Vol. 37, No. 4

ECONOMIC REVIEW

2001 Quarter 4

FEDERAL RESERVE BANK
OF CLEVELAND

ECONOMIC REVIEW

2001 Quarter 4
Vol. 37, No. 4

Economic Review is published quarterly by the Research Department of the Federal Reserve Bank of Cleveland. To receive copies or to be placed on the mailing list, e-mail your request to: *4d.subscriptions@clev.frb.org* or fax it to 216/579/3050.

Economic Review is available electronically through the Cleveland Fed's site on the World Wide Web: *http://www.clev.frb.org.*

We welcome your comments, questions, and suggestions. E-mail us at *editor@clev.frb.org.*

ISSN 0013-0281

Estimates of Scale and Cost Efficiency for Federal Reserve Currency Operations

by James Bohn, Diana Hancock, and Paul Bauer

James Bohn is an associate at the Brattle Group, in Cambridge, Mass., Diana Hancock is the chief of monetary and financial studies at the Board of Governors of the Federal Reserve System, and Paul Bauer is an economic advisor at the Federal Reserve Bank of Cleveland. The authors thank Jon Cameron, Eric Severin-Lossin, and James Thomson for their valuable comments and suggestions. All errors remain those of the authors.

Introduction

It is the conventional wisdom that a government monopoly in the provision of currency is an appropriate public policy. In fact, although Congress required the Federal Reserve to establish a system of fees for its payment services (check clearing and collection, wire transfers, automated clearinghouse transfers, and securities safekeeping) to encourage competition and to promote payment system efficiency, it did *not* require any charges for cash services of a "governmental nature," such as the disbursement and receipt of new or fit coin and currency.[1] Accordingly, provisions in the Monetary Control Act of 1980 require a schedule of fees only for cash services such as coin wrapping and transportation. Congress felt that it could safely open such services to competition from the private sector because the provision of these services is not essential to the Federal Reserve in its pursuit of its primary responsibility "to provide the nation with currency and coin of high quality" nor with the Federal Reserve's ability to "expand or contract the amount of currency and coin in response to the public's demand."[2]

Recently, some researchers have developed stylized theoretical models to defend this conventional wisdom, while others have begun to question it. Two themes, which do not necessarily depend upon one another, have emerged to support the view that a currency monopoly is economically efficient. One theme is that currency provision is a natural monopoly.[3] This natural monopoly arises in part because there are increasing returns to scale that result from the cross-sectional independence of demands for individual redemptions of currency.[4] Increasing returns to scale could also arise because fixed costs are high—sophisticated technology for soil measurement and

■ **1** During the course of congressional consideration of the Monetary Control Act of 1980, Senator Proxmire indicated that the Federal Reserve would not be required to charge for these services. (Remarks of Senator William Proxmire, *Congressional Record* (daily edition), March 27, 1980, p. S 3168.)

■ **2** See U.S. Congress (1980, S3168, March 27).

■ **3** See King (1983, pp. 132–33).

■ **4** Miller and Orr (1966) provide a cash-management model that possesses this increasing-returns-to-scale property.

counterfeit note detection is needed to process currency, and transporting currency is expensive. The second theme focuses on the impracticality of interest-bearing currency. It is argued that the computation, information, and transaction costs of collecting interest on currency can easily exceed the value of the interest that might be collected, particularly for low-denomination notes.[5] In this environment, it is argued that competitive currency providers would have an incentive to take actions that increase the "quality" of currency (for example, by making expenditures that would reduce counterfeiting or improve the fitness of notes). To the extent that such in-kind benefits are valued by consumers at less than their cost of production, seigniorage (the difference between the nominal interest rate and the cost of production) would be wastefully dissipated.[6]

The view that a competitive currency market would be more efficient than a monopoly provider is based on the argument that competitive providers would choose to offer higher-quality currency services—at a higher cost—only when the additional cost is justified by the additional value of output.[7] Implicitly, it is assumed that competitive firms are not legally restricted from freely employing whatever technology is most cost efficient.[8] That is, the monopolist and the competitive firm have *identical* costs for providing a given quantity and quality of currency services. Under such circumstances, nonprice competition improves social efficiency because a monopolist has no incentive to provide an appropriate level of currency services with the valued quality enhancements.

It is unlikely, however, that a monopolist and a competitive firm would have *identical* costs for providing a given quality and quantity of currency services. This is because monopolists tend to pursue the objective of cost minimization less effectively than do competitive firms, being more likely to produce at an inefficient level of output (scale inefficiency) or to produce a given level of output with more resources than are required (cost inefficiency). Thus, competition among providers of currency services may be beneficial because such providers may be induced to be more scale and cost efficient than a monopoly provider.

Although an empirical understanding of currency cost considerations would enlighten the foregoing discussion, there is a dearth of research on such topics. First, only two studies have analyzed the scale economies for currency provision, Zimmerman (1981) and Dotsey (1991). Neither of those studies is likely to shed light on potential scale economies today, however, because they were based on data from periods (1977 and 1988–90, respectively) when the Federal Reserve used relatively unsophisticated technologies to determine the fitness level of notes, detect counterfeits, and destroy unfit notes. New currency processing technologies may have increased potential scale economies. Second, because it is a complex task to estimate the costs of improving currency quality, such studies have been performed even more sporadically. Currency-quality cost estimates require quality measurements (such as soil-detection standards) and calculations that measure changes in currency handling costs that might arise from changes to these qualities. These calculations are complicated even more because handling costs also depend on a number of other intertwined factors. For example, if currency quality is improved (by improving the quality of existing notes or adding additional anticounterfeiting features, for example), handling costs may rise due to factors such as a higher destruction rate for unfit currency, additional printing costs for the new notes, a shorter life span for these notes, and possibly the need to process additional volume.[9] Finally, there are no studies that estimate the cost efficiency of Federal Reserve currency operations. For Federal Reserve payment services that do compete with private sector providers, cost-efficiency estimates are available (Bauer and Hancock

■ **5** See Fama (1981, 1983), White (1987), and Sumner (1993).

■ **6** See, for example, Fama (1983) and Sumner (1993, 2000).

■ **7** See White and Boudreaux (2000, p. 152).

■ **8** Lacker (1996) has discussed the case in which competitive private issuers have an industry cost curve everywhere higher than that of a monopoly issuer because of legal restrictions on the competitors' technology choice.

■ **9** Although the following example moves currency quality in the opposite direction, in 1991 the Federal Reserve System estimated that it would cost about $1.2 million per year (in 1990 dollars) if it lowered the minimum quality threshold for recirculated notes by one step on a 16-step measurement scale, but such costs would be offset by lower annual printing costs of about $7.6 million per year, which would be absorbed by the Bureau of Engraving and Printing. Over a 10-year period, the discounted net present value of savings from a one-step decrease in the minimum quality threshold for U.S. currency was estimated to be $82.9 million. See Board of Governors of the Federal Reserve System (1991, pp. 1–14).

[1993], Bauer and Ferrier [1996]). Such estimates are comparable to those for private-sector financial institutions reported elsewhere (such as Bauer, Berger, and Humphrey [1993] and Berger, Hancock, and Humphrey [1993]).

This paper explores the Federal Reserve's cost of providing currency services, using quarterly data from 1991 to 1996. Whereas previous studies (Zimmerman [1981] and Dotsey [1991]) were concerned only with whether a Federal Reserve currency processing and handling facility was producing an efficient level of output (scale efficiency), this study also considers cost inefficiency (deviations from the cost frontier), which is sometimes interpreted as X-inefficiency. Cost inefficiency occurs when actual costs are higher than the minimum possible for a given level of output, either because more of each input is employed than is required by the production process or because inputs are employed in suboptimal proportions.

We construct and compare several different models of currency operations to test the robustness of our results. A unique aspect of this study is that we have detailed information about the equipment that was used at each facility. This information included the number of processing machines of each type, an implementation schedule for new equipment installations, and usage statistics that indicated the amount of "downtime" each piece of equipment experienced at each facility. These data allow us to account in our cost-function specifications for the unavoidable transition costs associated with the introduction of a new technology. Such adjustment costs could significantly affect the scale and cost efficiency of each facility as it deployed the new technology.[10] In addition, because the cost functions we estimate depend not only on costs associated with keeping fit currency in circulation but also on the costs associated with destroying unfit currency, the parameter estimates could possibly be used to forecast changes in processing costs associated with changing the quality of currency in circulation.[11]

The next section describes the currency operations of the Federal Reserve System. Section II discusses the cost-function specifications we use and the econometric techniques we employ to estimate the cost frontier and facility-level returns to scale and cost efficiencies. Section III describes the data for output and input quantities as well as input prices. Section IV presents our estimates of returns to scale, marginal costs of different outputs, and facility-level cost efficiencies for Federal Reserve currency processing and handling facilities. Section V concludes and discusses some policy implications of our empirical findings.

I. Background

The Federal Reserve's involvement in the provision of paper currency dates back to its founding in 1913. The provisions of the Federal Reserve Act relating to currency were designed to provide for an "elastic currency" that could expand and contract as the public's demand for cash increased and decreased.[12] Prior to that time, the kinds of currency in circulation had proved incapable of meeting the needs for additional amounts that developed from time to time on a seasonal and cyclical basis as well as in periods of financial crisis.

The first Federal Reserve notes were issued in late 1914. By statute, these notes were obligations of the United States, a first lien on the assets of the issuing Federal Reserve Bank, backed 100 percent by discounted commercial notes and bills and a 40 percent gold reserve.[13] These notes were redeemable in gold or lawful money.

In 1920, the Appropriations Act authorized the Secretary of the Treasury to transfer to the Federal Reserve Banks and branches various functions that were performed in connection with the issue, exchange, and replacement of U.S. paper currency and coins and the receipt for redemption of national bank notes and Federal Reserve Bank notes.[14] This assumption of duties by the Federal Reserve Banks and branches led to an improvement in the overall quality of U.S. currency and an improvement with respect to the supply of notes of desired denominations.[15]

10 See Hancock, Humphrey, and Wilcox (1999).

11 As noted above, such cost estimates would play a role in estimating the social costs that would be incurred to improve currency quality.

12 See Booth (1989).

13 In 1934, the Gold Reserve Act stopped redemption of currency for gold. In 1968, the requirement of gold reserves against Federal Reserve notes was eliminated. Also in 1968, the redemption of currency for silver was discontinued.

14 See U.S. Treasury Department, *Circular No. 55.*

15 These improvements were noted in the 1921 *Annual Report* of the Board of Governors of the Federal Reserve System.

In 1980, Congress again expanded the role of Federal Reserve Banks with respect to the distribution of currency and coin. The Monetary Control Act of 1980 (MCA) authorized Federal Reserve Banks and branches to distribute available supplies of coin and currency to depository institutions, including not only member and nonmember banks, but also savings banks, savings and loan associations, and credit unions. This change dramatically increased the number of end points served by the Reserve Banks. In addition, the MCA required that transportation of coin and currency as well as coin-wrapping services be provided according to a schedule of fees established by the Board of Governors of the Federal Reserve System.

The growing number of institutions with access to Federal Reserve cash services, in conjunction with the limited facilities of Reserve Banks, led to the adoption of "uniform cash service standards" in 1984. These standards defined normal service to each depository institution as once per week; indicated that Reserve Banks would accept deposits of reusable currency and coin when a depository institution accumulates a surplus that cannot be reasonably stored or disposed of by direct exchange with other depository institutions; and minimized or eliminated, where practicable, "cross-shipments" (the deposit of excess fit currency and reorder of the same denomination within five business days).[16]

Today, Federal Reserve Banks and branches have significant currency and coin responsibilities. Most Federal Reserve Banks and branches maintain facilities for currency processing, handling, and distribution. These facilities are specially constructed, high-security areas that receive new notes from the Bureau of Engraving and Printing and used notes from depository institutions with excess currency holdings. Used notes are deposited in the form of *straps* (bundles of one hundred notes) and *blocks* (bundles of ten straps). Straps and blocks are manually counted in the receiving area, catalogued, and stored in a vault for a short period of time, generally between 10 and 15 days, although deposits of $1 notes often exceed 30 days.

Used notes are counted and verified as genuine on high-speed currency sorters. This equipment is fairly sophisticated and performs many tasks. First, the packaging material on each strap is removed. Second, sensors determine which notes are fit for circulation. A note may be deemed unfit because of its physical condition: It may be torn or have holes in it, be too soiled, or no longer have a sufficiently crisp texture. Unfit notes are destroyed using on-line shredders that are attached to the high-speed equipment. Fit notes are repackaged by the sorter into straps and blocks for workers to return to the vault. Third, counterfeit notes or those that cannot be read by the high-speed equipment are sent to workers who manually examine each note and pass it through a low-speed machine that, along with the high-speed machines, reconciles the account of the depositing depository institution.

Over the last 20 years, the Federal Reserve has adopted increasingly sophisticated machines that are better able to determine the fitness level of notes and detect counterfeits. The first generation of high-speed currency sorters, Currency Verification, Counting, and Sorting (CVCS) machines, were installed during the 1980s. These machines were gradually replaced by a second generation of high-speed currency sorters, Banknote Processing System (BPS) 3000 machines, which were installed during the 1990s. This second generation of machines was subsequently upgraded with new software and hardware to improve performance.

Currency, whether fit or new, enters circulation when withdrawn by a depository institution. Depository institutions cannot specifically request new currency from the Federal Reserve. Rather, orders are filled with the first available currency from the vault. Federal Reserve offices handle a considerable amount of currency each year (table 1). In 1999, for example, more than 7 billion notes were destroyed and more than 9.5 billion new notes were put into circulation.

II. Estimation Techniques

A variety of econometric techniques have been developed for estimating cost frontiers. Such techniques employ a specific flexible functional form for the cost function, and each one imposes some additional assumptions about the statistical properties of the inefficiency terms. In this paper, we use two different

16 These standards have since been modified. See the *Federal Register* notice dated April 30, 1996, for a detailed discussion of the uniform cash access policies, which became effective May 1, 1998.

TABLE 1

Currency in Circulation, New Notes Issued, and Notes Destroyed, 1999 (millions of pieces)

Dollar denomination	Notes in circulation	New notes issued[a]	Notes destroyed
1	7,536	3,865	3,767
2	602	25	4
5	1,799	866	694
10	1,620	788	616
20	5,804	2,744	1,736
50	1,294	426	183
100	3,862	879	257
Total	**22,516**	**9,593**	**7,257**

a. Does not include additions to inventory at Reserve Banks.

SOURCE: *Annual Report, Budget Review,* Board of Governors of the Federal Reserve System (2000).

functional forms for the cost function. This allows us to examine the robustness of our results to different assumptions with respect to the functional form of the cost function. In addition, we use two different frontier estimation techniques to ascertain how robust our facility-level efficiency rankings are to alternative assumptions about the statistical properties of the inefficiency terms.

The first functional form for the cost function we consider is the translog,

$$(1)\quad \ln C_{it} = \beta_0 + \sum_{j=1}^{J} \beta_j \ln y_{jit} + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk} \ln y_{jit} \ln y_{kit} + \sum_{l=1}^{L} \delta_l \ln w_{lit} + \sum_{j=1}^{J}\sum_{l=1}^{L} \delta_{jl} \ln y_{jit} \ln w_{lit} + \frac{1}{2}\sum_{l=1}^{L}\sum_{m=1}^{M} \gamma_{lm} \ln w_{lit} \ln w_{mit} + \sum_{n=1}^{N} \chi_n \ln e_{nit} + \sum_{t=1992}^{1996} \varphi_t D_t + \sum_{q=1}^{4} \eta_q Q_q + \sum_{j=1}^{3} \phi_j T_{jit} + u_{it} + v_{it},$$

where i and t indicate that an observation is for facility i at time t, y is a vector of output quantities, w is a vector of input prices, and e is a vector of N environmental variables (these are described in more detail later). D is a set of indicator variables that equals one in a particular year and zero otherwise and is meant to allow for technical change. Q is a set of indicator variables that equals one in a particular quarter and zero otherwise and is meant to allow for seasonal effects. The T_{it} terms allow for adjustment costs associated with the installation of second-generation equipment (BPS 3000 machines). Since such costs were not necessarily linear over time, we define three adjustment-cost variables, T_{1it}, T_{2it}, and T_{3it}. The first two, T_{1it} (T_{2it}), are indicator variables that equal one in the first (second) quarter that a machine was installed, and zero otherwise. The third, T_{3it}, is a time-trend variable that begins three quarters after the new machine was installed. Finally, u and v represent deviations from the cost frontier due to facility-level cost efficiency ($u \geq 0$) and statistical noise (v).[17]

The translog cost function is a second-order approximation to any function about a point of expansion and has been used extensively in recent years. Unfortunately, as one moves away from the point of expansion, the approximation becomes less precise.[18]

Because we derive our estimates of scale economies and marginal costs from the relationship between outputs and cost, we must be careful to specify the relationship with a sufficient amount of flexibility to ensure that the true underlying relationship between cost and output can be revealed. To explore whether the translog is flexible enough, we also employ a "hybrid-translog" cost function. This cost function includes the terms of the translog model (equation [1]) together with first-, second- and third-order trigonometric terms of the Fourier functional form.[19] More formally, the hybrid-translog takes the form

17 Estimates of u_i, the facility-specific cost-efficiency term, need to be interpreted with care. These estimates are conditional on assumptions about the functional form for the cost function and its error terms.

18 Generally, the expansion point used is the mean of the data.

19 The hybrid-translog functional form has been used by Gallant (1982) and Berger, Leusner, and Mingo (1997).

$$\text{(2)}\quad \ln C_{it} = \beta_0 + \sum_{j=1}^{J} \beta_j \ln y_{jit}$$
$$+\frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{J} \beta_{jk} \ln y_{jit} \ln y_{kit} + \sum_{l=1}^{L} \delta_l \ln w_{lit}$$
$$+ \sum_{j=1}^{J}\sum_{l=1}^{L} \delta_{jl} \ln y_{jit} \ln w_{lit}$$
$$+\frac{1}{2}\sum_{l=1}^{L}\sum_{m=1}^{M} \gamma_{lm} \ln w_{lit} \ln w_{mit}$$
$$+\sum_{n=1}^{N} \chi_n \ln e_{nit}$$
$$+\sum_{t=1992}^{1996} \varphi_t D_t + \sum_{q=1}^{4} \eta_q Q_q$$
$$+\sum_{j=1}^{3} \phi_j T_{jit}$$
$$+\sum_{j=1}^{J} [\psi_j \cos z_{jit} + \omega_j \sin z_{jit}]$$
$$+\sum_{j=1}^{J}\sum_{k=1}^{J} [\psi_{jk} \cos(z_{jit}+z_{kit}) + \omega_{jk} \sin(z_{jit}+z_{kit})]$$
$$+\sum_{j=1}^{J}\sum_{k=1}^{J}\sum_{l=1}^{J} [\psi_{jkl} \cos(z_{jit}+z_{kit}+z_{lit})$$
$$+ \omega_{jkl} \sin(z_{jit}+z_{kit}+z_{lit})] + u_{it} + v_{it},$$

where the z_{jit} are the output quantities normalized over the range $(0, 2\pi)$. Equations (1) and (2) are estimated along with corresponding equations for input shares, with the usual restrictions on symmetry and linear homogeneity imposed.

Although the currency processing machines themselves are quite intricate, the production technology for currency processing, as revealed by the cost function, may be relatively simple. If this hypothesis is correct, then the translog functional form should adequately characterize currency processing. One risk associated with using the hybrid-translog functional form is that the data may be overfitted. That is, statistical noise in the data may be absorbed into the cost function. By estimating both functional forms, we can obtain a deeper understanding of the relationship between output and costs.

We use two different econometric techniques to measure facility-level cost efficiencies for Federal Reserve currency processing and handling operations because empirical measurements of cost efficiency in the financial services industry often vary significantly depending on the methodology employed.[20] By estimating the model two ways, we can explore how sensitive our results are to differing assumptions. The generalized least squares (GLS) estimation technique uses the longitudinal aspect of the data and avoids assuming a specific distribution for the inefficiency term.[21] Basically, this technique uses repeated observations over time to identify firm-specific, time-invariant inefficiencies, that is, $u_{it}=u_i$. Iterative, seemingly unrelated regression techniques are used to estimate the system of cost and input share equations using longitudinal data. The inefficiency terms are calculated by using the average of the residuals for each facility i, $\hat{\alpha}_i$. The most efficient currency handling facility in the sample is assumed to be fully efficient, and the inefficiency of every other facility i is measured by the proportionate increase in predicted costs above the predicted cost of the most efficient facility, or $\hat{\alpha}_i - \min_j \hat{\alpha}_j$.

For GLS estimators to be consistent, the density of the inefficiency disturbances must be nonzero in the neighborhood of $(0, \omega)$ for some $\omega > 0$. In other words, as the number of facilities in the sample increases, the probability that a facility lies near the frontier approaches one. Our efficiency measure is calculated using the log difference in average residuals between the facility on the cost frontier and the actual average residual of facility i, or $\exp(-(\hat{\alpha}_i - \min_j \hat{\alpha}_j)$. This measure is bounded between zero and one, with the most efficient facility having an efficiency value equal to one.

The maximum likelihood technique (MLE) identifies inefficiency primarily by the skewness in residuals, rather than by persistence over time.[22] Statistical noise, v, is assumed to be normally distributed, while inefficiency, u, is assumed to be half-normally distributed. Our MLE measure of inefficiency is calculated as the conditional mean, $E(u|u+v)$.[23] Because the MLE technique requires a priori assumptions concerning the distribution of the inefficiency terms while the GLS technique does not, we tend to favor the GLS results. Both cost-

20 See, for example, Ferrier and Lovell (1990), Bauer, Berger, and Humphrey (1993), and Berger (1993).

21 See Schmidt and Sickels (1984).

22 The maximum likelihood technique we employ is based on Bauer, Ferrier, and Lovell (1987).

23 This technique was developed by Jondrow, Lovell, Materov, and Schmidt (1982).

efficiency measures are presented below to ascertain how robust our facility-level efficiency rankings are to alternative assumptions about the statistical properties of the inefficiency terms.

In practice, managers at each facility may have little ability to vary the amount of equipment and building inputs used over the short run. Adding vault space, for example, would require considerable modifications to a facility because of the high security required for such space. Also, high-speed currency processing machines are relatively unique, and when they were last purchased in 1990, enough were purchased to meet anticipated demand through 2003. If an input, l, is, for all practical purposes, fixed, then it is appropriate to use the input quantity, q_{lit}, rather than the input price, w_{lit}, in the cost function. This substitution would be appropriate regardless of the functional form chosen for the cost function or the estimation technique employed. Returns-to-scale measures, of course, depend on whether inputs are fixed or variable. When all inputs are variable, returns to scale (RTS) can be measured using the cost elasticity ($\sum_l \delta \ln C / \delta \ln q_l$). When K inputs are fixed over the short run, the fixed-input version of the RTS measure is:

$$(3)\quad \text{RTS} = \frac{\sum_{j=1}^{J} [\delta \ln C(y_{it}, w_{it}, q_{kit}) / \delta \ln y_{jit}]}{1 - \sum_{k=1}^{K} [\delta \ln C(y_{it}, w_{it}, q_{kit}) / \delta \ln q_{kit}]}$$

If the RTS measure is greater than one, then costs increase more than proportionately with output, which implies that the facility is operating in a region where there are diseconomies of scale. If, however, the RTS measure is less than one, then the facility is operating in a region where there are increasing returns to scale. Only when the RTS measure equals one, is the facility operating with constant returns to scale.

III. Data

We collected quarterly data (1991 to 1996, inclusive) on total costs, output volumes, input prices, input quantities, and environmental variables for 37 Federal Reserve currency processing and handling facilities. We chose this time period because it was a fairly tranquil period for currency operations. Before 1991, some Federal Reserve facilities used an assortment of currency processing machines in addition to the first generation of high-speed processors, the CVCSs. After 1996, new note designs were introduced—the \$50 note in October 1997 and the \$20 note in September 1998.[24] Also, as the century date change drew near, the amount of currency outstanding increased substantially.[25] In all, the sample included 886 observations.[26]

The primary data source is annual functional cost accounting records from the Federal Reserve's Planning and Control System (PACS), which are collected to monitor costs and improve resource allocation within the System. These data are supplemented by other cost data, machine counts, and usage statistics; data from occasional Federal Reserve surveys; and price index information from the Bureau of Economic Analysis and the Bureau of Labor Statistics.

Total costs for each currency facility, C_{it}, include the direct costs that arise from high-speed currency processing; off-line currency verification, destruction, and cancellation; and currency paying and receiving operations. Associated with these cost-generating activities are three distinct and measurable outputs. We include in the output vector, y_{it}, the number of fit notes generated by the high-speed currency processing operations, y_{fit}; the number of notes destroyed either on-line by the high-speed machines or off-line at the reconciliation stations, y_{dit}[27]; and the total number of transactions with depository institutions, that is, the sum of the number of incoming shipments of currency received at facility i and the number of outgoing orders for currency filled by facility i, y_{nit}.

Inputs used in currency operations are classified into buildings, B, labor, L, equipment, E, and materials, M. Buildings' share of total currency costs averages only about 16 percent

24 The new design for the \$100 note was introduced in March 1996. Because the average life of a \$100 note is about 8.5 years, however, the number of \$100 notes received by the Federal Reserve in 1996 (about 1 billion notes) was relatively small compared to the number of smaller-denomination notes received (about 22.6 billion notes). Thus, the introduction of the new design for the \$100 note is unlikely to have materially affected the Federal Reserve's currency operations in 1996.

25 Total currency in circulation increased about 22 percent in 1999, although it had only grown about 7 percent per annum in the previous four years.

26 Although the Helena, Montana, facility operated during the two quarters when the new BPS 3000 machines were installed, there are no reported volumes for this period.

27 Off-line destruction is undoubtedly more labor intensive and consequently more expensive, but these items make up less than half of 1 percent of notes destroyed.

because the interest expenses associated with the acquisition of buildings are not included in the cost-accounting framework. For the price of building services, w_B, we use an annual square-foot replacement cost, adjusted by the depreciation rate, for the location of each building with currency operations.[28] The quantity of building services, q_B, is proxied by the actual number of square feet of space occupied by currency operations at each facility.

Over the 1991–96 period, approximately 52 percent of total costs for currency operations are attributable to labor expenses. The price of labor, w_L, is constructed using data on expenditures for labor, including salaries, retirement and other benefits, and the number of employee hours spent in currency operations.

Equipment expenditures consist of depreciation and maintenance expenses. For the price of equipment, w_E, we use the price index for currency processing equipment reported in the *Producer Price Index Detailed Report*. For the quantities of equipment in operation at each facility during each quarter, q_{CVCS} and q_{BPS}, we use actual machine counts for CVCS and BPS 3000 machines, respectively.

Expenditures on materials, including currency straps and packaging, computer support, printing and duplicating, and other centrally provided support costs, account for about 11 percent of total costs. The price for material inputs, w_M, is constructed from material expenditure shares and various price indexes for components of materials from the *Survey of Current Business*[29] and the *Producer Price Index Detailed Report*, using a Tornquist approximation to a Divisia index.[30]

Environmental variables are used to control for the "quality" of incoming shipments and the operating environment at each facility. To measure the quality of incoming shipments we use several proxies: the number of depositor errors per 100,000 notes deposited; the proportion of incoming notes that are $1 notes; the proportion of notes that are rejected by high-speed equipment; and the mean shipment size. A higher value for any of these proxies is expected to increase costs at the facility. Depositor errors occur when a strap has more or fewer than 100 notes of the same denomination. When such errors occur, the depositor's reserve account must be reconciled to reflect the correct deposit balance.[31] Small-denomination notes, such as $1 notes, have a relatively short life compared to large-denomination notes. For example, the average life of a $1 note is 1.5 years, while the average life of a $50 note is 5 years. Consequently, we can use the proportion of incoming $1 notes in a shipment to proxy for the shipment's overall quality. The proportion of notes rejected by high-speed equipment would directly affect the costs of currency handling and processing because rejects are handled manually by an operator on a low-speed machine. Also, a large average shipment size may create a temporary backlog for processing work.

The operating environment is controlled for by variables that capture usage intensity of the equipment. Usage intensity is proxied by the proportion of scheduled operating time that processing equipment is "down"; the proportion of notes processed on BPS 3000 machines; and the mean throughput of processing equipment (measured in notes per hour). In addition, indicator variables are used to account for temporary cost increases that resulted from an unusual event that would affect only one Federal Reserve processing facility. One indicator variable allows for higher costs during the period when the New York office was testing the new BPS 3000 machines (between 1992:IIIQ and 1994:IVQ). Because that office was the first to use the BPS 3000 machines, it is reasonable to assume that its transition costs may have been higher than those who adopted the machines later. Another indicator variable is used to accommodate higher costs during the period when the new Dallas branch was constructed (1995:IQ through 1996:IVQ).

IV. Empirical Results

The two forms of the cost function, the translog and the hybrid-translog, are each specified with different combinations of fixed and variable inputs. In one specification, all inputs are variable, in another, equipment inputs are fixed, and in another, equipment

28 For these purposes, data on replacement costs for buildings are taken from the 17th edition of *Means Square Foot Costs* (Means, 1996).

29 See U.S. Department of Commerce (1990–97).

30 The Tornquist index is constructed using the rates of growth in the prices in each category. These rates of growth are weighted by the average proportionate shares of materials expenses to each category over adjoining periods. The base year is 1990.

31 Note that the Reserve Banks absorb the differences on $1 note deposits, so there are no depositor errors reported for $1 notes.

TABLE 2

Average Returns to Scale for Federal Reserve Currency Operations

Cost-function specification	GLS/translog	GLS/hybrid-translog	MLE/translog
	Functional form/estimation technique		
Arithmetic mean			
All inputs variable	0.79	0.78	0.77
Equipment fixed	0.81	0.76	0.80
Equipment and buildings fixed	0.83	0.78	0.81
Weighted mean			
All inputs variable	0.93	0.91	0.91
Equipment fixed	0.97	0.86	0.93
Equipment and buildings fixed	0.98	0.78	0.94

SOURCE: Authors' calculations. Weighted means use the number of notes processed as the weighting measure.

and buildings are fixed. In all specifications, labor and materials inputs are treated as variable. Cost frontiers are estimated using the GLS technique for both cost-function specifications and the MLE technique for the tranlog cost-function specification. These frontiers are denoted as GLS/translog, GLS/hybrid-translog, and MLE/translog.[32]

Returns to Scale

For every cost-function specification we estimate, the average-cost curve implied by the cost frontier is U-shaped with a fairly flat portion at the bottom of the U. The implied average cost curve descends rapidly up to 100 million notes per quarter and then is fairly flat for most of the remaining observed output range.[33] This shape is fairly typical for the average cost curve of paper-based payment technologies.[34]

The minimum efficient scale (MES) for currency depends on whether equipment and buildings are treated as fixed inputs. For example, using the GLS/translog model, the MES is 260 million notes per quarter when all inputs are variable, 680 million notes per quarter when equipment is fixed, and 250 million notes per quarter when equipment and buildings are fixed.[35] While this range for the MES is quite large, the difference in average costs for the largest and smallest MES output levels is not economically meaningful because the implied average cost function is so flat through the region of 250 million notes per quarter to 650 million notes per quarter.

Table 2 presents average RTS measures for Federal Reserve System currency and handling operations using several cost-function specifications and the two frontier-estimation techniques. In this table, both the arithmetic mean and a weighted mean of facility RTS measures are presented in the top and bottom panels, respectively. Weights each quarter vary across Federal Reserve facilities and are equal to the volume of notes processed at each facility. Strikingly, neither the cost-function specification nor the frontier-estimation technique greatly affects the average RTS measures. The measures are fairly robust to whether equipment and buildings are considered fixed. And, with one exception—the GLS/hybrid-translog model specified with fixed buildings and equipment—such measures are fairly robust to whether a translog or hybrid-translog functional form is used. Regardless of the cost-function specification or frontier-estimation technique chosen, the mean of the unweighted RTS is smaller than the mean of the weighted RTS measure. This happens because, even though there are substantial unexploited potential scale economies at many facilities in the System, the vast majority of notes are processed and handled at the facilities that have already achieved constant returns to scale.

Focusing on the first and last quarters of the estimation period, 1991:IQ and 1996:IVQ, respectively, we calculate facility-specific returns-to-scale measures. Table 3 presents these returns-to-scale measures for each office that is estimated using the GLS/translog model with all inputs specified as variable. The

32 We did not estimate an MLE/hybrid-translog model because the computation time required to achieve convergence would have been unacceptably long on account of the large number of additional parameters required for this model.

33 Most facilities with output less than 100 million notes per quarter have only one high-speed currency processing machine.

34 Zimmerman (1981) and Dotsey (1991) report cost-function estimates for Federal Reserve currency operations that are consistent with a U-shaped average cost curve. Humphrey (1981), Bauer and Hancock (1993), and Bauer and Ferrier (1996) also report cost-function estimates for Federal Reserve check processing operations that are consistent with a U-shaped average cost curve.

35 MES estimates from the GLS/translog model are representative of MES estimates from the GLS/hybrid-translog and MLE/translog models.

TABLE 3

Returns-to-Scale Measures for Federal Reserve Currency Processing and Handling Facilities

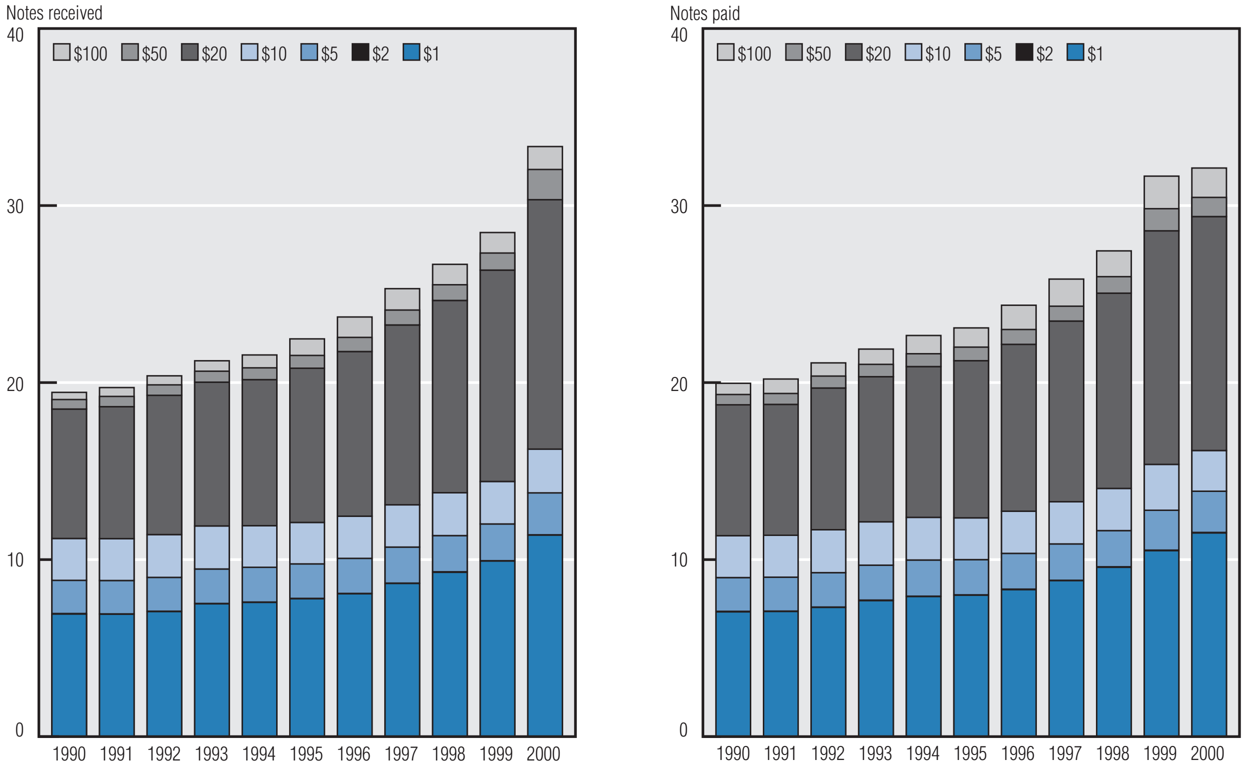
Federal Reserve office	1991:IQ estimate	1991:IQ t-test	1996:IVQ estimate	1996:IVQ t-test
FR 1	0.800	3.199	0.837	2.652
FR 2	0.933	1.109	0.937	1.046
FR 3	0.691	4.337	0.826	2.801
FR 4	1.006	–0.080	1.030	0.366
FR 5	0.843	1.760	0.594	4.031
FR 6	1.066	0.822	1.052	0.723
FR 7	0.990	0.133	1.122	1.529
FR 8	0.765	4.087	0.821	3.189
FR 9	0.793	3.105	0.874	2.135
FR 10	0.761	3.777	0.804	3.425
FR 11	0.712	4.621	0.810	3.346
FR 12	0.838	2.804	0.815	2.304
FR 13	0.497	6.692	0.619	5.579
FR 14	0.404	6.648	0.493	6.417
FR 15	0.621	5.255	0.795	3.649
FR 16	0.735	3.351	0.687	3.324
FR 17	0.613	5.903	0.680	5.252
FR 18	0.484	6.708	0.689	5.021
FR 19	1.208	2.404	1.141	1.502
FR 20	0.641	5.240	0.701	4.860
FR 21	0.687	4.866	0.734	4.313
FR 22	0.744	2.621	0.676	2.534
FR 23	0.778	2.953	0.787	2.575
FR 24	0.724	4.671	0.715	4.923
FR 25	0.779	3.863	0.815	2.817
FR 26	1.038	0.393	1.042	0.347
FR 27	0.657	5.399	0.690	4.859
FR 28	0.460	6.682	0.583	5.596
FR 29	0.916	1.145	0.946	0.796
FR 30	0.831	2.997	0.787	3.595
FR 31	0.572	5.441	0.554	6.461
FR 32	0.973	0.359	0.922	1.286
FR 33	0.572	6.323	0.509	6.572
FR 34	0.650	5.361	0.759	3.658
FR 35	1.062	0.848	1.013	0.166
FR 36	0.592	5.504	0.707	3.829
FR 37	0.648	5.152	0.754	4.349

SOURCE: Authors' calculations. The t-test statistic is for a "two-tail" test of the hypothesis that the estimated coefficient is equal to 1.

t-statistics in the table test the hypothesis that the RTS estimate differs from one. At the beginning of the sample, 27 facilities have statistically significant increasing returns to scale, 9 facilities exhibit constant returns to scale, and 1 facility has significant decreasing returns to scale. At the end of the sample, 28 facilities have statistically significant increasing returns to scale and 10 exhibit constant returns to scale. Although the number of notes received by the Federal Reserve System increased by about 20 percent over this period (figure 1),

FIGURE 1

Federal Reserve System Notes Paid and Received Annually (in thousands of notes)



SOURCE: Board of Governors of the Federal Reserve System.

the introduction of higher-speed processing machines enabled each facility to achieve similar returns to scale as before their introduction. Facilities with insufficient processing volume to achieve scale efficiency in 1991 continued to have insufficient processing volume to achieve scale efficiency in 1996. It remains the case, however, that scale efficiency for Federal Reserve currency operations improved on average because no facility operated with statistically significant decreasing returns to scale, and far fewer facilities operated with an RTS measure less than 0.70 by the end of the estimation period.

Our RTS findings suggest that there was considerable excess capacity in Federal Reserve currency operations. This excess capacity has been used to accommodate the public's growing demand for currency, which increases demand for currency processing, including the destruction of unfit currency. However, to the extent that use of the new Sacagawea dollar coins substitute for $1 notes, more Federal Reserve facilities may be operating with increasing returns to scale in the future unless some smaller facilities are consolidated. Of course, the higher unit costs of operating with increasing returns to scale must be weighed against the costs of being unable to meet demand shocks for currency processing and handling. Also, the cost savings of consolidating facilities would have to be balanced against the higher costs of transporting currency deposits and withdrawals over longer distances.[36] Alternatively, the overall level of scale efficiency

■ **36** Bauer et al. (2000) considered whether Federal Reserve System currency processing costs could be lowered by reallocating currency volume across facilities or by consolidating facilities. They found most cost savings could be achieved without closing any existing currency processing facilities. In addition, they argued that a new facility located in Phoenix, Arizona, would help to lower System currency processing costs.

TABLE 4

Marginal Costs of Production for Fit Currency, Destroyed Currency, and Cash Shipments

	Functional form/estimation technique		
	GLS/translog	GLS/hybrid-translog	MLE/translog
Arithmetic mean			
Fit currency (Cents per note)	0.358	0.344	0.398
Destroyed currency (Cents per note)	0.161	0.145	0.175
Cash shipments (Dollars per shipment)	22.63	30.43	14.09
Weighted mean			
Fit currency (Cents per note)	0.428	0.369	0.521
Destroyed currency (Cents per note)	0.234	0.200	0.235
Cash shipments (Dollars per shipment)	38.82	58.15	21.11

SOURCE: Authors' calculations. Weighted means use the number of notes processed as the weighting measure.

TABLE 5

The Effects of Environmental Variables on the Cost of Currency Operations

Environmental variable	Sample mean	Estimate
Quality of incoming shipments		
The number of depositor errors per $100,000 notes deposited	13.152	0.0317**
The proportion of incoming notes that were of $1 denomination	0.370	–0.024
The proportion of notes rejected by high-speed equipment	0.022	0.156***
The mean shipment size (in notes)	70,757	0.345***
Usage intensity of equipment		
The proportion of scheduled operating time that processing equipment was "down"	0.041	0.083***
The proportion of notes processed on BPS 3000 machines	0.223	0.068***
The mean throughput of processing equipment (in notes per hour)	66,030	0.234*

SOURCE: Authors' calculations. Estimates significantly different from zero at the 10, 5, and 1 percent significance level are indicated with *, **, and ***, respectively. All significance tests used a two-tailed t-test.

achieved by Federal Reserve currency operations could potentially be improved over time by adjusting standards for normal service levels or by changing fees for nonstandard access or nonstandard cash services.[37]

Marginal Costs

Recently, some observers have argued that the Federal Reserve should explicitly charge for currency services because outgoing currency from the Reserve Banks is of higher quality than the incoming currency from depository institutions.[38] Such observers typically argue that some currency users would be willing to pay for a higher-quality, "superfit" currency because they manage newer technologies that would perform better with higher-quality currency (such as ATMs or vending machines).

With the view to improve economic efficiency, potential prices for currency services would likely depend on the marginal costs of production for fit currency, for destroying currency, and for currency shipments. We estimate such marginal costs at each facility by using cost frontiers that assume all inputs are variable in the short run. In table 4, both the arithmetic mean and a weighted mean of Federal Reserve facility-level marginal cost estimates for fit currency, destroyed currency, and cash shipments are presented.

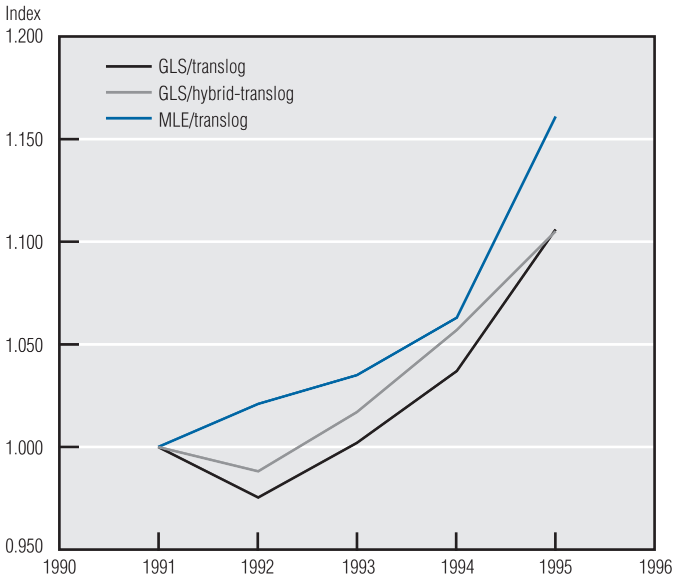
As with the weighted RTS measures presented in table 2, the weights used in table 4 each quarter vary across Federal Reserve facilities and are equal to the volume of notes processed at each facility. The average Federal Reserve System marginal cost for fit currency, destroyed currency, and cash shipments is higher when each facility's marginal cost is weighted by its output volume. This happens because the marginal cost is less than the average cost when a facility operates with increasing returns to scale, as did most facilities. The average marginal costs for fit and destroyed currency, unweighted or weighted, are higher when the MLE technique is used

37 Although the uniform cash access policy established normal access for a depository institution as once per week, Reserve Banks can charge for more frequent nonstandard access. In addition, Reserve Banks can and do charge for services such as nonstandard packages and nonstandard packaging of same-day express cash orders.

38 See Supel and Todd (1984) and Lacker (1993).

FIGURE 2

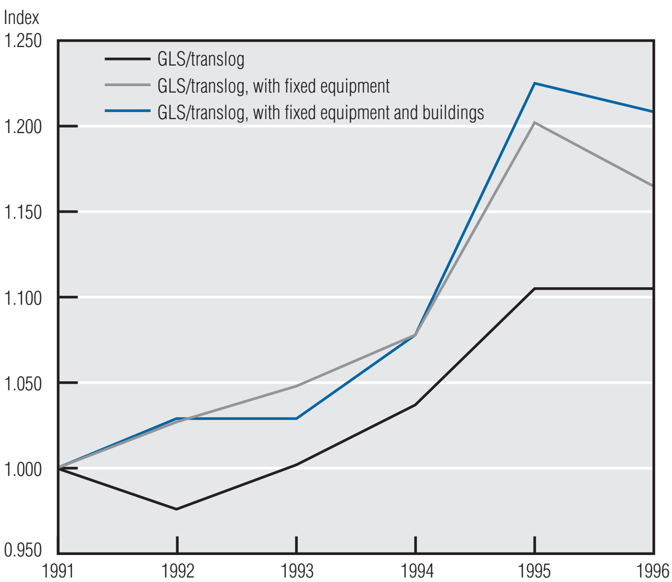
Year-Indicator Effects for Various Estimation Techniques



SOURCE: Authors' calculations.

FIGURE 3

Year-Indicator Effects for Various Model Specifications



SOURCE: Authors' calculations.

than when the GLS technique is used, but of similar magnitude.

The average Federal Reserve marginal cost for fit currency is typically less than one-half of a cent per note during the period studied. The average marginal cost for destroyed currency is about half that. Marginal cost estimates for cash shipments, which are reported at the bottom lines of the top and bottom panels of table 4, are very sensitive to both the functional form of the cost function and the econometric technique used to estimate the cost frontier: Depending on which forms and techniques are used, the arithmetic mean marginal cost for Federal Reserve cash shipments ranges from $14 to $30, and the weighted mean ranges from $21 to $58.

Environmental Factors

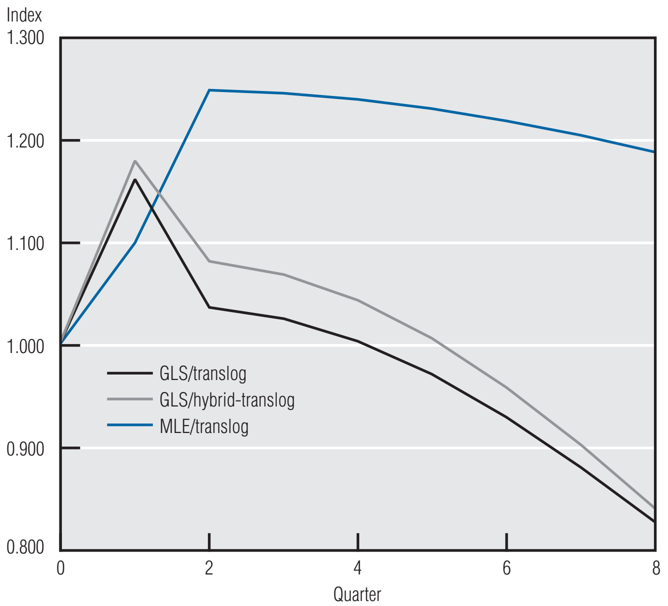
In every specification for the cost frontier we estimated, environmental factors are important cost determinants. Table 5 presents the effects of environmental variables on the cost of production when the cost frontier estimated is the GLS/translog and all inputs are considered variable.[39]

The quality of incoming currency deposits significantly influences Federal Reserve costs. The number of depositor errors per 100,000 notes deposited, the proportion of notes that are rejected by high-speed equipment, and the mean shipment size each significantly raises Federal Reserve currency handling and processing costs at the 5 percent level of significance. A 1 percent increase in the number of depositor errors increases costs by 0.03 percent, a 1 percent increase in the proportion of notes rejected by high-speed equipment increases costs by 0.16 percent, and a 1 percent increase in mean shipment size increases costs by 0.35 percent. Only the proportion of incoming notes that are of $1 denomination is not significant among the variables that measure incoming currency quality.

■ **39** The environmental effects we present in table 5 are representative of the others we derive using alternative cost frontiers.

FIGURE 4

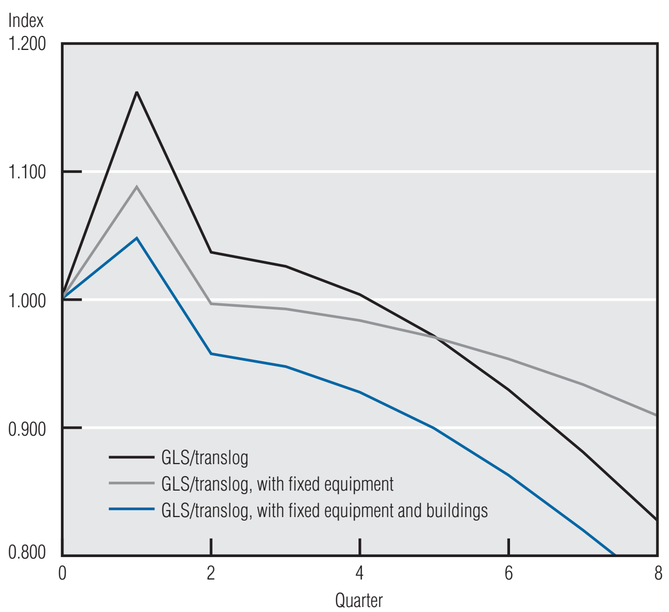
Learning-by-Doing Effect for Various Estimation Techniques



SOURCE: Authors' calculations.

FIGURE 5

Learning-by-Doing Effect for Various Model Specifications



SOURCE: Authors' calculations.

Operating environment variables also affect the cost of providing currency services. The proportion of scheduled operating time that processing equipment is down and the proportion of notes processed on BPS 3000 machines each significantly raises costs at the 1 percent level of significance. In addition, the mean-throughput-of-processing-equipment variable is significant at the 10 percent level.

Time Adjustment Factors

The last environmental coefficients we consider are the variables included to control for technical change in general and the introduction of the new machines in particular. The estimated coefficients for these variables are fairly consistent across the various specifications. Figure 2 plots the effects of the year-indicator variable for the GLS/translog, GLS/hybrid-translog, and MLE/translog models. Figure 3 plots the GLS/translog model with all inputs variable, with equipment fixed, and with equipment and building inputs fixed. Although it appears that the cost function has shifted up 10 to 15 percent from 1991 to 1996, it should be noted that because we include variables that control for the transition to new machines, it is not possible to interpret these coefficients as a technical change index, as is often done. The reason is that the full effect of technical change must include the effects of these other variables.

Recall that we include an indicator variable for the first quarter during which a processing site employed the new machines, another indicator variable for the second quarter of use, and a time trend to capture the effects in later quarters. The combined effect of these three variables, which we call the learning-by-doing effect, is plotted in figure 4 for the GLS/translog, GLS/hybrid-translog, and MLE/translog models. As one might suspect, new machines disrupt the previous work environment, and all the models report higher costs in the quarter in which the new machines are introduced. The models do differ as to what happens after that. Our preferred estimation technique, GLS, estimates that, after spiking sharply in the quarter when the new machines are introduced, costs fairly rapidly decline. After just two years, the learning-by-doing effect of the new machines would be more than enough to offset the year-indicator variables. When the model is estimated using MLE,

TABLE 6

Facility-Level Efficiency Estimate Statistics for Federal Reserve Currency Operations

	Functional form/estimation technique		
Type of specification	GLS/translog	GLS/hybrid-translog	MLE/translog
All inputs variable			
Arithmetic mean	0.815	0.844	0.893
Weighted mean	0.810	0.838	0.888
Minimum	0.596	0.633	0.770
Maximum	1.00	1.00	0.952
Median	0.822	0.848	0.897
Equipment inputs fixed			
Arithmetic mean	0.825	0.863	0.902
Weighted mean	0.819	0.850	0.896
Minimum	0.623	0.662	0.800
Maximum	1.00	1.00	0.946
Median	0.834	0.861	0.903
Equipment and building inputs fixed			
Arithmetic mean	0.844	0.861	0.917
Weighted mean	0.836	0.866	0.911
Minimum	0.716	0.722	0.871
Maximum	1.00	1.00	0.958
Median	0.839	0.855	0.918

SOURCE: Authors' calculations. The weighted mean level of cost efficiency weights each facility-level cost-efficiency estimate by the mean number of notes processed per quarter at that facility during the 1991:IQ to 1996:IVQ period.

the learning-by-doing effect is much smaller, and consequently, the decline in costs is much slower. As figure 5 reveals, the GLS estimates of the learning-by-doing effect demonstrate the same pattern whether capital or buildings are treated as variable or fixed inputs.

Cost Efficiency

The estimates for cost efficiency using the alternative functional forms, estimation techniques, and specifications are presented in table 6. The GLS/translog model yields the lowest average cost-efficiency estimates for Federal Reserve System currency operations. Depending on whether equipment and building inputs are fixed, the arithmetic average of facility-level cost-efficiency estimates ranges between 0.81 and 0.84, and the minimum cost-efficiency estimates range between 0.60 and 0.72. These statistics suggest that during the period studied, the average facility operated at more than 80 percent of the efficiency of the "best-practice" facility, and the worst performer in the system could have substantially improved its efficiency.

Estimates of average cost-efficiency levels are higher when a hybrid-translog functional form is used for the cost function. This is not surprising because the hybrid-translog attributes a larger proportion of the variance in cost to the structural component of the model. This attribution results in higher average cost-efficiency estimates. In addition, estimates of average cost-efficiency levels are higher when the MLE technique is used to estimate the frontier.

The weighted means of the cost-efficiency estimates are lower than the corresponding arithmetic means. The weighted means are lower because larger facilities in the System typically have lower cost-efficiency estimates than do smaller facilities.[40]

In general, cost-function specifications with fixed inputs yield higher average levels of cost efficiency, regardless of the functional form for the cost function or the technique used to estimate the cost frontier. This consistency across the top, middle, and bottom panels of table 6 suggests that managers view equipment and building inputs as largely fixed. Because changes to such inputs require substantial lead times, this view seems quite plausible.

If facility-specific efficiency measures are to be useful for managerial or policymaking purposes, they should rank facilities consistently and be robust to different specifications of the cost function or to different econometric techniques. In tables 7, 8, and 9, we present facility-specific cost-efficiency estimates for all of our models. Table 7 reports estimates when all inputs are variable, table 8 when equipment inputs are fixed, and table 9 when equipment and buildings are fixed inputs. Each table shows the results of all combinations of the functional form of the cost function and the two estimation techniques we consider. In addition, for each model, cost-efficiency estimates are ranked from most efficient (cost-efficiency estimate equals one) to least efficient

40 Currency processing and handling facilities with the largest output levels tend to be Reserve Bank head offices. Such facilities have additional administrative staff for the District. Even after incorporating a head office indicator variable into the cost-function models, facilities with the largest output volume tend to have lower levels of cost efficiency than smaller facilities.

TABLE 7

Cost-Efficiency Estimates and Relative Rankings, Cost Model with All Inputs Variable

Federal Reserve facility	GLS/translog cost-efficiency estimate	Rank	GLS/hybrid-translog cost-efficiency estimate	Rank	MLE/translog cost-efficiency estimate	Rank
FR 1	1.000	1	0.989	3	0.943	2
FR 2	0.991	2	1.000	1	0.941	3
FR 3	0.982	3	0.992	2	0.952	1
FR 4	0.958	4	0.971	4	0.931	4
FR 5	0.900	5	0.967	5	0.916	8
FR 6	0.872	6	0.910	6	0.899	15
FR 7	0.871	7	0.863	13	0.898	17
FR 8	0.871	8	0.849	17	0.926	5
FR 9	0.868	9	0.900	8	0.924	6
FR 10	0.859	10	0.890	9	0.920	7
FR 11	0.859	11	0.874	11	0.896	21
FR 12	0.855	12	0.848	18	0.916	9
FR 13	0.854	13	0.849	16	0.893	24
FR 14	0.854	14	0.874	10	0.912	10
FR 15	0.845	15	0.901	7	0.899	16
FR 16	0.837	16	0.854	14	0.882	26
FR 17	0.833	17	0.838	20	0.895	22
FR 18	0.831	18	0.821	24	0.893	23
FR 19	0.822	19	0.874	12	0.910	11
FR 20	0.810	20	0.853	15	0.896	20
FR 21	0.809	21	0.826	23	0.881	28
FR 22	0.808	22	0.848	19	0.900	14
FR 23	0.799	23	0.821	25	0.879	29
FR 24	0.786	24	0.817	26	0.897	19
FR 25	0.770	25	0.827	22	0.901	13
FR 26	0.764	26	0.802	29	0.891	25
FR 27	0.760	27	0.808	27	0.872	30
FR 28	0.756	28	0.801	30	0.882	27
FR 29	0.744	29	0.835	21	0.897	18
FR 30	0.728	30	0.804	28	0.906	12
FR 31	0.726	31	0.780	32	0.867	31
FR 32	0.722	32	0.712	36	0.839	35
FR 33	0.714	33	0.726	34	0.866	32
FR 34	0.706	34	0.759	33	0.838	36
FR 35	0.706	35	0.784	31	0.865	33
FR 36	0.693	36	0.723	35	0.852	34
FR 37	0.596	37	0.633	37	0.770	37

SOURCE: Authors' calculations. Facility with rank 1 in each column is the "best-practice" facility using the specification/econometric technique specified. Cost-efficiency estimates are for the full period from 1991:IQ to 1996:IVQ.

TABLE 8

Cost-Efficiency Estimates and Relative Rankings, Cost Model with Equipment Fixed

Federal Reserve facility	GLS/translog cost-efficiency estimate	Rank	GLS/hybrid-translog cost-efficiency estimate	Rank	MLE/translog cost-efficiency estimate	Rank
FR 1	0.921	4	0.957	5	0.931	4
FR 2	0.974	2	0.975	3	0.941	3
FR 3	0.943	3	0.969	4	0.946	1
FR 4	1.000	1	1.000	1	0.943	2
FR 5	0.911	5	0.983	2	0.927	5
FR 6	0.890	6	0.918	7	0.917	13
FR 7	0.862	12	0.891	11	0.908	17
FR 8	0.863	11	0.861	19	0.925	7
FR 9	0.861	14	0.918	6	0.924	8
FR 10	0.867	9	0.904	9	0.920	11
FR 11	0.865	10	0.882	13	0.902	21
FR 12	0.840	18	0.868	16	0.918	12
FR 13	0.815	21	0.851	23	0.883	31
FR 14	0.862	13	0.876	14	0.915	15
FR 15	0.846	16	0.905	8	0.908	16
FR 16	0.875	7	0.868	17	0.899	23
FR 17	0.809	24	0.856	22	0.903	20
FR 18	0.814	22	0.829	28	0.899	24
FR 19	0.844	17	0.885	12	0.916	14
FR 20	0.809	25	0.867	18	0.899	22
FR 21	0.828	20	0.858	21	0.894	28
FR 22	0.856	15	0.897	10	0.921	10
FR 23	0.812	23	0.807	31	0.896	26
FR 24	0.834	19	0.861	20	0.923	9
FR 25	0.788	26	0.842	25	0.894	27
FR 26	0.783	27	0.833	27	0.897	25
FR 27	0.732	33	0.801	32	0.864	33
FR 28	0.755	29	0.823	29	0.892	29
FR 29	0.781	28	0.851	24	0.903	19
FR 30	0.870	8	0.875	15	0.925	6
FR 31	0.750	31	0.810	30	0.891	30
FR 32	0.725	34	0.744	36	0.853	36
FR 33	0.712	35	0.769	35	0.874	32
FR 34	0.709	36	0.788	33	0.862	34
FR 35	0.754	30	0.842	26	0.905	18
FR 36	0.734	32	0.786	34	0.861	35
FR 37	0.623	37	0.662	37	0.800	37

SOURCE: Authors' calculations. Facility with rank 1 in each column is the "best practice" facility using the specification/econometric technique specified. Cost-efficiency estimates are for the full period from 1991:IQ to 1996:IVQ.

TABLE 9

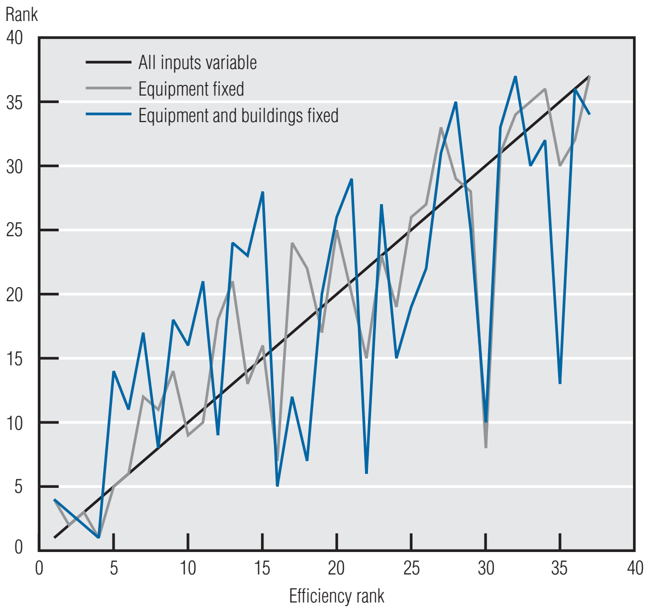
Cost-Efficiency Estimates and Relative Rankings, Cost Model with Fixed Equipment and Buildings

Federal Reserve facility	GLS/translog cost-efficiency estimate	Rank	GLS/hybrid-translog cost-efficiency estimate	Rank	MLE/translog cost-efficiency estimate	Rank
FR 1	0.963	4	0.961	3	0.943	4
FR 2	0.982	3	0.958	4	0.958	1
FR 3	0.994	2	0.977	2	0.944	3
FR 4	1.000	1	1.000	1	0.946	2
FR 5	0.858	14	0.911	6	0.931	11
FR 6	0.875	11	0.902	7	0.932	9
FR 7	0.844	17	0.854	20	0.919	17
FR 8	0.897	8	0.878	12	0.931	10
FR 9	0.844	18	0.864	16	0.927	14
FR 10	0.847	16	0.864	17	0.925	15
FR 11	0.823	21	0.845	25	0.919	18
FR 12	0.890	9	0.886	11	0.930	12
FR 13	0.817	24	0.845	26	0.905	29
FR 14	0.820	23	0.849	23	0.918	19
FR 15	0.801	28	0.831	27	0.918	21
FR 16	0.927	5	0.891	8	0.933	6
FR 17	0.863	12	0.861	18	0.918	20
FR 18	0.902	7	0.890	9	0.938	5
FR 19	0.837	20	0.853	21	0.914	23
FR 20	0.807	26	0.825	28	0.913	24
FR 21	0.800	29	0.814	32	0.910	28
FR 22	0.909	6	0.922	5	0.927	13
FR 23	0.803	27	0.803	34	0.902	31
FR 24	0.851	15	0.849	22	0.932	7
FR 25	0.839	19	0.875	13	0.910	27
FR 26	0.821	22	0.855	19	0.911	26
FR 27	0.796	31	0.846	24	0.896	32
FR 28	0.747	35	0.807	33	0.886	33
FR 29	0.811	25	0.874	15	0.912	25
FR 30	0.881	10	0.874	14	0.932	8
FR 31	0.753	33	0.819	29	0.903	30
FR 32	0.716	37	0.722	37	0.877	36
FR 33	0.796	30	0.816	30	0.916	22
FR 34	0.766	32	0.815	31	0.884	34
FR 35	0.860	13	0.888	10	0.924	16
FR 36	0.744	36	0.770	35	0.871	37
FR 37	0.747	34	0.760	36	0.882	35

SOURCE: Authors' calculations. Facility with rank 1 in each column is the "best practice" facility using the specification/econometric technique specified. Cost-efficiency estimates are for the full period from 1991:IQ to 1996:IVQ.

FIGURE 6

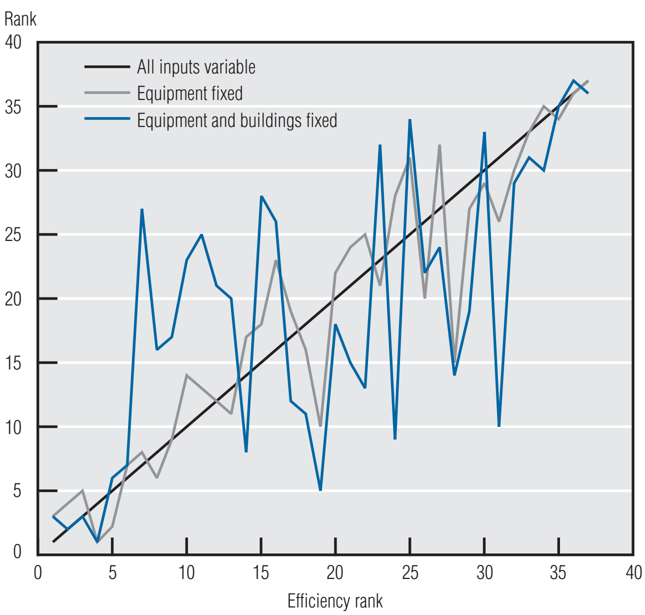
Cost-Efficiency Rankings, GLS/Translog



SOURCE: Authors' calculations.

FIGURE 7

Cost-Efficiency Rankings, GLS/Hybrid-translog



SOURCE: Authors' calculations.

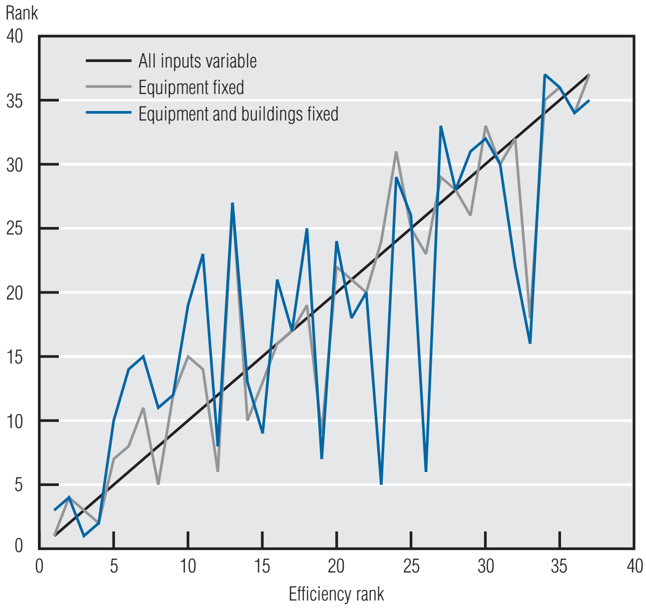
(cost-efficiency equals the minimum cost-efficiency estimate). Facilities are identified using their cost-efficiency estimate rank for the GLS/translog model that considered all inputs variable (table 7, second column).

Although the relative ranks of cost-efficiency estimates for individual facilities vary considerably across our different models, a number of facilities are consistently ranked among the most and least efficient in the Federal Reserve System. Facilities denoted by FR1 through FR4 are consistently ranked among the five most efficient facilities. FR1 through FR6 are consistently among the top 10 efficient facilities. Consistently ranked among the least efficient facilities are facilities denoted by FR32, FR34, FR36, and FR37. Interestingly, both the most efficient and least efficient groups contain head offices and branches. Rankings in the middle range of efficiency rankings are more volatile because, for facilities in that range, a small change in efficiency leads to a large change in ranking, and efficiency estimates for facilities in the middle range are very close. The consistency of the relative rankings of the most and least efficient offices and the considerable disagreement about the rankings of facilities in the middle of the range of rankings is illustrated in figures 6 through 8.

We calculate Spearman and Kendall Tau-b rank-order correlations, reported in tables 7, 8, and 9, to determine the association between cost-efficiency estimates. The Spearman rank-order correlation measure is concerned with differences in *absolute* rankings, putting the highest weight on facilities at the extremes, whereas the Kendall Tau-b rank-order correlation measure captures the differences in *relative* rankings. These rank-order correlations are presented in tables 10 and 11 along with rank-order correlations between cost-efficiency estimates and two accounting-based performance measures—the number of notes processed on high-speed equipment per man-hour (a labor productivity measure) and the number of notes processed per dollar of cost (a unit cost measure). These accounting-based measures of performance do not take into account differences in environmental variables, nor do they account for the effect of differences in the scale of operations between facilities, but they are commonly used as performance measures.

FIGURE 8

Cost-Efficiency Rankings, MLE/Translog



SOURCE: Authors' calculations.

The Spearman correlation coefficients presented in table 10 are generally greater than 0.70, and each is statistically significant at the 5 percent level. Overall, such correlation coefficients are larger when the models' assumptions are more similar or when the same econometric technique is used to estimate the cost frontier. Not surprisingly, the weakest relationships are between cost-frontier-based and accounting-based measures of facility-level efficiency.

The Kendall Tau-b correlation coefficients presented in table 11 are smaller than the corresponding Spearman correlation coefficients presented in table 10. This implies that the efficiency rankings do not perform as well for facilities in the middle cost-efficiency range. As with the Spearman correlation coefficients, every Kendall Tau-b correlation coefficient is statistically significant at the 5 percent level, which suggests an extremely high degree of concordance between our cost-efficiency measures. Again, consistency is greatest between cost-efficiency measures when the models' assumptions are most similar or where the same econometric technique is used to estimate the cost frontier. Consistency is weakest between cost-frontier-based and accounting-based measures of facility-level efficiency. Indeed, the Kendall correlation coefficients between cost-frontier-based and accounting-based measures of facility-level efficiency are all less than 0.50. This suggests that econometric measures of site-level efficiency may provide useful insights for the identification of "best-practice" currency processing and handling facilities within the Federal Reserve System, which accounting measures of performance may overlook.

Some Federal Reserve districts may focus on cost efficiency to a greater degree than others. Because five of the six consistently most efficient facilities are located in just two of the twelve Federal Reserve districts, we test for differences in the estimates of the mean level of cost efficiency of facilities in those two Federal Reserve districts against the remaining ten districts. We find that the average efficiency of the facilities in the two districts exceeds that of the remainder of the Federal Reserve System at the 1 percent significance level using a one-sided t-test.

We were curious to see whether cost-efficiency estimates show any pattern across services. To do so, we compare our baseline estimates of currency cost efficiency with those for check as reported by Bauer and Ferrier (1996). The Spearman correlation coefficient is 0.309 and is statistically significant only at the 10 percent confidence level. The Kendall Tau-b correlation coefficient is lower, 0.232, but is statistically significant at the 5 percent confidence level. These results suggest that senior officers who oversee currency operations well also tend to oversee check operations well. These correlations also suggest that high cost efficiency in one area is not achieved by shifting costs to other operations.

TABLE 10

Spearman Correlation Coefficients Between Site-Level Measures of Efficiency

		Cost-frontier-based cost-efficiency measures									Accounting-based performance measures	
		All inputs variable			Equipment fixed			Equipment and buildings fixed				
		GLS/ translog	GLS/ hybrid-translog	MLE/ translog	GLS/ translog	GLS/ hybrid-translog	MLE/ translog	GLS/ translog	GLS/ hybrid-translog	MLE/ translog	Labor Productivity	Unit Cost
All inputs variable	GLS/translog	1.0	0.938	0.825	0.881	0.865	0.743	0.690	0.645	0.712	0.598	0.639
	GLS/hybrid-translog		1.0	0.848	0.881	0.934	0.772	0.649	0.653	0.684	0.612	0.667
	MLE/translog			1.0	0.845	0.881	0.910	0.737	0.743	0.750	0.630	0.645
Equipment fixed	GLS/translog				1.0	0.925	0.864	0.794	0.730	0.836	0.578	0.554
	GLS/hybrid-translog					1.0	0.892	0.717	0.716	0.766	0.574	0.598
	MLE/translog						1.0	0.825	0.798	0.867	0.595	0.564
Equipment and buildings fixed	GLS/translog							1.0	0.936	0.936	0.575	0.463
	GLS/hybrid-translog								1.0	0.871	0.561	0.483
	MLE/translog									1.0	0.535	0.465
Other	Labor productivity										1.0	0.810
	Unit cost											1.0

SOURCE: Authors' calculations.

TABLE 11

Kendall tau-b Correlation Coefficients Between Site-Level Measures of Efficiency

		Cost-frontier-based cost-efficiency measures									Accounting-based performance measures	
		All inputs variable			Equipment fixed			Equipment and buildings fixed				
		GLS/ translog	GLS/ hybrid-translog	MLE/ translog	GLS/ translog	GLS/ hybrid-translog	MLE/ translog	GLS/ translog	GLS/ hybrid-translog	MLE/ translog	Labor Productivity	Unit Cost
All inputs variable	GLS/translog	1.0	0.804	0.670	0.742	0.691	0.592	0.526	0.496	0.565	0.411	0.456
	GLS/hybrid-translog		1.0	0.697	0.727	0.802	0.625	0.486	0.480	0.544	0.426	0.471
	MLE/translog			1.0	0.670	0.703	0.778	0.562	0.574	0.571	0.465	0.486
Equipment fixed	GLS/translog				1.0	0.781	0.689	0.616	0.568	0.661	0.429	0.414
	GLS/hybrid-translog					1.0	0.727	0.541	0.534	0.598	0.396	0.423
	MLE/translog						1.0	0.664	0.662	0.715	0.429	0.429
Equipment and buildings fixed	GLS/translog							1.0	0.790	0.787	0.399	0.324
	GLS/hybrid-translog								1.0	0.697	0.399	0.354
	MLE/translog									1.0	0.396	0.345
Other	Labor productivity										1.0	0.637
	Unit cost											1.0

SOURCE: Authors' calculations.

V. Conclusion

It costs more than half a billion dollars each year to meet the demands for currency by depository institutions, businesses, and consumers.[41] Yet very little research has been devoted to understanding the factors that affect such costs. This paper has attempted to fill a portion of this gap by considering the scale and cost efficiency of Federal Reserve currency operations.

Our finding that there are limited scale economies for Federal Reserve currency operations suggests that currency services are not a natural monopoly. As with other paper-based payments technologies, the average cost curve implied by the cost frontier for currency operations is U-shaped with a fairly wide, flat portion at the bottom of the U. Indeed, the Federal Reserve System processes the vast majority of notes at facilities that have near-constant returns to scale. By the end of 1996, 10 Federal Reserve facilities were already operating at constant returns to scale, and the volume of notes processed (figure 1) has continued to increase since that time. Going forward, information on facility-specific marginal costs and returns-to-scale measures could potentially be used to improve resource allocations. For example, such information could be used to set fees for some currency services or to tailor cash service standards that define normal service levels to each depository institution. From a policymaking perspective, the technology for currency handling and processing does not appear to have the declining unit costs that would give rise to a market failure that is sufficient to justify or to sustain a monopoly currency provider. Of course, there may be other reasons for preserving the current arrangement given the Federal Reserve's critical role in supplying and maintaining the integrity of currency in the United States.

Our finding that the average facility operates at more than 80 percent of the efficiency of the "best-practice" facility is comparable to cost-efficiency estimates reported elsewhere for private-sector financial institutions.[42] Just like its for-profit services,[43] the Federal Reserve could potentially reduce costs by having the worst-performing facilities adopt the procedures and operations of the best-performing facilities.

Our finding of significant concordance between cost-efficiency estimates for currency processing and handling services, which are not required by the MCA to recover economic costs in the marketplace, and check processing services, which are, suggests that the Federal Reserve may have realized some of the benefits that a more competitive currency market might have delivered. By encouraging the formation of a competitive market for check services, the MCA also gave Reserve Banks a greater incentive to control costs and improve resource allocation. Thus, the MCA appears to have generated spillover benefits by creating a management culture that increased operational efficiency even for currency services in which the Federal Reserve maintained its monopoly. This spillover benefit from the Federal Reserve's participation in competitive payments markets to services of a more purely governmental nature has previously been ignored in studies that have examined the Federal Reserve's role in the payments system.[44]

■ **41** In 2000, for example, the Federal Reserve spent about $133 million on high-speed currency operations; $4.6 million on currency cancellation, verification, and destruction; $65 million on paying and receiving activities; and $456 million for printing of new Federal Reserve notes, shipment of new notes by the Bureau of Printing and Engraving (BPE), intra-System shipments of fit notes, counterfeit deterrence research, the return of currency pallets to the BPE, and reimbursement to the U.S. Treasury Office of Currency Standards. Federal Reserve outlays do not include the costs of currency operations borne by depository institutions.

■ **42** See, for example, Bauer, Berger, and Humphrey (1993) and Berger, Hancock, and Humphrey (1993).

■ **43** See Bauer and Hancock (1993) and Bauer and Ferrier (1996).

■ **44** See, for example, Green and Todd (2001).

References

Bauer, Paul W., and Gary D. Ferrier. "Scale Economies, Cost Efficiencies, and Technological Change in Federal Reserve Payments Processing," *Journal of Money, Credit, and Banking*, Part 2, vol. 28, no. 4 (November 1996), pp. 1004–39.

Bauer, Paul W., and Diana Hancock. "The Efficiency of the Federal Reserve in Providing Check Processing Services," *Journal of Banking and Finance*, vol. 17, no. 2–3 (April 1993), pp. 287–311.

Bauer, Paul W., Allen N. Berger, and David B. Humphrey. "Efficiency and Productivity Growth in U.S. Banking," in Harold O. Fried, C.A. Knox Lovell, and Shelton S. Schmidt (eds.), *The Measurement of Productive Efficiency: Techniques and Applications.* New York: Oxford University Press, 1993, pp. 386–413.

Bauer, Paul W., Gary D. Ferrier, and C. Knox Lovell. "A Technique for Estimating a Cost System that Allows for Inefficiency," Federal Reserve Bank of Cleveland, Working Paper no. 87–04, 1987.

Bauer, Paul W., Apostolos Burnetas, Viswanath CVSA, and Gregory Reynolds. "Optimal Use of Scale Economies in the Federal Reserve's Currency Infrastructure," Federal Reserve Bank of Cleveland, *Economic Review*, vol. 36, no. 3 (Quarter 3, 2000), pp. 13–27.

Berger, Allen N. "'Distribution-free' Estimates of Efficiency in the U.S. Banking Industry and Tests of the Standard Distributional Assumptions," *Journal of Productivity Analysis*, vol. 4, no. 3 (September 1993), pp. 261–92.

Berger, Allen N., Diana Hancock, and David B. Humphrey. "Bank Efficiency Derived from the Profit Function," *Journal of Banking and Finance*, vol. 17, no. 2–3 (April 1993), pp. 317–48.

Berger, Allen N., John H. Leusner, and John J. Mingo. "The Efficiency of Bank Branches," *Journal of Monetary Economics*, vol. 40, no. 1 (September 1997), pp. 141–62.

Booth, George. *Currency and Coin Responsibilities of the Federal Reserve: A Historical Perspective.* Cleveland, Oh.: Federal Reserve Bank of Cleveland, 1989.

Board of Governors of the Federal Reserve System. *Annual Report: Budget Review.* Washington, D.C.: Board of Governors, 2000.

Board of Governors of the Federal Reserve System. *A Comprehensive Assessment of U.S. Currency Quality, Age, and Cost Relationships.* Washington, D.C.: Board of Governors, September 1991.

Board of Governors of the Federal Reserve System. *Eight Annual Report of the Federal Reserve Board*, 1921. Washington, D.C.: Board of Governors, 1922.

Dotsey, Michael. "Cost Function for Federal Reserve Currency Handling," in *A Comprehensive Assessment of U.S. Currency Quality, Age, and Cost Relationships.* Washington, D.C.: Board of Governors of the Federal Reserve System, September 1991.

Fama, Eugene F. "Commodity and Fiduciary Standards," unpublished manuscript, December 1981.

Fama, Eugene F. "Financial Intermediation and Price Level Control," *Journal of Monetary Economics*, vol. 12, no. 1 (July 1983), pp. 7–28.

Ferrier, Gary D., and C.A. Knox Lovell. "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence," *Journal of Econometrics*, vol. 46, no. 1–2 (October–November 1990), pp. 229–45.

Gallant, A. Ronald. "Unbiased Determination of Production Technologies," *Journal of Econometrics*, vol. 20, no. 2, (November 1982), pp. 285–323.

Green, Edward J., and Richard B. Todd. "Thoughts on the Fed's Role in the Payments System," Federal Reserve Bank of Minneapolis, *2000 Annual Report*, April 2001, pp. 1–25.

Hancock, Diana, David B. Humphrey, and **James A. Wilcox.** "Cost Reductions in Electronic Payments: The Roles of Consolidation, Economies of Scale, and Technical Change," *Journal of Banking and Finance*, vol. 23, no. 2–4 (February 1999), pp. 391–421.

Humphrey, David B. "Economies to Scale in Federal Reserve Check Processing Operation," *Journal of Econometrics*, vol. 15, no. 1 (January 1981), pp. 155–73.

Jondrow, James, C.A. Knox Lovell, Ivan Materov, and Peter Schmidt. "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model," *Journal of Econometrics*, vol. 19, no. 2–3 (August 1982) pp. 233–38.

King, Robert G. "On the Economics of Private Money," *Journal of Monetary Economics*, vol. 12, no. 1 (July 1983), 127–58.

Lacker, Jeffrey M. "Stored Value Cards: Costly Substitutes for Government Currency," Federal Reserve Bank of Richmond, *Economic Quarterly*, vol. 82, no. 3 (Summer 1996), 1–25.

Means, R.S. *Means Square Foot Costs.* 17th edition. Kingston, Mass.: R.S. Means Company, 1996.

Miller, Merton H., and Daniel Orr. "A Model of the Demand for Money by Firms," *Quarterly Journal of Economics*, vol. 80, no. 3 (August 1966), 413–35.

Proxmire, William. "Remarks on the proposed Monetary Control Act," *Congressional Record* (daily edition), March 27, 1980, p. S3168.

Schmidt, Peter, and Robin C. Sickels. "Production Frontiers and Panel Data," Journal of *Business and Economic Statistics*, vol. 2 (1984), pp. 367–74.

Sumner, Scott. "Privatizing the Mint," *Journal of Money, Credit, and Banking*, vol. 25, no. 1 (February 1993), 13–29.

Sumner, Scott. "Is Nonprice Competition in Currency Inefficient? A Comment," *Journal of Money, Credit, and Banking*, vol. 32, no. 1 (February 2000), 146–49.

Supel, Thomas M., and Richard M. Todd. "Should Currency be Priced Like Cars?" Federal Reserve Bank of Minneapolis, *Quarterly Review*, vol. 8, no. 2, (Spring 1984), pp. 3–14.

U.S. Department of Commerce, Office of Business Economics. *Survey of Current Business*, Washington, D.C.: U.S. Government Printing Office, 1990–97.

U.S. Department of Labor, Bureau of Labor Statistics. *Producer Price Index Detailed Report*, Washington, D.C.: Government Printing Office, 1990–97.

U.S. Department of the Treasury. *Circular No.55.* Washington, D.C.: Government Printing Office, 1920.

White, Lawrence H. "Accounting for Non-interest Bearing Currency: A Critique of the Legal Restrictions Theory of Money," *Journal of Money, Credit and Banking*, vol. 19, no. 4 (November 1987), pp. 448–56.

White, Lawrence H., and Donald J. Boudreaux. "Is Nonprice Competition in Currency Inefficient?" *Journal of Money, Credit, and Banking*, vol. 30, no. 2 (May 1998), 252–60.

White, Lawrence H., and Donald J. Boudreaux. "Is Nonprice Competition in Currency Inefficient? A Reply," *Journal of Money, Credit, and Banking*, vol. 32, no. 1 (February 2000), 150–3.

Zimmerman, Gary C. "The Pricing of Federal Reserve Services under MCA," Federal Reserve Bank of San Francisco, *Economic Review* (Winter 1981), pp. 22–40.

The Employment of Nations—A Primer

by Richard Rogerson

Richard Rogerson is the Rondthaler Professor of Economics at Arizona State University and a research associate at the Federal Reserve Bank of Cleveland. He thanks Nathan Porter and Yasuo Terajima for research assistance and the National Science Foundation for financial support.

Introduction

One of the striking macroeconomic developments of the last 30 years has been the marked rise in European unemployment in comparison to that in the United States. As of 2000, U.S. unemployment was basically unchanged from 1970 at around 4 percent, whereas over the same period, the unemployment rate in the European Union almost tripled, increasing from around 3 percent to almost 9 percent. Not surprisingly, a relatively large literature has emerged that both documents various aspects of this differential evolution of unemployment rates and tries to account for it.[1]

The premise of this paper is that our understanding of this and other related phenomena will likely be facilitated by placing them in a broader context. In particular, since economics is often defined as the study of how scarce resources are allocated, I follow the standard economic approach of approaching the labor market from the perspective of documenting differences in resource allocations. Because labor is a key input in the market production of goods and services, it follows that how much time is allocated to the market production of goods and services is likely to be an important feature of the resource allocation achieved by an economy. While economists have long recognized variation in labor input as a (if not *the*) central element in business cycle fluctuations, in contexts other than business cycles, variation in labor input has received much less attention. The objective of this paper, then, is to document the empirical properties of the low-frequency component of labor input in a cross section of industrialized countries over the period 1960–95. This is done using both aggregate data as well as data that are disaggregated by age and sex.

While the larger issue of interest is that of the allocation of time, this study will concern itself almost exclusively with employment patterns. As such, it abstracts from differences in such things as workweeks and vacation days. Obviously, it would be of interest to

■ **1** The literature is too large to reference exhaustively. See, however, Bertola and Ichino (1995), Millard and Mortensen (1997), Ljungqvist and Sargent (1998), Blanchard and Wolfers (2000), and Blanchard (2000).

supplement the analysis provided here with information on these additional aspects of time allocation, but due to data limitations, I do not undertake it in any great detail here. I also restrict attention to a study of industrialized countries, since there are reasons to believe that countries at very different stages of development face considerations which make their time-allocation problems less comparable. Specifically, countries with large rural or agricultural populations may behave quite differently.

The paper presents its findings in the form of nine stylized facts. I will not present them all in this introduction, but a few are worth emphasizing. First, large and persistent differences in employment-to-population ratios are common throughout the period. While differences have become larger since 1970, persistent differences exist throughout the entire sample and are in no way a novel feature of the post-1970 world. Second, countries move about considerably in the distribution of employment-to-population ratios. Some countries move up, and some countries move down. Third, and perhaps most important, a comparison of aggregate and disaggregate data shows that there is substantial variation in disaggregated labor input that is obscured when examining aggregate data. When employment-to-population ratios change in a persistent fashion, the changes are distributed across groups disaggregated by sex and age in a very disproportionate fashion. Because changes across demographic groups display so much variation, it seems natural to think that they contain a great deal of information that will be helpful in sorting out the causes of the aggregate changes.

The paper is unapologetically atheoretical. There is no attempt to discuss what the facts uncovered have to say about various models of the labor market or various explanations for differences in labor market outcomes. Rather, the objective here is to simply lay out the facts that a general theory of low-frequency movements in employment should be able to account for.

An outline of the paper follows. The next section gives a brief description of data sources. Section II examines the aggregate data and presents the main stylized facts that follow from them, while section III does the same for the data disaggregated by age and sex. Section IV discusses some additional measures of labor input, and section V concludes.

I. Data

The measure of labor input that I use in this study is employment to population. For the aggregate analysis, I use the ratio of total employment to the population of individuals between the ages of 15 and 64.[2]

All of the data used in this analysis come from Organisation for Economic Co-operation and Development (OECD) sources, which in turn are based on surveys carried out by individual member countries. Aggregate data on employment relative to the population of individuals aged 15–64 are taken from various issues of the OECD publication *Historical Statistics*. Data on employment rates disaggregated by sex and age are taken from various issues of the OECD publication *Labor Force Statistics*. This publication actually provides information on participation rates and unemployment rates, from which I have computed the employment-to-population ratios. The disaggregated data are not available for as many countries or for as long a time, so the time period and countries analyzed differ for the two exercises. Section IV of the paper, which considers some additional measures of labor input, relies on data that appear in various issues of the OECD publication *Employment Outlook*.

As is true for any study that relies on cross-country data from country-level sources, an important caveat that must be mentioned is the possibility that the data are not strictly comparable. Survey procedures may differ from country to country, as may classification procedures. Additionally, there are occasional changes in the accounting methods for some countries.[3] For examining cross-country differences in employment, these accounting issues do not appear to be very significant. The OECD publications listed above provide documentation of the various country surveys and definitions used to determine employment,

2 Note that this measure includes employed individuals above 65 in the numerator but not in the denominator. It is obviously debatable whether this measure is preferable to the employment-to-population ratio for all individuals or the employment-to-population ratio for all individuals above the age of 15. Since much of the analysis is also carried out with data that are disaggregated by age and the basic findings there are similar, this is probably not an important issue for the purposes of this paper.

3 For the most part, these did not appear to be too serious.

and there is a high degree of uniformity of the criteria. Basically, in order to be counted as employed, an individual has to have either worked at least one hour for pay during the reference period, had a job from which he or she was absent (due to sickness, vacation, strike, weather, and so on), been self-employed, or been an unpaid employee in a family business.[4] Having said this, I will take the employment data at face value, and no additional space will be devoted to this potentially important measurement issue. However, in section IV, when data are presented on hours of work and part-time versus full-time employment, measurement issues are likely to be substantial, and the appropriate caveats will again be raised.

II. Facts about Employment Rates I: The Aggregate Data

This section focuses on patterns found in the aggregate data. I examine 18 countries over the period 1960–95.[5] Because we are interested in low-frequency rather than high-frequency movements, I present five-year averages at five-year intervals.[6] In each case, the indicated year is the center of the five-year period used to construct the average; that is, data for 1985 represent the average for the period 1983–87.[7] Table 1 presents the data on aggregate employment relative to the total population for individuals between the ages of 15 and 64 for 18 countries plus some summary statistics: the mean, standard deviation, and the 85:15 ratio for the cross-sectional distribution in each year. The 85:15 ratio is the ratio of the highest to the lowest value for the employment ratio after having excluded the top two and bottom two values.[8] I report this rather than the ratio of maximum to minimum values in order to downplay the possible role of extreme values. Note also that the mean value reported is the simple mean of the cross-country observations and does not weight countries by their size.

Some of the important features of these data are described next.

Fact 1. The average employment rate remained roughly constant over this period.

As table 1 indicates, there is no evidence of a secular trend in the average employment rate across countries. The average values from 1980 onward are slightly lower than the average values before 1975, but by less than one-half of a percentage point. The average value does fluctuate somewhat over time, and the fluctuations seem larger in the post-1975 period.

Fact 2. Differences in employment rates across countries are large.

No matter which of the cross sections we look at, the variation in employment rates is strikingly large. For the sample as a whole, the maximum values exceed 80 percent, whereas the lowest values are less than 50 percent. The standard deviations are also large, though it can be difficult to gauge what constitutes large in this context. The 85:15 ratio is perhaps more informative in this regard. Note that this ratio exceeds 1.20 for each cross section. To understand the significance of this value, note that in the largest postwar recession in the United States (1982), the ratio of employment to population at the peak was only about 1.03 times larger than it was at the trough. It is particularly significant that the large difference in employment rates across countries is not a phenomenon that emerged after the early 1970s. Cross-country differences were large even in the 1960–70 period.

4 We note that cross-country studies of unemployment are potentially more problematic because of differences in criteria across countries, particularly in what constitutes searching for employment. However, the OECD does publish a series of "standardized" unemployment rates that attempts to correct for cross-country differences in measurement.

5 See Blanchard and Wolfers (2000) for a presentation of the low-frequency movements in unemployment rates over this period.

6 An alternative procedure to isolate the low-frequency component would be to use the trend component generated by applying the Hodrick–Prescott filter to the data. The results of using this alternative procedure are very similar and hence are not reported.

7 The exception to this is 1960. Due to data limitations, 1960 simply refers to the average for 1960–62.

8 More accurately, this is really the 83:17 ratio.

TABLE 1

Aggregate Employment/Total Population, Ages 15–64

	1960	1965	1970	1975	1980	1985	1990	1995
Canada	58.9	60.9	61.3	63.4	65.5	66.1	68.9	67.3
United States	62.2	62.9	64.5	64.4	66.8	68.5	72.4	73.4
Japan	74.3	71.5	71.1	69.7	70.3	70.6	72.5	74.4
Australia	64.6	66.6	68.8	67.5	65.5	64.7	67.6	67.4
New Zealand	62.9	63.7	64.1	64.8	64.2	62.3	66.8	69.2
Belgium	59.5	60.5	61.3	61.1	58.4	54.5	56.6	56.3
Denmark	70.0	72.2	74.3	73.3	73.1	75.3	76.8	74.1
France	68.6	66.4	65.8	65.4	63.5	59.4	60.1	58.9
Germany	70.1	69.6	68.6	65.7	63.6	62.5	65.8	65.2
Ireland	64.4	64.3	62.0	58.0	57.0	51.7	52.7	55.2
Italy	62.6	58.8	56.3	55.6	55.7	54.0	55.0	52.5
Netherlands	61.1	59.9	57.3	54.5	53.6	53.3	61.5	65.3
Norway	63.6	63.0	64.1	68.9	73.8	76.0	74.3	74.2
Portugal	59.2	59.8	62.2	64.5	63.8	64.6	68.9	67.3
Spain	59.9	59.8	60.1	58.1	50.2	45.4	48.9	46.4
Sweden	73.0	71.9	72.7	76.3	78.7	79.3	79.8	70.4
Switzerland	78.2	78.7	77.5	75.3	73.7	74.6	81.7	79.9
United Kingdom	70.8	71.4	70.4	70.8	68.7	65.7	70.7	68.6
Mean	65.7	65.7	65.7	65.4	64.8	63.8	66.7	65.9
Standard deviation	5.81	5.71	5.91	6.38	7.67	9.45	9.30	8.82
85:15 ratio	1.23	1.20	1.21	1.26	1.32	1.41	1.40	1.34

SOURCES: Organisation for Economic Co-operation and Development, *Historical Statistics,* various issues; and author's calculations.

Fact 3. The spread of employment rates across countries changes substantially across time.

Whether one looks at the standard deviation or the 85:15 ratio, the spread of employment rates across countries has exhibited substantial change over time. The distribution became significantly more spread out during the first part of the sample period, and though it subsequently became more compressed, there has still been a significant spreading out over the period as a whole.

Fact 4. Differences in employment rates across countries are persistent.

To see this, I report the correlation matrix for the country-level observations across time in table 2.

TABLE 2

Correlation Matrix for County-Level Observations across Time

	1960	1965	1970	1975	1980	1985	1990	1995
1960	1							
1965	.95	1						
1970	.88	.96	1					
1975	.76	.84	.93	1				
1980	.66	.72	.81	.94	1			
1985	.59	.65	.75	.89	.98	1		
1990	.60	.68	.77	.87	.93	.97	1	
1995	.53	.62	.69	.76	.84	.89	.95	1

SOURCES: Organisation for Economic Co-operation and Development, *Historical Statistics,* various issues; and author's calculations.

Note that at one lag (five years), the average correlation coefficient exceeds 0.95. Obviously, this indicates a great deal of persistence.[9] Note also, however, that the average persistence decreases as the number of lags increases. At four periods (20 years) the average correlation drops to 0.70. It is also interesting to note that these correlations are not driven by changes that occurred before and after the oil shock of the mid-1970s. For example, the correlation between 1980 and subsequent years looks very similar to the pattern of correlations between 1960 and subsequent years. The fact that at longer horizons the correlations decrease considerably suggests our next fact.

Fact 5. There is considerable mobility within the distribution.

There are several ways to motivate this fact. One is to simply look at the changes in individual countries over the entire sample. Table 3 reports changes over three horizons—1960 to 1995, 1960 to 1970, and 1985 to 1995. This split is of interest because the latter two lie on either side of the 1970s' oil shocks.

From the perspective of mobility, the key is dispersion in changes. The dispersion of changes over each period in table 3 is large. Over the 1960–95 period, some countries see their employment rates climb by more than 10 percentage points, while others see their employment rates fall by more than the same amount. Moreover, there is very little correlation between a country's starting value in 1960 and the subsequent change; the correlation is only –0.18. It is important to emphasize that the situation is not characterized as one in which countries simply differ in the extent to which their employment rates decrease—almost half of the countries experience an increase in their employment rates over this period, and as we know from above, there is virtually no secular trend in the average employment rate across these countries.

9 Keep in mind that our observations are themselves five-year averages, which obviously induce some persistence relative to what would be found using annual data.

TABLE 3

Changes in Employment Rates

	Δ (1960–95)	Δ (1960–70)	Δ (1985–95)
Canada	+8.4	+2.4	+1.2
United States	+11.2	+2.3	+4.9
Japan	+0.1	–3.2	+3.8
Australia	+2.8	+4.2	+2.7
New Zealand	+6.3	+1.2	+6.9
Belgium	–3.2	+1.8	+1.8
Denmark	+4.1	+4.3	–1.2
France	–9.7	–2.8	–0.5
Germany	–4.9	–1.5	+2.7
Ireland	–9.2	–2.4	+3.5
Italy	–10.1	–6.3	–1.5
Netherlands	+4.2	–3.8	+12.0
Norway	+10.6	+0.5	–1.8
Portugal	+8.1	+3.0	+2.7
Spain	–13.5	+0.2	+1.0
Sweden	–2.6	–0.3	–8.9
Switzerland	+1.7	–0.7	+5.3
United Kingdom	–2.2	–0.4	+2.9

SOURCES: Organisation for Economic Co-operation and Development, *Historical Statistics,* various issues; and author's calculations.

As suggested earlier, it is significant that there is substantial mobility even abstracting from the 1970–85 period. In the period 1960–70, the range of changes exceeds 10 percentage points, and the same holds true for the period 1985–95 as well. It is also of interest to note the heterogeneity of experiences for countries that experience similar changes over the full sample. Specifically, France, Italy, and Spain all experience drops in their employment rates of around 10 percent. However, Italy experiences over half of this drop in the first 10 years, while France experiences roughly a third in the first decade, and Spain experiences virtually no decline in this period. It should also be noted that there is virtually zero correlation between changes in the first decade and changes in the final decade.

Lastly, to highlight the range of mobility experiences found in the data, consider the evolutions of the Netherlands and Sweden. Consider first the Netherlands. Beginning in 1965, the employment rate in the Netherlands begins a steady decline, losing almost 8 percentage points and bottoming out in 1985. Subsequently, however, the rate increases by more than 10 percentage points, and the Netherlands climbs significantly in the distribution. Next consider Sweden. Between 1960 and 1970, Sweden's employment rate is relatively constant, after which it increases substantially, raising Sweden to the top of the distribution between 1980 and 1990. Subsequently, however, the employment rate falls back to a value very near its level in the 1960–70 period.

TABLE 4

Correlation between Employment and Unemployment Rates

	Trend components	Cyclical components
Canada	.96	–.92
United States	.53	–.92
Japan	–.30	–.60
Australia	–.42	–.90
New Zealand	–.97	–.90
Belgium	–.89	–.76
Denmark	.73	–.88
France	–.99	–.77
Germany	–.87	–.77
Ireland	–.98	–.85
Italy	–.88	–.72
Netherlands	–.32	–.52
Norway	.79	–.81
Portugal	.86	–.77
Spain	–.98	–.89
Sweden	.10	–.93
Switzerland	–.36	–.57
United Kingdom	–.94	–.97

SOURCES: Organisation for Economic Co-operation and Development, *Historical Statistics,* various issues; and author's calculations.

A Digression on Unemployment versus Employment

The previous analysis has focused on employment-to-population ratios as the measure of labor input. As noted earlier, many studies focus instead on unemployment rates. I have argued that from the perspective of understanding differences in resource allocations, it is differences in labor input that are of primary interest. But is the choice really substantive, or is it the case that movements in employment and unemployment are close to mirror images of each other, in which case the issue is largely irrelevant? In this section I present some evidence on the matter. It turns out that while movements in employment and unemployment are highly negatively correlated at cyclical frequencies, this is not necessarily true at lower frequencies. To investigate this I examined the series for employment-to-population ratios used earlier and the standardized unemployment rate series produced by the OECD for the period 1960–95. For each country I apply the Hodrick–Prescott filter to each series and compute the trend and cyclical components. For each country I compare the behavior of the two trend series and the two cyclical series, in each case computing the correlation between the two series. Table 4 displays the correlations.

The two columns of table 4 tell quite a different story. The first column shows that there is really no tendency for trend increases in the employment rate to be associated with either trend increases or decreases in the unemployment rate; while there are many values that are close to negative one, there are also several that are close to positive one, as well as several that are not close to either of the extreme values. Reflecting this, the average of the values in the first column is –0.27. On the other hand, the values in the second column are all negative, are all greater than 0.5 in absolute value, and all have an average value of –0.80.

The message from this is that when studying low-frequency movements in the labor market, one may get a very different picture depending upon the measure one uses. Having said this, it would appear that in at least one context—that of the "European unemployment problem" mentioned in the introduction—the differences are likely to be quantitative rather than qualitative in nature, at least in terms of changes over this period. The main reason for this is that many of the European countries that experienced large decreases in employment over the sample period, including Belgium, France, Germany, Italy, Ireland, and Spain, are countries for which the correlation between employment and unemployment is close to negative one.

III. Facts about Employment Rates II: The Disaggregated Data

This section examines employment rates from the perspective of data that are disaggregated by age and sex. As mentioned earlier, data availability limits the scope of the analysis in terms of time period and the set of countries that can be examined. I note first that using the available data, one can establish the equivalent to facts 2, 4, and 5 for the disaggregated data as well: there are large differences in the rates across countries, and they are persistent, but there is also substantial mobility. Hence, in this section I focus on additional findings that the disaggregated data present us with. A basic issue to explore here is the extent to which the aggregate data capture the differences that exist across countries at a point in time and the changes that take place across countries over time.

We begin by examining the cross sections for 1995 for a set of 15 countries. For the time-series analysis we are restricted to a much smaller set of countries, but it is of interest to examine one cross section for a much larger set. In this case, the data represent a three-year average, taken over the period 1994–96. Table 5 presents the data and several summary statistics. The 80:20 ratio represents the ratio of highest to lowest values after removing the two highest and lowest values for each category.

Several features of the data in table 5 are worth remarking on. First, the basic shape of the life-cycle employment rate profile is the same across all countries for both men and women. The basic shape is that of an inverted U. The peak in most cases occurs for the 35–44 age group, though in some cases it occurs for the 45–54 age group. However, although the basic shape is the same across countries, there are important quantitative differences in the profiles across countries. In particular, the disaggregated data show that the large differences in aggregate data do not reflect an aggregation phenomenon; that is, it is not the case that aggregate differences are accounted for by different age distributions across countries that have very similar life-cycle profiles. Examination of tables 5(a) and 5(b) lead to the following conclusions.

Fact 6.1. Differences in employment rates are much larger for women than for men.

Fact 6.2. Differences in employment rates are much larger for young and old individuals than for prime-aged individuals.

These are pretty much self-evident from table 5. For example, the 80:20 ratios are U-shaped for both men and women, and the values are much higher for women than for men with the lone exception of the over-65 group. It follows that many of the differences in aggregate data are driven by differences among women and men who are not of prime age. Specifically, the tables show that the highest correlation between disaggregate and aggregate values occurs for prime-aged women—the correlation of employment rates for women in the two age groups 35–44 and 45–54 is 0.86. In contrast, the correlation coefficient between aggregate employment rates and those for men aged 35–44 is only 0.33 and 0.41 for those aged 45–54. In fact, excluding workers over the age of 65, the lowest value of this correlation occurs for males aged 35–44. However, even though differences are smallest for prime-aged males, it is worth noting that the data still support the following conclusion.

TABLE 5

Employment Rates over the Life Cycle, 1995

A. Men

	15–24	25–34	35–44	45–54	55–64	65+
Canada	52.2	81.7	84.6	82.8	54.4	10.1
United States	58.0	88.0	88.6	85.7	63.7	16.2
Japan	44.9	94.3	96.1	95.7	80.9	36.4
Korea	27.6	90.5	95.2	92.3	78.8	41.6
New Zealand	61.3	86.2	87.7	87.1	62.9	9.9
Belgium	29.3	86.9	89.0	82.5	34.8	2.3
Denmark	68.5	87.0	88.5	86.1	60.5	4.3
France	22.0	83.7	88.7	86.3	38.7	2.6
Germany	51.4	83.6	90.4	87.4	47.4	4.2
Ireland	37.7	81.1	81.7	77.2	59.0	15.0
Italy	26.4	71.1	88.6	90.2	55.5	6.0
Portugal	42.7	86.4	91.9	86.6	58.9	21.8
Spain	26.6	72.2	83.2	80.3	48.8	2.8
Sweden	41.3	79.0	85.0	86.4	65.0	13.2
United Kingdom	61.2	84.2	86.0	85.8	56.6	7.5
Mean	43.4	83.7	88.3	86.2	57.7	12.9
Standard deviation	15.0	6.2	4.0	4.5	12.5	12.1
Coefficient of variation	.35	.07	.05	.05	.22	.93
80:20 ratio	2.30	1.11	1.09	1.09	1.37	7.79
Correlation with aggregate aged 15–64	.66	.59	.33	.41	.45	.30
Correlation with men aged 35–44	–.06	.68	1	.86	.44	.65

B. Women

	15–24	25–34	35–44	45–54	55–64	65+
Canada	51.6	69.4	71.8	66.4	33.9	3.3
United States	55.2	70.4	73.8	72.2	47.5	8.5
Japan	44.5	57.6	63.9	68.1	47.5	15.5
Korea	39.3	47.6	61.6	59.6	49.9	20.5
New Zealand	55.4	62.0	70.0	73.6	38.3	3.1
Belgium	24.0	69.9	63.1	45.5	12.5	0.9
Denmark	60.7	72.6	80.9	73.7	37.8	1.2
France	18.3	65.2	69.3	67.3	28.6	1.3
Germany	48.6	65.8	67.7	62.2	24.5	1.5
Ireland	34.0	62.1	45.7	36.2	20.3	2.7
Italy	21.6	46.9	51.8	47.4	20.2	1.7
Portugal	33.5	73.0	74.5	61.0	34.1	9.7
Spain	21.4	44.9	43.1	32.0	17.6	1.3
Sweden	41.9	73.9	83.0	85.3	59.9	4.6
United Kingdom	57.5	66.2	71.9	71.7	38.9	3.1
Mean	40.5	63.2	66.1	61.5	34.1	5.3
Standard deviation	14.4	9.8	11.7	15.0	13.6	5.9
Coefficient of variation	.36	.15	.18	.24	.40	1.11
80:20 ratio	2.56	1.53	1.44	1.62	2.35	7.46
Correlation with aggregate aged 15–64	.75	.63	.86	.86	.78	.32
Correlation with females aged 35–44	.60	.79	1	.92	.64	.06

SOURCES: Organisation for Economic Co-operation and Development, *Labor Force Statistics,* various issues; and author's calculations.

Fact 7. Differences in employment rates across countries, even for prime-aged males, are large.

To see this, simply recall the comment made earlier to put the 80:20 ratios in perspective. In postwar time series data for the United States, a value of 1.03 is big. Hence, values of 1.1 must also be viewed as large.

Fact 8. Differences in disaggregated employment rates are not proportional to differences in aggregate data.

Another way to phrase this is to say that when two countries have very different aggregate employment rates, the corresponding life-cycle profiles of employment rates are not simply shifted up or down in a parallel fashion, even controlling for sex. The relatively low correlations between prime-aged employment rates and the rates for other age groups (controlling for sex) indicate this. A few examples serve to illustrate the significance of this point. The three largest economies of continental Europe, Germany, Italy, and France, all have employment rates at the aggregate level (or even for males) that are substantially lower than that of the United States in 1995. However, each of these countries has prime-aged employment rates for men that actually exceed the corresponding value for the United States! Yet, in sharp contrast, for the 55–64 age group the employment rate in the United States is roughly double that in these other countries. Spain has the lowest employment rates in the aggregate level, and this continues to hold when the data are disaggregated by sex. Canada, on the other hand, has a relatively high aggregate employment rate. Yet the employment rates for prime-aged males are roughly similar in these two countries. On the other hand, male youths in Canada are almost twice as likely to be employed as their Spanish counterparts, as is roughly true for females of all age groups in the two countries.

TABLE 6

Changes in Employment Rates: 1972 to 1995

A. Men

	15–24	25–34	35–44	45–54	55–64	65+
Canada	–4.7	–9.9	–8.2	–7.7	–23.3	–7.4
United States	–8.6	–3.1	–4.5	–3.6	–12.1	–5.5
Japan	–10.3	–2.4	–0.9	–0.1	–4.1	–9.7
France	–33.7	–12.4	–9.6	–7.7	–32.4	–16.3
Germany	–18.5	–10.8	–7.1	–6.8	–26.6	–10.2
Italy	–16.0	–23.6	—	–5.0	–13.2	–1.4
Spain	–43.2	–21.1	–13.0	–14.0	–32.3	–22.5
Sweden	–22.8	–12.5	–8.6	–6.6	–16.3	–10.8

A. Women

	15–24	25–34	35–44	45–54	55–64	65+
Canada	6.2	27.2	30.4	25.7	4.4	1.0
United States	7.5	24.4	23.8	18.7	7.3	0.2
Japan	–4.8	13.4	5.6	7.9	3.6	–0.6
France	–25.0	10.0	19.6	16.6	–8.1	–5.6
Germany	–13.3	15.0	17.9	13.3	–1.3	–4.3
Italy	–7.6	15.6	—	17.5	11.3	0.3
Spain	–29.4	16.9	18.5	5.5	–5.7	–6.5
Sweden	–14.9	11.3	12.9	16.5	15.4	–2.9

SOURCES: Organisation for Economic Co-operation and Development, *Labor Force Statistics,* various issues; and author's calculations.

Having examined the differences in life-cycle employment rates from a single cross section, I next turn to a look at the evolution of life-cycle profiles over time and across countries. For this exercise, data availability limits us to a set of eight countries—Canada, the United States, Japan, France, Germany, Italy, Spain, and Sweden, and the period 1972–95. The appendix contains a complete set of tables for the evolution of these life cycles over time. For the sake of illustration I focus on changes between 1972 and 1995. The data for 1995 are a three-year average centered on 1995, whereas the data for 1972 are simply an average of 1972 and 1973.[10] Table 6 presents changes in life-cycle employment profiles, disaggregated by age and sex for eight countries.[11]

A look at table 6 indicates a striking pattern for this period—in all countries there was a huge reallocation of employment away from males and toward females. In the table for males, every entry is negative. In the table for women, the entries are all positive for ages 25–54. A closer look reveals the following fact.

10 Data for 1971 are not available for all countries, so the two-year average is used rather than further reducing the number of countries in the sample. It also seemed preferable to not use 1974 since this marks the beginning of the oil price shocks.

11 Note that there are two missing values for Italy.

TABLE 7

Changes in Normalized Life-Cycle Profiles, 1972–1973 to 1994–1996

A. Men	15–24	25–34	45–54	55–64	Scale	Absolute deviations
Canada	1.00	.97	1.00	.77	.91	.26
United States	.91	1.01	1.01	.88	.95	.23
Japan	.82	.98	1.01	.96	.99	.26
France	.43	.95	1.01	.60	.91	1.03
Germany	.79	.96	1.00	.69	.93	.56
Italy	.69	.83	1.05	.82	.91	.71
Spain	.44	.89	.98	.69	.86	1.00
Sweden	.71	.95	1.02	.88	.91	.48
Standard deviation	.21	.06	.02	.12	.04	

B. Women	15–24	25–34	45–54	55–64	Scale	Absolute deviations
Canada	.66	.95	.94	.66	1.73	.79
United States	.78	1.04	.95	.80	1.48	.51
Japan	.82	1.19	1.03	.99	1.10	.41
France	.30	.85	.95	.56	1.39	1.34
Germany	.58	.95	.94	.70	1.36	.83
Italy	.44	.90	.95	1.36	1.67	1.07
Spain	.24	.92	.69	.43	1.75	1.72
Sweden	.62	1.00	1.05	1.14	1.18	.57
Standard deviation	.21	.11	.11	.31	.24	

SOURCES: Organisation for Economic Co-operation and Development, *Labor Force Statistics,* various issues; and author's calculations.

Fact 9. Changes in aggregate employment rates are associated with large changes in the shape of life-cycle employment-rate profiles.

This fact is really a time-series equivalent to fact 8. There we saw that if two countries have very different aggregate employment rates, the life-cycle profiles of employment rates are not simply parallel transformations of each other. What table 6 shows is that when the employment rate changes in a country, the life-cycle employment-rate profiles do not shift in a parallel fashion. Consider a few examples. Aggregate employment and male employment rates fall significantly over this period in France, Germany, Italy, and Spain. Yet the change is disproportionately accounted for by changes in the employment rates of young and old workers. The changes would look even more skewed if we considered them relative to the starting values—in some countries, youth employment rates are falling by more than half. Similarly, female employment rates are actually increasing over this period in France, Germany, and Spain, yet despite this, there are massive decreases in the employment rates for female youths.

I next present this same information in a manner that can better highlight the relative contribution of parallel movements of the life-cycle profile against changes in the shape of the profile. I adopt the following procedure to normalize the shape of the life-cycle profile. Normalize the employment rate for the 35–44 age group to one, and then express all other values relative to this value. This profile of relative values plus the actual employment rate for the 35–44 group completely describes the whole profile. This procedure is carried out for both the 1972 and 1995 data. Then, to summarize the changes in life-cycle profiles over time, I compute the ratio of the relative values and the ratio of the actual values for the 35–44 age group. I do this for both sexes for each of the countries in our sample, and table 7 presents the results. The column labeled scale indicates the ratio of the employment rate for the 35–44 group in 1995 relative to its value in 1972–73. The columns with age ranges show the ratios of the relative life-cycle employment profiles. The column labeled absolute deviations gives the sum of the absolute deviations of the four life-cycle points from one. Note that if the shape of the profile stayed the same but was shifted up or down proportionately, each of the first four columns would have a value of one, and the sum of absolute deviations would be zero. Hence, the last column is a measure of how much the shape of the profile is changing. Table 7 does not report data for the over-65 age group since its relatively low employment rates make it relatively unimportant from our perspective.

Consider the results for men first. It is striking that the values in the third column (ages 45–54) are so close to one. This implies that the shape of the life-cycle employment profile changes very little for prime-aged males. For the other age groups, the changes are much larger. Note that the scale factor is less than one for each country. Note also that the vast majority of the values in the table are less than one as well. Comparing the standard deviations of the various columns gives us a way to ascertain the extent to which the changes in scale dominate the changes in shape. For three of the four age groups, the standard deviation of the shape ratios exceeds the standard deviation of the scale factors.

Now consider the case of women. First note the magnitude of the scale factors—they are all much larger than one, indicating that in all countries the employment rate for women aged 35–44 increased. However, the rest of the entries are typically less than one, indicating that although the employment rates were increasing for most age groups in most countries, the increase was less than proportionate to the increase for prime-aged women. Once again, however, with the exception of Spain, the entries for the 45–54 age group are fairly close to one. Interestingly, however, for women the standard deviation in scale factors exceeds that of three of the four changes in shape. This is the opposite of what was found for men.

Note the asymmetry of the changes for men and women. For men, the changes for prime-aged workers are much smaller proportionately than are the changes for other age groups. For women, the opposite is true. Changes are proportionately largest for prime-aged women. An important development is that in 1994–96, women's life-cycle profiles increasingly resemble those of their male counterparts, whereas in the 1972–73 period, women's profiles tended to be relatively flat.

A Closer Look at Sweden and the Netherlands

To illustrate the range of experiences that exists across countries, we now take a closer look at Sweden and the Netherlands for the period 1972–96. Recall from the analysis of aggregate data that these two countries follow quite different paths over this period. Sweden experiences a significant increase in its employment rate over the first part of the period and then witnesses a decline at the end that brings it back roughly to where it began. In contrast, the Netherlands experiences a large decrease in its employment rate over the first part but subsequently experiences a large increase and ultimately ends the period with a higher employment rate. What happened to the life-cycle employment profiles for these countries over this period? That is the issue I turn to next. I examine the profiles at the dates 1972, 1980, 1985, 1990, and 1995. In each case, the data represent a three-year average with the given year as the midpoint of the period, except for the initial point, which is a two-year average based on 1972–73. Data limitations require that we focus on three age groups: 15–24, 25–54, and 55–64. For each year, I present normalized profiles that give the relative employment rates for the two extreme age groups relative to the 25–54 age group. I also report a scale factor that gives the employment rate of the 25–54 group relative to its value in 1972. Table 8 gives the results. These data are aggregated for men and women.

TABLE 8

Changing Profiles in Sweden and the Netherlands

A. Sweden

	ER 15–24 / *ER* 25–54	*ER* 55–64 / *ER* 25–54	Scale
1972	.76	.79	1.00
1980	.75	.75	1.10
1985	.68	.72	1.13
1990	.71	.76	1.14
1995	.50	.76	1.03

B. Netherlands

	ER 15–24 / *ER* 25–54	*ER* 55–64 / *ER* 25–54	Scale
1972	.93	.77	1.00
1980	.71	.56	1.06
1985	.64	.44	1.04
1990	.76	.41	1.21
1995	.75	.40	1.28

SOURCES: Organisation for Economic Co-operation and Development, *Labor Force Statistics,* various issues; and author's calculations.

The differences across the two countries are rather striking. First consider the case of Sweden. As the scale column indicates, the employment rate for prime-aged workers mimics the behavior of the aggregate employment rate: a substantial increase between 1972 and 1980, relative constancy over the 1980s, and then a decline back to its earlier value. Moreover, until the final year, the relative life-cycle profile of employment rates changes little. However, when the aggregate employment rate falls in 1995, the shape of the profile changes a good deal, as the employment rate for young workers falls disproportionately. Next, consider the case of the Netherlands. What is most striking in the Netherlands is that the behavior of the prime-aged employment rate does not really mimic at all the behavior of the aggregate employment rate. Whereas the aggregate employment rate drops continuously between 1970 and 1985, the prime-aged-employment rate experiences a mild increase between 1972 and 1985. Subsequently, it does mirror the large increase between 1985 and 1995 found in the aggregate rate. In contrast to the case of Sweden, however, the shape of the life-cycle employment profile changes substantially. Over the period 1972–85, there are massive relative downward shifts for both young and old workers. Employment rates for these groups fall relative to those for prime-aged individuals by roughly a third. Moreover, over the period 1985–95, although the employment rate for young individuals partially recovers, the relative rate for older individuals falls slightly.

A Closer Look at Prime-Aged Males

Prime-aged males are a group that attracts considerable attention in cross-country comparisons of labor market outcomes. One reason for this interest is that they correspond to a group whose main activity is presumed to be market work. They are too old to be in school, too young to be retired, and, even though social norms are changing, they typically do not have primary responsibility for child care or other family situations. Although fact 7 explicitly deals with differences among prime-aged males in the 1995 cross section (noting

TABLE 9

Changes in Employment Rates: 1972 to 1995

Males 35–44	1972	1980	1985	1990	1995
Canada	93.0	91.9	88.2	88.0	84.6
United States	93.1	91.8	90.3	90.2	88.6
Japan	97.0	96.5	95.7	96.6	96.1
France	97.3	95.2	92.9	92.1	88.7
Germany	97.5	95.9	92.6	91.4	90.4
Italy	—	97.1	95.0	93.1	88.6
Spain	96.2	90.6	85.0	88.9	83.2
Sweden	93.6	95.6	95.1	94.7	85.0
Mean	95.4	94.3	91.9	91.9	88.2
90:10 ratio	1.05	1.05	1.08	1.07	1.07

SOURCES: Organisation for Economic Co-operation and Development, *Labor Force Statistics,* various issues; and author's calculations.

that differences across countries are large), it is of interest to look at the evolution of this group in our panel of eight countries. Table 9, which also appears in the appendix, shows the employment-to-population ratios for males aged 35–44 over the period 1972–95.

Three points are worth making about the data in table 9. First, as already mentioned, one of the striking facts for this age group is the negative secular trend in all countries, even for those in which aggregate employment rose during the period. Second, while it is true that cross-country differences among other demographic groups tend to be much larger, the differences for this group are still substantial. Third, the patterns found for this demographic group do not reflect the patterns in aggregate data. For example, in table 9, Germany appears to be a high-employment country and Canada appears to be a country with low employment, but aggregate data suggest just the opposite.

IV. Other Measures of Labor Input

This paper began by arguing that it is of interest to understand differences in labor input, both across time within a given economy and at a given point in time across different economies.

Until now, however, all of the data analysis has pertained to employment ratios. In fact, employment is but one, albeit important, component of labor input. Other components include hours of actual work per employed person and work effort. Differences in actual hours of work per employed person can be further subdivided by differences in normal weekly hours of work, overtime hours, the extent of multiple job holding, or paid vacation and sick days. While there seem to be no attempts to officially document work effort by national statistical agencies, most countries do attempt to measure hours of work.

Unfortunately, strictly comparable cross-country time-series data on hours of work per employed person do not exist. Differences in procedures across countries with regard to such matters as whether the category "hours of work" refers to hours paid versus hours actually worked limit the appropriateness of cross-country comparisons. Having made this qualification, I present in table 10 data reported by

TABLE 10

Annual Hours of Work per Employed Worker

	1970	1975	1979	1983	1990	1996	h_{1970}/h_{1996}
Canada	1890	1837	1832	1780	1788	1784	1.06
United States	1889	1832	1845	1808	1819	1839	1.03
Japan	2201	2112	2126	2095	2031	1892	1.16
France	1962	1865	1806	1712	1657	1608	1.22
Germany	1949	1801	1696	1657	1598	1511	1.29
Italy	1969	1841	1722	1699	1674	1636	1.20
Norway	1766	1653	1514	1485	1432	1407	1.25
Spain	—	—	2022	1912	1824	1810	—
Sweden	1641	1516	1516	1518	1546	1623	1.01
Mean	1908	1807	1757	1719	1693	1663	1.15
Standard deviation	163.1	172.2	197.8	189.9	185.0	165.6	.108
90:10 ratio	1.11	1.13	1.22	1.19	1.18	1.22	1.22

SOURCES: Organisation for Co-operation and Economic Development; *Employment Outlook,* various issues; and author's calculations.

the OECD for annual hours of work for a set of nine countries over the period 1970–96.[12] All of the summary statistics are for the set of eight countries not including Spain, since values for Spain are not available before 1979. I chose to include the available data for Spain because in our earlier comparisons Spain was typically the country with the lowest labor input as measured by employment. However, in table 10 we see that as of 1996, annual hours per employed worker in Spain were some 10–15 percent higher than their corresponding values in France, Germany, and Italy. Assuming that this comparison is appropriate, it indicates the care that one must take in extrapolating from cross-country differences in employment (and hence unemployment) to differences in labor input.

There are several patterns worth noting in table 10. First, in all countries the tendency since 1970 has been for annual hours of work per employed person to decrease, though the range of decreases is very large. Second, as already discussed, comparisons at a point in time across countries may be misleading due to differences in how the data are collected. However, assuming that the effects of these differences are roughly constant over time, the data suggest very large relative movements in labor input across countries. As the last column in the table indicates, annual hours per worker fell in Germany by more than 25 percent relative to the United States.[13] One issue to keep in mind when interpreting these differences is the fact that in all countries there is a tendency for the workweek in manufacturing to decrease as a country becomes richer. While this decrease for the United States occurred prior to the 1970s, in many other countries it occurred after 1970.

■ **12** These data are taken from various issues of the OECD *Employment Outlook.* The OECD reports these measures with a strong warning that they should not be used for cross-country comparisons at a point in time. The value for Sweden in 1996 is not directly comparable to the earlier values because of survey changes. It seems that the effect of the change is to increase the value of hours worked in 1996 by around 3 percent.

■ **13** Although cross-country, point-in-time comparisons are not recommended with these data, there are some indications that such comparisons may be meaningful in some cases. In their study of the auto industry, Fuss and Waverman (1992) document differences in annual hours of work per auto-industry employee in Canada, the United States, Japan, and Germany between 1961 and 1981. In 1961, the values were Canada 1970, United States 2042, Japan 2495, and Germany 2007. In 1981 the values were Canada 1857, United States 1923, Japan 2200, and Germany 1602.

TABLE 11

Full- and Part-Time Employment Ratios (1970 and 1995)

	Full-time		Part-time	
	1970	**1995**	**1970**	**1995**
Canada	55.4	54.4	6.0	12.9
United States	54.4	59.4	10.1	13.0
Japan	61.2	53.3	9.9	16.2
Australia	60.6	49.0	8.2	17.2
New Zealand	56.9	45.3	7.2	13.1
Belgium	59.0	47.2	2.3	9.1
France	61.9	50.1	3.9	8.8
Germany	61.7	54.4	6.9	10.2
Italy	52.7	45.1	3.6	5.8
United Kingdom	59.1	53.5	11.3	15.9
Mean	58.3	51.2	6.9	12.2
90:10 ratio	1.13	1.20	2.81	1.84

SOURCES: Organisation for Co-operation and Economic Development; *Employment Outlook*, various issues; and author's calculations.

How do these changes in hours correlate with the changes in employment ratios documented earlier? The correlation between the change in hours between 1970 and 1996 and the change in aggregate employment-to-population ratios between 1970 and 1995 for this set of eight countries is 0.36. Norway is somewhat of an outlier in this regard, and if Norway is excluded, the correlation increases to 0.77. This suggests that the relative changes in aggregate labor input are probably substantially larger than are the relative changes in employment-to-population ratios.

A closely related issue that often comes up in this context is cross-country differences in the extent of part-time employment.[14] The measures of hours presented above do include part-time employment, and hence do control for these differences. However, it may also be of interest to directly examine data on full- versus part-time employment. Here again, however, a major caveat is necessary since cross-country measures are not directly comparable due to differences in definitions of part-time employment.

As a crude attempt to decompose the previously examined employment-to-population ratios into full-time and part-time components so that I can compare relative changes in the proportion of the population employed full- and part-time, I compute full- and part-time employment-to-population ratios for 1970 and 1995. I use the data on the fraction of employment that is part-time in 1973 and 1997 in conjunction with the earlier data on employment-to-population ratios for individuals aged 15–64 for the years 1970 and 1995. Although the years do not match exactly, to the extent that changes in part-time versus full-time employment patterns have been occurring gradually through time, this comparison should give a good idea of the changes over the 1970–1995 period.

The values are in table 11.[15]

14 In particular, some researchers have argued that the dramatic improvement in employment in the Netherlands is entirely due to increases in part-time employment. See, for example, Nickel and van Ours (2000) and its discussion.

15 The 90:10 ratio in this table refers to the ratio of the second-highest value to the second-lowest value.

A few patterns emerge. First, note that in all countries the employment-to-population ratio has increased for part-time workers, and in every country except the United States the employment-to-population ratio for full-time employment has decreased. Second, it is in no way true that differences across countries in aggregate employment-to-population ratios are dominated by differences in the extent of part-time work. The 90:10 ratio for full-time employment exceeds 1.10 for both years. Moreover, the value increases over time, suggesting that differences in full-time-employment ratios have become larger. In contrast, differences in part-time-employment ratios have actually narrowed slightly, though based on the 90:10 measure, differences in part-time-employment ratios still exceed those for full-time-employment ratios. Third, from the perspective of the "European unemployment problem," the full-time-employment ratios suggest even larger relative changes than do the aggregate numbers, since in 1970, part-time work was much less prevalent in Europe but has since become increasingly common there relative to the United States.

V. Conclusion

This paper has studied the empirical properties of aggregate and disaggregate employment in a cross section of developed countries over the period 1960–95. It has documented several facts that a successful theory of employment should be able to account for. Though much information is presented here, there is much additional information that would give a richer picture of differences in time allocations across countries. First, we need better cross-country measures of hours of actual work, especially at a disaggregated level. Second, additional disaggregations would be useful, especially by family structure. Third, it is important to understand how people spend their time when not working in the market—are they taking care of other family members, in school, or what?

Three conclusions seem to bear repeating. First, large and persistent differences in employment ratios across countries seem to be pervasive. In particular, although the relative changes in unemployment between the United States and Europe over the last 30 years have been dramatic, one should not be misled into thinking that in this regard the world is dramatically different after 1970 than before. From the perspective of employment-to-population ratios, there are large cross-country differences before 1970 as well as after, and there were substantial changes in cross-country relative employment ratios in the period before 1970 as well as after. Second, the changes found in data disaggregated by sex and age do not at all mirror the changes found in aggregate data. Changes tend to be concentrated among the young and the old and among women. Any successful theory of cross-country changes in employment must successfully account for this concentration. Third, the patterns found in aggregate data for a cross section of countries do not carry over to all demographic groups within those countries.

Appendix—Disaggregated Changes in Employment Rates

This appendix presents the complete set of tables for disaggregated changes in employment rates between 1972 and 1995 for a set of eight countries. The data for 1972 represent an average for 1972 and 1973, whereas for each other year, the data represent a three-year average centered on the given year; for example, for 1980 the data are an average of 1979–81. The 90:10 ratio is the ratio of the second-highest to the second-lowest value.

TABLE 12

Disaggregated Changes in Employment Rates

A. Males 15–24

	1972	1980	1985	1990	1995
Canada	56.9	62.5	58.3	60.7	52.9
United States	66.8	64.0	62.7	62.8	60.9
Japan	55.2	41.6	40.7	41.6	45.3
France	55.7	46.3	38.0	33.6	25.5
Germany	69.9	59.4	57.0	59.5	54.0
Italy	42.2	38.7	34.3	33.8	31.3
Spain	69.8	54.0	40.4	42.5	31.4
Sweden	64.1	66.8	61.4	64.4	40.8
Mean	60.1	54.2	49.1	49.9	42.8
90:10 ratio	1.26	1.54	1.62	1.86	1.73

B. Males 25–34

	1972	1980	1985	1990	1995
Canada	91.6	89.7	83.9	84.8	81.6
United States	91.6	89.5	88.1	88.6	88.0
Japan	96.7	95.2	94.4	95.4	94.3
France	96.1	93.2	88.6	88.4	83.7
Germany	94.4	88.8	82.0	85.0	83.6
Italy	94.7	87.6	82.5	79.2	71.1
Spain	94.3	85.8	76.3	81.4	72.2
Sweden	91.5	92.8	91.1	90.5	79.0
Mean	93.9	90.3	85.7	86.7	81.7
90:10 ratio	1.05	1.06	1.11	1.11	1.22

C. Males 35–44

	1972	1980	1985	1990	1995
Canada	93.0	91.9	88.2	88.0	84.6
United States	93.1	91.8	90.3	90.2	88.6
Japan	97.0	96.5	95.7	96.6	96.1
France	97.3	95.2	92.9	92.1	88.7
Germany	97.5	95.9	92.6	91.4	90.4
Italy	—	97.1	95.0	93.1	88.6
Spain	96.2	90.6	85.0	88.9	83.2
Sweden	93.6	95.6	95.1	94.7	85.0
Mean	95.4	94.3	91.9	91.9	88.2
90:10 ratio	1.05	1.05	1.08	1.07	1.07

D. Males 45–54

	1972	1980	1985	1990	1995
Canada	90.5	88.8	85.2	85.8	82.8
United States	89.3	88.2	86.9	87.2	85.7
Japan	95.9	95.0	94.5	95.7	95.7
France	94.0	91.6	88.7	88.2	86.3
Germany	94.2	92.4	90.7	88.7	87.4
Italy	95.2	96.3	94.9	94.2	90.2
Spain	94.3	87.3	80.6	84.9	80.3
Sweden	93.0	93.5	93.4	93.3	86.4
Mean	93.3	91.6	89.4	89.8	86.9
90:10 ratio	1.05	1.07	1.11	1.10	1.09

E. Males 54–64

	1972	1980	1985	1990	1995
Canada	77.7	72.5	64.5	60.0	54.4
United States	75.8	69.5	64.9	64.7	63.7
Japan	85.0	81.7	79.0	80.5	80.9
France	71.1	64.5	46.5	42.9	38.7
Germany	74.0	63.4	54.4	51.3	47.4
Italy	—	72.0	65.6	63.1	55.5
Spain	82.1	71.6	59.2	56.7	48.8
Sweden	81.3	77.2	73.2	73.8	65.0
Mean	78.2	71.6	63.4	61.6	56.8
90:10 ratio	1.11	1.20	1.35	1.44	1.37

F. Males 65+

	1972	1980	1985	1990	1995
Canada	17.5	14.5	12.1	11.0	10.1
United States	21.7	18.5	15.5	15.7	16.2
Japan	46.1	40.1	36.2	36.3	36.4
France	16.3	8.2	5.3	3.8	2.6
Germany	14.4	6.8	5.0	4.4	4.2
Italy	7.4	7.9	5.3	5.2	6.0
Spain	25.3	12.3	6.1	3.9	2.8
Sweden	24.0	13.7	11.6	13.5	13.2
Mean	21.6	15.3	12.1	11.7	11.4
90:10 ratio	1.76	2.32	2.92	4.03	5.79

G. Females 15–24

	1972	1980	1985	1990	1995
Canada	45.4	55.2	55.9	59.1	51.2
United States	47.7	54.0	55.3	56.0	55.2
Japan	49.3	42.5	41.6	43.0	44.5
France	43.3	33.9	28.0	25.1	18.3
Germany	61.9	53.2	50.6	54.2	48.6
Italy	29.3	28.1	24.1	25.1	21.6
Spain	50.8	34.2	22.4	28.3	21.4
Sweden	56.8	65.2	61.8	64.8	41.9
Mean	48.1	45.8	42.5	44.5	37.8
90:10 ratio	1.31	1.63	2.32	2.35	2.39

H. Females 25–34

	1972	1980	1985	1990	1995
Canada	42.2	58.8	64.1	70.2	69.4
United States	46.0	60.7	65.6	69.0	70.4
Japan	44.2	47.2	50.2	54.8	57.6
France	55.2	63.5	64.4	65.7	65.2
Germany	50.8	57.5	56.3	62.5	65.8
Italy	31.3	47.1	47.3	50.0	46.9
Spain	28.0	32.2	36.3	44.3	44.9
Sweden	62.3	79.4	85.4	86.3	73.9
Mean	45.0	55.8	58.7	62.9	61.8
90:10 ratio	1.76	1.34	1.39	1.40	1.50

I. Females 35–44

	1972	1980	1985	1990	1995
Canada	41.4	58.1	64.5	72.2	71.8
United States	50.0	61.9	67.8	73.0	73.8
Japan	58.3	60.1	62.5	65.1	63.9
France	49.7	60.4	64.7	67.4	69.3
Germany	49.8	55.3	57.6	64.3	67.7
Italy	—	43.1	48.8	52.8	51.8
Spain	24.6	27.5	28.3	37.1	43.1
Sweden	70.1	83.5	88.9	91.4	83.0
Mean	49.1	56.2	60.4	65.4	65.6
90:10 ratio	1.41	1.44	1.39	1.38	1.42

J. Females 45–54

	1972	1980	1985	1990	1995
Canada	40.7	50.7	55.6	64.1	66.4
United States	51.5	57.2	61.3	68.6	72.2
Japan	60.2	61.0	63.4	67.6	68.1
France	50.7	54.5	58.0	60.9	67.3
Germany	48.9	49.6	50.3	57.6	62.2
Italy	29.9	36.8	40.5	45.6	47.4
Spain	26.5	26.5	24.4	28.1	32.0
Sweden	68.8	82.2	87.0	89.4	85.3
Mean	47.2	52.3	55.1	60.2	62.6
90:10 ratio	2.01	1.66	1.57	1.50	1.52

K. Females 55–64

	1972	1980	1985	1990	1995
Canada	29.5	32.2	31.1	32.9	33.9
United States	40.2	40.0	40.3	43.8	47.5
Japan	43.9	44.6	43.5	46.5	47.5
France	36.7	36.8	28.6	28.5	28.6
Germany	25.8	26.9	21.7	22.2	24.5
Italy	8.9	22.0	20.7	21.3	20.2
Spain	23.3	20.9	18.5	18.1	17.6
Sweden	44.5	54.7	57.6	64.3	60.0
Mean	34.8	34.8	32.8	34.7	35.0
90:10 ratio	1.70	2.03	2.10	2.18	2.35

L. Females 65+

	1972	1980	1985	1990	1995
Canada	4.3	4.3	4.0	3.7	3.3
United States	8.3	7.9	7.2	8.2	8.5
Japan	16.1	15.5	15.4	16.2	15.5
France	6.9	3.6	2.2	1.6	1.3
Germany	5.8	3.2	2.3	1.9	1.5
Italy	1.5	2.7	2.2	2.1	1.7
Spain	7.7	4.1	2.4	1.6	1.3
Sweden	7.5	3.9	3.3	5.3	4.6
Mean	7.3	5.7	4.9	5.1	4.7
90:10 ratio	1.93	2.47	3.27	5.12	6.53

SOURCES: Organisation for Economic Co-operation and Development, *Labor Force Statistics,* various issues; and author's calculations.

References

Bertola, Giuseppe, and Andrea Ichino. "Wage Inequality and Unemployment: The United States versus Europe," *NBER Macroeconomics Annual,* Cambridge: MIT Press, 1995, pp. 13–54.

Blanchard, Olivier. *The Lionel Robbins Lectures.* Presented at the London School of Economics, 2000.

Blanchard, Olivier, and Justin Wolfers. "The Role of Shocks and Institutions in the Rise of European Unemployment," *Economic Journal,* vol. 110, no. 462 (March 2000), pp. 1–33.

Ljungqvist, Lars, and Thomas J. Sargent. "The European Unemployment Dilemma," *Journal of Political Economy,* vol.106, no. 3 (June 1998), pp. 514–550.

Millard, Stephen P., and Dale T. Mortensen. "The Unemployment and Welfare Effects of Labour Market Policy: A Comparison of the USA and the UK," in *Unemployment Policy: How Should Governments Respond to Unemployment,* edited by Dennis Snower and Guillermo de la Dehesa, Oxford: Oxford University Press, 1997.

Mortensen, Dale T., and Chris A. Pissarides. "Unemployment Responses to 'Skill Biased' Technology Shocks: The Role of Labour Market Policy," *Economic Journal,* vol. 109, no. 455 (April 1999), pp. 242–65.

Nickell, Steve, and Jan van Ours. "The Netherlands and the United Kingdom: a European Unemployment Miracle?" *Economic Policy,* vol. 0, no. 30 (April 2000), pp. 135–80.

OECD Employment Outlook, various issues, Organisation for Economic Co-operation and Development.

OECD Historical Statistics, various issues, Organisation for Economic Co-operation and Development.

OECD Labor Force Statistics, various issues, Organisation for Economic Co-operation and Development.

Monetary Policy and Asset Prices with Imperfect Credit Markets

by Charles T. Carlstrom and Timothy S. Fuerst

Charles T. Carlstrom is an economic advisor at the Federal Reserve Bank of Cleveland. Timothy S. Fuerst is the Owens-Illinois Professor in the Department of Economics at Bowling Green State University and a research associate at the Federal Reserve Bank of Cleveland.

Introduction

In a world with perfect capital markets (the world of the Modigliani–Miller theorem), a firm's financial position—that is, its debt versus its equity level—is irrelevant to its decisions on production and investment. The reason is that perfect capital markets let information flow freely. If an entrepreneur has a good idea for a new product, she will be able to produce it regardless of her personal financial position because outside investors, well informed and readily perceiving an attractive profit opportunity, will provide whatever financing is needed.

The supposition that an entrepreneur's financial position is irrelevant has important implications for monetary policy. A worthy production activity will be funded whatever the entrepreneur's finances may be. Therefore, monetary policy need not respond to asset prices.[1] The Modigliani–Miller theorem, however, is not necessarily meant to be a description of reality. On the contrary, a voluminous empirical literature supplies evidence that a firm's financial position *does* affect its ability to operate. But does this departure from the Modigliani–Miller theorem provide a rationale for basing monetary policy partly on equity prices? As it stands, the theorem provides an important benchmark and enforces careful thinking about financial markets' workings and the imperfections that would create a world where a firm's financial position (hence equity prices) affects its ability to engage in production.

Many possible imperfections could generate such a world. This article focuses on failures of information. Suppose that only the entrepreneur knows every detail of the proposed project, while outside investors financing the project have no way of knowing exactly what the entrepreneur would do with their funds. Suppose further that outside investors have limited ability to punish an entrepreneur who runs off with their money or squanders it on a misguided production activity. In this scenario, external investors are likely to provide financing only if they know they will be able to

1 A response might be warranted if equity prices help forecast macro variables of interest such as output and inflation.

recoup their investment if the project turns sour. One way to ensure this is to restrict the amount of funds they provide to the size of the entrepreneur's financial position. That is, external financing will be no greater than the collateral that investors can seize after the fact.

We have just summarized a story in which a firm's financial position, which we will henceforth call "collateral" or "net worth," has a powerful effect on a its ability to produce. Increases in its equity price will increase this collateral—and with it a firm's ability to produce. Clearly, this is no Modigliani–Miller world, but what role does monetary policy have in it? Can policy help the economy respond to the fundamental shocks that buffet it? Should policy respond to asset prices in such a world?

This article uses a theoretical model to address these questions. The model is highly stylized to keep the analysis tractable, but its essential point will survive more complicated modeling environments.[2] A key conclusion is that there is a role for activist monetary policy. Imperfect information imposes a collateral constraint on this economy, and monetary policy can be useful in alleviating this constraint by responding to productivity shocks or exogenous changes in equity prices.

I. The Model

The theoretical model consists of households and entrepreneurs. We will discuss the decision problems of each in turn.

Households

Households are infinitely lived, discounting the future at rate β. Their period-by-period utility function is given by

$$(1) \quad U(c_t, L_t) \equiv c_t - \frac{L_t^{1+\frac{1}{\tau}}}{1+\frac{1}{\tau}},$$

where c_t denotes consumption and L_t denotes work effort. We choose this form for convenience. Each period, the household decides how much to work at a real wage of w_t. The resulting labor supply relationship is given by

$$(2) \quad L_t = \left(\frac{w_t}{R_t}\right)^{\tau}.$$

Notice that labor supply responds positively to the real wage with elasticity τ. R_t denotes the gross nominal interest rate. Labor supply is negatively related to the nominal rate because we assume that households must use cash to facilitate their consumption purchases (a "cash-in-advance constraint").[3] Because the opportunity cost of holding cash is given by the gross nominal rate, higher nominal rates make it more difficult to turn labor income into consumption, thus discouraging labor supply. To put it another way, the gross nominal interest rate acts like a wage tax where $\frac{1}{R_t} = (1-t_w)$. The celebrated "Friedman rule"—that the net nominal interest rate should be zero (or $R=1$)—is based directly on the observation that a zero interest rate eliminates this implicit wage tax.[4]

A household must also make a decision about consumption versus saving. Households can save only by acquiring shares to a real asset that pays out (real) dividends of D_t consumption goods at the end of time t. It is helpful to think of this as an apple tree that produces D_t apples in time t. The tree trades at share price q_t at the beginning of the period (before the time-t dividend is paid). Under our assumption on household preferences, the equilibrium real share price is given by dividends' present discounted value (the assumption of linear utility implies that the discount rate on dividends is the constant β)

$$(3) \quad \bar{q}_t = E_t \sum_{j=0}^{\infty} \beta^j D_{t+j}.$$

■ 2 For example, empirical evidence suggests that collateral constraints have a stronger effect on small firms than on large ones (see Gertler and Gilchrist [1999]). We abstract from this heterogeneity and posit a single representative firm. Future efforts to quantify collateral effects should model heterogeneity more explicitly.

■ 3 See the appendix for a precise statement of the household's problem and the resulting first-order conditions.

■ 4 In our model, the first-best policy will be the Friedman rule. Our policy section includes analysis of a second-best problem where, for some unspecified reason, the monetary authority desires to keep the long-run average interest rate above zero, $R>1$.

If the share price were below this level ($q_t < \overline{q}_t$), then household demand for shares would be infinite; if the share price were above this level ($q_t > \overline{q}_t$), then household demand for shares would be negative infinity, that is, a desire to sell short. Thus, households will hold a finite and positive level of tree shares only if $\overline{q}_t$ is the equilibrium price. The dividend process is given by

$$D_{t+1} = (1-\rho_D)\, D_{ss} + \rho_D\, D_t + \varepsilon^D_{t+1}.$$

The symbol E_t denotes the rational forecast of future dividends; recall also that β is the rate of household time preference, which is also the real interest rate in this environment. Notice that the asset price depends only on the exogenous dividend process and that the share price is increasing in the current and future dividend levels. The exogenous discount process is an AR1, which means that next period's dividend is a weighted average of today's dividend (D_t) and the long-run average of dividends (D_{ss}) plus a random i.i.d. shock (ε^D_{t+1}).

Entrepreneurs

Entrepreneurs too are infinitely lived and have linear preferences over consumption. They are distinct from households in that they use a constant-returns-to-scale production technology in which labor produces consumption goods

(4) $y_t = A_t H_t,$

where A_t is the current level of productivity, and H_t denotes the number of workers employed at real wage w_t. Like dividends productivity, A_t is an exogenous AR1 random process given by

$$A_{t+1} = (1-\rho_A)\, A_{ss} + \rho_A\, A_t + \varepsilon^A_{t+1}.$$

The entrepreneur is constrained by a borrowing limit. In particular, she must be able to cover her entire wage bill with collateral accumulated in advance. We will denote this collateral as n_t (net worth). The loan constraint is thus

(5) $w_t H_t \leq n_t.$

Notice that all variables are in real terms.

Why is the firm so constrained? Many possible information stories would motivate such a constraint. We will assume the classic hold-up problem: Suppose that hired workers first supply their labor input but that output is subsequently produced if and only if the entrepreneur contributes her unique human capital to the process. This production sequence implies that the entrepreneur could force workers to accept lower wages ex post; otherwise, nothing would be produced. Workers, anticipating this hold-up possibility, will take steps to prevent it. This is harder than it sounds. For example, an equity-type arrangement in which worker and entrepreneur agree ex ante to split the production ex post will not work. After the worker has supplied his labor, the entrepreneur can refuse to provide her unique human capital unless the worker's share is made arbitrarily small. In that case, the worker's only choices are to accept this small share or to take nothing. The worker could seize the entrepreneur's existing assets, but then we are back to our collateral constraint. In fact, as Hart and Moore (1994) and Kiyotaki and Moore (1997) demonstrate, these hold-up problems can only be avoided completely if the entire wage bill is covered by existing collateral that workers could seize in case of default.[5]

We can easily enrich this story by assuming the existence of financial institutions that intermediate between workers and entrepreneurs. For example, suppose that such intermediaries provide within-period financing to entrepreneurs, who use it to pay workers. An intermediary, however, is concerned about the hold-up problem, so it limits its lending to the firm's net worth. This returns us to the collateral constraint described in equation (5).[6]

We assume in what follows that the loan constraint is binding, so that labor demand is given by

(6) $H_t = \left(\dfrac{n_t}{w_t}\right).$

Notice that labor demand varies inversely (with a unit elasticity) to the real wage but is positively affected by the level of net worth. Firms that have more collateral can employ

■ **5** This implicitly assumes a one-period problem so that an entrepreneur who withholds her labor cannot be punished by being deprived of future income.

■ **6** Kiyotaki and Moore (1997) use a similar constraint. See Hart and Moore (1994) for more discussion of the hold-up problem.

more workers because hold-up problems are less severe. The binding collateral constraint implies that $A_t > w_t$, that is, the firm would like to hire more workers but is collateral-constrained.

An entrepreneur's only source of net worth is previously acquired ownership of apple trees. If we let e_{t-1} denote the number of tree shares acquired at the beginning of time $t-1$, then net worth at time t is given by

$$(7) \quad n_t = e_{t-1} q_t ,$$

so that the loan constraint is given by

$$(8) \quad w_t H_t \le e_{t-1} q_t .$$

As noted above, the assumption of a binding loan constraint implies that the firm's marginal profits per worker employed are $(A_t - w_t)$. These profits motivate the entrepreneur to acquire more net worth. We will need to limit this accumulation tendency so that collateral remains relevant. The entrepreneur's budget constraint is given by

$$(9) \quad c_t^e + e_t q_t = e_{t-1} q_t + e_t D_t + H_t (A_t - w_t).$$

The right side of the budget constraint equation is the entrepreneur's income in period t, which consists of her revenue from the sale of existing trees $(e_{t-1} q_t)$, dividends from new tree purchases $(e_t D_t)$, and profits $(H_t [A_t - w_t])$. The left side represents her potential purchases in period t. With her revenue, she purchases either consumption (c_t^e) or new tree shares $(e_t q_t)$. Using the binding loan constraint, we can rewrite this as

$$(10) \quad c_t^e + e_t (q_t - D_t) = e_{t-1} q_t \frac{A_t}{w_t},$$

Because of the profit opportunities from net worth $(A_t > w_t)$, the entrepreneur would like to accumulate trees until the constraint no longer binds (trees are more valuable to collateral-constrained entrepreneurs than they are to households). To prevent this, we will assume that entrepreneurs must consume a fraction of their net income each period

$$(11) \quad c_t^e = (1-\gamma) e_{t-1} q_t \frac{A_t}{w_t},$$

so that entrepreneurial tree holdings evolve as

$$(12) \quad e_t (q_t - D_t) = \gamma e_{t-1} q_t \frac{A_t}{w_t}.$$

Below we will choose $\gamma < 1$ to offset the high return to internal funds, thus keeping the entrepreneur's collateral constrained in equilibrium. This forced-consumption-savings decision implies that households will price trees so that in equilibrium $q_t = \bar{q}_t$.

Equilibrium

In this theoretical model, there are two active markets, the market for apple trees and the labor market (the money market and bond market are discussed in the appendix). We normalize the supply of tree shares to unity so that the asset market clears with $e_t + s_t = 1$. The equilibrium tree price is given by (3). As for the labor market, equating labor supply with labor demand $(L_t = H_t)$ and solving for the real wage yields

$$(13) \quad w_t = n_t^{\frac{1}{1+\tau}} R_t^{\frac{\tau}{\tau+1}}.$$

The equilibrium real wage is increasing in net worth because higher net worth increases labor demand. The wage is also increasing in the nominal interest rate because a higher nominal rate decreases labor supply. Equilibrium employment is given by

$$(14) \quad L_t = \left(\frac{n_t}{R_t} \right)^{\frac{\tau}{1+\tau}}.$$

For the reasons already noted, employment responds positively to net worth and negatively to the nominal rate.

Log-Linearizing the Model

Because the model is relatively simple, it is convenient to express the equilibrium in terms of log deviations. In what follows, the ~ represents a percent deviation from the steady state.

$$(15) \quad \tilde{L}_t = \frac{\tau}{1+\tau} (\tilde{n}_t - \tilde{R}_t)$$

$$(16) \quad \tilde{n}_t = \tilde{q}_t + \tilde{e}_{t-1}$$

$$(17) \quad E_t \tilde{n}_{t+1} = \frac{\tau}{1+\tau} (\tilde{n}_t - \tilde{R}_t) + \tilde{A}_t,$$

where (17) comes from (12) and the asset price (3). Using (16) to eliminate n_t, we can rewrite (15) and (17) in terms of e_t as

$$(18) \quad \tilde{L}_t = \frac{\tau}{1+\tau}(\tilde{q}_t + \tilde{e}_{t-1} - \tilde{R}_t)$$

$$(19) \quad \tilde{e}_t = \frac{\tau}{1+\tau}(\tilde{e}_{t-1} - \tilde{R}_t) + \tilde{A}_t + \left(\frac{\tau}{1+\tau} - \rho_D\right)\tilde{q}_t .$$

To calculate (19), we have also used the ability to express the share price (3) as

$$(20) \quad \tilde{q}_t = \tilde{D}_t \left(\frac{1-\beta}{1-\beta\rho_D}\right),$$

where ρ_D is the autocorrelation in the dividend process. To sum up, the model consists of equations (18)–(20). There is one predetermined variable, e_{t-1}, and there are three exogenous shocks: A_t, D_t, and R_t.

II. The Experiments

Before turning to the question of monetary policy, it is useful to sharpen one's economic intuition about the model by considering several experiments.

First Experiment: A Shock to Productivity (A_t)

Suppose that we hold all other variables constant and consider only shocks to productivity. Then we have

$$(21) \quad \tilde{L}_t = \frac{\tau}{1+\tau}\tilde{e}_{t-1}$$

$$(22) \quad \tilde{e}_t = \frac{\tau}{1+\tau}\tilde{e}_{t-1} + \tilde{A}_t .$$

By combining, we obtain

$$(23) \quad \tilde{L}_{t+1} = \frac{\tau}{1+\tau}(\tilde{L}_t + \tilde{A}_t).$$

Notice that contemporaneous employment does not respond to shocks to productivity, A_t (see [21]). This is a manifestation of the collateral constraint. When productivity is high, the firm would like to expand employment but it cannot because it must finance current activity with current collateral. Thus, the collateral constraint limits the firm's ability to respond to shocks.

There is, however, a delayed response. A positive shock to A_t has no effect on current employment, but it increases e_t and, through it, tomorrow's net worth (see [22]). Hence, employment responds with a lag to productivity shocks.

This lagged response generates persistence to a temporary shock. That is, even if the shock to A_t lasts only one period, the effect on employment, L_t, and thus on output, lasts much longer and only dies out at the rate given by $\tau/(1+\tau)$. If the shock to productivity is serially correlated, this effect remains, so that the collateral constraint prolongs the effect of the productivity shock.

Second Experiment: A Shock to Dividends

Proceeding as before, we have:

$$(24) \quad \tilde{L}_t = \frac{\tau}{1+\tau}\tilde{e}_{t-1} + \left(\frac{\tau}{1+\tau}\right)\left(\frac{1-\beta}{1-\beta\rho_D}\right)\tilde{D}_t$$

$$(25) \quad \tilde{e}_t = \frac{\tau}{1+\tau}\tilde{e}_{t-1} + \left(\frac{\tau}{1+\tau} - \rho\right)\left(\frac{1-\beta}{1-\beta\rho_D}\right)\tilde{D}_t .$$

By combining, we obtain

$$\tilde{L}_t = \frac{\tau}{1+\tau}(\tilde{L}_{t-1} + \tilde{\varepsilon}_t^D).$$

Recall that ε_t^D is the innovation in the dividend process. The most remarkable observation is that employment responds positively to dividend shocks, even though these shocks have no effect on either worker productivity or labor supply. Instead, dividends affect employment solely through the collateral constraint. Because trees are used as collateral, and a dividend shock drives up their price, the collateral constraint is relaxed and the firm can expand employment. Once again, these effects are highly persistent.

Third Experiment: A Monetary Policy Shock

We will assume that monetary policy is given by directives for the gross nominal interest rate, R_t. The implied path for the money supply can be backed out of the money demand relationship (see the appendix).

Proceeding as before, we have

$$(26)\quad \tilde{L}_t = \frac{\tau}{1+\tau}(\tilde{e}_{t-1} - \tilde{R}_t)$$

$$(27)\quad \tilde{e}_t = \frac{\tau}{1+\tau}(\tilde{e}_{t-1} - \tilde{R}_t).$$

By combining, we obtain

$$(28)\quad \tilde{L}_t = \frac{\tau}{1+\tau}(\tilde{L}_{t-1} - \tilde{R}_t).$$

There are two differences between the interest rate shock and the productivity shock. First, the interest rate shock has an immediate effect on employment because it alters labor supply contemporaneously. Second, its effect is negative because the higher interest rate lowers the households' desire to work. As in the previous cases, the shock has a persistent effect through the collateral constraint.

III. Monetary Policy

Optimal Monetary Policy

What is the nominal interest rate's optimal response to productivity and dividend shocks? To answer such a question, we need a welfare criterion. The most natural choice in the present context is the sum of household and entrepreneurial utility, which is given by

$$(29)\quad V_t \equiv c_t + c_t^e - \frac{L_t^{1+\frac{1}{\tau}}}{1+\frac{1}{\tau}} = A_t L_t + D_t - \frac{L_t^{1+\frac{1}{\tau}}}{1+\frac{1}{\tau}},$$

where the equality follows from the fact that total time-t consumption must equal the total supply of time-t consumption goods, which comes from the goods produced using the entrepreneur's production technology, and dividends produced by the apple tree. The only choice variable in V_t is employment. Maximizing V_t with respect to L_t yields the optimality condition

$$(30)\quad L_t = A_t^{\tau}.$$

We will call this solution the "first-best" outcome because the welfare criterion can go no higher. The first-best has two natural features. First, employment responds positively to productivity shocks. When productivity is high, it is efficient for employment to respond positively. Second, the first-best employment does not respond to dividend or share prices. The welfare criterion V_t is increasing in D_t, but these shocks have no effect on labor productivity; thus, it is efficient for employment not to respond to these shocks.

Is the first-best achievable? If there were no collateral constraint, we would have $w_t = A_t$, and the first-best could be achieved by setting $R_t = 1$, that is, by setting the net nominal rate to zero. This is the celebrated Friedman rule. It is optimal in this model because the cash-in-advance constraint on consumption distorts the labor margin.

But in a world with agency costs, the first-best is impossible because employment is given by (14), which, as noted above, is rendered too low ($A_t > w_t$) by the collateral constraint, Furthermore, according to (14), employment fluctuates with net worth and not with the level of productivity. Compared to the first-best outcome, these employment responses are dreadful. Contemporaneous employment does not respond to productivity, even though it is efficient to do so; employment, however, does respond to share prices which, in an efficient world, should not affect it. In short, the collateral constraint causes the economy to under-respond to productivity shocks and to over-respond to dividend shocks.

The advantage of the Friedman rule is that it minimizes the distortion on labor from the cash-in-advance-constraint.[7] The disadvantage is that a pegged zero nominal interest rate precludes the monetary authority's responding to shocks to make employment respond efficiently. It turns out that the benefit of a lower nominal interest rate always wins out in this

■ **7** Recall that because cash must be held to facilitate transactions, higher nominal rates discourage labor supply in (2).

environment—the first-best policy is simply to set the nominal interest rate to zero (that is, $R=1$) and leave it there. But what happens if the monetary authority does not set the long-run interest rate to zero but keeps it positive for some unspecified reason?[8] Can monetary policy improve on this economy's ability to respond to shocks in this world? Yes. To illustrate, let us consider a second-best exercise.

Optimal Policy in Log-Deviations

We take the steady state of the economy as given and use monetary policy so that the economy responds to shocks efficiently. Optimal employment (in log deviations) is given by

$$(35)\quad \tilde{L}_t = \tau \tilde{A}_t .$$

To find the optimal (second-best) interest rate policy, we can impose equation (35) in the system (18)–(19), and back out the implied interest rate. This exercise yields

$$(36)\quad \tilde{R}_t = \tilde{q}_t + \tilde{e}_{t-1} - (1+\tau)\tilde{A}_t$$

$$(37)\quad \tilde{e}_t = (1+\tau)\tilde{A}_t - \rho_D \tilde{q}_t .$$

By combining, we obtain

$$\tilde{R}_t = \varepsilon_t^D - (1+\tau)\left[\varepsilon_t^A + (\rho_A - 1)\tilde{A}_{t-1}\right].$$

What are the properties of this (second-best) optimal monetary policy?[9] When there is a positive shock to productivity A_t, the central bank should lower the nominal interest rate so that employment can expand efficiently. A constant-interest-rate policy does not allow this because of the collateral constraint, but a procyclical interest rate policy overcomes the collateral constraint and allows the economy to respond appropriately.

Suppose that productivity shocks are autocorrelated with coefficient ρ_A. A positive technology shock of 1 percent calls for an immediate interest rate decline of $(1+\tau)$ percent, but then an increase to $(1+\tau)(1-\rho_A)$. The increase is needed to prevent over-expansion of employment, because net worth rises with the initial interest rate decline.

In contrast, if there is a shock to share prices that drives up net worth, n_t, the central bank should increase the interest rate enough to keep employment constant. It is inefficient for employment to respond to these dividend shocks, and the central bank can ensure no response by raising the nominal rate in response. Notice, however, that even if a shock to share prices (dividends) is autocorrelated ($\rho_D > 0$), the optimal interest rate response is iid.

IV. Conclusion

This article addresses the question of how monetary policy should be conducted in a world where asset prices affect real activity directly because of binding collateral constraints, that is, a world in which the Modigliani–Miller theorem does not hold. How should monetary policy be conducted in such a world? Should it respond to asset prices? How should it respond to productivity movements? In this environment, there is a welfare-improving role for a monetary policy that responds actively to asset price and productivity shocks. This activist interest rate policy allows the economy to respond to shocks in a Pareto efficient manner. By assumption, monetary policy cannot eliminate the long-run impact of the information constraint, but it can improve welfare by smoothing the fluctuations in this constraint.

Our results are stark because all firms in the economy are subject to this hold-up problem. One can imagine an environment in which small firms are the ones most subject to agency costs. This will change the quantitative—but not the qualitative—predictions of the model.

This article uses a monetary model with flexible nominal prices. In contrast, Bernanke and Gertler (1999) analyze a similar question in a model with sticky prices. They conclude that as long as monetary policy responds aggressively to inflation, there is no rationale

8 For example, a positive nominal interest rate may be set to give the government inflation-tax revenues.

9 Optimal monetary policy refers to how the central bank should change the interest rate in response to technology shocks and share prices. Money growth is endogenous and, as discussed in the appendix, can be backed out of the money demand relationship, that is, the cash-in-advance constraint.

for a direct response to asset prices. They reach this conclusion because, in their model, asset price shocks directly increase aggregate demand and thus the price level. Hence, a policy that responds aggressively to inflation is automatically responding to asset prices. The model described in this article creates no direct link between inflation and asset prices, so the central bank must respond directly to the latter. It suggests that, to the extent that asset prices do not immediately lead to price inflation, there may be a role for a monetary policy response to asset price movements.

Appendix 1

The household's maximization problem is given by

$$\text{Max} \quad \mathrm{E}\sum_{t=0}^{\infty} \beta^{t} \left\{ c_t - \frac{L_t^{1+\frac{1}{\tau}}}{1+\frac{1}{\tau}} \right\}$$

$$\text{s.t.} \quad \frac{M_{t-1}+X_t}{P_t} + s_{t-1}q_t + s_t D_t + \frac{R_{t-1}B_{t-1}-B_t}{P_t} - s_t q_t - c_t \geq 0$$

$$\frac{M_{t-1}+X_t}{P_t} + s_{t-1}q_t + s_t D_t + w_t L_t + \frac{R_{t-1}B_{t-1}-B_t}{P_t} - s_t q_t - c_t - \frac{M_t}{P_t} \geq 0,$$

where B_t denotes bond holdings (in zero net supply), and households are assumed to receive lump-sum monetary injections, $X_t = \frac{M_t^s}{M_{t-1}^s} - 1$, at the beginning of the period (M_t^s denotes the per capita money supply at time t). Notice that the bond and tree markets open either simultaneous to or before the consumption market. The first constraint is the cash-in-advance constraint: The cash remaining after leaving the bond and tree markets is the cash that can be used to purchase consumption. The second is the intertemporal budget constraint.

After minor simplification, household optimization is defined by the binding cash constraint and the following Euler equations:

$$\text{(A1)} \quad 1 = \beta R_t E_t(P_t/P_{t+1})$$

$$\text{(A2)} \quad L\frac{1}{\tau} = w_t \beta E_t\left(\frac{P_t}{P_{t+1}}\right)$$

$$s_t = \infty \text{ if } q_t < \bar{q}_t$$

$$s_t \text{ indeterminate if } q_t = \bar{q}_t$$

$$s_t = 0 \text{ if } q_t > \bar{q}_t, \text{ where}$$

$$\bar{q}_t = D_t + E_t \beta \bar{q}_{t+1}.$$

Substituting (A1) into (A2), we have

$$L_t = \left(\frac{w_t}{R_t}\right)^{\tau},$$

which is equation (2) in the text. Along with the equilibrium conditions given in the text, we also have $B_t = 0$ and $M_t^s = M_t$. Since we are following an interest rate policy, the implied inflation behavior is given by (A1). The supporting money growth process can then be backed out of the binding cash-in-advance constraint.

References

Bernanke, Ben, and Mark Gertler. "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, vol. 79, no. 1 (1989), pp. 14–31.

Bernanke, Ben, and Mark Gertler. "Monetary Policy and Asset Price Volatility," Federal Reserve Bank of Kansas City *Economic Review*, vol. 84 (1999), pp. 17–52.

Bernanke, Ben, Mark Gertler, and Simon Gilchrist. "The Financial Accelerator in a Quantitative Business Cycle Framework," in John Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*. Volume 1C. New York: Elsevier Science, 2000, pp. 1341–93.

Carlstrom, Charles T., and Timothy S. Fuerst. "Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis," Federal Reserve Bank of Cleveland, Working Paper no. 9602, 1996.

Carlstrom, Charles T., and Timothy S. Fuerst. "Agency Costs, Net Worth, and Business Fluctuations: A Computable General Equilibrium Analysis," *American Economic Review*, vol. 87, no. 5 (1997), pp. 893–910.

Carlstrom, Charles T., and Timothy S. Fuerst. "Agency Costs and Business Cycles," *Economic Theory*, vol. 12, no. 3 (1998), pp. 583–97.

Carlstrom, Charles T., and Timothy S. Fuerst. "Monetary Shocks, Agency Costs and Business Cycles," *Carnegie-Rochester Series on Public Policy,* vol. 54 (2001), pp. 1–27.

Cooley, Thomas F., and Kwanghee Nam. "Asymmetric Information, Financial Intermediation, and Business Cycles," *Economic Theory,* vol. 12, no. 3 (1998), pp. 599–620.

Cooley, Thomas F., and Vincenzo Quadrini. "Monetary Policy and the Financial Decisions of Firms," New York University, unpublished manuscript, 1997.

Fisher, Jonas D.M. "Credit Market Imperfections and the Heterogeneous Response of Firms to Monetary Shocks," *Journal of Money, Credit, and Banking*, vol. 31, no. 2 (1999), pp. 187–211.

Fuerst, Timothy S. "Monetary and Financial Interactions in the Business Cycle," *Journal of Money, Credit, and Banking*, vol. 27, no. 4 (1995), pp. 1321–38.

Gertler, Mark, and Simon Gilchrist. "Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms," *Quarterly Journal of Economics*, vol. 109, no. 2 (1994), pp. 309–40.

Hart, Oliver, and John Moore. "A Theory of Debt Based on the Inalienability of Human Capital," *Quarterly Journal of Economics*, vol. 109, no. 4 (1994), pp. 841–79.

Kiyotaki, Nobuhiro, and John Moore. "Credit Cycles," *Journal of Political Economy*, vol. 105, no. 2 (1997), pp. 211–48.