

Efficiency and Economies of Scale of Large Canadian Banks*

Jason Allen
Department of Economics
Queen's University

and

Ying Liu
Department of Monetary and Financial Analysis
Bank of Canada

July 12, 2004

ABSTRACT

This paper measures economies of scale of Canada's six largest banks and their cost efficiency over time. Using a unique panel of data spanning from 1983q1 to 2003q3, we estimate pooled translog cost functions and derive measures for relative efficiency and economies of scale. The disaggregation of the data allows us to model Canadian banks as producing multiple outputs, including non-traditional activities. Given the long sample period of our dataset, we are also able to incorporate technological and regulatory changes in the banks' cost functions, as well as time-varying bank-specific effects. Our model leads us to reject constant returns to scale. These findings suggest there are potential scale benefits in the Canadian banking industry. We also find that technological and regulatory changes have had significant positive effects on the banks' cost structure.

Keywords: banking, economies of scale, panel data

JEL classification: G21, D24, C33

*Correspondence: Department of Economics, Queen's University, Kingston, ON, Canada, K7L 3N6. e-mail: allenj@qed.econ.queensu.ca. The authors thank Allan Crawford, Geoffrey Dunbar, Walter Engert, Kim Huynh, Ian Keay, Jeremy Lise, Darcey McVanel, and Greg Tkacz, as well as participants at a Bank of Canada workshop and the Canadian Economic Association meetings, 2004. We are especially grateful for comments provided by Allan Gregory and Erwin Diewert. All errors are our own. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada. Jason Allen thanks the Bank of Canada for their hospitality and facilities while writing this paper.

I. Introduction

A key question in the bank merger debate is whether a merged bank can achieve higher efficiency and reduce its production cost. In Canada the merger debate is a topical one. Some banks have argued that a merger will lead to cost savings for consumers. This paper looks at the efficacy of this hypothesis by measuring the economies of scale and efficiency in the Canadian banking industry. These measures provide important insights not only to managers in their process of making operational decisions, but also to policy makers in their debate on regulatory issues such as mergers. Using a proprietary panel of data of the six largest Canadian banks from 1983q1 to 2003q3, we estimate pooled translog cost functions and derive measures for relative efficiency and economies of scale. Economies of scale allow us to assess statistically whether “larger is better”, based on existing technology. Relative efficiency allows us to compare banks cost minimizing behaviour over time.

The disaggregation of the data is critical and allows us to model Canadian banks as producing multiple outputs. Included as bank output are non-interest related activities such as deposit account services, security underwriting and wealth management. Because of data limitations, non-interest activities have rarely been studied in the literature. We proxy these activities by an asset-equivalent-measure of non-interest income derived in Boyd and Gertler (1994).

The analytical framework is the flexible translog cost function, assuming the intermediation approach in bank production (c.f. Murray and White (1983))¹. That is, banks use deposits, labour, and capital as inputs to produce outputs such as loans. Three econometric models are estimated: (i) a time-varying fixed-effect panel model, (ii) stochastic cost efficiency frontier model estimated by maximum likelihood (ML), and (iii) a system of seemingly unrelated regressions (SUR) using GLS. We also include a fixed effect in the SUR model to represent inefficiency. Measures of economies of scale are calculated from the derivatives of cost with respect to output. We also generate measures of relative cost efficiency among banks. The results are general and do not depend on a particular estimator.

Given the relatively long time series component to our panel, we are able to identify possible technological changes over time. Here, technological change is a broad term which includes financial innovations, changes in the competitive nature of banks, and demographically led changes in household portfolios. Freedman and Goodlet (1998) note that banks have been undergoing significant technologi-

¹Other functional forms such as the normalized quadratic could be considered. The translog function seems to be the standard in the financial intermediation literature.

cal changes in recent years which affect the way services are provided, the instruments used to provide services, and the nature of the financial service providers. These changes include the adoption of electronic processing of transactions, the development of new instruments and products, the spread of ATMs, and internet banking and better risk management techniques. We model technological changes in two mutually exclusive ways: by imposing common technological trends on the cost function or by allowing technological changes to affect bank-specific progress.

Another important factor that may have affected the banks' cost structure over time are regulatory changes. Calmès and Liu (2003) argue that changes to the Bank Act in 1987, 1992 and 1997 may have encouraged the trend towards direct financing, i.e. financing done at financial markets rather than through financial intermediaries. At the same time, banks have been increasingly involved in non-traditional, typically market-oriented, activities. We model the potential impact of such changes in the sources of a bank's income on its cost structure by introducing regulatory variables.

Our results suggest that we can reject constant returns to scale. Depending on the model, our results suggest that banks can reduce average cost by five to twenty percent by doubling their output. We also conclude that both technological and regulatory changes have had positive effects on the banks' cost structure.

The paper is organized as follows: Section II introduces the literature on efficiency and economies of scale of financial institutions. Section III presents our model. Section IV discusses the econometric issues and the models that we estimate. Section V provides a detailed description of the data. Section VI presents our estimation results. Section VII summarizes our findings.

II. Literature Review

The study of economies of scale in financial institutions has had a long history. While studies have been done for different types of financial institutions in different countries, very few have focused on Canadian institutions. In general, most studies find only small economies of scale in a firm's cost structure. In those studies that find evidence of increasing returns to scale, the measured economies of scale seem to be stronger in small to medium sized firms than for large firms.² However, more recent

²See Ferrier and C.A.K.Lovell (1990) for U.S. banks, Rime and Stiroh (2003) for Swiss universal banks, Rezvanian and Mehdiian (2002) for Singaporean banks.

studies find stronger evidence of increasing returns to scale in large U.S. commercial banks in the 1990s (Berger and Mester (1999), and Stiroh (2000)).

Two recent studies on Canadian financial institutions address economies of scale assuming a Cobb-Douglas cost function. Using a panel data set of 25 Canadian trust companies for the years 1985-1988, Breslaw and McIntosh (1997) show that the scale function of these trust companies is convex with respect to firm size. Assuming a Cournot oligopoly and using a time-series data set of the “Big Five” banks from 1976 to 1996, McIntosh (2002) finds significant increasing returns to scale among these banks.³

An obvious limitation of the Cobb-Douglas framework is that banks are assumed to produce one (composite) output with the same inputs and technology. This assumption is debatable (and rejected in our sample) given that banks are diversified in their business lines. The Cobb-Douglas framework also has an overly-restrictive functional form. Lawrence (1989) demonstrates that the non-rejection of the Cobb-Douglas technology results from an *ad hoc* specification that excludes the possibility of multiple product cost complementarity. Applying this to British Columbia Credit Unions, Murray and White (1983) show that none of the restrictive production conditions commonly imposed by researchers using the Cobb-Douglas framework provides a valid representation of the technology of the firms they studied. Instead, the authors propose to use a translog specification which captures the heterogeneous nature of a bank’s intermediation activity. Using data from 1976 to 1977, Murray and White (1983) find that most of the credit unions in the sample experience significant increasing returns to scale.

The translog cost function was first proposed by Christensen and Lau (1971). Schmidt and Lovell (1979) show that under the cost minimization assumption, a firm’s stochastic production frontier can be written as a cost function. Diewert and Kopp (1982) show that any frontier cost function, such as the translog function, can be derived without knowing its underlying production function. The translog specification is often used to provide a numerical efficiency value, called X-efficiency, and ranking of firms. The X-efficiency of a bank is measured as its cost level compared to that of the best-practice banks of similar size (the frontier firm), controlled for type of banking activity and the input prices it faces. Inferences regarding scale economies of banks are drawn from the derivative of a bank’s cost with respect to its output. The specification is often applied to a set of cross-sectional data on banks

³The five banks are: Royal Bank Financial Group, Bank of Montreal, Canadian Imperial Bank of Commerce, TD Bank Financial Group, and Bank of Nova Scotia

and estimated for several years. The parameters for economies of scale are averaged over the sample years. Berger and Humphrey (1997) provide a detailed survey on the literature.

While most studies using the translog cost function examine U.S. or European banks, to our knowledge, no study has applied it to a panel of Canadian banks. Nathan and Neave (1992) estimate a translog cost function for a cross-section of Canadian banks. Results on economies of scale depend on whether or not deposits are assumed to be inputs. There are limited studies on Canadian banks because of the number of banks and partly due to the unavailability of individual bank data to the public.⁴ However our panel has a significantly long time dimension and a sufficient cross-sectional dimension to conduct a detailed study of the Canadian banking industry.

A major difficulty in dealing with larger time dimensions relative to the size of the panel is that the assumption of constant firm effects over time is most likely violated. Cornwell and Sickles (1997) propose to solve the problem by replacing the coefficients on firm effects with a flexibly parameterized function of time. Applying this to eight U.S. airlines from 1970q1 to 1981q4, the authors construct an efficiency frontier for the firms and find that the firms become more efficient over time. We follow the approach of Cornwell and Sickles (1997) to allow bank specific effects to vary over time.

A further contribution of this study is the inclusion of non-traditional activities. Most studies measure bank output by their traditional activities, such as loan-generation and security investment. However, banks have been moving into non-traditional activities such as depositor services, underwriting, and wealth management. Excluding these activities will result in a misspecified cost function and potentially lead to incorrect inference about economies of scale.

Clark and Siems (2002) apply the asset equivalent measure of off-balance-sheet (OBS) activities proposed by Boyd and Gertler (1994) to measure the impact of off-balance sheet (OBS) activities on the efficiency measure of banks. The authors find that OBS activities are important determinants in explaining cost efficiency. The idea of an asset equivalent measure is to capitalize the bank's noninterest income to proxy the assets required to produce such revenue. Boyd and Gertler (1994) also argue that the credit equivalent measure proposed by the Basel Committee often underestimates the off-balance-sheet assets of banks.⁵ We follow the framework of Clark and Siems (2002).

⁴As of July 2003, there were 17 domestic chartered banks and 51 foreign banks or subsidiaries in Canada.

⁵Under the current reporting requirement, OBS activities are allocated into four broad risk categories that carry conversion factors of 100 percent, 50 percent, 20 percent and 0 percent.

III. Models

A. Cost Minimization

We assume a bank i , ($i = 1, \dots, N$) is a cost minimizing entity who produces output $Q = (Q_1, \dots, Q_m) \in \mathbb{R}_+^m$ using inputs $X = (X_1, \dots, X_k) \in \mathbb{R}_+^k$ at prices $W = (W_1, \dots, W_k) \in \mathbb{R}_+^k$ subject to a production constraint, $F(Q, X)$.⁶

$$\min C = \sum_{j=1}^k W_j X_j(Q, W)$$

subject to a production constraint

$$F(Q, X) = 0$$

Possible environmental variables and proxies for technological change are included in $G = (G_1, \dots, G_L) \in \mathbb{R}_+^L$. Consistent with most of the firm-efficiency literature we estimate a multi-product translog cost function.⁷ The function is assumed to be positive for all positive prices and outputs, homogeneous of degree one, monotonic, and concave in prices. The total cost function, $C(Q, W)$ can be defined in terms of the unit cost function $c(w)$ as follows: $c(q, w) = q + c(w)$. A second order Taylor expansion around the log of output and prices gives the following cost function. All lower case variables are in logarithms.

$$\begin{aligned} c(q, w) = & \alpha_0 + \sum_{l=1}^m \alpha_k q_l + \sum_{j=1}^k \beta_j w_j + \frac{1}{2} \sum_{l=1}^m \sum_{j=1}^m \sigma_{lj} q_l q_j + \\ & \sum_{l=1}^m \sum_{j=1}^k \gamma_{lj} q_l w_j + \frac{1}{2} \sum_{l=1}^k \sum_{j=1}^k \delta_{lj} w_l w_j + \sum_{l=1}^L \theta_l G_l + \xi + \varepsilon \end{aligned} \quad (1)$$

We do not subscript the variables across time or cross-section explicitly because subscripts indicate the methodology employed. The assumption on the distribution of ξ and ε is also dependent on the methodology. Economic theory imposes certain restrictions on the parameters: the cost function is

⁶Shaffer (1993) tests for banking competition in Canada between 1965 and 1989 and cannot reject input price-taking behaviour. Nathan and Neave (1989) find that the Canadian financial system does not exhibit monopoly power.

⁷The Fourier functional form has been shown by Mitchell and Onvural (1996) to perform better than the translog function when there is a wide variety of banks. This is clearly not the case in our sample and thus we estimate the simpler of the two models.

homogeneous if: $\sum_j^k \beta_j = 1$, $\sum_j^k \gamma_{lj} = 0$, and $\sum_j^k \delta_{lj} = 0$. For simplicity we also impose $\sigma_{lj} = 0$. Imposing these restrictions gives,

$$\begin{aligned} \log(c/w_1) = & \sum_{j=2}^k \beta_j \log(w_j/w_1) + \sum_{l=1}^{k-1} \sum_{j=l+1}^k \delta_{lj} [w_l w_j - 0.5 * (w_l^2 + w_j^2)] \\ & + \sum_{l=1}^m \sum_{j=2}^k \gamma_{lj} q_l \log(w_j/w_1) + \alpha_0 + \sum_{l=1}^m \alpha_l q_l + \sum_{l=1}^L \theta_l G_l + \xi + \varepsilon \end{aligned} \quad (2)$$

It is important to note that we gain estimation efficiency by including input cost share equations in the model. The cost share of input j is given by S_j and is derived using Shephard's lemma. The cost share equations add information without adding parameters to the multivariate regression. Since shares must sum to unity, $j - 1$ input share equations are specified for a system of j shares. Results are invariant to which share is dropped.

$$S_j = \frac{\partial \log(C)}{\partial \log(w_j)} = \beta_j + \sum_l \delta_{lj} w_l + \sum_l \gamma_{lj} q_l, \quad j = 1, \dots, m \quad (3)$$

Imposing symmetry, $\gamma_{lj} = \gamma_{jl}$, gives,

$$S_j = \beta_j + \sum_{l=2}^m \delta_{lj} \log(W_l/W_j) - \sum_{l=1}^m (\gamma_{l2} + \gamma_{l3}) q_l \quad (4)$$

B. Economies of Scale and Technological Changes

One important focus in this study is the economies of scale of the Canadian banking industry. Economies of scale are measured as,

$$\zeta = \left(\sum_{l=1}^m \alpha_l + \sum_{l=1}^m \sum_{i=2}^m \gamma_{li} \log(W_l/W_j) \right)^{-1}$$

There are increasing returns to scale if $\zeta > 1$, constant returns to scale if $\zeta = 1$, and decreasing returns to scale if $\zeta < 1$. Economies of scale inform us of the cost savings/dissavings when a bank increases its output.

We think it is essential to capture technological changes over time. We achieve this in two ways. First, we assume technological change affects the cost function directly, i.e. banks are subject to the

same technological shocks over time. We proxy such shocks by including a quadratic time polynomial. The rate of technological change is given by

$$T^* = -\frac{\partial C}{\partial t} \quad (5)$$

and change is progressive if $T^* > 0$. Change is regressive if $T^* < 0$. Since it is hard to pinpoint when banks adopt those technologies and when the effects on their cost structure are fully realized, such a nonparametric specification of technological change has the advantage of not requiring knowledge of the date of those changes.

Second, if technological changes affect banks differently over time, then there should be time-varying effects in the bank-specific terms. Such time-variant effects will affect the measure of efficiency. To illustrate, let us redefine equation (2) as,

$$y_{it} = X'_{it}\beta + \xi_i + u_{it} \quad (6)$$

where $\xi_i = \alpha_0 + \varepsilon_i$ and u_{it} , ε_i are white noise. For the cost function $X_{it} = [\log(W_j/W_1), (W_iW_j - 0.5 * (W_i^2 + W_j^2)), q_l \log(W_j/W_1), q_l, G_l]'$, and $\beta = [\beta_j, \gamma_l, \delta_l, \alpha_l, \theta_l]'$ for each bank i .

Assuming time-invariant inefficiency, the cost frontier intercept is given by the minimum of the firm-specific effects, ξ_i , and inefficiency is the difference between the frontier, or the “best-practice” firm, and the firm effects. Such deviation from the “best-practice” firm can come from differences in management skills, human inertia, and adoption of technology.

$$\hat{\alpha}_0 = \min_j(\hat{\xi}_j) \quad \text{and} \quad \hat{\varepsilon}_i = \hat{\xi}_i - \hat{\alpha}_0$$

Assuming that cost efficiency is constant over time is implausible given the long time dimension of the panel. We allow time-varying cost efficiency. Consider a fixed effects approach. For time-varying firm inefficiency,

$$\hat{u}_{it} = \Omega_{i1} + \Omega_{i2}t + \Omega_{i3}t^2$$

and the time-varying fixed effect is

$$\hat{\xi}_{it} = \hat{\Omega}_{i1} + \hat{\Omega}_{i2}t + \hat{\Omega}_{i3}t^2$$

This specification allows cost efficiency to be time-varying as well as different for each bank. The time-invariant case is nested if the same firm is selected for all t .

$$\hat{\alpha}_t = \min_j(\hat{\xi}_{jt}) \quad \text{and} \quad \hat{\varepsilon}_{it} = \hat{\xi}_{it} - \hat{\alpha}_t$$

Cost efficiency is derived as

$$CE_{it} = \exp\{-\hat{\varepsilon}_{it}\}$$

Two components of efficiency can be distinguished: technical efficiency, the ability to obtain maximum output from a given set of inputs, and allocative efficiency, the skill to use the inputs in optimal proportions, given their respective prices and the production technology. The two can be combined to provide a measure of economic efficiency, or, when cost instead of production is considered, cost efficiency. Measures of cost efficiency allow us to rank the banks over time from the most to the least cost-efficient. The most efficient bank has a measure of cost efficiency equal to one and less efficient banks have measures below one. Measures of cost efficiency are consistent as $T \rightarrow \infty$ if the model is estimated by maximum likelihood and consistent if $T \rightarrow \infty$ and $I \rightarrow \infty$ if estimated by generalized least squares. The particular estimators are discussed in section IV.

IV. Methodology

In this section we discuss alternative ways used to estimate the cost function. Valid estimation techniques depend on the assumptions we are willing to make.⁸ The unique panel that we have consists of six banks and eighty-three quarterly observations. The availability of a long time-series and short cross-section leads to several natural parametric estimators.

A. Fixed Effects Model

Recall equation (6), a generic unobserved effects model.

$$y_{it} = X'_{it}\beta + \xi_i + u_{it}$$

⁸As McIntosh (2002) notes, valid inference requires stationarity. Phillips and Moon (1999) show the advantage of pooling the cross-section and time series components to get consistent estimates of the long-run coefficients.

The fixed effects model assumes that we can capture differences across banks in the constant term, ξ_i . These effects can be correlated with X_{it} . However, we do need $E[u_{it}|X_{it}, \xi_i] = 0$. An alternative assumption would be to assume ξ_i comes from some distribution. However, in our case the cross-section draws are not random and it would be inappropriate to estimate the model using a random effects estimator.⁹ The fixed effect estimator of β and the individual effect ($\xi_i + \mu_{it}$) are consistent as $T \rightarrow \infty$. The large-T component also allows us to evaluate the cost inefficiency apart from the statistical noise.

B. Stochastic Frontier Model

The stochastic frontier methodology decomposes the fixed effect in equation (6) into a constant α_0 and a firm-specific inefficiency variable, ε_i . This framework allows us to calculate the efficiency level of each Canadian bank relative to the “best practice” firm in the sample, rather than to the absolutely efficient firm. Inefficiencies are assumed to follow the truncated normal distribution, while the random errors follow a standard normal distribution. The logic is that the inefficiencies must have a truncated distribution because inefficiencies cannot be negative. Both the inefficiencies and the errors are assumed to be orthogonal to the explanatory variables in the model. Relative to the fixed effects estimator, the truncated normal assumption may be overly restrictive because it clusters near full efficiency. We estimate this model by the maximum likelihood (ML), which is consistent under the following

- $u_{it} \sim iidN(0, \sigma_u^2)$
- $\xi_i \sim iidN^+(\mu, \sigma_\xi^2)$
- u_{it} and ξ_i are independently distributed.

The log likelihood is given by (see Kumbhakar (2002) for details)

$$\log L = -\frac{I(T-1)}{2} \log \sigma_u^2 - \frac{I}{2} \log(\sigma_u^2 + T\sigma_\xi^2) + \sum_i \log \left[1 - \Phi \left(-\frac{\mu_i^*}{\sigma^*} \right) \right] - \left(\frac{e'e}{2\sigma_u^2} \right) + \frac{1}{2} \sum_i \left(\frac{\mu_i^*}{\sigma^*} \right)^2$$

where $\mu_i^* = \frac{T\sigma_\xi \bar{\varepsilon}_i}{(\sigma_u^2 + T\sigma_\xi^2)}$, $\sigma^* = \left(\frac{\sigma_\xi^2 \sigma_u^2}{(\sigma_u^2 + T\sigma_\xi^2)} \right)^{1/2}$, and $\bar{\varepsilon}_i = (1/T) \sum_i (\xi_i + u_{it})$.

⁹For completeness we did estimate the model assuming random effects. The Breusch and Pagan Lagrange multiplier test for random effects rejects that assumption.

C. Seemingly Unrelated Regressions (SUR)

The cost specifications so far have not taken into account the share equations. Imposing cross-equation restrictions can improve estimator efficiency. The system of equation is estimated by ML.

$$\begin{aligned}C_{it} &= A_{it}\Theta + \xi_{it} + u_{it} \\S_{Kit} &= B_{it}\Gamma + \eta_{Ki} + v_{Kit} \\S_{Lit} &= D_{it}\Phi + \eta_{Li} + v_{Lit} \quad i = 1, \dots, N \quad t = 1, \dots, T\end{aligned}\tag{7}$$

The share equation with SUR estimation provides sufficient structure to identify the coefficients. Consistent estimation of each equation separately leads to poorly identified parameter estimates due to multicollinearity. As expected, the assumption of zero contemporaneous correlation is strongly rejected.

V. Data

The data are quarterly observations of the six largest banks in Canada starting from 1983q1 to 2003q3 deflated by the GDP deflator (1997=100). Our choice on the number of banks is limited by data availability. The six banks are: Royal Bank Financial Group, Bank of Montreal, Canadian Imperial Bank of Commerce, TD Bank Financial Group, Bank of Nova Scotia and National Bank. They are the only banks for which data are available in the entire sample period. The majority of the rest of the domestic chartered banks did not start reporting until after 2000. The big six account for approximately 90 percent of the Canadian banking industry in terms of total assets. The data set is from the chartered banks' consolidated monthly balance sheet and quarterly consolidated statement of income collected by the Office of the Superintendent of Financial Institutions Canada (OSFI). The consolidated monthly balance sheet data at the aggregate level are published in Tables C1 and C2 in the Bank of Canada Banking and Financial Statistics. Large categories of the consolidated statement of income at the aggregate level are available in Table K2 in the same publication with an annual frequency.¹⁰ All balance sheet data are end-of-month values and are converted to quarterly series by taking the quarterly average. Appendix A provides a detailed definition of all variables used in this study.

¹⁰Disaggregate data are confidential.

Since the definition of a bank's inputs and outputs is an on-going debate and we have opted for the intermediation approach. This is the approach commonly used in the conventional cost function literature. In the intermediation approach a bank is assumed to use labour, capital and deposits to produce earning assets. Deposits are treated as an input. On the other hand, the production approach postulates that banks also provide value-added in their deposit services. Deposits are treated as an output under the production approach.

There are three input prices. L is the hourly wage of a bank's full-time equivalent employees. K is capital cost, measured by the expense on premises, and computer and equipment divided by the total stock of premises and fixed assets on the bank's balance sheet. D is the price of deposits, measured by the total interest expense on total deposits divided by total deposits.

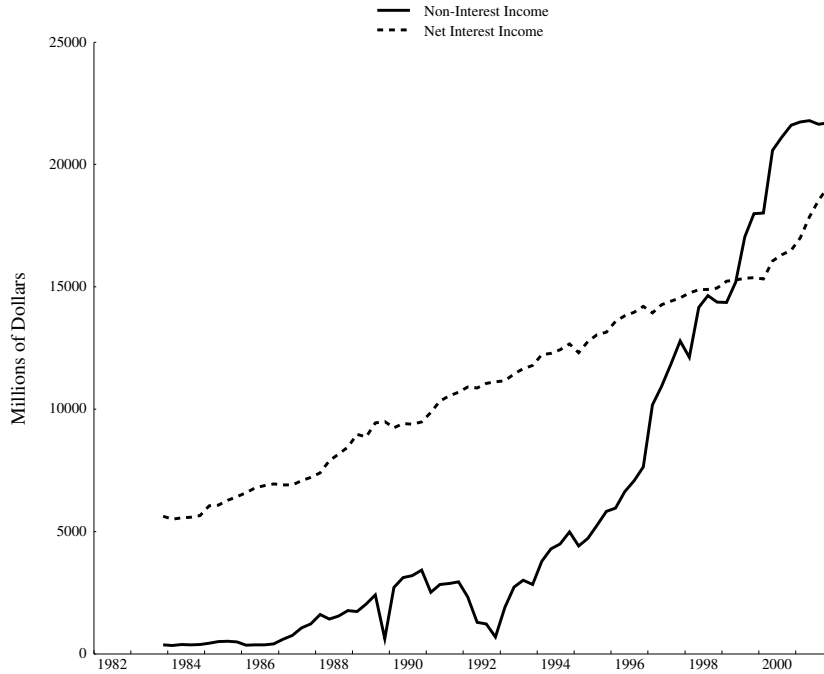
We define five output categories: Y_1 , consumer loans; Y_2 , non-mortgage loans; Y_3 , mortgage loans; Y_4 , security investment and Y_5 , non-traditional banking activities. The first four categories are taken from the asset side of a bank's balance sheet.¹¹

Non-traditional banking activities are often overlooked in the literature of bank efficiency, and scale economies. They are non-interest related activities such as depositor services, underwriting, foreign exchange trading and wealth management. Figure (1) shows that non-interest income has increased substantially since the late 1980s and has exceeded net interest income. The chart shows that banks have been shifting away from traditional lending and investment activities to non-traditional activities, a trend that is also observed in U.S. banks (c.f. Stiroh (2000)).

Nontraditional activities can be divided into three categories according to their underlying asset or liability position. The first type is from on-balance-sheet assets that are captured in the output measures mentioned above. Examples include mortgage loan fees, gains and losses from trading and investment activities, and insurance income. The second type is from the liability-side of the balance sheet, such as deposit account fees and payroll processing fees. Because of data limitations, this type of output is often omitted in studies using an intermediary approach. The third type of nontraditional activity comes from off-balance-sheet activities like securities underwriting, wealth management, the provision of loan guarantees and letters of credit. These activities are also unavailable under the conventional approach to measuring bank output.

¹¹All assets are reported in book value, except securities that are categorized under the trading account.

Figure 1. Canadian Chartered Bank Total Net Interest Income vs. Noninterest Income



Rogers (1998) suggests including non-interest income in the cost function as a measure of non-traditional activities. However, the resulting cost function mixes stocks (asset) and flows (revenue). A more internally consistent approach is introduced by Boyd and Gertler (1994). They introduce an asset-equivalent measure of these non-traditional activities. The authors assume that all non-interest income is generated from off-balance-sheet assets. Assuming that these non-traditional activities yield the same rate of return on assets (ROA) as traditional activities, the assets that are required to produce non-interest income can be calculated by dividing non-interest income by the ROA of traditional activities. Their definition of bank profits is as follows:

$$\pi = II - IE - PROV - NE + NII$$

where II is total interest income; IE is total interest expense; $PROV$ is loan loss provisioning; NE is non-interest expense, and NII is non-interest income. Boyd and Gertler (1994) assume that (i) non-

interest income is generated by some hypothetical off-balance-sheet asset A_o and that (ii) using the same capital and liabilities, A_o generates the same rate of return as on-balance-sheet asset A_b . The asset-equivalent measure (AEM) of non-traditional activities is given by,

$$AEM = A_o = A_b[NII/II - IE - PROV]$$

This measure can be viewed as the hypothetical asset holdings that would be required to generate non-interest income. It has the convenience of using available data and it is easy to understand. There are some limitations. AEM includes elements that have been captured by other measures of output. As mentioned, some components of non-interest income are also generated from on-balance sheet assets. Ideally, we would like to be able to subtract those components from non-interest income. Unfortunately, disaggregate data of non-interest income are only available from 1997 onwards.¹² We consider this the best available proxy for such activities. A second deficiency is the assumption that off-balance-sheet assets yield the same rate of return as on-balance-sheet assets. This ignores the risk aspect of the two types of assets. It is also possible that, if banks can achieve diversification effects, the two returns could be negatively correlated. Calmès and Liu (2003) have shown these effects to be minimal.

A bank's cost function can also be influenced by exogenous factors, such as changes in the regulatory environment. However, we should note that regulatory changes may not have been aiming at reducing costs or improving efficiency. Rather, they are often the product of a convolution of forces, such as technological advances, demographic changes and global trends. However, by changing the regulatory environment in which banks operate, such changes may not have helped banks in their cost-minimizing/profit-maximizing objectives.

Three notable changes to the Bank Act took place in our sample period. The most influential one occurred in 1987, when Canadian banks were permitted to invest in corporate securities, as well as distribute government bonds. Banks are now allowed to purchase control of investment dealers and invest in the securities business. As a result, banks have substantially increased their financial-market based activities. On the demand side, as bank customers began to invest in the financial market directly through their banks, the amount of direct financing (for example, financing done through financial markets as opposed to through financial intermediaries) also increased.

¹²Non-interest income generated from on-balance-sheet assets account for approximately 20 percent of total non-interest income from 1997 to 2003.

The situation evolved further when banks were given the right in 1992 to enter the trust business through the establishment or acquisition of trust companies in Canada. Over the subsequent years, the major banks have bought up most of the trust companies in Canada. They were also allowed to offer a number of in-house activities such as portfolio management and investment advice. These changes may have attracted a larger fraction of depositors to invest in financial markets directly through their banks. Finally, various changes to update and refine the amendments made in 1992 took place in 1997.

Figure (1) shows a gradual launch of the banks' non-interest income in 1987, reflecting a trend towards more non-loan-oriented activities. This trend grew further after the amendment in 1992 and continued throughout the 1990s.

Besides regulatory changes which permitted banks to diversify their business mix, developments in legal reserve or capital requirements aimed at ensuring the financial soundness of banks may have affected the banks' cost structure. In 1989, Canadian banks adopted the minimum capital requirement proposed by the Basel Committee on Banking Supervision. This may have occasionally affected a bank's output decisions if capital requirements were binding. In addition, the removal of legal reserve requirement in 1991 and the complete phasing out of reserve requirement by banks in 1994 is likely to have an impact on the banks' production decisions.

To investigate whether such institutional changes in a bank's revenue source would have a significant statistical impact on a bank's cost structure, we include three regulatory dummies, 1987Q2, 1989Q1, 1991Q1, 1992Q1, 1994Q1 and 1997Q1, in our cost function. The dummies are zero before these dates and one afterwards.

VI. Findings

Equations (2) and (4) are estimated by SUR. Equation (2) is also estimated by fixed effects-GLS and by single-equation MLE. For each method we estimate two models. The first model, labelled Model REG, includes statistically significant regulatory changes. The second model, labelled Model T (for time dummies), includes a time polynomial which is assumed to proxy technological change.¹³ Complete parameter estimates of each model are presented in the appendix.

¹³When both types of dummies are included in the model, the cost function is non-concave everywhere.

A. Model Evaluation

Before we discuss the results of economies of scale and relative efficiency, we conduct two model evaluation checks. The first one is on the own- and cross-price elasticity of the cost function. Table I presents the own- and cross-price elasticities for the SUR specification. All own-price elasticities are negative and their absolute values are less than one, implying that all inputs are price-inelastic. Capital has the highest own-price elasticity among the three inputs. This is interesting because capital expense makes up the smallest share of input. Deposits, by far the largest component in the input mix, have the lowest price sensitivity. This is reasonable as banks are likely to shift any price changes in its sources of funding to their customers by charging a higher rate of interest on loans or demanding a higher rate of return on its securities. However, changes in capital expenses may be much harder to absorb. The cross-price elasticities between labour and capital, and between labour and deposits are roughly equal, suggesting similar substitutability between the two pairs of inputs. The cross-price elasticity between capital and deposits is negative but statistically insignificant at conventional levels.

Table I
Own- and Cross-Price Elasticities

	Capital	Labor	Deposits
Capital	-0.8709 (0.0159)		
Labor	0.6862 (0.1241)	-0.8000 (0.0970)	
Deposits	-0.2746 (0.2862)	0.7709 (0.1442)	-0.3391 (0.1233)

Standard deviations are in parentheses

The second check on the model specification is testing restrictions on the production structure. Table II shows the results on testing restrictions of homotheticity and homogeneity against the unrestricted cost function. The production function is homothetic if $\gamma_{lj} = 0, \forall l, j$. and homogeneous of degree one if $\sum_l \alpha_l = 1$. The homothetic or homogenous production structure restrictions typically imposed by researchers using a Cobb-Douglas framework is strongly rejected for the Canadian banking sector.¹⁴

¹⁴We only report results for the stochastic frontier model. Results for the fixed effects model are similar.

Table II
Tests of Production Structure

Model	Log Likelihood	Test Statistic	Degrees of Freedom	P-value
Unrestricted model				
Model REG	999.56		31	
Model T	1021.96		31	
Homotheticity				
Model REG	886.31	226.38	16	0.0000
Model T	901.21	241.50	16	0.0000
Homogeneity				
Model REG	978.95	41.22	24	0.0000
Model T	990.28	63.36	24	0.0000

B. Economies of Scale

Results on economies of scale are presented in Table III. We test the null hypothesis of constant returns to scale (CRS) and present test statistics and the p-values of the test. The likelihood ratio test of CRS is distributed chi-squared with one degree of freedom under the null hypothesis. Overall, the results from all six models are in favour of increasing returns to scale. All six economies of scale measures are greater than one and statistically significant. The economies of scale measures are smaller under Model T than under Model REG. Evaluated at the sample mean, the measured economies of scale under Model T range between six percent and thirteen percent. That is, a 1 percent increase in output will raise production cost by 0.87 percent to 0.94 percent, depending on the methodology. The implied cost savings are even higher under Model REG, ranging between eighteen to twenty percent. This magnitude is stronger than those typically found in the literature on large banks. For example, Stiroh (2000) finds significant economies of scale in medium to large sized US Banking Holding Companies of about five percent.

As a robust check, we also evaluate the economies of scale for the sub-periods 1983-1986, 1987-1996 and 1997-2003, to reflect changes in the Bank Act that are found to be significant in our model. All six models produce significant economies of scale in the three sub-periods. The magnitude of average cost savings associated with a higher output, however, changes over time. It seems to be consistent across all models that the degree of economies of scale dropped in 1987-1996, and reached the highest in 1997-2003. As a further robust check, we exclude non-traditional activities as a bank

output. The measured economies of scale is still significant but slightly smaller, ranging from three to eighteen percent, depending on the model.

Table III
Economies of Scale

Model	ζ	$H_0 : \zeta = 1$	
		Statistic	P-value
SUR			
Model REG	1.181	67.47	0.0000
Model T	1.064	9.48	0.0021
single-FEGLS			
Model REG	1.197	63.52	0.0000
Model T	1.126	19.73	0.0000
single-MLE:			
Model REG	1.176	50.08	0.0000
Model T	1.083	7.50	0.0062

^aThe restriction imposed on equation (2) is actually $\zeta^{-1} = 1$ and $\sum_j \delta_{lj} = 0 \forall l$ since returns to scale is defined as $\frac{\partial C}{\partial q_l} = \sum_l \alpha_l + \sum_j \delta_{lj} \log(\bar{P}_j / \bar{P}_1)$ where $\bar{\cdot}$ is the sample mean

C. Technological and Regulatory Changes

Recall that our proxy for technological change is given by equation (5). Tests regarding T^* are conducted using likelihood ratio tests under the null hypothesis that $T^* = 0$. We find strongly significant evidence of increasing cost-efficiency, on the order of 1% per quarter over the sample period in all three models.

We have included six regulatory dummies in our cost function: 1987, 1989, 1991, 1992, 1994 and 1997. As expected, the dummies for 1987 and 1997 are highly significant in the cost function, suggesting that regulatory changes to the Bank Act in those years to allow banks to be more diversified did have an impact on the banks cost structure. However, the other dummies were not significant and thus were not included in the regressions. All models have negative estimated coefficients for the dummies, implying that regulatory changes led to a decrease in costs. The results are similar across models, suggesting roughly a five percent cost savings following either regulatory change.

D. Relative Efficiency

Measures of time-invariant and time-varying relative efficiency are calculated from the fixed effects model and stochastic frontier model for the big six Canadian banks. Because of confidentiality we cannot provide point estimates. However, the range of inefficiency of the six banks is approximately twenty percent. These results are similar to those found in the literature on U.S. banks and bank holding companies.

Our calculations of time-varying cost-efficiency from the stochastic frontier model are almost identical to those from the fixed firm-effect model. Overall, there is little change in the ranking of relative efficiency in our sample period. The average of the measures of inefficiency is very similar to those from the time-invariant model. The dispersion of inefficiency between banks seems to have narrowed over time.

VII. Conclusion

We have applied the flexible translog cost function framework to study the efficiency, economies of scale and scope of the six largest Canadian banks. Using a unique panel dataset from 1983q1 to 2003q3, we estimate three econometric models based on this framework. Given the long time dimension of the dataset, we add a time trend and a time trend squared in the cost function to capture any technological change over time. We find that banks have experienced significant technology progress over the past 20 years. We also find that regulatory changes have helped to reduce the production cost of banks

Overall, we reject the hypothesis of constant returns to scale. That is, banks can enjoy cost savings by increasing their scale of production. This finding is similar to recent studies on large U.S. banks but our measured economies of scale is substantially larger than those found in the U.S. studies.

Measures of efficiency are derived from the efficiency frontier estimated by the stochastic frontier model. On average, the inefficiency of Canadian banks is around ten percent, close to what is typically found in the literature on U.S. banks using the cost function approach. The ranking of efficiency also suggests that larger banks seem to be more cost-efficient than smaller banks. Given that scale economies are already accounted for in our model, such heterogeneous effects may come from differences in other factors such as management skills and adoption of technology. Our time-varying

fixed-effect panel specification allows us to trace the changes in efficiency levels over time. The results suggest that the relative efficiency level of banks has remained unchanged, and the relative efficiency dispersion among banks seems to have narrowed.

VIII. Tables

Table IV: **SUR: Model REG**

Variable	Equations		
	Equation $\log(c)$	Equation S_k	Equation S_l
D_1	-0.09742** (0.01046)		
D_2	-0.09887** (0.01094)		
α_1	-0.23412** (0.08385)		
α_2	0.42090** (0.07851)		
α_3	0.05055 (0.05574)		
α_4	0.36156** (0.10125)		
α_5	0.17454** (0.05351)		
β_1		-0.24733** (0.03077)	
β_2	-1.21306** (0.06539)		-1.21306** (0.06539)
β_3	2.46039** (0.07454)		
δ_{12}	0.03101** (0.00171)	0.03101** (0.00171)	0.03101** (0.00171)
δ_{13}	-0.04859** (0.00132)	-0.04859** (0.00132)	

Continued on next page...

... table IV continued

Variable	Equations	
δ_{23}	-0.12333** (0.00305)	-0.12333** (0.00305)
γ_{11}	0.00335 (0.00286)	
γ_{12}	0.03074** (0.00597)	0.03074** (0.00597)
γ_{13}	-0.03408** (0.00704)	
γ_{21}	-0.0102** (0.00261)	
γ_{22}	-0.00618 (0.0054)	-0.00618 (0.0054)
γ_{23}	0.01637** (0.00625)	
γ_{31}	0.01437** (0.00192)	
γ_{32}	0.00496 (0.004)	0.00496 (0.004)
γ_{33}	-0.01933** (0.00472)	
γ_{41}	0.0036 (0.00352)	
γ_{42}	-0.01254 [†] (0.00725)	-0.01254 [†] (0.00725)
γ_{43}	0.00894 (0.00839)	
γ_{51}	-0.01593** (0.0018)	

Continued on next page...

... table IV continued

Variable	Equations	
γ_{52}	-0.01085** (0.0037)	-0.01085** (0.0037)
γ_{53}	0.02678** (0.00442)	
Log-likelihood	3656.7	

Table V
Fixed Effects

Variable	Model REG		Model T	
	Coefficient	Std. Error	Coefficient	Std. Error
α_1	-1.37128*	(0.54683)	-1.09263*	(0.52506)
α_2	3.13151**	(0.4852)	2.59261**	(0.46146)
α_3	-0.92259*	(0.35976)	-1.28914**	(0.34680)
α_4	0.38785	(0.61813)	0.56971	(0.59401)
α_5	-0.56202 [†]	(0.29758)	-0.18192	(0.28848)
β_2	4.20985**	(0.82177)	4.15805**	(0.79775)
β_3	-0.20342	(0.30928)	-0.65401*	(0.29332)
δ_{12}	0.33198**	(0.05356)	0.33826**	(0.05099)
δ_{13}	-0.17716**	(0.02256)	-0.18761**	(0.02154)
δ_{23}	0.03414	(0.02387)	0.04373 [†]	(0.02266)
γ_{12}	0.12243**	(0.0456)	0.09445*	(0.04385)
γ_{13}	-0.04647 [†]	(0.02543)	-0.07026**	(0.02379)
γ_{22}	-0.21692**	(0.0404)	-0.18215**	(0.03829)
γ_{23}	0.14395**	(0.01901)	0.11644**	(0.01948)
γ_{32}	0.07832**	(0.02987)	0.12281**	(0.02894)
γ_{33}	-0.09132**	(0.01827)	-0.07344**	(0.01781)
γ_{42}	-0.0243	(0.0517)	-0.03226	(0.04965)
γ_{43}	-0.04092	(0.02559)	0.00149	(0.0248)
γ_{52}	0.06105*	(0.02389)	0.03021	(0.0232)
γ_{53}	0.06413**	(0.01625)	0.06698**	(0.01552)
α_0	-20.84538*	(8.50474)	-20.92548*	(8.20029)
D1	-0.04903**	(0.01065)		
D2	-0.05208**	(0.01087)		
θ_1			-0.00944**	(0.00102)
θ_2			0.00005**	(0.00001)
N	498		498	
R ²	0.9907		0.9914	
F _(27,470)	2265.9		2464.4	

Table VI
Stochastic Frontier Model

Variable	Model REG		Model T	
	Coefficient	Std. Error	Coefficient	Std. Error
α_1	-1.3446*	(0.53479)	-1.01653*	(0.51572)
α_2	3.22947**	(0.47598)	2.73162**	(0.45644)
α_3	-0.89810*	(0.35182)	-1.31962**	(0.33982)
α_4	0.44000	(0.60459)	0.59944	(0.58132)
α_5	-0.62384*	(0.29171)	-0.21008	(0.2821)
β_2	4.25144**	(0.80359)	4.12266**	(0.77941)
β_3	-0.23031	(0.30256)	-0.70444*	(0.28788)
δ_{12}	0.31716**	(0.05271)	0.31407**	(0.05082)
δ_{13}	-0.17143**	(0.02217)	-0.17941**	(0.02136)
δ_{23}	0.03165	(0.02336)	0.03974 [†]	(0.02226)
γ_{12}	0.12183**	(0.04459)	0.09109*	(0.04299)
γ_{13}	-0.03984	(0.02499)	-0.05723*	(0.02401)
γ_{22}	-0.22353**	(0.03958)	-0.19136**	(0.03774)
γ_{23}	0.14899**	(0.0187)	0.12593**	(0.0194)
γ_{32}	0.07591**	(0.02922)	0.12573**	(0.02836)
γ_{33}	-0.09451**	(0.01791)	-0.08028**	(0.0177)
γ_{42}	-0.02919	(0.05057)	-0.03567	(0.04858)
γ_{43}	-0.04268 [†]	(0.02503)	-0.00366	(0.0244)
γ_{52}	0.06525**	(0.02339)	0.03103	(0.02269)
γ_{53}	0.05976**	(0.01599)	0.0605**	(0.0153)
α_0	-23.08205**	(8.34915)	-23.49573**	(8.10225)
D1	-0.0491**	(0.01042)		
D2	-0.05591**	(0.01075)		
θ_1			-0.00987**	(0.001)
θ_2			0.00005**	(0.00001)
N	498		498	
Log-likelihood	938		961.3	
$\chi^2_{(22)}$	51834		55467	

References

- Berger, Allan N., and David B. Humphrey (1997) 'Efficiency of financial institutions: International survey and directions for future research.' *European Journal of Operational Research* 98, 175–212
- Berger, Allan N., and Loretta J. Mester (1999) 'What explains the dramatic changes in cost and profit performance of the u.s. banking industry?' Wharton School Center for Financial Institutions, University of Pennsylvania, working paper 99-10
- Boyd, J., and M. Gertler (1994) 'Are banks dead? or are the reports greatly exaggerated.' Federal Reserve Bank of Minneapolis Quarterly Review
- Breslaw, Jon A., and James McIntosh (1997) 'Scale efficiency in canadian trust companies.' *Journal of Productivity Analysis* 8, 281–292
- Calmès, C., and Y. Liu (2003) 'Financial structure change and banking riskiness: the case of canada.' Bank of Canada Working Paper, forthcoming
- Christensen, L.R., D.W.Jorgenson, and L.J. Lau (1971) 'Conjugate duality and the transcendental logarithmic production function.' *Econometrica* 39, 255–256
- Clark, Jeffrey A., and Thomas F. Siems (2002) 'X-efficiency in banking: Looking beyond the balance sheet.' *Journal of Money, Credit, and Banking* 34, 987–1013
- Cornwell, C., P. Schmidt, and R.C. Sickles (1997) 'Production frontiers with cross-sectional and time-series variation in efficiency levels.' *Journal of Econometrics* 46, 185–200
- Diewert, W.E., and R.J. Kopp (1982) 'The decomposition of frontier cost function deviations into measures of technical and allocative efficiency.' *Journal of Econometrics* 19, 319–331
- Ferrier, G.D., and C.A.K.Lovell (1990) 'Measuring cost efficiency in banking.' *Journal of Econometrics* 46, 229–245
- Freedman, C, and C. Goodlet (1998) 'The financial services sector: Past changes and future prospects.' Bank of Canada Technical Report 82
- Kumbhakar, S.C (2002) *Stochastic Frontier Analysis*, first ed. (Cambridge University Press)
- Lawrence, Colin (1989) 'Banking costs, generalized functional forms, and estimation of economies of scale and scope.' *Journal of Money, Credit and Banking* 21, 368–379

- McIntosh, James (2002) 'A welfare analysis of canadian chartered bank mergers.' *Canadian Journal of Economics* 35, 457–475
- Mitchell, K., and N.M. 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
- Murray, John D., and Robert W. White (1983) 'Economies of scale and economies of scope in multi-product financial institutions: A study of british columbia credit unions.' *Journal of Finance* 3, 887–902
- Nathan, Alli, and T. Neave (1989) 'Competition and contestability in canada's financial system: Empirical results.' *Canadian Journal of Economics* 3, 576–594
- (1992) 'Operating efficiency of canadian banks.' *Journal of Financial Services Research* 6, 265–276
- Phillips, P.C.B., and H.R. Moon (1999) 'Linear regression limit theory for nonstationary panel data.' *Econometrica* 67, 1057–1111
- Rezvanian, Rasoul, and Seyed Mehdian (2002) 'An examination of cost structure and production performance of commercial banks in singapore.' *Journal of Banking and Finance* 26, 78–98
- Rime, Bertrand, and Kevin J. Stiroh (2003) 'The performance of universal banks: Evidence from switzerland.' *Journal of Banking and Finance, forthcoming*
- Rogers, Kevin E (1998) 'Nontraditional activities and the efficiency of us commercial banks.' *Journal of Banking and Finance* 22, 467–482
- Schmidt, P., and C.A.K. Lovell (1979) 'Estimating technical and allocative inefficiency relative to stochastic production and cost frontier.' *Journal of Econometrics* 9, 343–366
- Shaffer, S. (1993) 'A test of competition in canadian banking.' *Journal of Money, Credit, and Banking* 25, 49–61
- Stiroh, Kevin J (2000) 'How did bank holding companies prosper in the (1990)s?' *Journal of Banking and Finance* 24, 1703–1747

Appendix A. Data Description

Table VII
Data Description

Variable Definitions		
Y1	Consumer Loans	Dollar value of personal loans for non-business purposes
Y2	Non-mortgage loans	Dollar value of secured call and other loans to investment dealers and brokers + loans to regulated financial institutions + loans to domestic and foreign governments + lease receivables + reverse repurchase agreements + loans to individuals and others for business purposes
Y3	Mortgage loans	Dollar value of residential and non-residential mortgage loans
Y4	Other	Dollar value of other financial assets on a bank's balance sheet
Y5	OBS	Asset-equivalent measure of off-balance-sheet activities
L	Price of Labor	Total salaries, pensions and other staff benefits divided by the number of full-time equivalent employees and hours in a year
K	Price of capital	Rental expense on real estate and depreciation on premises, furniture, fixture, computer and equipment divided by total stock of land, buildings, and equipment, less accumulated depreciation
D	Price of deposits	Total interest expense on deposits divided by the total dollar amount of deposits
C	Variable costs	interest cost + labour expenses + capital costs