

Electronic Payments and ATMs: Changing Technology and Cost Efficiency in Banking*

Santiago Carbo Valverde
Department of Economic Theory and History
University of Granada, Spain

David B. Humphrey
Department of Finance
Florida State University, U.S.A.

Rafael Lopez del Paso
Department of Economic Theory and History
University of Granada, Spain

Presented at SUERF Colloquium
October 14-16, 2004, in Madrid, Spain

Abstract:

Two major technical changes in banking are the expanded use of ATMs to deliver depositor services and the shift to electronic payments. Over 1992-2000, these two changes in bank production are estimated to have reduced operating cost by 37%, saving 4.5 billion euros in Spain (0.7% of GDP). As these trends continue, further savings may be realized. These results are robust to using composite, translog, or Fourier functional forms to estimate an output characteristics cost function. Other European countries as well as the U.S. may have experienced similar savings.

* Support for this research was provided by the Fundaci\u00f3n de las Cajas de Ahorros Confederadas para la Investigaci\u00f3n Econ\u00f3mica y Social.

Corresponding Author: Dave Humphrey, Department of Finance, Florida State University, Tallahassee, FL, USA, 32306-1042; Phone: (850) 644-7899; dhumphr@garnet.acns.fsu.edu

Electronic Payments and ATMs: Changing Technology and Cost Efficiency in Banking

1. Introduction.

Of the numerous technical changes in banking, two stand out as having a large impact on operating costs. These are the expanded use of ATMs to supplement and replace expensive branch offices in delivering an important subset of depositor services and the concurrent substitution of lower cost electronic payments for paper-based transactions. Both of these newer banking technologies intensively use computers and telecommunication facilities and have benefited from cost reductions and efficiency improvements in these important inputs. Our purpose is to quantify the overall effect these newer banking technologies in service delivery and payment method have had on banking costs. We also illustrate the portion of the overall cost change attributed to expanded ATM use versus the shift to electronic payments. This is implemented using a statistical model relating operating costs to certain physical characteristics of service delivery and payment levels and mix for Spanish savings and commercial banks over 1992-2000.

The usual approach for identifying cost savings from technical change in banking has been to specify a time-specific indicator variable or, less often, tie technical change to the use of certain inputs (e.g., less labor and more capital). Our approach is quite different. Instead of relying on a time dummy or presuming that technical advances are embodied in or augment the use of certain inputs, we relate operating expenses to five characteristics of banking service output that reflect known differences in cost. As the service mix shifts to these lower cost characteristics, unit operating expenses should fall and reflect the savings associated with the spread of new technology within the industry. By specifying five output characteristics associated with technical change we are able to not only provide an overall estimate of the cost savings from new technology versus scale effects, but also show more clearly how these various components have contributed to this change.

In what follows, Section 2 provides background on how service delivery methods and use of different payment instruments have changed over the past decade in 11 Euro-using countries (which includes Spain) plus the U.K. Our cost model is specified in Section 3. Although we rely on a composite cost function for our analysis (Pulley and Braunstein, 1992; Pulley and Humphrey, 1993), we demonstrate the robustness of our results by specifying and estimating the more commonly used translog and Fourier cost models.

Section 4 presents and discusses our estimates of the cost effect of new technology for both savings and commercial banks in Spain. We find that banks have apparently saved 37% in unit (or average) operating cost between 1992 and 2000 which translates into 4.5 billion euros for the banking system as a whole. As larger institutions have progressed further in shifting from branch offices to ATMs for dispensing cash and also process higher volumes of lower cost electronic payments, these institutions have benefited the most from the reduction in unit operating expenses. Section 5 notes our translog and Fourier cost function results, illustrating the robustness of our composite cost model, while Section 6 summarizes our main conclusions. Since the same service delivery and payment trends shown to have benefited Spain are observed for other European countries, it is likely that similar gains may have occurred in these countries as well.

2. Changes in Service Delivery and Payment Mix.

All European countries (including Spain) deliver banking services using ATMs as well as branch

offices and have provided electronic as well as paper-based payment methods. While the mix of these delivery and payment methods often differ markedly among countries, all have consistently expanded the supply of ATMs relative to branches and have increased the share of non-cash transactions which are electronic. For the services they deliver, ATMs are considerably cheaper than branches and an electronic payment only costs about one-third to one-half as much as a paper-based transaction. Thus it not surprising to find that the shift to ATMs and electronic payments appears to be associated with significant reductions in operating cost as a percent of bank asset value during the 1990s. Indeed, Table 1 suggests that this has been the case for 11 Euro-using countries in Europe plus the U.K.¹

Table 1: Service Delivery, Payment, and Operating Cost in Euroland + UK (1992-1999)

	1992	1994	1996	1998	2000	% Change
ATM/BR	.62	.76	.93	1.12	1.30	110%
Non-Cash/POP	97	105	112	125	136	40
Ele/Non-Cash	.56	.63	.68	.75	.79	41
OC/TA	.020	.018	.017	.016	.016	-21
Spain OC/TA	.030	.027	.024	.022	.019	-37

Sources: OECD, ECB, Bank of Spain, and own calculations.

ATMs and branch offices generate labor and capital costs associated with service delivery or front office expenses. The rapid expansion of ATMs in Europe indicates that, for the range of services provided (cash withdrawal, account transfer, balance inquiry), ATMs have replaced the traditional banking office for a large and growing segment of depositors. Evidence of this shift can be seen in Table 1 (row 1) which shows that the ratio of ATMs to branch offices (BR) rose by 110% during the 1990s.² While the number of branch offices used to deliver banking services can differ considerably across countries, this primarily reflects differences in the average size of banking offices in a country.³ Regardless of the number of branches or their average size, the number of bank employees per 10,000 inhabitants fell in all but Germany and the U.K. during the 1990s, suggesting that ATM use has conserved on bank labor costs as well as capital.⁴

¹All ratios in Table 1 treat the 12 countries as if they were a single entity (i.e., they are the sum of the numerator divided by the sum of the denominator). To make the value data in the table comparable across countries, the OECD translated domestic currency values into euros using exchange rates so that changes over time here can differ from national data in domestic currency units. Due to insufficient data, Greece--a more recent Euro-using country--could not be included.

²During this period, the number of branch offices per 10,000 inhabitants in different Euro-using countries and the U.K. either grew slowly or fell. Since any reduction in the number of banking offices was slight, the primary reason for the rise in the ATM/branch ratio was the rapid rise in ATMs.

³For example, Spain provides 9 to 10 offices per 10,000 inhabitants while Germany, France, Italy, and Portugal provide only 4 to 5. The offset to providing many offices is that there were only around 6 workers per banking office in Spain, which is less than half the 12 to 13 workers per office in Portugal and Italy and about a third as many workers in France (15) or Germany (18).

⁴The reduction in workers per 10,000 of population over 1992-2000 was around 5% for Spain, Italy, and Portugal with 12% for France, and 39% for Finland.

Another major contributor to banking costs is associated with the number and mix of payment transactions. Payments have to be processed and debited/credited to accounts and so generate the vast majority of back-office capital (including computer) and labor expenses. As seen in Table 1 (row 2), the number of non-cash (check, giro, card) transactions per person (POP) per year has risen from 97 in 1992 to 136 in 2000, a 40% increase.⁵ Importantly, it is also seen that the share of non-cash transactions that are electronic, and thus cheaper to process, has also risen by 41% so that by 2000 fully 79% of all non-cash transactions were electronic. Electronic debit card or giro payments for point-of-sale transactions, consumer bill payments, and employee disbursement are typically cheaper than their paper-based alternatives (a check or paper giro transaction). For these types of transactions survey information and cost estimates suggest that an electronic payment costs only one-third to one-half as much as a comparable paper-based transaction (Flatraaker and Robinson, 1995; Wells, 1996; Humphrey, Kim, and Vale, 2001; Humphrey, Willeson, Lindblom, and Bergendahl, 2003).⁶

The large expansion of ATMs relative to branch offices combined with the shift to electronic payments would be expected to lower bank unit operating costs. As shown in the fourth row of Table 1, this seems to be the case. Operating costs as a percent of asset value fell by 21% over 1992-2000 for the 11 Euro-using countries plus the U.K.⁷ For Spain, the operating cost ratio reduction has been 37%.⁸ In what follows, we attempt to determine the effect on operating cost of the shift to electronic payments and ATM use in Spain. This requires a statistical analysis which relates savings and commercial bank operating costs to bank-specific information on ATMs, branch offices, labor and capital input prices, as well as national information on the transaction volume of different types of payment instruments.

3. Using Output Characteristics to Determine Cost Effects of Technical Change and Scale.

Costs in banking are primarily incurred by providing payment processing, deposit safekeeping, cash access, and loan initiation and monitoring services through a geographically diversified set of general and specialized branch offices as well as ATMs. While deposit safekeeping and loan services are specific to branch offices, ATMs substitute with branches for cash withdrawal, balance inquiry, and account transfer services. Some initial payment processing may occur at branch offices but most is incurred in separate dedicated facilities associated with the bank or

⁵While the trend is upward, the levels of non-cash use across countries can be quite different. The total number of non-cash transactions per person ranges at the lower end from 42 to 84 payments a year for Italy, Spain, and Portugal while at the higher end it is 156 to 178 annually for the U.K., Finland, Germany, and France.

⁶This is largely due to the fact that electronic payments experience greater scale economies than paper-based transactions (since the fixed cost component is much more important than the variable one). In addition, advances in computer and telecommunications technology have lowered the absolute cost of processing electronic payments at all scales of operation.

⁷Specifically, the reductions were 17% for Italy, 19% for Germany, 38% for the U.K. and 42% for Finland. Reductions close to 25% were experienced for France, and Portugal.

⁸OECD data would show a smaller change in this ratio due to the variation in the peseta/ecu exchange rate used to translate earlier data in domestic currency units into euros and because of coverage differences (i.e., OECD data includes credit co-operatives, leasing companies, and other credit institutions). Our values in pesetas prior to the adoption of the euro were translated into euros using the fixed conversion rate of 1.0 euro = Pta 166.386 and so are unbiased from changes in exchange rates before the adoption of the euro.

outsourced to non-bank processors.⁹ All of these services involve labor, physical capital, and materials operating expenses and our purpose is to determine statistically the association of operating costs to the processing of check, giro, and card payment transactions as well as the use of ATMs and branch offices. Interest rates on deposits and loans do not really affect the production of payment services and ATMs but can influence the tradeoff between the use of branch offices to collect (lower cost) deposit funds and sell/service mutual funds relative to (more expensive) interbank purchased monies over the interest rate cycle. For Spain, this effect is minor: the aggregate ratio of produced deposits plus mutual funds to assets was very stable over 1992-2000 consistent with little substitution.

The common approach taken in academic studies regarding technical change does not directly measure the actual flow of payment or other banking services as we try to do here. Rather, it is assumed that this service flow is proportional to the value of the stock of bank deposits, securities, and loans in the balance sheet. Inferences on how costs may vary from changing technology and scale of operation are obtained by relating total operating and interest expenses across banks and over time to the value of their deposits, loans, and security holdings (or some other combination of balance sheet positions). As information does not normally exist regarding the adoption of specific technical and other cost-saving innovations in banking, the default is to assume that unknown technical change occurs linearly (or quadratically) with the passage of time and/or is somehow associated with (embodied in) the value of particular inputs.

Our approach directly relates bank operating (not interest) costs to measurable physical characteristics of banking output associated with service delivery and payment processing levels and mix.¹⁰ This achieves two goals. First, the number of bank branches and ATMs--but not necessarily their mix--is directly associated with the size of a bank and its capital and materials operating cost as is the number--but not necessarily the mix--of transactions being processed on behalf of bank customers. When mix is constant and technology is not improved, levels of these activities reflect bank size and hence scale economies.¹¹ Second, changes in the mix of ATMs to branches or in the mix of electronic to paper-based transactions over time, along with improvements in their associated technology, represent an alternative and more specific way to identify the cost effect of technical change in banking.¹²

⁹The initial contact for consumer and some business loans typically involves a customer's local branch office but further processing of loan documents, loan origination, and monitoring services are often handled by larger and more specialized branches or dedicated loan production offices in centralized locations in larger cities.

¹⁰The service delivery and payment functions are largely separable. The primary interaction would be consumers and businesses depositing (a declining number of) checks at a branch office and, on a one-time basis, filling out documents to pay recurring bills by electronic giro or applying for a debit/credit card. After establishing a giro account, bill payments occur automatically, as do all card payments, without branch or ATM intervention unless problems arise.

¹¹The number of branch offices is a good proxy for the stock of loans outstanding, deposits raised, or value of total assets as the R^2 s here all range from .78 to .79.

¹²To circumvent the impossibility of separating technical change from scale effects with only time-series data, it has been common practice to use panel data so that the cross-section component identifies scale while the time-series component identifies technical change. Note that in addition to cross-section and time-series components in our panel data set, we use differences in level and mix to assist in the decomposition between technical change and scale.

3.1 A Composite Cost Function.

Our data consist of an unbalanced panel of 93 commercial and savings banks over 1992-2000 in Spain observed at 6-month intervals (giving 1,541 observations).¹³ Bank-specific information on operating cost, numbers of ATMs, branch offices, and labor and capital input prices were combined with aggregate (national) data on the number of check, giro, and card payments and used in a non-linear, functionally separable, composite cost function.¹⁴ The composite model can approximate better the scope-type joint cost effects that are associated with altering how banking services are delivered and how payments are processed. This is because the level of banking output in a composite function is not in logs, although input prices are. By keeping output in absolutes, we specify a direct relationship between output and operating costs that is likely more accurate--for prediction purposes when one or more outputs are small--than if the log of output is related to the log of operating cost.¹⁵ As well, by specifying the log of input prices, it is possible to impose the theoretical condition of linear homogeneity in input prices in estimation.¹⁶

The composite cost function (1), in its output/input price separable quadratic form, is estimated jointly with $n-1$ cost share equations. The Box-Cox (1964) transformation is represented by a superscripted parameter in parenthesis (ϕ) where $OC^{(\phi)} = (OC^\phi - 1)/\phi$ for $\phi \neq 0$ and $OC^{(\phi)} = \ln OC$ for $\phi = 0$ in:

¹³The panel includes all savings banks, all but the very smallest commercial banks (which were excluded due mostly to missing ATM data), but no cooperative banks (who also had missing data). This accounts for 77% (80%) of all assets (operating cost) in the Spanish banking system in 1992 and 92% (90%) in 2000. The excluded cooperative banks only account for five percentage points of the banking system's operating costs while the excluded commercial banks account for the remaining five percentage points in 2000.

¹⁴Use of a balanced panel by (a) backward aggregation of merging institutions before they actually merge or (b) including only acquiring banks could have biased our results. With (a), combining data on banks before they merge and realize possibly lower costs associated with their larger post-merger size would tend to understate scale benefits in pre-merger years. With (b), the sample would exclude acquired banks before they merged and distort measured cost/payment volume relationships over time. For these reasons we use an unbalanced panel.

¹⁵As illustrated in Pulley and Braunstein (1992), this can occur when one or more outputs is less than 5-10% of total output. This occurs for ATMs (as a percent of ATMs plus branches) for some banks early in our sample and for checks (as a percent of check, giro, and card transactions) for some later years.

¹⁶A similar function (CES-quadratic) was used by Rölller (1990) to determine scope effects of local and long-distance telephone costs for the Bell System while Pulley and Humphrey (1993) used a composite form to assess the cost effects of separating risky loan assets from deposit liabilities into two separate banks, funding the former with uninsured CDs and investing the latter in safe assets.

$$\begin{aligned}
OC^{(\phi)} &= f^{(\phi)}(\underline{Q}, \underline{\ln P}) \\
&= \{[\alpha_0 + \sum_{i=1}^5 \alpha_i Q'_i + 1/2 \sum_{i=1}^5 \sum_{j=1}^5 \alpha_{ij} Q'_i Q'_j] \cdot \exp[\beta_0 \\
&\quad + \sum_{k=1}^2 \beta_k \ln P_k + 1/2 \sum_{k=1}^2 \sum_{m=1}^2 \beta_{k,m} \ln P_k \ln P_m]\}^{(\phi)} \\
S_k &= \beta_k + \sum_{m=1}^2 \beta_{k,m} \ln P_m
\end{aligned}$$

where:

- OC = total operating expenses, composed of labor, capital, and materials costs;¹⁷
- $Q'_{i,j}$ i,j = five output characteristics composed of two service delivery alternatives--automated teller machines (*ATM*) and bank branches (*BR*)--along with three payment processing alternatives--the number of checks (*CHECK*), giro payments (*GIRO*), and debit and credit card transactions (*CARD*).¹⁸ Service delivery data are available by bank but payment transactions data are not (so data for all banks are used instead) In (1), $Q' = Q - 1$;
- $P_{k,m}$ k,m = two input prices referring to the average labor cost per employee and an approximation to the price of physical capital and materials represented by capital depreciation expenditures divided by the value of physical capital; and
- S_k = the cost shares for the labor input (the capital/materials input share is deleted to avoid singularity).

It is expected that operating costs not directly associated with the mode of service delivery or type of payment will be represented in the intercept term. The composite function is non-linear and is estimated iteratively. Following Pulley and Braunstein (1992), let $\underline{D} = \underline{Q}$ and $GM^{\phi-1}$ be the geometric mean of operating cost OC , then the separable quadratic form of the composite model is estimated from the pseudo model (2):¹⁹

¹⁷OC is in nominal terms. The specification of bank-specific input prices accounts for inflation effects on costs more accurately than use of standard inflation indicators (e.g., cost-of-living index or GDP deflator).

¹⁸Giro transfers are typically electronic in Spain and debit cards accounted for about 55% of all card transactions (a share that only rises about 1 percentage point a year).

¹⁹Pulley and Braunstein (1992) note that it is generally not feasible to estimate both α_0 and β_0 intercepts. As we are more interested in output than input prices, and on the basis of fit, we set $\beta_0 = 0$ and retain α_0 in estimation.

$$\begin{aligned}
D &= [-(OC^{(\phi)}/GM^{\phi-1}) + f^{(\phi)}(\underline{Q}, \underline{\ln P})/GM^{\phi-1}] \\
&= [-\{(OC^{\phi} - 1)/\phi GM^{\phi-1}\} + (\{\alpha_0 + \sum_{i=1}^5 \alpha_i Q'_i \\
&\quad + 1/2 \sum_{i=1}^5 \sum_{j=1}^5 \alpha_{ij} Q'_i Q'_j\} \cdot \exp[\beta_0 + \sum_{k=1}^2 \beta_k \ln P_k \\
&\quad + 1/2 \sum_{k=1}^2 \sum_{m=1}^2 \beta_{k,m} \ln P_k \ln P_m]\}^{\phi} - 1)/\phi GM^{\phi-1}] \\
S_k &= \beta_k + \sum_{m=1}^2 \beta_{k,m} \ln P_m
\end{aligned}$$

3.2 Alternative Translog and Fourier Cost Functions.

To illustrate the robustness of our results, we also estimate translog and Fourier cost functions. A translog function may generate biased results, compared to the composite form, when levels of some outputs are small and outputs are specified in logs. Even so, as these two additional functions are often used in cost analyses, it is useful to compare their results with those from our composite form.

The translog cost function (3) is estimated jointly with n-1 cost share equations:

$$\begin{aligned}
\ln OC &= \alpha_0 + \sum_{i=1}^5 \alpha_i \ln Q_i + 1/2 \sum_{i=1}^5 \sum_{j=1}^5 \alpha_{ij} \ln Q_i \ln Q_j + \sum_{i=1}^5 \sum_{k=1}^2 \delta_{i,k} \\
&\quad \ln Q_i \ln P_k + \sum_{k=1}^2 \beta_k \ln P_k + 1/2 \sum_{k=1}^2 \sum_{m=1}^2 \beta_{k,m} \ln P_k \ln P_m \\
S_k &= \beta_k + \sum_{m=1}^2 \beta_{k,m} \ln P_m + \sum_{i=1}^5 \delta_{i,k} \ln Q_i
\end{aligned}$$

where the variables have been defined above.

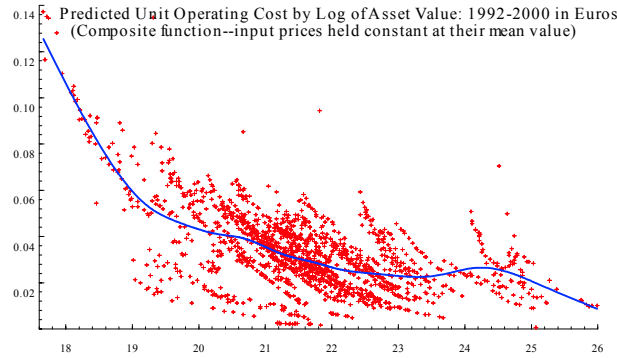
The Fourier form we use adds sin and cos terms to the translog cost function. As our main concern is to allow for greater flexibility in the local identification of output effects on operating costs, the sin and cos terms are applied to the output (Q) measure. The Fourier form is a globally flexible approximation since the respective \sin and \cos terms are mutually orthogonal over the $[0, 2\pi]$ interval. The Fourier function (4) is estimated jointly with the cost shares:

$\ln TC =$ Translog Cost Function

$$\begin{aligned}
& + \sum_{n=1}^5 [\tau_n \cos(\ln Q_n^*) + \omega_n \sin(\ln Q_n^*)] \\
& + \sum_{n=1}^5 \sum_{q=n}^5 [\tau_{nq} \cos(\ln Q_n^* + \ln Q_q^*) + \omega_{nq} \sin(\ln Q_n^* + \ln Q_q^*)] \\
& + \sum_{n=1}^5 [\tau_{nnn} \cos(\ln Q_n^* + \ln Q_n^* + \ln Q_n^*) \\
& + \omega_{nnn} \sin(\ln Q_n^* + \ln Q_n^* + \ln Q_n^*)] \\
S_k & = \beta_k + \sum_{m=1}^2 \beta_{k,m} \ln P_m + \sum_{i=1}^5 \delta_{i,k} \ln Q_i
\end{aligned}$$

The new terms are $\ln Q^* = \ln Q \cdot YQ + ZQ$, $YQ = (0.8 \cdot 2\pi) / (\max \ln Q - \min \ln Q)$, $ZQ = 0.2\pi - \min \ln Q \cdot YQ$, and $\pi = 3.141593\dots$, so that $\ln Q^*$ is essentially expressed in radians.²⁰

Figure 1



4. Cost effects from Changes in Service Delivery and Payment Levels and Mix.

4.1 Composite Function Results.

Predicted unit operating cost from the composite function for 93 savings and commercial banks over 1992-2000 is shown in Figure 1.²¹ While the levels and mix of ATMs, branch offices, and check, giro, and card payment volumes are allowed to vary, input prices are held constant at their mean values. As ϕ in the composite form is .20, the estimated model is closer to a specification which includes the log of output as well as input prices (when $\phi = 0.0$) than it is to a

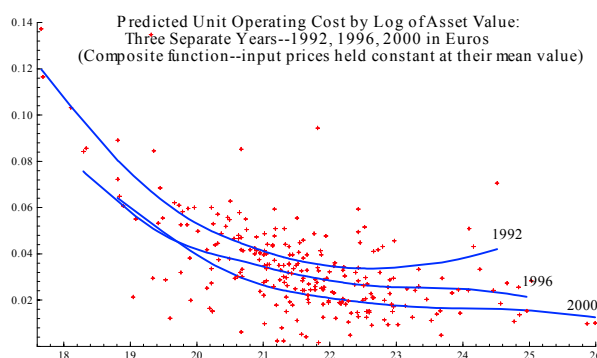
²⁰See Mitchell and Onvural (1996) and Berger and Mester (1997). Our Fourier specification follows Berger and Mester.

²¹Unit operating cost is the ratio of operating cost to asset value and is a measure of average operating cost.

specification with output in absolutes and prices in logs (when $\phi = 1.0$).²² Even so, the estimated model is significantly different from either of these alternatives since ϕ is significantly different from zero or one.²³

The curve fitted to the scattergram in Figure 1 is a cubic spline and illustrates how unit operating cost generally varies by bank asset size over time.²⁴ This figure combines both technical change (time-series) and scale (cross-section) effects associated with front office service delivery and back office payment processing cost changes. The distinction between technical change (a shift in the unit operating cost curve between years) and scale effects (moving along a unit operating cost curve for a single year) is illustrated in Figure 2. Here separate predicted unit operating cost curves are shown for 1992, 1996, and 2000. Scale economies exist since unit cost falls as (the log of) asset size increases on the X-axis. As well, the operating cost curves shift down over time showing that unit operating expenses are falling as technical change progresses with the substitution of ATMs for branch offices, the replacement of checks (and cash) with giro and card electronic payments, and technological improvements associated with all five of these output characteristics.

Figure 2



Looking at all banks together where unit operating cost reflects the ratio of the sum of predicted operating expense across all banks divided by the sum of observed asset values, this aggregate ratio is .033, .023, and .018, respectively, for 1992, 1996, and 2000 in Figure 2

²²The estimated parameters of the composite function underlying this figure are presented in an Appendix. Likelihood ratio tests of setting the 11 parameters associated with ATM and branch variables, or the 15 parameters associated with check, giro, and card payment transactions, equal to zero were $-2\ln \lambda = 3873$ and 25.7, respectively. The ATM and branch variables varied by bank and over time and were significant at the .01 level while the three payment transaction variables were significant at the .05 level. Payment data by bank are not available in any country so these data only vary over time, which accounts for their lower significance level.

²³For more on these two alternative specifications which depend on the value of ϕ , see Pulley and Braunstein (1992) or Pulley and Humphrey (1993).

²⁴Bank size on the X-axis is indicated by the natural log of asset value. Taking the log improves comparability among the numerous smaller and less numerous very large banks. As ATMs and the production functions for processing payments are essentially identical across types of financial institutions, use of a single cost function covering both savings and commercial banks is justified. This accords well with our purpose of illustrating the efficiency gains from the shift to ATMs and electronic payments for the entire banking sector (rather than a subset of the industry).

indicating a predicted 45% reduction.²⁵ This is very similar to the observed ratios in Table 2 for Spain (.030, .024, and .019) which indicated that for the Spanish banking system as a whole, unit operating cost has fallen by 37% over 1992-2000.²⁶ Since total bank operating cost in 1992 was 12.1 billion euros, this suggests that operating expenses would have been 4.5 billion euros (.37 times 12.1 billion euros) higher in 2000 than they were if there were no technical change or scale effects to reduce operating costs from their ratio to assets in 1992. This savings equals 0.7 % of GDP in 2000 and is equivalent to having unit operating cost at banks fall by 4% a year due to changes in service delivery and payment costs.²⁷

4.2 Service Delivery Costs: ATMs and Branch offices.

Predicted service delivery costs represent operating expenses associated with installed ATMs and branch offices, holding check, giro, and card payment volumes and input prices constant at their mean values. These predicted delivery expenses, expressed as a ratio to total assets for each bank, fell by 45% between 1992 and 2000.²⁸ Internal estimates from a confidential industry source indicate that an additional ATM costs around 27,500 euros while an additional branch costs 112,500 euros. In our sample, the number of ATMs expanded by 142% (rising from 17,300 in 1992 to 41,800 in 2000) while the number of branch offices grew by 22% (expanding from 28,200 to 34,300). Thus most of the reduction in the service delivery cost to asset ratio is due to using more ATMs relative to branches.

The extent to which this cost ratio falls as the ratio of ATMs to branch offices rises is shown in Figure 3. The average ATM/branch ratio rose from 0.6 in 1992 to 1.2 in 2000. Reference to Figure 3 indicates that unit costs may continue to fall up to the point where the ATM/branch ratio approaches 2.0. Thus there seems to be additional scope for further operating cost to asset ratio reductions with a higher ratio of ATMs to branch offices in the future.

The relative importance of changes in delivery costs versus payment expenses over 1992-2000 can be illustrated by looking at the level of operating cost, rather than its ratio to assets. If only ATMs and branch offices had changed from 1992 going forward to 2000 (i.e., holding payment transactions and input prices at their 1992 levels), then the rise in predicted operating cost from delivery expenses alone is 2.7 billion euros. However, doing the same for payment transactions (i.e., holding ATMs, branches, and input prices at their 1992 levels), results in a fall in predicted operating cost from payment activities alone of 1.9 billion euros. Finally, letting

²⁵The aggregate ratio gives a larger weight to larger banks that typically have lower unit operating cost.

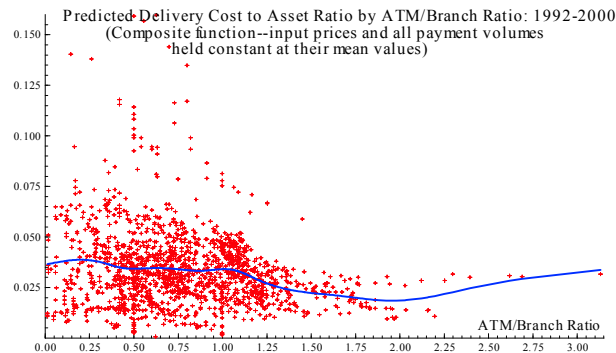
²⁶If some of the decline in the operating cost/asset ratio is due to banks substituting purchased funds (which generate interest costs) for produced deposits or servicing off-balance sheet mutual funds (which generate operating expenses), then we should see a reduction in the ratio of produced deposits (demand, savings, and time deposits) plus mutual funds to assets over 1992-2000. As this ratio rises slightly from 78.0% in 1992 to 79.5% in 2000, the reduction in the operating cost/asset ratio is attributed to cost effects.

²⁷The 12.1 billion euro figure is the sum of operating cost for all banks in Spain in 1992 reported to the Bank of Spain. GDP in 1992 was 609.3 billion euros.

²⁸If the ratio of delivery expenses to asset value were observed in the raw data (as total operating costs are), this ratio would be similar to a measure of average delivery cost per euro of assets. However, as noted by Baumol, Panzer, and Willig (1982), in a multiproduct cost function where output characteristics are not fully functionally separable, it is not possible to determine accurately the level of average cost. This is because the predicted value of delivery expenses, for example, will include the mean values of input prices as well as check, giro, and card expenses. While this does not affect our ability to determine changes in delivery costs over 1992-2000, it does artificially raise the level of predicted delivery expenses.

only input prices change after 1992 raises operating costs by 2.4 billion euros.²⁹ Overall, this suggests that the observed rise in total operating expenses was pretty much equally driven by the rise in input prices (where the average price of labor rose by 40%) as it was from the expansion of ATMs and branches. Although the volume of our three payment instruments rose by 64%, the joint effect of payment scale economies and the shift to cheaper electronic payments seemingly led to an absolute reduction in total payment costs.³⁰

Figure 3



4.3 Processing Costs: Check, Giro, and Card Transactions.

Predicted payment processing costs represent operating expenses associated with the level and composition of check, giro, and card transactions, holding ATMs, branch offices, and input prices constant at their mean values. These predicted payment costs, divided by the total number of check, giro, and card transactions made each year, give unit payment costs that fall both over time (top of Figure 4) and by total payment volume (bottom of Figure 4). Predicted unit payment cost fell by 48% between 1992 and 2000.³¹

Not all payment costs are falling. Indeed, the reduction in unit payment expense seen in Figure 4 is composed of rising check average costs and falling giro and card average costs over time. These changes are seen in Figure 5 where the level of predicted unit cost for check, giro, and card transactions have been "normalized" at their mean to reflect internal industry estimates of the level of average cost for each of these payment instruments. As a result, and only in this case, both the levels and the changes shown in Figure 5 are likely good approximations to cost accounting values--if such values existed--for these three payment instruments. According to confidential industry sources, the average cost of a check is 0.275 euros, a giro transaction is 0.0775 euros, and a card payment is 0.075 euros. These values were used to adjust or normalize the level of the curves shown in Figure 5 and suggest that an electronic giro (card) payment costs

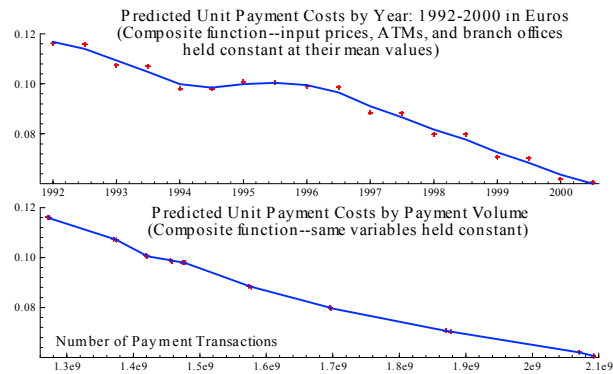
²⁹As these three operating cost categories are not fully functionally separable from one another, they will not add up to the change in actual or predicted overall operating costs over this period when two of the three cost categories are being successively held constant at their level in 1992.

³⁰Although total observed operating expenses rose by 63%, total assets expanded by 161% (so the ratio of operating cost to assets fell by 37%).

³¹Again, the predicted unit cost here is not a standard measure of average cost. In addition to the predicted cost of three payment instruments, the mean cost of ATMs, branch offices, and input prices are also included. This raises the level but does not affect the predicted changes over 1992-2000.

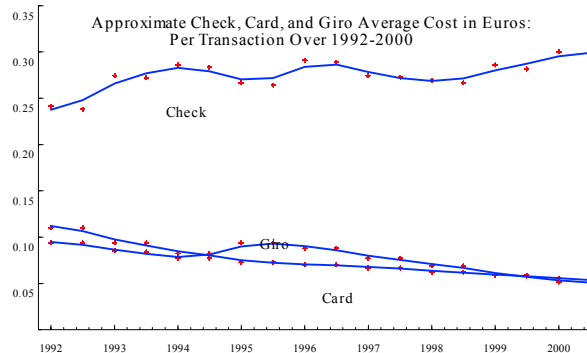
28% (27%) as much as a check. This corresponds to limited survey information available for other countries (Humphrey, Willeson, Lindblom, and Bergendahl, 2003) and to more detailed cost information available for Norway (Gresvik and Øvre, 2002).

Figure 4



Over 1992-2000, giro and card payments expanded by 85% and 81%, respectively, while checks fell by 17%. As a result, the share of checks in all non-cash payment transactions fell from .19 in 1992 to only .10 in 2000. Giro transactions accounted for a .56 share in 2000 while cards were .34. Scale economies in the processing of electronic giro and debit card payments help explain the reduction in the average per transaction payment expenses in Figures 4 and 5 while the scale benefit works in reverse (to offset some of this benefit) as the number of checks processed falls.

Figure 5



5. Translog and Fourier Function Results.

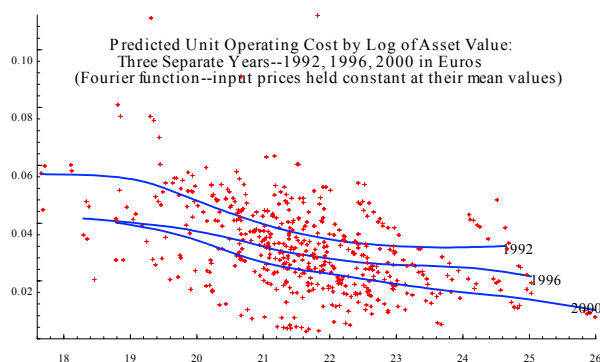
The operating cost results using the easier to estimate translog model (3) are very close to those presented above for the composite cost function. Indeed, Figures 1 and 2 for the composite form--showing predicted unit operating cost over 1992-2000 by bank asset size and for three years separately--are so close to those using the translog that it is difficult to tell them apart. Consequently, the predicted change in the operating cost to asset ratio between 1992 and 2000 was -45% for the composite form and -37% with the translog. Changes in predicted unit delivery and payment expenses over 1992-2000 were also within a few percentage points of each other.³²

³²The strong correspondence between the translog and composite results is not unexpected since ϕ in the composite form is .20, indicating that this estimated model is closer to a translog specification which

The Fourier cost model (4) adds sin and cos terms for the five output characteristics to the standard translog form. The purpose is to capture more than just the quadratic nonlinearity which would be captured with the translog specification. Since bank-specific data are available for ATM and branch offices in our panel data set, so that for these two output characteristics we have both cross-section and time-series variation, the Fourier form may improve the fit compared to the translog form. However, no country has publicly available bank-specific data on the volumes of check, giro, and card transactions. Thus our payment data have no cross-section variation, only time-series variation at the national level. As payment volumes experience low variance over time, it turns out that the quadratic specification in the translog portion of the Fourier form is sufficient to locally identify all of the nonlinearity in these data. For this reason, all the parameters specified in the full Fourier model in (4) could not be estimated and the cos and sin terms in the final estimated model only refer to two (ATMs and branches) rather than the full five output characteristics.³³

The main differences in our results for the Fourier form, compared to the composite, are for the smallest and largest banks, not for the average bank. The simplest way to see this is to compare Figure 6, which illustrates predicted unit operating cost for 1992, 1996, and 2000 using the Fourier form, with Figure 2, which shows the same three cost curves for the composite function. While the middle segments of these two figures which reflect the average bank are

Figure 6



largely congruent, the large operating cost scale economies for the smallest banks shown for the composite (or translog) form are considerably smaller when the Fourier form is used. As well, the slight scale diseconomies evident for the largest banks in 1992 with the composite (or translog) form are shown as constant unit cost with the Fourier form in that year. Even so, these differences have little effect on the predicted changes in unit operating cost between 1992 and 2000. With the composite form, the change in the predicted operating cost to asset ratio was -45% while with the Fourier model it was -46%, with similarly close correspondence for changes in unit delivery and payment costs. In sum, the results presented above for the composite function seem to be robust to the use of alternative cost function forms.

includes the log of output as well as input prices (as would be the case if $\phi = 0.0$).

³³As the sin and cos terms for three output characteristics (check, giro, and card transactions) could not be estimated, all of the single and double summations shown in (4) are over two output characteristics (ATMs and branches), not five.

6. Summary and Conclusions.

Two common trends among banking systems in developed countries have had a large impact on operating cost, and hence on service level and price, to users. One trend has been the expansion of lower cost and more convenient ATMs, relative to branch offices, to deliver cash, account transfer, and balance inquiry services to depositors. A second trend has been the ongoing replacement of paper-based payment instruments (checks and paper giro payments) with lower cost electronic alternatives (debit cards and electronic giro payments). Indeed, these five banking output characteristics related to service delivery and payment (and deposit account) processing make up the bulk of bank operating costs.³⁴

The effect on cost from these five activities incorporate both changing technology and scale influences. A statistical model based on these output characteristics relates operating cost to service delivery and payment levels and mix to determine how changes in these characteristics have affected operating costs at Spanish savings and commercial banks over 1992-2000. We find that the average Spanish bank has apparently saved 37% in unit (or average) operating cost between 1992 and 2000, or about 4.5 billion euros for the banking system as a whole (0.7% of GDP in 2000). As larger institutions have progressed further in shifting from branch offices to ATMs for dispensing cash and also process higher volumes of lower cost electronic payments, these institutions have benefited the most from the associated reduction in unit operating expenses. In determining the average effect on operating cost from changes in service delivery methods and the level and mix of payment volumes, it does not matter much whether a composite, translog, or Fourier cost model is used (although for very small and very large banks there are some differences).

With respect to the future, it appears that if ATMs were expanded further (relative to branch offices) additional operating cost could be saved. At present, the ATM/branch ratio is 1.2 but costs appear to continue to fall for institutions with ratios up to around 2.0. It is also evident that additional operating expense could be saved with a further shift to electronic payments since they generally only cost one-third to one-half as much as its paper-based non-cash alternative. It would not be surprising if similar savings in operating cost were found in other European countries (or even the U.S.) as these countries have often experienced similar changes in service delivery and payment composition.

³⁴While banks also provide loan origination and monitoring services, asset liquidity management with security holdings, and trust and safekeeping services, these are performed using branch offices (an included variable) and the labor input component is small relative to that associated with deposit service delivery and payment activities.

Bibliography

- Baumol, W., J. Panzer, and R. Willig (1982). *Contestable Markets and the Theory of Industrial Structure*. Harcourt Brace Jovanovich, San Diego.
- Berger, A., and L. Mester (1997). "Inside the black box: What explains differences in the efficiencies of financial institutions?", *Journal of Banking and Finance*, 21: 895-947.
- Box, G., and D. Cox (1964). "An analysis of transformations", *Journal of the Royal Statistical Society, Series B*, 26: 211-246.
- Flatraaker, D.-I., and P. Robinson (1995). Income, costs and pricing in the payment system, Norges Bank *Economic Bulletin*, 66: 321-332.
- Gresvik, O., and G. Øwre (2002), Banks' Costs and Income in the Payment System in 2001, Norges Bank *Economic Bulletin*, 73: 125-33.
- Humphrey, D., M. Kim, and B. Vale (2001). "Realizing the gains from electronic payments: Costs, pricing, and payment choice", *Journal of Money, Credit and Banking*, 33: 216-234.
- Humphrey, D., M. Willeson, T. Lindblom, and G. Bergendahl (2003). What does it cost to make a payment?, *Review of Network Economics*, 2: 159-174.
- Mitchell, K., and N. Onvural (1996). "Economies of scale and scope at large commercial banks: Evidence from the Fourier flexible functional form", *Journal of Money, Credit and Banking*, 28: 178-199.
- Pulley, L., and Y. Braunstein (1992). "A composite cost function for multiproduct firms with an application to economies of scope in banking", *Review of Economics and Statistics*, 74: 221-230.
- Pulley, L., and D. Humphrey (1993). "The role of fixed costs and cost complementarities in determining scope economies and the cost of narrow banking proposals", *Journal of Business*, 66: 437-462.
- Röller, L.-H. (1990). "Proper quadratic cost functions with an application to the Bell System", *Review of Economics and Statistics*, 72: 202-210.
- Wells, K. (1996). "Are Checks Overused?", *Federal Reserve Bank of Minneapolis Quarterly Review*, 20: 2-12.

Appendix: Parameter Estimates for the Composite Cost Function.

Number of observations = 1541. Log likelihood = 1552.14. Standard Errors computed from heteroscedastic-consistent matrix (Robust-White). Durbin-Watson = 1.953. Likelihood ratio tests of setting the 11 parameters associated with ATM and branch variables, or the 15 parameters associated with check, giro, and card payment transactions, equal to zero were $-2\ln \lambda = 3873$ and 25.7, respectively. The ATM and branch variables varied by bank and over time and were significant at the .01 level while the three payment transaction variables were significant at the .05 level. Payment data by bank are not available in any country so these data only vary over time, which accounts for their lower significance level.

Parameter Estimate	t-statistic
ϕ	.2023 8.92
α_0	-50038 -.40
α_1	.376E-03 .42
α_2	-.400E-04 -.34
α_3	.792E-04 .37
α_4	-24.60 -.32
α_5	164.0 2.48
α_{11}	-.164E-11 -.46
α_{22}	.313E-13 .48
α_{33}	-.342E-14 -.01
α_{44}	.0574 3.84
α_{55}	.1133 8.10
α_{12}	.361E-12 .46
α_{13}	-.479E-12 -.45
α_{14}	.473E-06 .84
α_{15}	-.662E-06 -1.39
α_{23}	-.705E-12 -.25
α_{24}	.199E-07 .45
α_{25}	-.516E-07 -1.23
α_{34}	-.317E-07 -.36
α_{35}	-.360E-07 -.47
α_{45}	-.1517 -5.12
β_1	.4144 37.30
β_{11}	.0415 17.11