During the war, despite adequate income, many consumer goods were either in short supply or rationed. The weakness of the seasonality is evident in both the growth rates and autocorrelation function of the seasonally unadjusted data shown in Figure 7. Nevertheless, the data do display some seasonality, with small spikes in the autocorrelation function for seasonally unadjusted data at lags of 6, 12, and 24 months. Both the X11 and X12 packages return factors that annihilate the 12-month seasonal correlation. Yet, these patterns are distinctly different between decades, which suggests that the estimates are imprecise.

Further results for the period between 1933 and 1949, based on the X12 program, are shown in Figure 8. For data between January 1918 and December 1949, the program's regARIMA outlier identification algorithms suggest 25 intervention terms to handle outliers; the dates associated with these terms are marked with vertical lines in Figure 8. The months marked in the figure are November 1929; December 1930; October 1931; March and May 1933; March 1934; February 1935; February, May, and August 1936; February and April 1937; March and July 1938; April, September, November, and December 1939; August 1940; October 1941; August 1942; October 1943; October 1948; and, April and September, 1949. Note that each of these intervention variables may affect observations for more than one month, as is evident in Figure 6.

In our judgment, based on estimates from both the X11 and X12–regARIMA programs, seasonal variation between 1933 and 1949 was more than likely altered sharply by extraordinary economic events. Any residual seasonality is both too uncertain and unstable to allow usable estimates of seasonal adjustment factors. Hence, we include these years in our final series *without* seasonal adjustment.

Estimated seasonal adjustment factors for the last five decades of our data, between

Figure 7

Seasonal Adjustment of Adjusted Monetary Base, 1933-1949

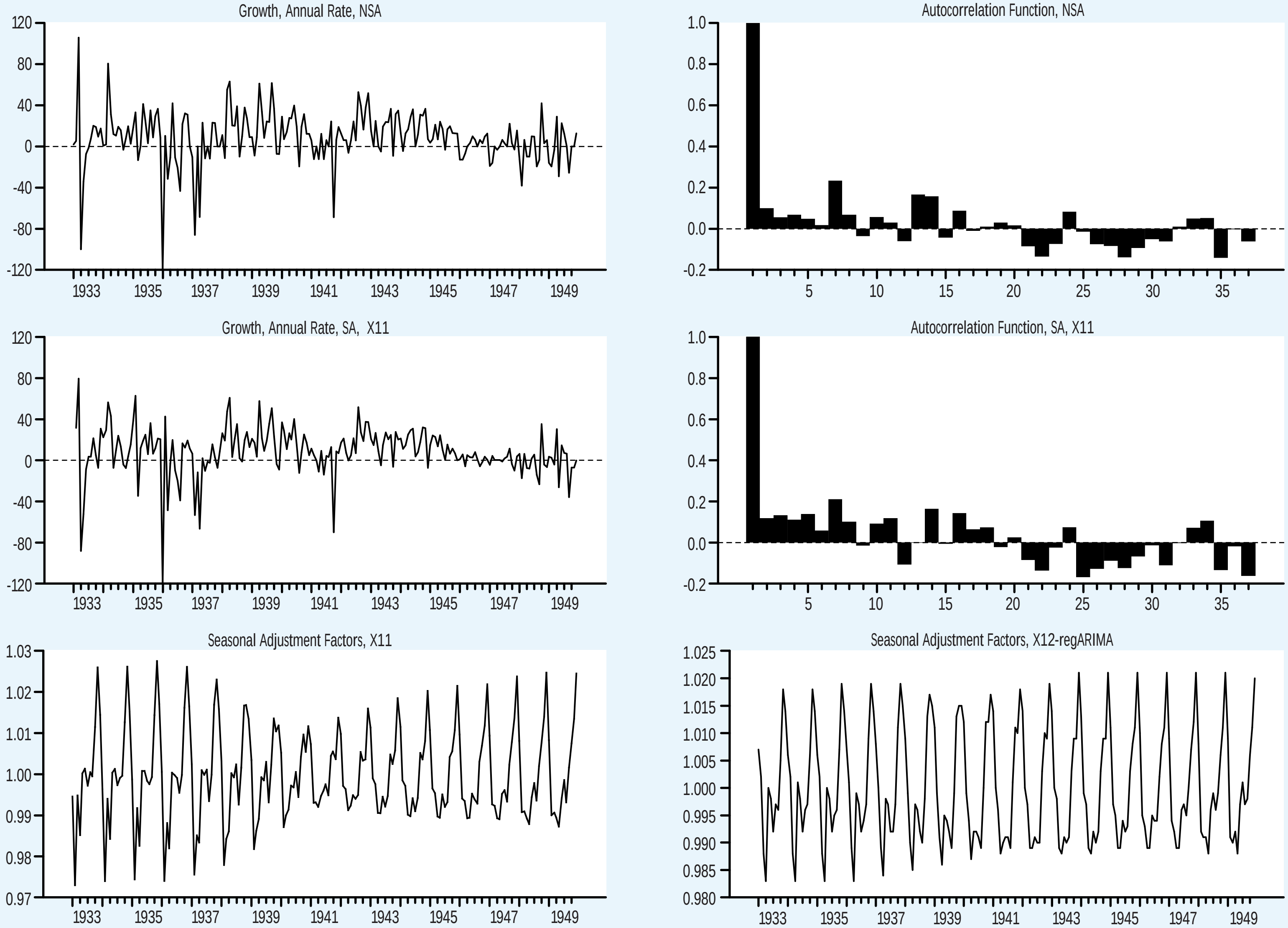
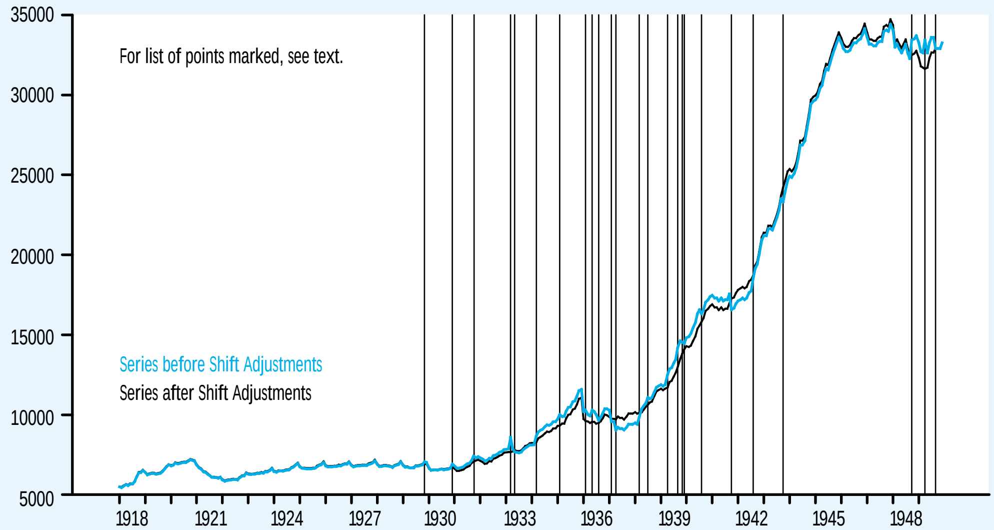


Figure 8

Outlier Analysis of Adjusted Monetary Base Using X12-regARIMA

January 1918-December 1949, Millions of Dollars, NSA



1950 and 1997, are shown in Figure 9. Growth rates also are shown for both the seasonally adjusted and unadjusted data. As our earlier statistical tests suggest, seasonal patterns during this period are relatively steady. Seasonal adjustment factors from both the X11 and X12 programs display some time variation but, except for a diminution during the 1990s, are relatively stable. The X12-regARIMA outlier analysis suggests interventions in December 1950, January 1951, June 1953, August 1954, February 1958, October 1962, April 1980, and January 1981. The diminution during the 1990s likely reflects the increasing amount of U.S. currency held outside the United States (Anderson and Rasche, 1998).

SUMMARY AND CONCLUSIONS

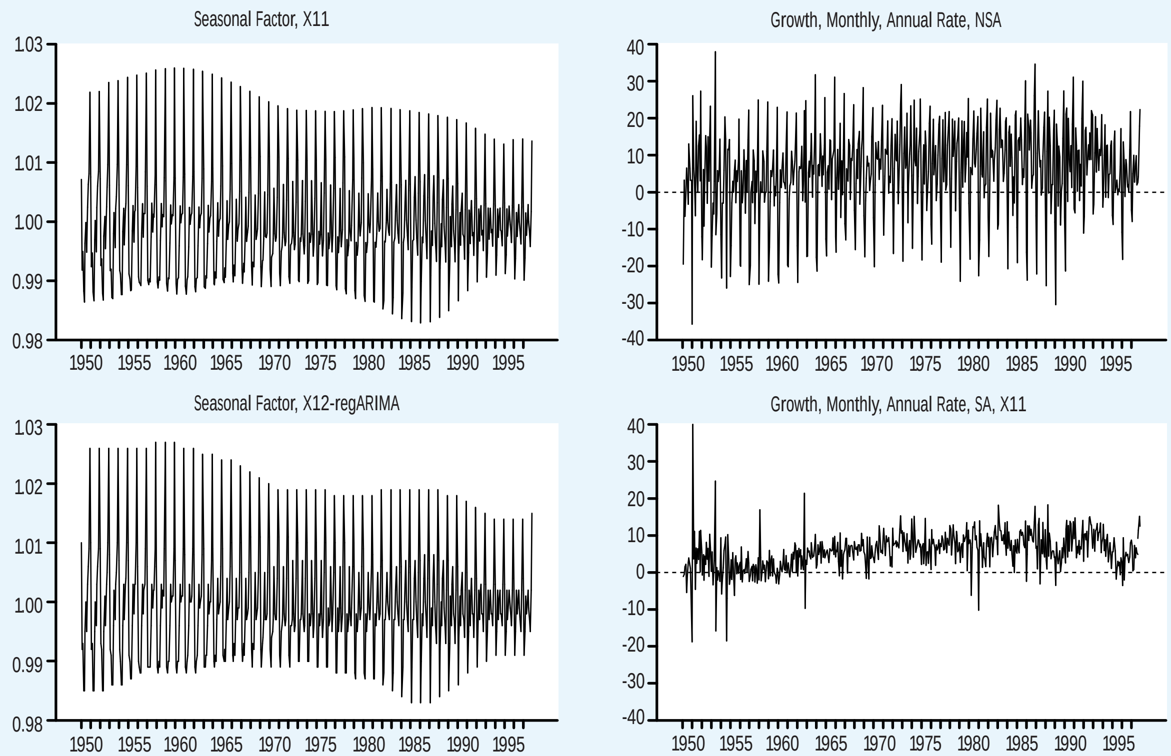
This study has extended the adjusted monetary base published by the Federal Reserve Bank of St. Louis to include the period between 1918 and 1935. It seems unlikely that this measure of the monetary base can be pushed any earlier, due to a scarcity of data and the major changes in the structure of reserve requirements that were implemented in 1917. Month-average data necessary to measure the monetary base, for example, begin August 1917 (see the shaded insert "Why Does RAM Begin in 1917?"). Readers are cautioned that attempts to splice our data to other series (such as Friedman and Schwartz's) must somehow adjust for the differences between daily-average and end-of-period data.

Our previous research has provided data on the adjusted monetary base beginning January 1936. In this analysis, we have introduced an additional RAM adjustment, RAM(1922), to measure the effects of changes in statutory reserve requirements between 1917 and 1935. We chained our previously estimated adjusted monetary base to these new data in August 1935, providing a consistently measured 80 years of the adjusted monetary base for the United States.

An examination of the seasonal properties of the adjusted monetary base suggests that estimates of stable seasonal

Figure 9

Seasonal Adjustment, 1950-1997



factors are readily obtained between 1918 and 1932 and between 1950 and 1997. But, our analysis suggests that seasonal variation in the adjusted monetary base nearly vanished during the Great Depression and World War II: These were not years of business as usual for the economy. We include those data in our final series on a *seasonally unadjusted* basis, to avoid inducing spurious seasonal patterns in the final *seasonally adjusted* data.

REFERENCES

Andersen, Leonall C., and Jerry L. Jordan. "The Monetary Base—Explanation and Analytical Use," this *Review* (August 1968), pp. 7-11.

Anderson, Richard G., and Robert H. Rasche. "A Revised Measure of the St. Louis Adjusted Monetary Base," this *Review* (March/April 1996) (1996a), pp. 3-13.

________ and ________ "Measuring the Adjusted Monetary Base in an Era of Financial Change," this *Review* (November/December 1996) (1996b), pp. 3-37.

________ and ________ "The Domestic Adjusted Monetary Base," mimeo, Federal Reserve Bank of St. Louis (July 1998).

Balbach, Anatol B., and Albert E. Burger. "Derivation of the Monetary Base," this *Review* (November 1976), pp. 2-8.

Barsky, Robert B., and Jeffrey A. Miron. "The Seasonal Cycle and the Business Cycle," *Journal of Political Economy* (June 1989), pp. 503-34.

Beaulieu, J. Joseph, Jeffrey K. MacKie-Mason, and Jeffrey A. Miron. "Why Do Countries and Industries with Large Seasonal Cycles Also Have Large Business Cycles," *Quarterly Journal of Economics* (May 1992), pp. 621-56.

Board of Governors of the Federal Reserve System. *Banking and Monetary Statistics 1914–1941* (November 1943).

________. *Twenty-Second Annual Report of the Board of Governors of the Federal Reserve System, Covering Operations for the Year 1935* (May 1936).

________. *Twenty-Third Annual Report of the Board of Governors of the Federal Reserve System, Covering Operations for the Year 1936* (June 1937).

________. *Fifty-Ninth Annual Report of the Board of Governors of the Federal Reserve System, 1972* (May 1973).

________. *Federal Reserve Bulletin*, various issues.

Box, George, and George C. Tiao. "Intervention Analysis with Applications to Economic and Environmental Problems," *Journal of the American Statistical Association* (March 1975), pp. 70-79.

Brunner, Karl, and Allan H. Meltzer. "The Federal Reserve's Attachment to the Free Reserve Concept," Originally presented to the U.S. Subcommittee on Domestic Finance on May 7, 1964. Reprinted in *Monetary Economics*, Karl Brunner and Allan H. Meltzer, eds., Basil Blackwell (1989), pp. 21-96.

Burger, Albert E., and Robert H. Rasche. "Revision of the Monetary Base," this *Review* (July 1977), pp. 13–28.

den Butter, F.A.G., and M.M.G. Fase. *Seasonal Adjustment as a Practical Problem*, North Holland (1991).

Deutsche Bundesbank. "Review of the monetary target and restructuring of the minimum reserve regulations," *Monthly Report* (July 1995), pp. 17-35.

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, and Bor-Chung Chen. "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program," *Journal of Business and Economic Statistics* (April 1998), pp. 127-52.

Friedman, Benjamin M. "Targets and Instruments of Monetary Policy," *Handbook of Monetary Economics*, vol. 2, Benjamin M. Friedman and Frank H. Hahn, eds., Elsevier Science Publishers (1990), pp. 1186–230.

Friedman, Milton, and Anna J. Schwartz. *A Monetary History of the United States*, *1867-1960*, Princeton University Press (1963).

Harvey, Andrew, and Andrew Scott. "Seasonality in Dynamic Regression Models," *Economic Journal* (November 1994), pp. 1324–45.

McCallum, Bennett T. "Issues in the Design of Monetary Policy Rules," *NBER Working Paper No. 6016* (April 1997).

________ "Some Issues Concerning Interest Rate Pegging, Price Level Determinancy and the Real Bills Doctrine," *Journal of Monetary Economics* (January 1986), pp. 135-60.

McCallum, Bennett T., and Monica Hargraves. "A Monetary Impulse Measure for Medium-Term Policy Analysis," *Staff Studies for the World Economic Outlook*, International Monetary Fund (September 1995), pp. 52-70.

Moody's. *Bank Directory 1928* (1928).

Patinkin, Don. "Financial Intermediaries and the Logical Structure of Monetary Theory: A Review Article," *American Economic Review* (March 1961), pp. 95-116.

Poole, William. "Optimal Choice of Monetary Policy Instruments in a Simple Stochastic Macro Model," *Quarterly Journal of Economics* (May 1970), pp. 197-216.

Rand McNally and Co., *Bank Directory 1919* (1919).

Tatom, John. "Issues in Measuring An Adjusted Monetary Base," this *Review* (December 1980), pp. 11-29.

Wallis, Kenneth. "Seasonal Adjustment and Relations Between Variables," *Journal of the American Statistical Association* (March 1974), pp. 18-31.

Robert Dittmar is a mathematician and William T. Gavin is vice president and research coordinator for the Federal Reserve Bank of St. Louis. Finn E. Kydland is a professor of economics at Carnegie Mellon University and visiting scholar of the Federal Reserve Bank of St. Louis.
Daniel R. Steiner provided research assistance.

The Inflation-Output Variability Tradeoff and Price-Level Targets

Robert Dittmar, William T. Gavin, and Finn E. Kydland

The stated long-term goal of monetary policy in the United States and around the world is price stability. Eight countries in the world now have explicit targets for inflation. Many more, including the United States, appear to operate as if they have implicit targets for inflation. There is an ongoing debate about how strictly one should try to target inflation. The idea is that if one tried to keep inflation too close to a target, there would be a significant increase in the variability of output and interest rates.

There is a subtle but important distinction to be made about the difference between targeting inflation in the short run (say, every month) and targeting a particular average inflation rate over many months. By targeting a long-moving average of zero inflation, or a horizontal price-level path, the central bank would have an operational target for price stability, but would not be required to keep inflation on an exact path each month or quarter. Objections to price-level targeting usually assume that any economic disturbance that caused the price level to deviate from the target would require the central bank to react immediately, and harshly, to get the price level back on track. But, there is no reason for this. Whether targeting inflation more closely in the long-run would lead to more or less short-run variability of inflation and output depends on how the economy works and how the central bank runs monetary policy.

By price-level targeting we mean that the central bank announces a path for the price level. It may be flat or it may be changing at a rate of x percent per year. For $x = 0$, the price level path will be horizontal. In any case, the notion of a price-level target means that the central bank will target a long-run average inflation rate, setting objectives that correct for past deviations from the target. Technically, we define a price-level-targeting regime as one in which the logarithm of the price level has a deterministic trend. An inflation-targeting regime is one in which the logarithm of the price level has a unit root and follows a stochastic trend. Results in this paper apply to a price-level target whether the average inflation rate is zero or not.

Taylor (1979) introduced the idea of using the inflation/output variability tradeoff to examine alternative monetary policy rules. Using a rational expectations model with staggered wage contracts, he explained why the choice facing policymakers in a dynamic setting involves the tradeoff between output variability and inflation variability. In his rational expectations framework there is no long-run tradeoff between levels of output and inflation. Policymakers can, however, choose alternative points along an inflation/output variability frontier by varying the relative weight they put on inflation versus output stabilization.[1]

Using a simplified version of Taylor's framework, Svensson (1997b) shows that, for a given level of output variability, the short-run variability of inflation depends on the amount of persistence in the output gap and on whether the central bank targets an inflation rate or a path for a price index. He shows that if the output gap is persistent enough, the central bank should target a

[1] For a more detailed description of the intuition underlying the inflation/output variability tradeoff, see Taylor (1994).

path for the price level. Svensson also explains why a price-level target can be used as a commitment mechanism to eliminate the inflation bias that results when a central bank tries to target an unrealistically high level of output.[2] In this paper, we explain how the inflation-output variability tradeoff changes if the central bank chooses to target a predetermined path for the price level rather than an inflation rate. Our analysis is more transparent than Svensson's because we do not try to distinguish between cases of commitment and discretion, nor do we consider the case where the central bank tries to achieve an unrealistic objective for output. We assume that the central bank cannot commit credibly to more than one period at a time. Since the central bank does not try to achieve unrealistically high levels of output, the steady state inflation rates are the same for both inflation and price-level targeting regimes.

INFLATION VERSUS PRICE-LEVEL TARGETING IN A SIMPLE PHILLIPS CURVE MODEL

The basic model described here is from Svensson (1997a, 1997b). The model is consistent with a wide range of sticky-price models in which monetary policy can have important real effects. The model has three main elements: a multiperiod objective function for the central bank, an aggregate supply equation, and a rational expectations assumption.

The central bank minimizes an intertemporal quadratic loss function:

$$(1) \quad L = \sum_{t=0}^{\infty} \beta^t \left(\lambda y_t^{\,2} + (\pi_t - \pi^*)^2 \right),$$

where y_t is the deviation of output from the target level (which we assume is the underlying trend in real output) and $(\pi_t - \pi^*)$ is the deviation of inflation from the central bank's inflation target. The central bank discounts future variability in the output gap and inflation by the factor β. The parameter, λ, relates the central bank's preference for output stability to its preference for inflation stability.

The economy is modeled as a short-run aggregate supply curve with persistence in the output gap:

$$(2) \quad y_t = \rho y_{t-1} + \alpha(\pi_t - \pi_t^e) + \varepsilon_t.$$

The introduction of a lagged output gap in this equation is important for comparing inflation and price-level targeting. Conceptually, the lag will be introduced any time friction prevents instantaneous and complete adjustment of output to unexpected changes in the price level. This friction could be induced by wage contracts, menu costs, transaction costs, incomplete markets, capital adjustment costs, etc. The slope of the short-run Phillips Curve is given by α which determines the response of the output gap to unexpected inflation $(\pi_t - \pi_t^e)$.

With this aggregate supply curve and rational expectations, that is, $\pi_t^e = E_{t-1}\pi_t$, the central bank's optimization problem implies a tradeoff between output and inflation variability. Minimizing this loss function—subject to the aggregate supply curve—leads to a rule for inflation that is contingent on the size of the output gap:

$$(3) \quad \pi_t^A = p_t^A - p_{t-1} = \pi^* - \frac{\alpha\lambda\rho}{1-\beta\rho^2} y_{t-1} - \frac{\alpha\lambda}{1-\beta\rho^2+\alpha^2\lambda} \varepsilon_t,$$

where the superscript A indicates that the variable is determined by the inflation-targeting rule and p is the logarithm of the price level. The inflation rate set in each period is equal to the inflation target with countercyclical adjustments proportional to the lagged output gap and the current technology shock. Following Svensson, we assume the central bank can control inflation directly.[3] Details of the solution procedure are presented in the appendix.

If the central bank cares about deviations of the price level rather than the

[2] Gavin and Stockman (1991) explain why a society that cares about inflation (not price level) stability may still prefer a price level target if the source of inflation shocks is unobservable to the public.

[3] An appendix in Svensson (1997b) shows that introducing money with a control error in the inflation equation would not change his results.

inflation rate, the natural logarithm of the price level will replace the inflation rate in the loss function. We reformulate the objective function as below:

$$(4) \qquad L = \sum_{t=0}^{\infty} \beta^t \left(\lambda y_t^{\,2} + (p_t - p_t^*)^2 \right),$$

where the target path for the price level may be constant or may be rising at a constant rate.

The central bank's rule for achieving the target path is given by:

$$(5) \qquad p_t^B = p_t^* - \frac{\alpha\lambda\rho}{1-\beta\rho^2}\, y_{t-1} - \frac{\alpha\lambda}{1-\beta\rho^2+\alpha^2\lambda}\, \varepsilon_t ,$$

implying the following rule for the inflation rate:

$$(6) \qquad \pi_t^B = p_t^B - p_{t-1} = \pi^* - \frac{\alpha\lambda\rho}{1-\beta\rho^2}(y_{t-1} - y_{t-2}) - \frac{\alpha\lambda}{1-\beta\rho^2+\alpha^2\lambda}(\varepsilon_t - \varepsilon_{t-1}),$$

where we have used the assumption that the price-level target, p_t^*, is given by $p_t^* = \pi^* + p_{t-1}^*$. The superscript *B* indicates that the variable is determined by the price-level targeting rule. With the price-level target, the central bank's reaction function, Equation 6, has three elements on the right-hand side. The first is the steady-state inflation embodied in the target path for the price level. The second and third are proportional, countercyclical adjustments to the change in the output gap from period $t-2$ to period $t-1$ and the change in the technology shock from period $t-1$ to period t, respectively.

The tradeoff between inflation and output is qualitatively different under the two different regimes, inflation targeting and price-level targeting. In an inflation-targeting regime, the bank sets inflation, π_t^A, as shown in Equation 3. With rational expectations, the model's Phillips Curve implies that output is given by:

$$(7) \qquad y_t = \rho\, y_{t-1} + \frac{1-\beta\rho^2}{1-\beta\rho^2+\alpha^2\lambda}\,\varepsilon_t \; .$$

As the relative weight on output variability, λ, gets large, the coefficient on the error term tends to zero as does the variance of the output gap. If the variance of ε_t is σ_ε^2, then the above decision rule for y_t implies that the unconditional variance of the output gap is:

$$(8) \qquad \sigma_y^2 = \frac{(1-\beta\rho^2)^2}{(1-\rho^2)(1-\beta\rho^2+\alpha^2\lambda)^2}\,\sigma_\varepsilon^2 \; .$$

After noting that ε_t is uncorrelated with y_{t-1}, we can use the decision rule for π_t to calculate the unconditional variance of inflation as:

$$(9) \qquad \sigma_\pi^2 = \left(\frac{\alpha\lambda\rho}{1-\beta\rho^2}\right)^2 \sigma_y^2 + \left(\frac{\alpha\lambda}{(1-\beta\rho^2+\alpha^2\lambda)}\right)^2 \sigma_\varepsilon^2 ,$$

which can be simplified to yield an expression only involving σ_ε^2, namely:

$$(10) \qquad \sigma_\pi^2 = \frac{\alpha^2\lambda^2}{(1-\rho^2)(1-\beta\rho^2+\alpha^2\lambda)^2}\,\sigma_\varepsilon^2 \; .$$

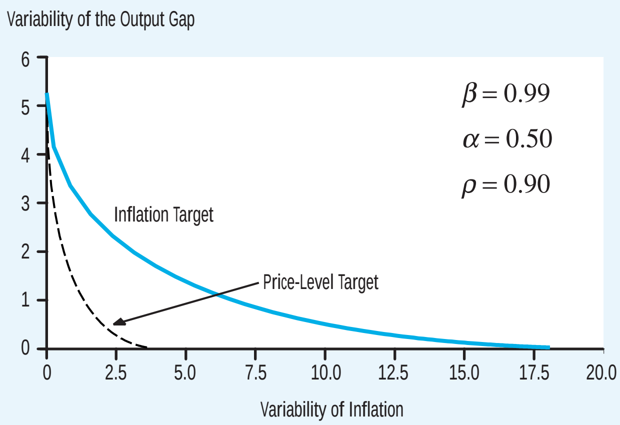
In a price-level-targeting regime, the central bank sets the inflation rate, π_t^B, as in Equation 5. Once again assuming rational expectations, $p_t^e = E_{t-1}p_t$, the following time series process for the output gap is derived from the model's Phillips Curve,

$$(11) \qquad y_t = \rho y_{t-1} + \frac{1-\beta\rho^2}{1-\beta\rho^2+\alpha^2\lambda}\,\varepsilon_t \; .$$

Note that this process for the output gap looks identical to Equation 7, which was derived in the inflation-targeting regime. The parameter λ, however, has a different

Figure 1

The Output-Inflation Variability Tradeoff



interpretation here, as the bank's preferences are different. The unconditional variance of the output gap as a function of this parameter is given by the same expression as noted in Equation 8. The unconditional variance of the inflation rate, however, is given by the following expression:

$$(12) \qquad \sigma_\pi^2 = \frac{2\alpha^2\lambda^2}{(1+\rho)(1-\beta\rho^2+\alpha^2\lambda)^2}\sigma_\varepsilon^2.$$

Regardless of whether the central bank is targeting inflation or the price level, a small weight on the output gap leads the bank to strive for keeping inflation or the price level close to its target. At the extreme, where the central bank places no weight on deviations of the output gap, the variance of the gap is determined by persistence in the output gap, ρ, and the variance of technology shocks. Here, the bank optimizes by fixing inflation, or the price level, at its target in every period. There is no inflation variability, no inflation uncertainty, and a simple autoregressive process for the output gap. Conversely, a large weight on the deviation of the output gap from the target would lead the bank to use the Phillips Curve constraint to closely control the output gap by letting inflation vary more.

We graphically display the difference between the inflation/output variability tradeoffs in the two regimes by first expressing the output gap variance and the inflation variance as functions of the preference parameter λ while holding the parameters of the Phillips Curve constant. For a given λ, the bank's decision rules can be used to calculate an unconditional variance for both inflation and the output gap (a single point in Figure 1). Varying the bank's preferences by varying λ will determine the location of the curve representing the trade off between σ_π^2 and σ_y^2.

A sample pair of variance tradeoff curves are displayed in Figure 1. For the chosen set of parameter values, the variance tradeoff under the price-level-targeting regime lies everywhere below that for the inflation-targeting regime. Thus, given this particular set of parameters, society would prefer the price-level-targeting regime.

More can be said about the relative position of these tradeoff curves. If we examine the expressions for the unconditional variances of the output gap and inflation derived above, we can fully describe the position of these curves in terms of the autoregressive parameter, ρ, in the Phillips Curve equation. Note that in either regime, if the bank places no weight on deviations of the output gap from target, then the bank simply sets the inflation rate, or the price level, equal to its target in every period. Thus, in the limit as the parameter λ approaches 0, the unconditional variance of inflation approaches 0, while the unconditional variance of output approaches that of the simple first-order autoregressive process $y_t = \rho y_{t-1} + \varepsilon_t$. Thus, the two tradeoff curves intersect the σ_y^2-axis at the same point.

If the central bank's weight on deviations of the output gap from target becomes large, then the central bank sets the output gap equal to its target and manipulates the inflation rate to reach this goal. Thus, as the parameter λ approaches infinity, the variance of output approaches 0. Examining the expressions for the unconditional variance of inflation shows that as λ approaches infinity, the variance of inflation under an inflation-targeting regime approaches $(\alpha^2(1-\rho^2))^{-1}$, and the variance of inflation under a price-level-targeting regime approaches

$2(\alpha^2(1+\rho))^{-1}$. Therefore, assuming that the tradeoff curves are convex for all parameter values, the tradeoff curves under price-level-targeting regimes will lie below those for inflation-targeting regimes as long as

$$2(\alpha^2(1+\rho))^{-1} < (\alpha^2(1-\rho^2))^{-1}, \quad (13)$$

or equivalently, $\rho > 1/2$.[4] Note that the relative position of the tradeoff curves does not depend on α, the slope of the short-run Phillips Curve, or on β, the central bank's discount factor.

We can gain some insight for the relative placement of the curves under the above condition by considering what happens as the auto-regressive parameter, ρ, approaches 1. As this happens, the output gap starts to behave more and more like a random walk. Under the inflation-targeting regime, the bank sets the inflation rate proportional to the output gap. Consequently, if the output gap behaves like a random walk, so will the inflation rate. Under the price-level-targeting regime, however, the bank sets the inflation rate proportional to the change in the output gap. Thus, even if the output gap becomes non-stationary as ρ approaches 1, the time path of the inflation rate remains stationary under such a regime.

EMPIRICAL EVIDENCE

The simple Phillips Curve model represents popular wisdom about the tradeoff between inflation and output variability. It is instructive to examine estimates of the persistence in the output gap. We use U.S. gross domestic product (GDP) data where we calculate three different measures of the output gap from three different measures of potential GDP:

- Congressional Budget Office (CBO) estimates.
- A quadratic time (QT) trend calculated using the logarithm of real GDP.
- A Hodrick-Prescott (HP) trend also calculated using logarithm real GDP.

Table 1

Output Gap Under Alternate Definitions of Trend GDP

	CBO	Quadratic Time	Hodrick-Prescott
CBO	2.7%		
Quadratic Time	0.84	3.1%	
Hodrick-Prescott	0.79	0.76	1.7%

Values on the diagonal are the standard deviations of the output gap variously measured. Values on off-diagonals are the correlation coefficients between the respective measures of the output gap. Data are quarterly U.S. GDP from 1949:Q1 to 1998:Q2. The quadratic time gap is calculated as the residual in the following regression:

$$y_t = constant + \hat{\beta}_1 Time + \hat{\beta}_2 Time^2 + \hat{e}_t$$

where y_t is the logarithm of GDP and $\hat{e}$ is the estimated residual. The Hodrick-Prescott gap is the deviation from trend calculated using the filter described in Prescott (1986).

Table 2

Estimates of Persistence in the Output Gap Using U.S. GDP Data

$$\Delta y_t = c + \rho y_{t-1} + \sum \omega_i \Delta y_{t-i} + e_t$$

Definition of Trend	Estimate of ρ	Standard Error
CBO	0.91	0.03
Quadratic Time	0.92	0.03
Hodrick-Prescott	0.79	0.05

Data are quarterly U.S. GDP from 1949:Q1 to 1998:Q2.

In each case, we calculate the output gap as the difference between the logarithm of real GDP and the alternate estimates of the trend. Table 1 shows the sample standard deviations and correlations between the different measures of the output gap. The estimate based on the quadratic time trend is the most variable and the most highly correlated with the CBO estimate. We assume the CBO estimate is closest to the data that the policymakers actually use.

Table 2 shows the estimates of the autoregressive parameter calculated for each measure of the output gap. The

[4] Svensson (1997b) derived a similar result for the discretion case.

Table 3

Estimates of Persistence in the Output Gap

(Using industrial Production to Measure Output)

$$\Delta y_t = c + \rho y_{t-1} + \sum \omega_i \Delta y_{t-i} + e_t$$

Country	Hodrick Prescott Filter		Quadratic Time Trend Filter	
	Estimate of ρ	Standard Error	Estimate of ρ	Standard Error
Belgium	0.58	0.08	0.90	0.04
Canada	0.77	0.05	0.94	0.02
France	0.41	0.10	0.89	0.05
Germany	0.69	0.06	0.93	0.03
Italy	0.53	0.09	0.90	0.04
Japan	0.82	0.04	0.97	0.01
Netherlands	0.51	0.09	0.97	0.02
Sweden	0.57	0.07	0.96	0.02
United Kingdom	0.60	0.07	0.89	0.04
United States	0.70	0.05	0.93	0.03

Data for the G-10 are quarterly averages of monthly industrial production from 1957:1 to 1997:12 published by the International Monetary Fund.

equation used to estimate ρ is shown at the top of Table 2. The properties of the distribution for this estimate were discussed in Dickey and Fuller (1981). By construction, the output gap is stationary so there is no prior reason to expect estimates of ρ to be close to unity. We find surprisingly high estimates of ρ using both the QT gap (0.92) and the CBO gap (0.91), however. The HP trend follows the actual series more closely than the other two series. The standard deviation is much smaller and the estimate of ρ is only 0.72. Even in this case, however, the estimate is still more than four standard deviations larger than 0.50. This confirms Svensson's result that if one believes in this output/inflation variability tradeoff, then setting a price-level target would most likely result in a more efficient set of options for the Fed than would an inflation target.

We also have estimated the persistence of the output gap in the G-10 countries. There is a lack of historical quarterly GDP data for the G-10, so we measured the persistence of the output gap in these countries by taking quarterly averages of industrial production and using both the HP and QT filters (see Table 3) to construct the output gap. Using the QT filter, ρ is estimated to be greater than 0.50 and highly significant in all the countries. Using the HP filter, the results are mixed. Only in one case is the point estimate below 0.50, but in over half of the cases, the estimate is within one standard deviation of 0.50.

CONCLUSION

In this paper we describe a popular model of monetary policy in which the central bank minimizes a discounted, multiperiod loss function that includes deviations of inflation and output from target levels. This minimization is constrained by a short-run tradeoff between inflation and real output. This simple model suggests that the choice between an inflation target and a price-level target depends on characteristics of real output. If the output gap is relatively persistent, then targeting the price level results in a better set of policy options for the central bank. We present evidence from the G-10 countries showing that conventionally measured output gaps are highly persistent. The policy implication of assuming rational expectations and this Phillips Curve model is that central banks should set objectives for a price level, not an inflation rate.

REFERENCES

Dickey, David A., and Wayne A. Fuller. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica 49* (June 1981), pp. 1057-72.

Gavin, William T., and Alan Stockman. "Why a Rule for Stable Prices May Dominate a Rule for Zero Inflation," *Economic Review*, Federal Reserve Bank of Cleveland, (1991 Quarter I), pp. 2-8.

Hodrick, Robert J., and Edward C. Prescott. "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit, and Banking*, 29 (February 1997), pp. 1-16.

Prescott, Edward C. "Theory Ahead of Business Cycle Measurement," *Carnegie-Rochester Conference Series on Public Policy*. New York: North-Holland, (Autumn1986).

Svensson, Lars E.O. "Optimal Inflation Targets, 'Conservative' Central Banks, and Linear Inflation Contracts," *American Economic Review 87* (March 1997a), pp. 98-114.

________. "Price Level Targeting vs. Inflation Targeting: A Free Lunch?" Institute for International Economic Studies, Stockholm University, August 1997b. An earlier version was published in August 1996 as NBER *Working Paper 5719*.

Taylor, John B. "The Inflation/Output Variability Tradeoff Revisited," in *Goals, Guidelines, and Constraints Facing Monetary Policymakers*, Jeffrey C. Fuhrer, ed., The Federal Reserve Bank of Boston Conference Series 38 (1994), pp. 21-38.

________. "Estimation and Control of a Macroeconomic Model with Rational Expectations," *Econometrica 47* (September 1979), pp. 1267-86.

Appendix

APPENDIX: SOLUTION OF THE CENTRAL BANK'S OPTIMIZATION PROBLEM

Since the central bank's objective under either the inflation targeting or price-level-targeting regime is quadratic and its constraints are linear, it is possible to guess that linear-decision rules solve the bank's optimization problem. We show that substituting the conjectured linear rules into the first-order conditions for the bank's optimization problem and equating coefficients will yield the decision rules described in the text. We treat inflation expectations as equilibrium variables are treated in a dynamic general equilibrium model. That is, we suppose that the bank bases its decisions at time t solely on the state variables y_{t-1} and ε_t while inflation expectations are left to be determined by a rational expectations condition.

Consider first the inflation-targeting regime. We form the bank's Lagrangian as:

$$\text{(A1)} \quad E_0\left\{\sum_{t=0}^{\infty}\beta^t\left[\begin{array}{l}\lambda y_t^2+\left(\pi_t-\pi^*\right)^2\\ -\mu_t\left(\begin{array}{l}y_t-\rho y_{t-1}\\ -\alpha\left(\pi_t-\pi_t^e\right)-\varepsilon_t\end{array}\right)\end{array}\right]\right\},$$

where the μ_t's are a sequence of random multipliers. The bank's first-order conditions take the form:

$$\text{(A2)} \quad 2\lambda y_t-\mu_t+\beta\rho E_t\mu_{t+1}=0\ ,$$

when taken with respect to the sequence of y_ts, and the form:

$$\text{(A3)} \quad 2\left(\pi_t-\pi^*\right)+\alpha\mu_t=0\ ,$$

when taken with respect to the sequence of π_ts. Eliminating the multipliers from these expressions gives the following Euler equation:

$$\text{(A4)} \quad \lambda y_t+\frac{1}{\alpha}\left(\pi_t-\pi^*\right) -\frac{\beta\rho}{\alpha}E_t\left(\pi_{t+1}-\pi^*\right)=0.$$

We now posit a linear decision rule for inflation of the form:

$$\text{(A5)} \quad \pi_t=A_1+A_2 y_{t-1}+A_3\varepsilon_t.$$

If expectations formed at time t-1 are rational then:

$$\text{(A6)} \quad \pi_t^e=A_1+A_2 y_{t-1}.$$

Hence, the constraint imposed by the aggregate supply relation (Equation 2 in the article) yields a decision rule for y_t directly of the form:

$$\text{(A7)} \quad y_t=\rho y_{t-1}+(\alpha A_3+1)\varepsilon_t.$$

Note that decision rules are invariant so that π_{t+1} can be determined by iterating on the rule for π_t to yield the following expression:

$$\text{(A8)} \quad \pi_{t+1}=A_1+A_2 y_t+A_3\varepsilon_t=A_1 +A_2\left(\rho y_{t-1}+\left(\alpha A_3+1\right)\varepsilon_t\right)+A_3\varepsilon_{t+1}.$$

Substituting Equations A5, A7, and A8 into the Euler Equation A4 above, taking expectations, and equating constant terms and coefficients on the states yields values for A_1, A_2, and A_3 in terms of parameters of the model.

Determining the bank's decision rules in the case of a price-level-targeting regime proceeds in a similar fashion. The only difference is the bank's price-level target changes over time, and hence p_t^* must enter as a state variable in the bank's decision rules. Since this target evolves in a deterministic manner as $p_t^*=p_{t-1}^*+\pi^*$, however, it is still possible to postulate a decision rule of the form $p_t = A_1 + A_2\, p_t^* + A_3\, y_{t-1} + A_4\, \varepsilon_t$, and iterate on it to calculate p_{t+1} in terms of time t states. After substituting the resultant linear rules into the bank's Euler equation and equating coefficients, we obtain the decision rule for price-level targeting given in the text.

Howard Wall is a senior economist with the Federal Reserve Bank of St. Louis. Eran Segev provided research assistance.

Using the Gravity Model to Estimate the Costs of Protection

Howard J. Wall

The United States, along with almost every other country in the world, maintains significant restrictions on the movement of goods across international borders. Although the recent Uruguay round of the General Agreement on Tariffs and Trade (GATT) resulted in a general lowering of tariffs and a broadening of goods and countries covered by the agreement, free trade remains elusive. Strong growth in the United States in recent years has kept protectionist pressures at bay, but recent calls for restrictions on steel and other imports suggest that protectionism in the United States has been dormant but is not dead.

When a country restricts imports, foreign producers are disadvantaged relative to their domestic competitors, and the volume of trade is reduced. This prevents both importing and exporting countries from realizing all of the gains from international trade, as resources will be diverted from industries where there are comparative advantages. The objective of this paper is to provide new estimates of the effects of protectionism on the volume of U.S. trade, and to obtain rough estimates of the resulting welfare effects.

Table 1 summarizes the variety of trade barriers imposed by the United States and its trading partners in the rest

Table 1

Summary of Trade Barriers in the United States and the ROW

Trade Barriers	ROW Against U.S. Goods[a]	United States Against ROW Goods[b]
Import policies	Tariffs and other import charges, quantitative restrictions, import licensing, customs restrictions	High tariffs on selected goods, quantitative restrictions, unilateralism (section 301, "super 301," and "special 301")
Administrative and other barriers	Standards, testing, labeling, and certification requirements	Onerous invoice requirements, user fees, merchandise processing fees, harbor maintenance tax, non-adherence to international product standards
Government procurement	"Buy national" policies and closed bidding	Buy America Act of 1933, state-level "buy local" legislation
Intellectual property	Inadequate patent, copyright, and trademark regimes	
Other	Bribery and corruption, tolerance of anti-competitive practices	Arbitrary anti-dumping legislation

[a] From 1998 National Trade Estimate Report on Foreign Trade Barriers, USTR.

[b] From Market Access Sectoral and Trade Barriers Database, European Commission Directorate-General I.

of the world (ROW).[1] Common to United States and ROW protectionism is the application of the traditional import policies of import tariffs and quantitative restrictions. The most important difference between U.S. and ROW policies is the prominence in the United States of unilateral actions via the so-called "section 301 family" of legislation. These unilateral actions, which are outside of multilateral arrangements in place, threaten and punish trading partners that the United States deems to be trading "unfairly." Where ROW protection stands out is in the tendency for developing and newly industrialized countries to have high levels of tariffs, red tape, and corruption, as well as little or no protection of intellectual property rights.

Whereas the theoretical calculations of the effects and costs of trade protection are well-established, the empirical estimates of the costs have been surprisingly small, especially considering the effort that economists spend decrying trade protection. For example, studies surveyed by Feenstra (1992) found the yearly cost of U.S. protection for years around 1985 to be $15.2 to 29.6 billion, or only 0.38 to 0.73 percent of gross domestic product (GDP). This is similar to a more recent study by Hufbauer and Elliot (1994), and to earlier studies surveyed by Baldwin (1984).

De Melo and Tarr (1992) argue that one reason for these small estimated costs is the use of a partial equilibrium method. In partial equilibrium models, the cost of protection in each of a large number of sectors is estimated separately, without regard to the cross-sector effects. The results for each protected sector are then simply summed to obtain the aggregate effect of protection. De Melo and Tarr propose an alternative general equilibrium model that takes explicit account of the consequences that the imposition of import protection in one sector has on other sectors in the economy. They find that the welfare cost of protection from quantitative restrictions alone was $25 to 29 billion in 1984.

A practical difficulty shared by previous approaches is their extremely high informational requirements. To estimate economy-wide costs of protection, one must know a great deal about each of the many sectors of the economy, and about the myriad of import policies and the avenues by which they can affect welfare. This is compounded by the use of unilateral antidumping actions, which can affect markets even when no actual duty or restriction is imposed.[2]

I outline an alternative estimation method that has a much lower informational requirement, while also having the advantages of general equilibrium approaches in estimating the effects of protection on the volume of trade. Specifically, I outline a gravity model of international trade that requires one to know for a cross-section of countries only their levels of bilateral trade, their GDPs, and a measure of the average level of trade protection.

Section II outlines the use of gravity models in international trade, and suggests a version of the model that allows trading relationships to differ across trading pairs. Section III estimates the gravity model using U.S. import and export data for 1994-96, and calculates the effect of U.S. and ROW protection of U.S. trade volumes. Section IV translates these estimates into rough calculations of the welfare costs of U.S. protection, and Section V concludes.

USING THE GRAVITY MODEL

The gravity model was first applied to international trade by Tinbergen (1962) and Pöynöhen (1963), but it has a long history in the social sciences. Since the latter half of the nineteenth century, it has been used to explain social flows, primarily migration, in terms of the "gravitational forces of human interaction." Its name is derived from its passing similarity to Newtonian physics, in that large economic entities such as countries or cities are said to exert pulling power on people or their products. The simplest form of the gravity model for international trade posits that the volume of exports between any two trading partners is an increasing function of their national incomes, and a decreasing function of the

[1] Detailed country-by-country descriptions of trade barriers are available at the web site of the United States Trade Representative <www.ustr.gov>. The State Department and World Trade Organization web sites <www.state.gov> and <www.wto.org> also have country reports of trade practices. For detailed descriptions of U.S. restrictions against the ROW, see the European Commission Directorate-General I's *Market Access Sectoral and Trade Barriers Database* <europa.eu.int/comm/dg01/dg1.htm>

[2] See Staiger and Wolak (1994) for a discussion and empirical evidence.

distance between them. Specifically, using Y_i and Y_j to denote national incomes, and D_{ij} to denote distance, the flow of goods from country *i* to country *j* is expressed in log-linear form as

$$(1) \qquad \ln X_{ij} = \alpha + \beta \ln Y_i + \gamma \ln Y_j - \delta \ln D_{ij},$$

where $\alpha, \beta, \gamma,$ and δ are positive constants. This is then estimated using a cross-section of trading countries taken across a single year or pooled over several years, typically measuring D_{ij} by the distance between the capital cities. It also is common to use dummy variables to capture contiguity effects, cultural and historical similarities, and regional integration.[3]

Although widely used because of its perceived empirical success (usually taken to mean a high R^2), the gravity model had lacked rigorous theoretical underpinnings, and was long criticized for being *ad hoc*. Recently though, Deardorff (1998) has shown that the gravity equation is consistent with several variants of the Ricardian and Heckschser-Ohlin models. This is in addition to earlier work by Anderson (1979) and Bergstrand (1985) who derived gravity equations from trade models with product differentiation and increasing returns to scale.

Although theoretical foundations have been established, the empirical application of the gravity model is still rather basic. As demonstrated by Cheng and Wall (1999), despite providing a high R^2, the standard estimation method tends to underestimate trade between high-volume traders, and overestimate it between low-volume traders. They attribute this to heterogeneity bias, which they address by relaxing the restriction that the intercept of the gravity equation must be the same for all trading partners.[4] Their fixed-effects method, which I will use in this study, assumes instead that there are fixed factors that can make the intercept of the gravity equation different for each trading pair.

The first of the two main benefits of the fixed-effects method is that it controls for omitted variables that are unobservable or difficult to measure. In terms of standard trade models, such variables might reflect the relative preference that an importing country's consumers have for an exporter's goods. For example, if U.S. consumers have a stronger preference for British-made goods over French-made goods, then all else equal, the United States will import more from Britain than from France. In addition to such considerations, other fixed factors such as historical links, cultural similarities, etc., that are difficult to quantify are captured by each trading-pair intercept.

The second advantage of the fixed-effects method is that fixed economic-distance variables are subsumed into the trading-pair intercept, instead of being proxied for by the geographic distance between the capital cities of the trading partners. This is particularly important for studies that include the United States, which has several economic centers on and between two distant coasts. For example, given that the West Coast of the United States is thousands of kilometers closer to Japan than is the East Coast, it is difficult to justify using the distance between Tokyo and Washington, D.C., to represent the trading distance between the two countries. Even correcting for this mismeasurement, it may not be a good measure of economic distance, because geographic distance ignores transport difficulties. For example, the geographic distance between New York and Moscow (7533 km) is shorter than that between Tokyo and Los Angeles (8816 km), but it is difficult to believe that Russia is economically nearer to the United States than is Japan.

THE EMPIRICAL RESULTS

I perform two least squares estimations of the gravity model, the first is under the restriction that the trading-pair intercepts are all equal, and the second relaxes this restriction. Both estimations retain the standard restriction that the coefficients on the other variables are the same for all countries. The restricted regression equation is

[3] See Oguledo and MacPhee (1994) for a summary of earlier models and results in the literature.

[4] Bayoumi and Eichengreen (1997) also allow for different intercepts, but with a different method.

Table 2

Regression Results for Gravity Model of U.S. Trade; Dependent Variable: Log of Bilateral Merchandise Exports

	Restricted Model	*Unrestricted Model*
Constant	-8.930 (-8.48)	—*
Origin GDP	0.922 (35.38)	0.446 (1.94)
Destination GDP	0.930 (29.90)	0.421 (1.83)
Distance	-0.942 (-10.03)	—
Trade policy index	-0.042 (-0.69)	-0.154 (-4.01)
R^2	0.750	0.953
Log-likelihood	-791.89	-364.86
Observations	510	510
Degrees of freedom	505	422

The *t*-statistics are in parentheses. *For space considerations the 85 trading-pair intercepts are not reported here.

$$(2)\qquad \ln X_{ijt} = \alpha + \beta \ln Y_{it} + \gamma \ln Y_{jt} + \delta \ln D_{ij} + \lambda T_{jt} + \varepsilon_{ijt}\ ,$$

where T_{jt} is the trade policy index for the importing country at time t, and ε_{ijt} is an error term. The unrestricted regression equation is

$$(3)\qquad \ln X_{ijt} = \alpha_{ij} + \beta \ln Y_{it} + \gamma \ln Y_{jt} + \lambda T_{jt} + \varepsilon_{ijt},$$

where α_{ij} is the trading-pair intercept, and the distance variable is subsumed into the intercept.

The data set is a panel of U.S. merchandise imports and exports to and from 85 countries for the years 1994-96. The 85 countries, listed in the appendix, are all the countries for which all variables are available for all three years. The trade data come from the Census Bureau's U.S. Import and Export History database, and the national income data are GDPs at market prices in U.S. dollars, taken from the World Bank's *World Tables*. Nominal GDP and trade data are converted into constant chained 1992 dollars. The distance variable is simply the great-circle distance between Washington, D.C., and the capital city of the trading partner.

Whereas it is relatively straightforward to gather data for bilateral trade, GDP, and distance, the extensive use of non-tariff and administrative barriers makes it difficult to quantify average levels of protection, and the oft-cited average tariff level is clearly inadequate. Instead, I use the trade policy index that is part of the Heritage Foundation's *Index of Economic Freedom*, the most recent of which is found in Johnson, Holmes, and Kirkpatrick (1998). When determining the score for a country, the authors considered average tariff levels along with descriptions from other sources of non-tariff policies, which are otherwise difficult or impossible to quantify.[5] The index rates countries on a scale of one to five, where numerical scores correspond respectively to levels of import protection: very low, low, moderate, high, and very high.[6] Because the information used to determine the index is collected during the year prior to its publication, the index is lagged in the estimation. Also, as the North American Free Trade Agreement (NAFTA) was in force during the sample period, the index takes the value of one for U.S. trade with Mexico and Canada.

[5] The 1996 scores for the 86 countries included in this study appear in the appendix.

[6] I would like to thank Bryan Johnson of the Heritage Foundation for providing me with the data tables.

Table 3

Effects of Protectionism on U.S. Merchandise Imports and Exports; 1996 (1992 dollars)

	Actual U.S. Trade ($ millions)	Effect of Protection ($ millions)	Effect as Percent of Trade	Effect as Percent of U.S. GDP
U.S. Imports (Non-NAFTA)				
From countries in sample	435,336	-67,190	-15.4	
From all countries	723,150	-111,611	-15.4	-1.66
U.S. Exports (Non-NAFTA)				
To countries in sample	295,761	-77,345	-26.2	
To all countries	498,754	-130,430	-26.2	-1.94

The least squares regression results summarized by Table 2 indicate that both versions of the model perform relatively well, and that the estimated coefficients have the expected signs. Note, however, that there are significant differences between the restricted and unrestricted versions. When the restriction on the intercept term is removed, the coefficients on the GDPs become much smaller, and that on trade policy becomes larger and statistically significant. A likelihood ratio test rejects the null hypothesis that the two models are statistically the same. The conclusion, therefore, is that the restriction on the intercepts cannot be supported statistically.

Focusing on the unrestricted model then, all else constant, a 10 percent increase in a country's national income tends to be associated with a 4.2 to 4.5 percent increase in the volume of merchandise trade between the country and its trading partners. Also, for each one-point increase in a country's trade policy index, merchandise imports tend to fall by 15.4 percent.

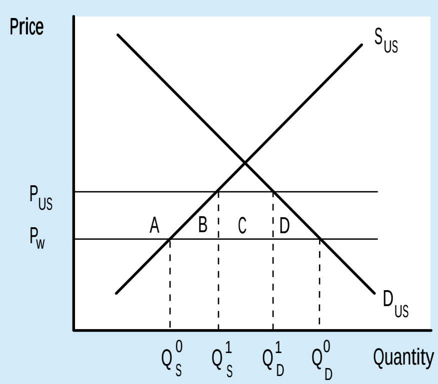
To calculate the total effect of U.S. protection on U.S. merchandise imports, and ROW protection on U.S. merchandise exports, I apply the results to the actual levels of trade and protection in the sample. I do so by taking the data for 1996, and calculating (i), the amount that the United States would have imported from each country if the United States had a trade policy index of one; and (ii), the amount that the United States would have exported to each country if every export market had a trade policy index of one. Because the United States already has free trade (or close to it) with Canada and Mexico, they are eliminated from the calculations. To extrapolate these calculations to the level of aggregate trade, I assume that the effect of protection as a percentage of trade is the same for non-NAFTA countries within and outside of the sample. Table 3 summarizes these calculations.

The United States imported over $723 billion in merchandise from non-NAFTA countries in 1996, but would have imported over $111 billion more if it had a policy of free trade. In percentage terms, U.S. protectionism decreased its merchandise imports from non-NAFTA countries by 15.4 percent, which amounted to about 1.66 percent of U.S. GDP. In the same year, the United States exported nearly half a trillion dollars of merchandise, but would have exported $130 billion more if the rest of the world had free trade. This was a 26.2 percent loss of U.S. merchandise exports to non-NAFTA countries, which was 1.94 percent of U.S. GDP. Including trade with Mexico and Canada, U.S. protection decreased its imports by 10.4 percent, whereas ROW protection decreased U.S. exports by 17.0 percent. This estimated effect of protection is much higher than found in previous

WELFARE COSTS OF IMPORT PROTECTION

Below is a brief description of the standard partial equilibrium dissection of the welfare effects of import protection under perfect competition. It begins with an analysis of import tariffs, which can be adapted easily to look at the effects of other forms of trade protection.

The figure below illustrates the market for a hypothetical good that the U.S. imports from the ROW, and S_{US} and D_{US} are the U.S. supply and demand curves. Under free trade, the good is imported at the world price, P_w. At this price, the United States consumes Q_D^0 and produces Q_S^0, and the difference, $Q_D^0 - Q_S^0$, is imported from the ROW. Assume that the United States levies a tariff of t per unit imports, and that the tariff does not affect the world price. After the tariff is levied, the price in the United States becomes $P_{US} = P_{w+t}$, causing consumption to fall to Q_D^1, and production to rise to Q_S^1. The tariff therefore decreases the level of imports to $Q_D^1 - Q_S^1$. Clearly, consumers are worse off because they pay a higher price and consume less of the good, whereas producers are better off because they produce more of the good at a higher price.



Imposition of the tariff means that consumer surplus is reduced by the area A+B+C+D. Part of this, area A, is transferred to firms as a gain in producer surplus, and another part, area C, goes to the government as tariff revenue. Because areas A and C are simply transfers within the United States, they do not represent a change in national welfare. However, parts of the consumer loss, areas B and D, are not transferred to anyone, and are therefore deadweight losses measuring the net decrease in national welfare due to the tariff. Area B is a deadweight production loss due to overproduction of the good, and area D is a deadweight consumption loss due to underconsumption of the good.

The figure also can be used to describe the welfare effects of a quantitative restriction (QR) such as an import quota or voluntary restraint agreement. Assume that the United States imposes a QR that limits imports to the same level as would result under the tariff described above. The price in the United States would rise to P_{US}, where the quantity supplied by U.S. and ROW producers would equal the quantity demanded by U.S. consumers. As with an import tariff, this reduces consumer surplus by A+B+C+D, with a gain in producer surplus of A, and deadweight losses of B and D. However, unlike the case of an import tariff, area C does not necessarily represent revenue collected by the government, as it measures the *quota rents* created by the difference between the U.S. price and the world price. If there is no government revenue-raising mechanism associated with the QR, then all quota rents are captured by ROW producers, and area C represents a net welfare loss to the economy. However, revenue might be raised through the sale of quota licenses, or by imposing an import tariff alongside the QR. Using θ to denote the government's share of the quota rents, the total net welfare loss from a quantitative restriction is therefore $B + D + (1-\theta)C$.

Although the above analysis focuses on import tariffs and QRs, it is readily adaptable to other commonly employed forms of trade protection. For example, administrative fees imposed on ROW producers have the same effects as tariffs. Also, the threat of section 301 actions can have the same welfare effects as a QR in which all quota rents are transferred overseas. This is because the mere threat of unilateral action can lead importers to raise their prices in the United States to avoid triggering anti-dumping cases.

studies such as Hufbauer and Elliot (1994) who use the standard method of adding up the partial equilibrium effects across protected industries. Their results for 1990 suggest that complete liberalization of U.S. trade would have lead to a 6 percent increase in imports.[7]

THE WELFARE COSTS OF U.S. PROTECTION

A disadvantage of the gravity model is that it is unsuitable for direct estimates of welfare costs. Studies such as Hufbauer and Elliot (1994) use industry-level data to estimate supply and demand functions, and therefore are readily useful for welfare calculations. However, as the gravity model is only a prediction of aggregate trade flows, without any information about the underlying supply and demand conditions, such welfare calculations are elusive.

To substitute for this, crude calculations can be obtained using Hufbauer and Elliot's results. According to their results, on average, a $1 decrease in imports due to import protection translates into a $2 decrease in consumer surplus. Also, of each $1 that consumers lose, $0.49 is transferred to producers, and $0.11 is deadweight loss. Applying these numbers to the estimates above, import protection in 1996 cost U.S. consumers $223.4 billion, or 3.3 percent of GDP. Of this, $109.1 billion was transferred to producers, and $24.5 billion was deadweight loss. The remainder is comprised of tariff revenue and quota rents. Subtracting the actual revenue collected in customs duties from this, the quota rents not captured by the U.S. government amounted to $72.8 billion.[8] If all of these quota rents were transferred to ROW producers, the net welfare cost of U.S. protection in 1996 was $97.3 billion, or 1.45 percent of GDP.

This estimate does not account for terms of trade effects, which occur because the size of the United States in the world market means that it can shift part of the burden of tariffs onto ROW producers. Using the Hufbauer and Elliot estimate of an average decrease of 9 percent in the world prices of protected goods, the terms of trade gain to the United States from its tariffs was only $1.5 billion, making the welfare cost of U.S. protection 1.43 percent of GDP in 1996. Note that the terms of trade effect is potentially much higher than this. However, because non-tariff protection is so prevalent, most of the quota rents created by U.S. protection are shifted to overseas producers, instead of to the U.S. government as tariff revenue.

Unfortunately, because there is no study of U.S. export markets analogous to Hufbauer and Elliot's study, such straightforward welfare estimates of the cost of ROW protection on the United States are not possible. Note though that because the effects of ROW protection on U.S. exports is similar in order of magnitude to the effects of U.S. protection on U.S. imports, it is tempting to conclude that the welfare costs are also of the same order of magnitude. However, recall that because of the prevalence of non-tariff barriers, much of the cost of U.S. protection is the transfer of quota rents to ROW producers. If ROW

[7] I obtain this figure by assuming that the total percentage effect on output in the 21 industries they examine, which represented about 10 percent of imports, was the same for all other industries in which protection was imposed. Also note that about one-third of imports in 1990 faced no import restriction.

[8] The *Statistical Abstract of the United States 1998* reports that real 1996 tariff revenue was $16.9 billion.

protection results in similar transfer to U.S. producers, the welfare effects of ROW protection on the United States would be mitigated significantly.

CONCLUSIONS

The objective of this paper was to provide new estimates of the effects of protectionism on the volume of U.S. trade, and to obtain rough estimates of the resulting welfare effects. In doing so, I outlined a new approach that uses a gravity model, which is capable of accounting for the general equilibrium effects while having a relatively low informational requirement. I also used a specific form of the gravity model, which allowed for trading-pair heterogeneity and which was statistically superior to the standard model. The method also included the use of a partly subjective trade policy index that accounts for forms of protection that are difficult to quantify, such as administrative barriers, unilateralism, procurement restrictions, corruption, etc.

Using this approach, I estimated that protectionism in the rest of the world meant that U.S. exports were 26.2 percent lower in 1996 than they would have been otherwise. I also estimated that U.S. protectionism decreased U.S. imports from non-NAFTA countries by 15.4 percent per year, which had a net welfare cost amounting to 1.45 percent of GDP in 1996. The primary source of this welfare loss was the transfer of quota rents overseas, rather than deadweight efficiency losses. Because the method I used takes into account general equilibrium effects and non-tariff and non-quota trade barriers, these estimates are much higher than those found in previous studies.

REFERENCES

Anderson, James E. "A Theoretical Foundation for the Gravity Equation," *American Economic Review* (March, 1979), pp. 106-16.

Baldwin, Robert E. "Trade Policies in Developed Countries," *Handbook of International Economics, Volume 1*, Ronald W. Jones and Peter B. Kenen, eds., North-Holland (1984), pp. 571-619.

Bayoumi, Tamim, and Barry Eichengreen. "Is Regionalism Simply a Diversion? Evidence from the Evolution of the EC and EFTA," *Regionalism versus Multilateral Trade Arrangements*, Takatoshi Ito and Anne O. Krueger, eds., University of Chicago Press, (1997).

Bergstrand, Jeffrey H. "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence," *Review of Economics and Statistics,* Vol. 67*,* (August 1985), pp. 474-81.

Cheng, I.-H., and H. J. Wall. "Controlling for Heterogeneity in Gravity Models of Trade," Federal Reserve Bank of St. Louis *Working Paper*, (1999).

Deardorff, Alan V. "Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World?" *The Regionalization of the World Economy*, Jeffrey A. Frankel, ed., University of Chicago Press (1998).

de Melo, Jamie, and David G. Tarr. *A General Equilibrium Analysis of U.S. Foreign Trade Policy*, Cambridge University Press (1992).

Feenstra, Robert C. "How Costly is Protectionism?" *Journal of Economic Perspectives*, Vol. 6 (1992), pp. 159-78.

Hufbauer, Gary C., and Karen A. Elliot. *Measuring the Costs of Protection in the United States*, Institute for International Economics (1994).

Johnson, Bryan. T., Kim R. Holmes, and Melanie Kirkpatrick. *1999 Index of Economic Freedom*, Heritage Foundation and Dow Jones and Company (1998).

Linnemann, Hans. *An Econometric Study of International Trade Flows*, North-Holland (1966).

Oguledo, Victor I., and Craig R. MacPhee. "Gravity Model: A Reformulation and an Application to Discriminatory Trade Arrangements," *Applied Economics*, Vol. 40 (1994), pp. 315-37.

Pöyhönen, Pentti. "A Tentative Model for the Volume of Trade Between Countries," *Weltwirtschaftliches Archive*, Vol. 90 (1963), pp. 93-100.

Staiger, Robert W., and Frank A. Wolak. "Measuring Industry-Specific Protection: Anti-Dumping in the United States," *Brookings Papers on Economic Activity: Microeconomics* (1994), pp. 51-118.

Tinbergen, Jan. *Shaping the World Economy–Suggestions for an International Economic Policy*, The Twentieth Century Fund (1962).

Alvin L. Marty is professor of economics at the Center for the Study of Business and Government at Baruch College, City University of New York (C.U.N.Y.). Rob Dittmar, an economist at the Federal Reserve Bank of St. Louis, provided mathematical expertise that was very helpful in clarifying the Lucas paper. Turin Bali, a Ph.D. candidate at C.U.N.Y., ran Box-Cox tests on the constant and semi-elastic demand schedules. Dan Steiner provided research assistance.

The Welfare Cost of Inflation: A Critique of Bailey and Lucas

Alvin L. Marty

THE CONCEPT OF THE OPTIMAL MONEY STOCK

The concept of the optimal stock of money introduced by Milton Friedman (1969) has stimulated much discussion and controversy. More recently, Lucas (1994) has provided new estimates of the welfare gain for the American economy from setting the optimal level of real cash balances. Before turning to a critique of Lucas, let us provide a rationale for this concept of optimality.

To society, real balances are produced at zero marginal costs. To an individual, however, there is a cost to holding cash that bears no interest rather than bonds, which yield a money rate of interest. An individual would willingly incur this cost only if real balances produced services on the margin equal to the foregone interest. A lower interest rate reduces this cost and increases the total services provided by real cash balances. One estimate of the consumption the consumer would forego—to get this potential maximum gain from holding cash—is the area under the money demand schedule between the quantity of cash held at the lowest possible rate of interest (perhaps zero) and the smaller quantity that would be held at a higher interest rate.[1]

Clearly any measure of this gain depends on the particular demand schedule used. Lucas utilizes a double log schedule rather than the semi-log schedule used by Friedman (1969) and Martin Bailey (1956). The Lucas' double log schedule yields greater welfare gains since the level of cash balances always increases as the interest rate approaches zero. In contrast, the semi-log schedule implies a finite level of cash is held at a zero nominal interest. In a later section we will provide a fuller discussion of these schedules.

LUCAS' EXPOSITION

What reduction in consumption would compensate for and measure the gain in utility from a larger stock of cash? An exact measure of this gain is derived by Lucas in the case of both the double log (the constant elasticity schedule) and semi-log demand schedules for real cash balances. He assumes a time-separable constant-relative-risk-averse utility function with consumption and the ratio of real balances to consumption as arguments. This function is compatible with steady growth of a non-durable output. A representative agent maximizes it subject to a budget constraint.

Such an infinite horizon model, (see Ramsey [1927] and Sidrauski [1967]), implies that we need not be concerned with changes in the real rate of interest as we vary the growth of nominal money balances. Note the contrast to Mundell (1963) and Tobin (1965) models in which the accumulation of non-interest bearing real balances competes with physical capital as vehicles for the savings of finite-lived individuals. In their models, a faster growth of nominal money raises the actual and anticipated rates of price change. The capital loss on real balances spurs the accumulation of physical capital. A new steady state is reached: Capital and per capita output are higher and the real rate of interest is lower. In the Ramsey-type model used by Lucas, the real rate is tied to the fixed utility discount factor of the representative agent. When the rate of

[1] Strictly speaking, this welfare measure utilizes a compensated demand schedule, which keeps the consumer at the same level of demand.

price change is fully anticipated, the money rate of interest adjusts to give borrowers and lenders the same fixed real rate. With a constant real rate, the demand schedule for real balances represents alternative steady states with the nominal interest rate adjusting one-for-one with the rate of price change.

Lucas approximates exact measures of the gain in utility by the consumer surplus under both the constant and semi-log schedules. For the range of interest rates in U.S. history, the approximations are very close to the exact compensating variation using these demand functions.[2]

THE NATURE OF THE DEMAND SCHEDULE: BAILEY VS. LUCAS

The constant elasticity schedule is $M/P = Ar^{-\eta} y^{\beta}$, where M is nominal money, P is the price level, y is real income and r the nominal interest rate (money rate). The income elasticity, β, is taken as unity and the estimated interest elasticity, η, is 0.5.[3] Although the demand for real balances increases without limit as the money rate approaches zero, the integral converges.

If other taxes distort (such as a tax on labor income), it is uncertain whether welfare improves if the authorities drive the money rate of interest towards zero. Other taxes would have to be raised to run a budget surplus and retire money at a rate equal to the real interest rate minus the output growth rate. Any welfare gain from higher real balances would have to be weighed against the welfare cost of increased taxes. Even in this case, however, Lucas estimated optimal money rate remains very close to zero.

As Lucas notes, Bailey (1956) approximated the exact welfare loss by the consumer surplus integral under the semi-log demand schedule, $M/P = e^{\gamma} e^{-aE}$, where e^{γ} is an index of real balances held at zero anticipated inflation.[4] Lucas prefers a constant elasticity demand function on the grounds that it fits the American data.

Before proceeding with a critique of Lucas, it is useful to discuss the properties of the constant elasticity function. Unlike the semi-log, the constant elasticity function does not generate a Laffer curve. Along a Laffer curve, as the inflation tax rate (the money rate) rises, the revenue increases. It reaches a maximum, then it declines. Using the Phelps (1973) and Auernheimer (1974) definition of the revenue as $(M/P)(r)$, the revenue $(R = rAr^{-\eta} = Ar^{(1-\eta)})$ continuously increases with inflation.[5]

Although the double log function may fit the American data that includes only moderate rates of inflation, it is not evident that the schedule should be extended—as Lucas does—to regions of hyperinflation, or for that matter, a deflation of prices approaching the real rate. Bali (1998) has run tests using the Box-Cox transformation to determine whether the constant or semi-elastic function best fits the data. Over the range of data in the United States (therefore not at hyperinflation or at rates of interest approaching zero), the double log performs better. For non-U.S. hyper-inflations, the semi-log fits better, however. We should be very cautious about extrapolating to non-observable ranges of the data. Since data are not available at rates of interest close to zero, we proceed to discuss Lucas' (1994) extension to hyperinflation where international data are available.

EXTENSION TO HYPERINFLATION

Cagan (1956) found that the semi-log fits the data for seven European countries during periods of hyperinflation. The semi-log generates a Laffer curve: The steady state maximum revenue is at

$$\frac{dM}{dt}\frac{1}{M}=\frac{1}{a}.$$

Cagan concluded that governments often inflated beyond this point, an overshooting paradox which he explained as follows. The revenue is the product of real balances (which depend on expected inflation) and the tax rate (the actual inflation). The authorities, therefore, could exploit a lag in expected inflation to temporarily get more revenue when, as Cagan assumed,

[2] For the details of these calculations, see Lucas (1994).

[3] It is crucial that the income elasticity is unity. If this were not the case, we could not use steady state analysis. If the income elasticity were less then one, real balances would fall as a ratio to real income and any welfare gain would become increasingly small.

[4] The semi-log schedule was used by Cagan in his classic paper (1956). Since the real rate was very small, as compared to the rate of inflation under hyperinflation, he used the expected rate of inflation as the opportunity cost of holding real balances.

[5] This result simply states that a monopolist facing a demand curve with a constant elasticity less than unity can always increase total revenue by raising relative price.

expectations are adaptive. If inflation is held constant, expectations catch up and the revenue falls below its steady state level. If the authorities attempt to exploit this lag repeatedly, a constant coefficient on adaptive expectations would imply serially-correlated errors, or the coefficient itself would be revised. Cagan's explanation rests on the joint hypothesis of adaptive expectations and the semi-log schedule.

Lucas (1994) sees no problem in extending the double-log to hyperinflation. Indeed, it clears up Cagan's overshooting paradox. With the double log, additional revenue always accrues at higher inflation. Moreover, no lag exists since Lucas assumes expectations are rational.

A constant growth of money (above Cagan's steady state maximum) would provide a controlled experiment allowing us to distinguish between these two composite hypotheses. If Cagan's joint hypothesis is correct, the revenue should first rise and then fall permanently. If Lucas' joint hypothesis is correct, the revenue should be unchanged. The real world does not provide us with neat experiments, however. Nor would the experiment allow us to distinguish between aspects of the composite hypothesis: With either demand function, no lag occurs if expectations are rational.

The German hyperinflation lasted for only a few years. The pattern of revenue was oscillating: It rose and fell in cycles depending on the fiscal needs of the authorities. In any case, the Box-Cox transformations come down squarely on the side of the semi-log. Is the Cagan overshooting paradox explainable by a regime change so that adaptive expectations are—for a time—rational? Or, even under rational expectations and the semi-log money demand function, do we have a case of time inconsistency? Here is a fascinating area of research for which data are available.

OPTIMAL TAX ON REAL BALANCES

We next turn to how the use of the double log rather than the semi-log affects the analysis of the optimal tax on cash (which well may be zero). Bailey used the average welfare-cost-per-dollar of revenue to measure the social costs of seignorage. Later, it was suggested that the marginal welfare-costs-per-dollar be set equal to the average of the distortions due to other taxes. General equilibrium theorists such as Ballard, Shoven and Walley (1985) have calculated the average value of these marginal welfare-costs-per-dollar of revenue for distortionary taxes. Including a tax on cash balances in the menu of taxes, the procedure for calculating the optimal tax rate is to set the marginal revenue from money creation equal to the average of the marginal distortions due to other taxes. This calculation accepts, in the spirit of the second best, the preexisting distortions caused by other taxes and assumes that any additional revenue occurring from the tax on cash is used for exhaustive expenditure rather than rebated to the private sector in the form of lower taxes. This procedure is one of many possible ways of handling second best optimal tax policy.

Given a constant real rate, the rate of price change sets the money rate of interest. For any demand function,

$$\frac{M}{P} = \phi(r),$$

the revenue is defined as the money rate (r) times the quantity of money demanded at that rate, $R = r\,\phi(r)$. The ratio of marginal welfare cost to the marginal increment to revenue is a function of the interest elasticity as given by

$$\frac{dW}{dr} \Big/ \frac{dR}{dr} = \frac{\eta}{1-\eta},$$

where W is the welfare loss. This result is a variant of Ramsey's inverse elasticity rule: Tax most heavily commodities in inelastic demand. It is a partial equilibrium result that implies the shifting of resources is from the taxed to the untaxed sector so that all cross elasticities are zero. In the special case that the optimal tax structure is proportional, real balances should be not taxed at all.[6] For the constant elasticity function,

[6] These optimal tax rules appear to have been recently rediscovered. For an earlier account see Marty (1976a, 1976b).

$$\frac{M}{P} = Ar^{-\eta} \,,$$

the average and marginal ratios are equal. The proof is simple: the welfare loss $W = \eta Ar^{(1-\eta)}/1-\eta$. Since the revenue is

$$r\frac{M}{P}, W/R \,,$$

(the average ratio) $= \eta/1-\eta$ (the marginal ratio).

With a constant elasticity, if dW/dR is greater than for other taxes, real balances shouldn't be taxed at all. But what if it is less? There appears to be no equilibrium solution. The double log highlights a problem with this approach, a failing that is independent of the demand function used.

To see this, take the semi-log function, set values for the real rate and the semi-elasticity, and then solve for the optimal money rate. This determines the optimal rate of price change that equates dW/dR for real balances to the value for other taxes.

The difficulty is that the authorities don't directly control the money rate of interest, they control the monetary base. Holding other taxes constant and varying the expenditure side makes no sense. In the double log case, if dW/dR for cash were lower than the average for other taxes, expenditures would increase without limit! The proper technique is to hold expenditures constant and optimally substitute one tax for another. No shortcut avoids estimating the welfare effects of altering other taxes.

THE COMPOSITION OF THE MONEY SUPPLY

We turn to the question of what is the proper definition of the money supply. Lucas defines the money supply as M1 and runs his welfare integrals from zero to a positive money rate—a procedure that adds high-powered money and interest bearing deposits and treats this composite as non-interest bearing. In this regard, it is relevant to return to Bailey's classic article and review his treatment of currency and deposits.

Bailey used Cagan's data to estimate the welfare cost of the inflation. He set the real rate at zero (its level is too small to play a role in hyperinflation) so that the cost of holding currency was the anticipated inflation (a proxy for the money rate). He assumed competitive banks pay interest on deposits but are subject to a sterile legal reserve requirement. A zero profit condition was imposed: The revenue from interest-bearing assets was completely dispersed in interest payments on deposits. Under these assumptions, competition would force banks to pay on deposits (D) the money rate, r, (i.e., the rate of inflation) on loans, L, times $1-z$, the reserve ratio. Zero profits are earned since $D(1-z)r = rL$. The cost of holding deposits is rz: the difference between r (the money rate on bonds) and the interest on deposits, $r(1-z)$. Since deposits are partially indexed against the inflation, Bailey assumed that at very high rates of inflation, the public uses only deposits and all currency would be held as bank reserves. Bailey's welfare integrals run from zero to rz reflecting the partial indexing of the deposit rate to inflation. Bailey then contrasts a currency-only economy with a bank-only economy. To get the same revenue and incur the same welfare cost as in a currency economy, the bank economy can inflate at a rate equal to that in the currency world multiplied by the reciprocal of the reserve ratio. The inflation rate is Π/z, but the cost of holding deposits is Πz so that the public holds the same real balances as in a currency-only world. Although the tax rate is Π/z, the tax base is zD so the tax revenue is the same as in a currency-only world. In contrast to Lucas, for Bailey there is a substantial reduction in welfare costs since the interest on deposits is partially indexed against inflation.

How satisfactory is Bailey's treatment of the different roles of currency and deposits? In the first place, there is no model of the general case where individuals hold both currency and deposits. Their distinct roles are not analyzed. Note, for example, the comparison of the currency and the banking economy implicitly assumes that the demand curve for currency is the same as that for deposits. Moreover, the zero-profit condition ignores any marginal costs of intermediation for which the banks would have to be compensated. Nor is this all. Bailey makes the implicit assumption—which may be correct—that individuals move from currency into deposits as the percentage point difference $\pi(1-z)$ between the yield on currency and deposits rises. At high inflation, this movement is complete and all currency is held as bank reserves.

We, like Lucas, make no claim to fully analyze the role of currency and deposits. The real issue is the polar choice between two very incomplete models, however. On the one hand, we can lump currency and deposits together and assume that both pay no interest. We can then run our integrals from zero to the money rate of interest. This is Lucas' choice. On the other hand, we could assume with Bailey that all high-powered money is held as reserves and run the welfare integrals from a money rate of zero to rz, the opportunity cost of holding interest bearing deposits. Lucas' choice overstates the welfare loss when we deviate from the Friedman rule. The assumption in Bailey, that all deposits pay interest, underestimates the welfare loss. Without a theory of banking in which the distinctive roles of currency and deposits are modeled, it remains uncertain how large is the degree of overestimation or underestimation.[7]

As an alternative to these two polar positions, one might tentatively try modeling the distinct roles of currency and deposits. Making the bold assumption that currency and deposits face the same demand function (including the elasticity), it would be possible to run separate integrals: For currency, the integration would run from a zero money rate to a positive rate: For deposits, it would run from zero to rz. The general expression for the welfare loss (W) is

$$W=\int_0^r \phi(x)dx - r\phi(r) + \int_0^{rZ}\phi(x)dx - rz\phi(rz).$$

With the constant elasticity schedule $M/P = Ar^{-\eta}$, the welfare loss is

$$W=\left[\frac{A\eta}{1-\eta}r^{1-\eta}\right]+\left[\frac{A\eta}{1-\eta}(rz)^{1-\eta}\right].$$

At $z = 0$, only currency would incur a welfare loss; as z approached 1, the welfare loss on deposits would increasingly approximate the loss on currency. All these models make the unsatisfactory assumption that the same demand schedule applies to currency and deposits. Moreover, the intermediate model assumes that the currency-deposit weights remain unchanged, which is unlikely during high rates of inflation.

Despite these defects, our preference—in an era of increasing financial deregulation of interest rates—is to go the intermediate route.

We came across Ed Prescott's comments on a paper by King and Wolman (1996) towards the completion of this paper. Prescott criticizes the authors' analysis:

> The theory has households holding non-interest bearing money, while the monetary aggregate used in the demand for money function is M1. Most of M1 is not non-interest-bearing debt held by households. Only a third of M1 is currency and half of that is probably held abroad. Another third is demand deposits held by businesses, which often earn interest *de facto*. Households do not use these demand deposits to economize on shopping time. The final third is demand deposits held by households that, at least in recent years, can pay interest.

This criticism is also applicable to Lucas' treatment.

CONCLUSION

Our discussion has shown skepticism about Lucas' extension of the double log schedule to both very low as well as very high money rates of interest (periods of hyperinflation). Since data are not available for the American economy at very low money rates, we have discussed the implications of using the double log rather than the

[7] Lucas is quite upfront about the need for an explicit theory of banking to distinguish between the separate roles of currency and deposits, but he leaves matters there.

more commonly used semi-log in hyperinflation. This is a fascinating area for research since data are available in hyperinflation periods.

Turning to the gains from going to very low rates of interest in the American economy, it would appear that Lucas overestimates these gains. If M1 is used as a relevant money supply, some correction should be made for the interest paid on portions of M1. (We contrasted Bailey's treatment with Lucas'.) Moreover, more than half of U.S. currency is held abroad by foreigners. Note that this affects any estimate of the welfare gains when either the monetary base or M1 is used. What remains is a cautionary tale. We should treat any estimates of the welfare costs with caution when the distinct roles of currency and deposits are not modeled.

REFERENCES

Auernheimer, Leonardo. "The Honest Government's Guide to the Revenue from the Creation of Money," *Journal of Political Economy* (May/June 1974), pp. 598-606.

Bailey, Martin J. "The Welfare Cost of Inflationary Finance," *Journal of Political Economy* (April 1956), pp. 93-110.

Bali, Turan. "Essays in Macroeconomics and Finance," unpublished Ph.D. dissertation, City University of New York, 1998.

Ballard, Charles L., John B. Shoven, and John Whalley. "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States," *The American Economic Review* (March 1985), pp. 128-38.

Cagan, Phillip. "The Monetary Dynamics of Hyperinflation," *Studies in the Quantity Theory of Money*, M. Friedman, ed., University of Chicago Press, 1956, pp. 25-117.

Friedman, Milton. *The Optimum Quantity of Money and Other Essays,* University of Chicago Press (1969).

King, Robert G., Alexander L. Wolman. "Inflation Targeting in a St. Louis Model of the 21st Century," this *Review* (May/June 1996), pp. 83-107.

Lucas, Robert E., Jr. "On the Welfare Cost of Inflation," manuscript, Department of Economics, University of Chicago, 1994.

Marty, Alvin L. "A Note on the Welfare Cost of Money Creation," *Journal of Monetary Economics* (January 1976a), pp.121-24.

________. "Real Cash Balance and the Optimal Tax Structure," in *Essays in Economic Analysis, Proceedings of the Association of University Teachers of Economics Annual Conference, Sheffield 1975,* Michael J. Artis and A. Robert Nobay, eds., Cambridge University Press, 1976b.

Mundell, Robert A. "Inflation and Real Interest," *Journal of Political Economy* (June 1963), pp. 280-83.

Phelps, Edmund S. "Inflation in the Theory of Public Finance," *Swedish Journal of Economics* (March 1973), pp. 67-82.

Ramsey, Frank P. "A Contribution to the Theory of Taxation," *Economic Journal* (1927), pp. 47-61.

Sidrauski, Miguel. "Rational Choice and Patterns of the Growth in a Monetary Economy," *American Economic Review* (May 1967), pp. 534-44.

Tobin, James. "Money and Economic Growth," *Econometrica* (October 1965), pp. 671-84.