Review of Income and Wealth
Series 50, Number 2, June 2004

THE ROLE OF INDUSTRIAL CLASSIFICATION IN THE MICRO-MACRO INTEGRATION: THE CASE OF THE BANKING BUSINESS IN THE 1997 NORTH AMERICAN INDUSTRIAL CLASSIFICATION SYSTEM

BY TAREK M. HARCHAOUI*

Statistics Canada

Although economic classification is not part of the Ruggles's prodigious contributions to the System of National Accounts, it is certainly meant to help achieve the integration and linking of macrodata with microdata. Unfortunately, economic classification is a component of the statistical infrastructure that often remains unquestioned by the existing industrial organization literature. This paper fills this gap using the banking business under the 1997 North American Industry Classification System (NAICS) as an example. More specifically, the paper ascertains the extent to which NAICS succeeds at combining the various activities performed by Canadian banks into homogeneous industries. Assuming that producing units within the same industry should display more similar cost structure than those in less similar industries, we find that NAICS—at least for the banking sector—is successful at identifying and grouping producing units into homogeneous economic activities. This result is particularly helpful for empirical research that relies on microdata to draw inferences on the structures, conduct and economic performance of the banking sector as whole.

1. INTRODUCTION

Dostoevsky apparently once remarked that all of Russian literature emerged from under Gogol's *Overcoat*. It is at least as true that all of the recent literature on the dynamics of firms and the extent to which it explains aggregate trends has emerged from the Ruggles's work on the integration and linking of macrodata with microdata (Ruggles and Ruggles, 1999). The Ruggles's pioneering work in this field produced a file that they called the Longitudinal Establishment Database, which later evolved into the Longitudinal Research Database maintained and updated by the Center of Economic Studies at the U.S. Bureau of Census. The database is now available to a wide range of researchers and has already given rise to numerous studies.[1] Similar developments have also taken place at Statistics Canada where longitudinal data have been exploited to address several public policy issues.[2]

Although the Ruggleses never touched on issues related to economic classification, this theme was somewhat implicit throughout their contributions on the micro-macro integration. For example, in their 1986 *Review of Income and Wealth*

Note: I am indebted to Georges Dionne, Helen Tice, Paul Warren and two anonymous referees for several comments and suggestions made on an earlier draft. The usual caveats apply.

*Correspondence to: Tarek M. Harchaoui, Statistics Canada, Microeconomic Analysis Division, R.H. Coats Building, 18-F, Ottawa, K1A OT6, Canada (harctar@statcan.ca).

[1]See Foster *et al*. (2001), for example.

[2]Statistics Canada's research based on longitudinal microdata may be accessed at http://www.statcan.ca/english/studies/eaupdate/

paper, Richard and Nancy Ruggles presented the economic classification as an important way to achieve this integration:

> Such a marriage of macro and micro data adjustments on both sides. In the first place, the universe of reporting units covered in the macrodata and the microdata should be the same. On the macro side, what is needed are sectors composed of well-defined and institutionally similar reporting units . . . On the micro side, the concern should be with comprehensiveness of coverage and the appropriateness of reporting units. This does not mean that every microdata set must correspond in coverage to the entire macro sector to which it relates, but rather that it should correspond to an identifiable segment of that sector, for which it is possible to establish control totals. It should, further, be composed of reporting units that can be related, in an identifiable way, to the units included in the macro account.

Since an industry's characteristics must be compiled from data reports of its constituent firms, assigning firms to their correct industry is critical. If firms are not properly separated into distinct industries, the firm data that compose the industry observation will be arbitrary. Hence statistical results and, along the way, the integration and the linking of macrodata with microdata, will be arbitrary as well.[3]

Despite its importance, the economic classification is part of the statistical infrastructure that rarely is discussed in its own right. Researchers typically restrict their interest to whether a particular activity relevant to their own studies can be found in the classification but the overall structure and design of the classification is usually taken as a given.

Recently, the concept of industrial classification has undertaken a new perspective. With the signing of the North American Free Trade Agreement (NAFTA) and increasing globalization, there is an increasing demand for internationally comparable data and particularly comparability between North American countries. In addition, the statistical agencies of the three countries have agreed to give special attention for service industries in general.

The banking business represents an interesting case in point for two primary reasons. First, the banking business is being dramatically redrawn and, therefore, poses many classification problems. Banks' activities have broadened considerably beyond the core of business of lending, deposit-accepting, and traditional service functions. In addition, new domestic and foreign competitors are entering the business, some doing so as "virtual" institutions. Second, the banking business poses some interesting challenges in the measurement of its output—another area where the Ruggleses have made an important contribution.

This paper investigates how well the hierarchical structure of the 1997 North American Industry Classification System (NAICS) separates banks' producing

[3]An example of how the SIC system is critical in empirical research can be found in Davis *et al.* (1996). Using plant-level data, they examine restructuring in the U.S. steel industry. This restructuring involved downsizing and, in some cases, exit of large integrated mills and the appearance and growth of special mini-mills. The two types of mills employ distinct technologies and operate under radically different cost structures. Analysis performed with aggregate data would not reveal the economically significant churning within the industry.

TABLE 1

STRUCTURE OF BANKING ACTIVITIES IN CANADA

	1996		1997		1998	
	$ M	%	$ M	%	$ M	%
Retail banking	24,099	67.3	27,749	67.2	28,166	67.7
Corporate and institutional finance	3,079	8.6	3,472	8.4	3,380	8.1
Electronic financial services	1,540	4.3	1,729	4.2	1,902	4.6
Investment banking and security dealing	6,624	18.5	7,852	19.0	7,503	18.0
Fiduciary services	466	1.3	479	1.2	661	1.6
Total consolidated	35,808	100.0	41,281	100.0	41,612	100.0

units into economically distinct industries. The maintained hypothesis is that there are technologies which are more similar among producing units that occupy the same industry than among those that are in more remotely connected industries. As a by-product, we also explore the user cost approach to the measurement of banking output and examine the extent to which it differs from the one suggested by the Ruggleses.

The remainder of the paper is organized as follows. Section 2 discusses the framework to ascertain the delineation in the banking business under the 1997 NAICS, the data sources and the measurement of banking output. Section 3 discusses the empirical results and provides some concluding remarks.

2. Assessing the Delineation of the Banking Business

2.1. *Set Up*

Although there are variations in the structure among the banks under the 1997 NAICS, their activities can conveniently be divided into retail banking, corporate and institutional finance, electronic financial services, investment banking and security dealing, and fiduciary services. As shown in Table 1, banks have broadened their activities beyond the core traditional business of lending and deposit-accepting. A notable feature has been their entry into the trust, mutual fund, insurance, and brokerage activities. The 1987 financial reforms allowed banks to purchase and create securities dealer subsidiaries, and the 1992 *Bank Act* granted banks additional powers, such as the right to engage in trust activities.[4]

The question is now how well the proposed delineation separates banks' producing units into economically distinct activities. The delineation of these activities is meaningful if, and only if, the producing units of an industry have a similar cost structure allowing them to react the same way to an exogenous shock. However, the similarities diminish within coarser industrial groups.

Thus, our maintained hypothesis is that there are technology structures which are more similar among producing units that occupy the same industry than among units that are in more remotely connected industries. To test this proposition, we use a variation of the diversification concept developed by Gollop and Monahan (1991), which allows us to quantify the extent to which an industry's

[4]For more details on the changes of the Canadian banking business, see Harchaoui (1998).

production units have similar technologies.[5] Although this concept is attractive, it does not indicate whether differences in the technological structures across producing units are statistically significant. Therefore, the results based on this heterogeneity index will be supplemented by statistical tests based on the econometric estimation of the cost function of a panel of producing units.

2.2. *A Framework for the Assessment of the Delineation*

The properties of a technology are captured in parameters defining the relationships among inputs, outputs and costs. Identical cost function parameters across producing units suggest homogeneous technologies while different parameters specify heterogeneous technologies. Identifying and measuring these parameters is the key to designing a statistical measure that can be used to assess the delineation of banks. It turns out that, under reasonable assumptions, the information required for identifying these technology parameters can be extracted from data commonly available in industrial accounts. Consider the following cost function of a production unit j defined as:

$$C_j(w, Q, t) \tag{1}$$

where w and Q represent, respectively, vectors of input prices and the output C_j represents the minimal cost incurred by the production unit j in order to produce the output Q under given market conditions. The simplest parameterization of this cost function under constant returns to scale is to assume that it has the Cobb–Douglas form

$$\ell n c_j = \sum_{i}^{I} \beta_{i,j} \ell n w_{i,j} + \lambda_{j,t} t, \tag{2}$$

where $c_j = \dfrac{C}{Q}$ is the total cost per unit of output.

Assuming competitive (input and output) markets, the Cobb–Douglas parameters $\beta_{i,j}$ represents the cost share s_{ij} of the input i

$$\beta_{i,j} = \frac{\partial \ell n c_j}{\partial \ell n w_{i,j}} = \frac{w_{i,j} \cdot X_{i,j}}{c_j} \equiv s_{i,j} \tag{3}$$

so that $\sum_{i}^{I} s_{i,j} = 1$ where $w_{i,j}$ is the price of the input i and $X_{i,j}$ the quantity of the input i used by the j-th production unit.

If one considers another producing unit, say h, which performs the same activity and uses a Cobb–Douglas technology, this technology will correspond to parameters $\beta_{h,f}$ and, accordingly, to input shares $s_{h,f}$. If both production units have the same technology, then one may expect to obtain $s_{j,f} = s_{h,f}$. Otherwise, none of these equalities would hold. Differences in input cost shares among producing units, which, therefore, quantify differences among parameter technologies, can be used to calibrate the extent of heterogeneity among producing units within an

[5] A similar approach was applied to the insurance business, see Harchaoui (2000).

industry. The heterogeneity index Δ developed by Gollop and Monaghan (1991) has the following form

(4) $$\Delta = \sum_{j}^{J} \mu_j \Delta_j$$

with

(5) $$\mu_j = \frac{c_j}{\sum_j c_j},$$

and

(6) $$\Delta_j = \sum_h \mu_h = \left(\frac{\sum_j |s_{i,j} - s_{i,h}|}{2} \right)$$

The symbol |·| refers to the absolute value. Dividing by two prevents double counting and ensures that the index Δ is bounded in the zero-one interval, $1 \leq \Delta \leq 1$. As differences among the parameters increase, Δ increases. As the differences decrease, the index Δ approaches zero. It turns out that the heterogeneity index is simply a weighted sum over differences in a unit cost function parameters describing the technology structures employed by producing units within an industry, where the weights m_j and m_n are defined as the shares of the *j*-th and *h*-th producing units in the industry's total cost. For any given difference in the input shares of the *j*-th and *h*-th producing units, the overall effect on industry Δ is determined by the relative importance of the *j*-th and *h*-th producing units. Therefore, input differences between large producing units have more impact on Δ than do input differences between small ones. The share variables m_j and m_h insure this result.

2.3. *Data Sources*

Year-end balance sheet and income statement data for the universe of deposit-accepting institutions were gathered from Statistics Canada's new Annual Survey of Deposit-Accepting Institutions for reference years 1996–98. This new survey, which provides a breakdown of deposit-taking institutions' consolidated operations (in Canada only) by type of activities, reflects the broadening of activities of these institutions as a result of the reforms that affected the financial sector in Canada since 1987. It is also based on the 1997 NAICS, an activity based industrial classification which allows the business sector data from Canada, Mexico and the United States to be compared. This data set is used to construct the variables used in the empirical part of the paper.

Total cost and input prices: Total operating costs are measured as the sum of labor compensation, purchased goods and services and capital compensation (or gross operating surplus). Labor compensation is the sum of wages and salaries, pension contributions by the employer and other paid benefits; purchased goods and services include the cost of energy, materials and services purchased by banks to perform their regular activities. Capital compensation includes profits before taxes, corporate income taxes, depreciation, amortization and provisions for losses.

It is computed residually as gross output minus labor compensation and purchased goods and services.

Labor inputs prices are measured as the ratio of labor compensation to the number of full-time equivalent employees. The rental price of capital is computed as the cost of capital adjusted for all kinds of taxes and deductions that affect the Canadian financial sector. Both of these prices change from one year to another and vary across producing units. Intermediate input prices are usually assumed to be constant across all firms in each year. This is not the case in this paper, where we estimated the implicit price of the intermediate inputs as follows. The total cost function $TC = f(w_L, Q, M, t)$ was estimated using the ordinary least squares method. The implicit price of the intermediate inputs M are then derived using the following result of duality theory $\frac{\partial TC}{\partial M} = \tilde{v}_m$. The results indicate that the implicit nominal price v_m varies significantly over time and across producing units. The estimated implicit price was then incorporated in the cost function $C = \Gamma(w_L, \tilde{w}_M, Q, t)$.

Output: There is controversy over whether deposit-accepting institutions (Q) provide both deposit and loan services (Q) or only one or the other. It has been suggested that deposits are intermediate inputs. However, for the purposes of this paper it will be assumed that banks produce both deposit and loan services.

Banks' activities can be viewed as providing three main categories of services: loan services, deposit services, and other services. When depositors buy demand deposits from deposit-accepting institutions, they are essentially bartering part of their expected interest entitlement away in return for the provision of other services. The interest rate that depositors are paid is assumed to be lower than otherwise by the amount of these free services. Essential to this view is the assumption that there is an intermediate interest rate between deposit and loan interest rates that represents the "pure" rate of interest. This rate is pure in the sense that it is the price to rent funds without any loan or deposit intermediation charges. Thus, the interest rate paid to depositors is the pure interest rate less the value of deposit-related intermediary services. For loans, the interest payments are assumed to include the pure rate plus explicit service charges related to the loan. Thus, we have:

(7)
$$u_d = r + f^d - r^d$$
$$u_l = r^l + f^l - r,$$

where

u_d = nominal rate of deposit services,
u_l = nominal rate of loan services,
r = pure rate,
i^l = nominal interest received on loans,
i^d = nominal interest paid on deposits,
f^l = explicit loan charges,
f^d = explicit deposit charges,

and nominal values of services are

(8) $$V_d = u_l \times D$$
$$V_l = u_d \times L,$$

where V_s = nominal values of services $s = d$ (deposits), 1(loans); D and L are the average balances of loans and deposits during the period.

Loan and deposit rates are determined on the basis of the ratio of interest paid or received to the corresponding average loan or deposit balances during the period (averages are used since they more accurately reflect deposit or loan values than beginning or end of period values) for each activity k. Thus:

$$i^l_{k,t} = \frac{I^\ell_{k,t}}{L_{k,t}} \quad \text{and} \quad i^d_{k,t} = \frac{I^d_{k,t}}{D_{k,t}},$$

where $I^s_{k,t}$ $(s = \ell, d)$ is the interest received or paid. The explicit loan and deposit service charges $f^s_{k,t}$ are estimated as

$$f^\ell_{k,t} = \frac{F^\ell_{k,t}}{L_{k,t}} \quad \text{and} \quad f^d_{k,t} = \frac{F^d_{k,t}}{D_{k,t}},$$

where $F^s_{k,t}$ $(s = \ell, d)$ is the service fee on loans and deposits. The pure rate is calculated as a weighted average rate of interest on loans and deposits. Using the financial statements of a deposit-taking institutions, we estimate the pure rate as a weighted average of all rates, where weights $z^\ell_{k,t}$, $z^d_{k,t}$ are the proportion of the specific stocks of loans and deposits outstanding of each activity z

$$i_t = \sum_k^K (r^\ell_{k,t} \times z^\ell_{k,t} + r^d_{k,t} \times z^d_{k,t})$$

This approach to the measurement of banking output, based on the user cost principle developed by Fixler and Zieschang (1991), is recommended by the System of National Accounts 1993 to calculate the imputed sectoral uses of financial services—termed Financial Intermediation Services Indirectly Measured or FISIM. However, it differs from the one proposed by Ruggles and Ruggles (1982), whereby interest received by banks is treated as the payment for financial services.[6] Conceptually the user cost of money is similar to the user cost of capital equipment. As applied to financial goods the user cost measures the net cost to the firm of providing the good. Specifically, the user cost for a financial product is the difference between its revenue and the sum of its explicit and implicit costs. The "net" aspect of the user cost is important to the pricing of financial services because it appropriately captures a bank's intermediary role, which is central to the provision of financial services. Observe that the user cost valuation of services is similar to the imputed value of services derived from the net interest technique used traditionally in national accounts.

[6]Similarly, for the insurance business, Richard Ruggles (1983) recommends the use of premiums as a measure of output. It seems that the approach recommended by the Ruggleses for the measurement of the output of the financial sector aims at eliminating the transfer portion that is considered by the current System of National Accounts.

3. Discussion and Concluding Remarks

The heterogeneity index defined in equation (4) was constructed for the most detailed level of banking activities reflected by the NAICS structure: retail banking, corporate and institutional finance, electronic banking, investment banking and security dealing, and fiduciary services. The index is based on a vector of three input shares for capital, labor and intermediate inputs constructed for each producing unit performing one of these activities.

Table 2 summarizes the application of the heterogeneity index Δ at the different level of refinement for the banking business for 1996 and 1998, a period during which the value of the heterogeneity index Δ at both aggregate and detailed levels was fairly stable. For both years, the value of the heterogeneity index for the consolidated banking activities under the 1980 Standard Industrial Classification (SIC) is significantly higher than its NAICS counterpart. This suggests that the 1980 SIC, which delineates the banking industry at the institutional level, is quite weak at classifying banks to the level necessary for industry division. This stems from the fact that under the 1980 SIC, banks correspond to regulated institutions, legally constituted, and often members of enterprises with a wide variety of activities. The SIC delineation, therefore, contains a number of unrelated activities, and like activities are not grouped within the same industry.

Separating banks into distinct producing units significantly decreases the level of heterogeneity. With 0.43, corporate and institutional finance displays the lowest level of heterogeneity followed closely by fiduciary services (0.44). Electronic banking and corporate and institutional finance both show a slightly higher heterogeneity index close to 0.50. Retail banking's heterogeneity index is 0.62, the highest amongst banks activities, thus indicating that a further refinement can be implemented in this industry. The aggregation of these activities gives a level of heterogeneity lower than the one corresponding to the SIC. Table 2 also indicates a great deal of variation across industries in terms of the structures of the technology.

The framework employed so far is useful to quantify how well NAICS separates producing units into distinct industries, but, like any non-parametric tool, it cannot say whether the results reported in Table 2 are statistically significant. One way to achieve this is to specify and estimate a translog cost function based on equation (1) with industry dummy variables. To estimate the model, we have pooled time-series cross section data for producing units for the most detailed banking industries over the 1996–98 period.

Using the likelihood ratio for the parameter restriction of the cost function, we tested several hypotheses related to cross industry differences in the technology employed by the various producing units:[7] (a) whether the production technology measured in terms of the capital-labor ratio differs across industries; (b) whether industries experience similarly technological change, an indication of the efficiency with which the resources are employed for a given level of output; and (c) whether industries react in the same way to changes in the quantity of the output produced. All of these hypotheses have been readily rejected, indicating

[7]The results are available on request.

TABLE 2

HETEROGENEITY INDEX OF BANKING ACTIVITIES IN CANADA

Industry Name	Year 1996	Year 1998
Consolidated activities under the 1980 SIC	0.75 (46)	0.77 (46)
Consolidated activities under the 1997 NAICS	0.58 (46)	0.60 (46)
Depository credit intermediation	0.60	0.62
Retail banking	0.61 (46)	0.64 (46)
Corporate and institutional finance	0.41 (35)	0.43 (35)
Electronic banking	0.47 (5)	0.48 (7)
Investment banking and security dealing	0.45 (23)	0.49 (27)
Fiduciary services	0.40 (11)	0.44 (16)

Note: The numbers in parentheses refer to the number of producing units. Under the consolidated activities, these numbers represent the number of banks.

that NAICS does a good job of separating producing units into homogeneous industries.

We also performed the likelihood ratio test of significance of changes in digit level refinement. First, we tested whether the parameters of retail banking and corporate and institutional finance are equal, which is another way of asking whether it makes sense to keep these activities separate. The rejection of this hypothesis constitutes a strong reason to provide a breakdown of these activities rather than to group them into one activity at the higher level of industry detail. The likelihood ratio tests also suggest a decisive rejection of the joint hypothesis that the coefficients of the different banking activities are statistically equal.

Second, we tested whether the banking activities delineated at the most detailed industry level are statistically different from the consolidated business of banks. Here again the hypothesis was strongly rejected. Therefore, it appears that NAICS is more effective at dividing banks into finer levels of industry detail than at dividing them into coarse industry groups.

These results are reassuring. They indeed suggest that the 1997 NAICS is particularly helpful as a method of separating activities of institutions into industries based on similar technology. Since these levels of refinement are those that most studies of industrial organization view as being close the "plant" level, NAICS performs satisfactorily at delineating industries.

REFERENCES

Davis, Steven, John Haltiwanger, and Scott Schuh, *Job Creation and Destruction*, MIT Press, Cambridge, 1996.

Fixler, Dennis and Kimberly D. Zieschang, "Measuring the Nominal Value of Financial Services in the National Income Accounts," *Economic Inquiry*, 24, 1991.

Foster, Lucia, John Haltiwanger, and C. J. Krizan, "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," in Edward Dean, Michael Harper, and Charles Hulten (eds), *New Directions in Productivity Analysis*, University of Chicago Press, 2001.

Gollop, Frank. M. and James L. Monahan, "A Generalized Index of Diversification: Trends in U.S. Manufacturing," *Review of Economics and Statistics*, 73, 318–30, 1991.

Harchaoui, Tarek, M., "Deposit-Accepting Institutions in 1996: The Structure and Economic Performance of their Activities," *Assurances*, 2, 203–13, July 1998 (special issue).
———, "Dealing with the Insurance Business in the Economic Accounts," in Georges Dionne (ed.), *Handbook of Insurance*, Kluwer Academic Publishers, 831–66, 2000.
Ruggles, Richard, "The United States National Income Accounts, 1947–1977," in Murray F. Foss (ed.), *The US National Income and product Accounts*, Income and Wealth Conference Volume, National Bureau of Economic Research, New York, 1983.
Ruggles, Richard and Nancy D. Ruggles, "Integrated Economic Accounts for the United States, 1947–78," *Survey of Current Business*, 62, 1–53, May 1982.
———, "The Integration of Macro and Micro Data for the Household Sector," *The Review of Income and Wealth*, 32, 245–76, 1986.
———, *Macro- and Microdata Analyses and their Integration*, Edward Elgar Publishing, 1999.