

Bank Efficiency in Colombia: A Brief Review of the Literature

Michel Janna G.¹
mjannaga@banrep.gov.co

I. Introduction

Research on bank performance and efficiency has advanced greatly in the past three decades. The large number of studies on the subject² worldwide is largely justified by the importance of a properly functioning financial system for the economy in general. Specifically, the financial system's role in allocating resources to productive sectors where liquidity is relatively scarce, its function as the engine of the payments system, and also the role it plays in promoting long-term growth are major factors motivating research into the efficiency of its productive structure.

In Colombia, too, bank efficiency has been the subject of research studies, though at a lesser extent than in developed countries. Between 1983 and 2004 barely a dozen studies were carried out on the financial system's cost structure. The present review will focus on a number of studies that have contributed to public debate on bank efficiency in Colombia.

When speaking of bank efficiency a distinction has to be made between two concepts: output efficiency and input efficiency. Output efficiency has to do with the likelihood that the banking firm is producing either optimal output levels (scale efficiency), or an optimal combination of several outputs (scope efficiency), or both. The level of inefficiency is measured by comparing the costs of the current output level with those of an optimal output level (the one that minimizes average costs).

On the other hand, input efficiency is related to the firm's capacity for using its inputs efficiently to produce a given quantity of output. Inefficiency in the use of inputs refers to: (1) the likelihood of using more inputs than necessary for producing a given level of output (technical inefficiency), and (2) the likelihood of using a wrong combination of inputs in such production (allocative inefficiency). These two types of efficiency in the use of inputs are called economic- or X-efficiency. X-efficiency is most commonly measured by estimating an efficiency frontier (a minimum cost function, for example) and for comparing how far each firm deviates from such "ideal behavior."

¹ An early (and less complete) version of this review can be found in Banco de la República (2003). As usual, the opinions expressed here are exclusively those of the author. I thank Juanita González for her collaboration.

² For a list of studies on bank efficiency in various countries, see Berger and Humphrey (1997).

The Colombian studies can be fitted into two large groups on both a chronological and a topical basis. The first group consists of papers published between 1983 and 1996 on measuring economies of scale in Colombia's financial sector. They include, notably, studies by Bernal and Herrera (1983), Suescún (1987) and Ferrufino (1991). A study by Suescún and Misas (1996) marks the transition between the studies on scale efficiency and those on economic efficiency or X-efficiency (also called input efficiency).

Since 1996 to present, research on Colombian banks' efficiency has focused almost exclusively on seeking measurements of economic efficiency. This may have been caused by the country's financial liberalization in the 1990s which substantially reduced the entry barriers that created distortions in the sector in terms of sunk costs and lack of competition. It thus became more interesting to study the banks' output structure in terms of their ability to use inputs in the best possible way, rather than simply reviewing the industry's position against its average cost curve. The studies of Castro (2001), Badel (2002), Janna (2003) and Estrada and Osorio (2004) make up the second group.

This brief review is organized as follows: in sections II and III I discuss literature related to scale economies and X-efficiency respectively (groups 1 and 2 defined above). In section IV I take a look at the time evolution of bank efficiency estimations for the Colombian case. Section V deals with studies of bank efficiency determinants. Other types of efficiency studied for the Colombian financial sector are included in section VI, and section VII shows some final remarks.

II. Scale Efficiency (scale and scope economies)

The literature on scale economies (SE) in Colombian banks began with the study by Bernal and Herrera (1983). The study sought to estimate a log-linear Cobb-Douglas cost function for the banking industry of the form:

$$\log C = \mathbf{b}_0 + \mathbf{b}_1 \log Y + \mathbf{b}_2 \log A \quad (1)$$

Where C represents overhead costs, Y is the bank output (measured as the stock of loan portfolio), and A is a measure of deposits structure (defined as the proportion of current accounts stock to total deposits stock). The authors find an estimate of output elasticity of costs (\mathbf{b}_1) that is less than one. This result suggests the existence of economies of scale for the year 1981 (Table 1).

The approach of Bernal and Herrera has, nevertheless, many problems. Since the cost function used is linear on output, the average cost curve estimated is not U-shaped but a decreasing monotonous function. According to this finding, the scale economies found would never be exhausted and there would be no optimal scale of production, because costs could always be saved by marginally increasing the level of output. Therefore, the study succeeds in estimating a measurement of

scale economies but fails to calculate a level of scale inefficiency, since each bank's current level of production cannot be compared with a unique optimal level (a minimum level in the average cost function).

In the same line, since the specification of the cost function ignores input prices, the underlying (and far fetched) assumptions are that each firm faces the same input costs, and even worse, that every bank uses the same amount and combination of inputs per unit of product. This assumption distorts the scale economies measurements since large costs in some banks are not exclusively due to scale production.

Suescún (1987) and Ferrufino (1991) updated and improved the estimations of Bernal and Herrera (1983) by using more flexible functional forms to model operating costs and by considering bank outputs other than the loan portfolio (Table 1). For example, Suescún (1987) uses a trans-logarithmic cost function (equation 2) which allows the scale economies to vary according to the level of production.

$$\ln C = \mathbf{b}_0 + \mathbf{b}_1 \ln Y + \frac{1}{2} \mathbf{b}_2 (\ln Y)^2 + \mathbf{b}_3 \ln w + \frac{1}{2} \mathbf{b}_4 (\ln w)^2 + \mathbf{b}_5 (\ln Y \ln w) + \mathbf{b}_6 \ln T + \frac{1}{2} \mathbf{b}_6 (\ln T)^2 + \mathbf{b}_8 (\ln Y \ln T) + \mathbf{b}_7 \ln S + \frac{1}{2} \mathbf{b}_7 (\ln S)^2 + \mathbf{b}_8 (\ln Y \ln S) \quad (2)$$

Where Y represents bank output (number of loan or deposit bank accounts), w is the labor cost measured as the total labor expenses over the number of workers, and T and S represent the average account size (loan or deposit) and the number of branches respectively.

Ferrufino (1991) uses the same specification and methodology of Suescún (1987), to put up to date the results for commercial banks and extend them to other financial intermediaries.

Although Suescún's and Ferrufino's results are similar for commercial banks, both of them, as does Bernal and Herrera, fail to obtain average cost curves with minimums for these banks. The literature reviewed above was therefore unable to determine an optimal output level.

The only study to overcome the constraint described above is by Suescún and Misas (1996), who used a trans-logarithmic specification of the banks' cost function (slightly different from equation 2) and modeled the banking system as a group of firms using physical capital and labor to produce stocks of loans (Table 2). The cost function also included the number of bank branches and a temporal trend to quantify the effect of technical change on costs.

The study measured scale inefficiency by comparing for each firm the difference between the unit costs of producing its observed loan stock and the corresponding

minimum average cost when the firm operates with its number of branches and the sector's average factor prices. The findings revealed that scale inefficiency was low, since operating at a socially optimal output level meant only a 3.2% saving of operating costs. Suescún and Misas (1996) also corroborated the existence of simple economies of scale (Table 1) but not economies of scale that take into account the opening of new branches (Named Broad economies of scale).

Table 1

Scale Economies in the Colombian banking system */

Study	Period	Type of output	SE **/	BSE **/
Bernal and Herrera (1983)	1981	Loan Portfolio	0.93	
Suescún (1987)	1983 & 1986	No. of loan & deposit accts.	0.71	0.83
Ferrufino (1991)	1986 - 1988	No. of loan accts.	0.61	0.82
		No. of loan & deposit accts.	0.67	
Suescún y Misas (1996)	1989 - 1995	Loan Portfolio	0.78	1.06
Castro (2001)	1994 - 1999	Loan Portfolio	0.76	
		Securities	0.18	

*/ Commercial banks only.

**/ Scale economies (SE). Broad Scale economies (BSE): SE that take into account output expansion by opening new branches.

III. Economic Efficiency

The idea behind X-efficiency measurement is pretty much standard. The most common procedure consists in determining a function that describes the optimal behavior or best practice of the industry. This can be done by estimating an efficiency frontier (of costs, profits or production), which allows to compare the stance of each firm with respect to the best practice.

In the specific case of cost efficiency, the objective is to estimate a cost function that describes the minimum cost for each combination of output and input prices, in other words to determine a functional form for the first term of the right hand side of equation 3. The remaining terms (u and v) correspond to deviations of observed costs from optimal costs. These deviations can be of two types: random shocks not controlled by the bank (v) or cost inefficiency (u). The challenge of X inefficiency measurement lies on a correct estimation of u and its severance from the stochastic term (u+v).

$$\ln C = \ln C(w, y) + u + v \quad (3)$$

Most of the X-efficiency studies in Colombia follow the latter approach. Nevertheless, they differ in the way the deviations of each bank from the efficient frontier are estimated.

Table 2

Different Characterizations of the Colombian Banking Activity

Study	Inputs considered	Outputs
Suescún and Misas (1996)	Physical capital, labor	Loan stock
Castro (2001)	Physical capital, deposits, labor	Loan & securities stocks
Badel (2002)	Financial (equity) capital, labor	Loan & securities stocks
Janna (2003)	Physical capital, deposits, labor	Loan & securities stocks
Estrada and Osorio (2004)	Physical and financial capital, deposits, labor	Loan & securities stocks, deposits with other intermediaries

The first measure of X-inefficiency in Colombian banking system was made by Suescún and Misas (1996). Applying the thick-frontier approach³, they used a semi-annual sample of 22 banks between 1989 and 1995 and found that X-inefficiency accounted for about 27% of the commercial banks' total operating costs (Table 3).

It is important to point out, however, that the study did not include financial costs within the banks' cost structure. Hence, the study's measurements of inefficiency ignored the greater part (about 66%) of total costs, tending to underestimate the degree of economic inefficiency.

Castro (2001) adopted an intermediation approach to characterize the productive activity of banks. According to this approach, banking consists of using deposits, physical capital and labor (three inputs) to produce stocks of loans and/or securities (two outputs). Thus, this approach takes into account financial costs as well as operating costs in estimating the cost function, thereby obtaining a more accurate measurement of input efficiency.

³ The "thick frontier" approach divides all financial institutions into different groups according to their historical performance, which makes it possible to separate "efficient" banks from "inefficient" ones. Once this is done, a cost frontier is estimated for each group. Cost differences between groups are considered inefficiencies (u), while each group's regression residuals are regarded as random noise (v). This classification into groups must be based on some arbitrary indicator. The most common indicator used for this practice is the ratio of overhead costs to assets. For this reason, the final result will rely heavily on the classification procedure.

The estimation methodology used unbalanced panel data from 30 banks, covering the period from 1994 to 1999. Using a *distribution free approach* (DFA)⁴, Castro found that the Colombian banking sector's relative cost efficiency averaged 55%, which suggests that, in the absence of X-inefficiency, the banks could reduce their costs by about 45% in producing the same quantities of loans and securities.

Badel (2002), for his part, sought to broaden the studies on X-efficiency for the Colombian banks through comparisons with other banking systems in Latin America. To this end, he estimated a cost function common to Colombian, Mexican and Costa Rican banks, using information from 54 banks for 1998-2000. Modifying the intermediation approach used by Castro (2001), Badel did not regard physical capital as a relevant input for producing stocks of loans and securities. Instead, he established financial (equity) capital as a fixed input, considering it an alternative source of financing for such outputs.

Badel's findings revealed that average efficiency was fairly homogeneous across countries, though there was high dispersion within each country. On average, the most efficient banks over the period under study were those of Costa Rica, with 77%, followed by Colombian banks (73%) and Mexican banks (66%). However, Colombian banks were found to be more efficient than the Mexican or Costa Rican ones during the last year of the period studied (2000).

Because of the methodology used, the studies by Suescún and Misas (1996), Castro (2001) and Badel (2002) could only measure inefficiency relative to the most efficient bank or group of banks. That is to say, their approaches assumed that the firm with the lowest cost per output represented the best possible practice in the industry, ignoring that the "most efficient" firm might also be wasting resources relative to an optimal cost frontier.

More recently, Janna (2003) and Estrada and Osorio (2004) have tried to overcome that constraint, so as to obtain estimators of absolute inefficiency rather than inefficiency relative to the best-practice bank.

Using a characterization of banking similar to Castro's (2001), and on information from 28 credit institutions from 1992 to 2002, Janna (2003) estimated a stochastic cost frontier⁵ for Colombia's banking system. He found that the system currently presents an efficiency indicator of around 43% (an average of 34% for the period studied), which suggests a lot of room for reducing costs (Table 3).

⁴ Under this approach, no particular assumption about the distribution of the inefficiency term is needed. This can be done by assuming that the average inefficiency of each firm (u) is persistent over time and that the random error average (v) is zero during the period considered. This approach, although simple to use, only allows calculating relative efficiency measurements of each bank to the most efficient firm (the one with the lowest value for u).

⁵ The stochastic frontier approach, also known as econometric approach, consists of severing the inefficient component (u) from the stochastic error (v) by means of assumptions regarding the probability distributions of both terms. Since it is necessary for the inefficient component to be non negative, truncated distributions of u are assumed.

Estrada and Osorio (2004), for their part, have used information from 1989 to 2003 for different financial intermediaries to construct a cost frontier for the entire financial system. Applying a frontier estimation similar to Janna's (2003), they have found that the efficiency indicator for the average bank is 28%, the lowest estimated so far by any of the studies.

Table 3

Economic efficiency

Study	Period	Approach**	Average X-Efficiency
Suescún and Misas (1996)	1989 - 1995	TFA	73%
Castro (2001)	1994 - 1999	DFA	49%
Badel (2002)*	1998 - 2000	DFA	73%
Janna (2003)	1992 - 2002	SFA	34%
Estrada and Osorio (2004)	1989 - 2003	SFA	28%

* The estimated cost frontier includes Costa Rican, Colombian and Mexican banks. The figure here is the average for Colombian banks.

** TFA: thick frontier approach; DFA: distribution free approach; SFA: stochastic frontier

IV. Time evolution of economic efficiency

Castro (2001), Badel (2002), Janna (2003) and Estrada and Osorio (2004) have all tried to measure to what extent bank efficiency has changed over time. Their findings are mostly similar and favor the conclusion that great advances were made regarding cost saving in the 1990's.

Castro tried to quantify the impact of various mergers and acquisitions on the efficiency of the institutions involved, in 1996-1999. His findings revealed that, on average, such reorganizations had a negative effect in terms of efficiency, though detailed analysis of each case provided mixed results. For example, privatization and acquisition by foreign agents subsequently improved bank efficiency, whereas nationalization had a negative effect. Between 1994 and 1999, average efficiency for the banking sector as a whole (including reorganized and other banks) improved by 10.3%.

Badel (2002), for his part, constructed time-varying indicators of efficiency to observe changes in the bank efficiency of each of the countries studied. He found that Colombian banks registered some improvement. In effect, though they were less efficient than Mexican or Costa Rican banks in the first half of 1998, by the second half of 2002 their indicator was the best of the three countries.

Janna (2003) sought to quantify Colombian banks' efficiency improvements between 1992 and 2002, so as to identify the major factors that had caused them. His findings revealed that the banks' average efficiency improved by 63% (or 17

percentage points) in those ten years but not uniformly over the period, for their progress was interrupted by the financial crisis of 1998-1999.

Janna also showed that efficiency progress between 1992 and 1998 stemmed largely from changes in general market conditions that affected the whole banking system (deregulation, economic cycle, market concentration), whereas improvements after 2000 resulted from each of the bank's control variables. He concludes therefore that this shift in efficiency-driving variables bears out the view that the crisis had a "disciplining effect" on the banks' cost management, since it detached progress in efficiency from improvements in environmental conditions and gave relevance to a number of variables under greater control by each bank.

V. Determinants of Economic Efficiency

Castro (2001) and Badel (2002) aimed to explain efficiency levels on the basis of each bank's particular variables (type of ownership, type of output, and other features of banks). In Castro's study, the simultaneous inclusion of all relevant variables in one regression provided no statistically significant coefficient. But regressions carried out with fewer variables provided some significant relationships (Table 4). In Badel's study, all the variables used were significant in a single regression.

Janna (2003) not only tried to explain banks' X-efficiency in terms of each bank's particular variables, as had Castro and Badel, but he also included some variables that described general market (environmental) conditions that had the same behavior for all banks. The inclusion of these industry-wide variables provided interesting results, revealing that regulatory burden, economic cycle and market concentration all had a negative effect on efficiency (Table 4).

Table 4

Determinants of colombian banks' X- Efficiency

Effects		Castro (2001)	Badel (2002)	Janna (2003)
Ownership	Foreign Public	Non significant Negative		Positive Non significant
Banks' Characteristics	Size Number of branches Return on assets Return on equity Capital ratio	Negative Positive Positive	Positive Negative Positive	Positive
Product differentiation	NPL Product quality Ratio of commercial loans	Negative Positive	Negative	Negative Positive
Environmental / market conditions	Regulatory burden Business Cycle Market concentration			Negative Negative Negative

VI. Other Types of Efficiency

Profit efficiency is another way of characterizing the behavior of a firm. As in the case of cost efficiency (economic efficiency), the idea of this measure is to gauge a bank's position against an optimal frontier of benefits (a function describing the greatest quantity of benefits that a financial entity can achieve subject to its industry characteristics).

In a market with perfect competition, cost efficiency and profit efficiency should be equal. But, in the presence of any type of market power, firms may exhibit optimal profit levels without operating at minimum cost. Hence, comparing indicators of profit efficiency and cost efficiency may cast light on the structure of the banking market.

The only measurement of profit efficiency for Colombia's financial sector is to be found in Estrada and Osorio (2004). Specifically, profit efficiency of commercial banks is estimated at around 82%; that is to say, if these banks operated optimally they could, on average, improve their profits in somewhat less than 20%. Comparing these authors' profit efficiency measurement with their cost-efficiency estimation reveals the first one to be greater than the latter. Similarly, no high relation is found to exist between individual intermediaries' cost-efficiency and profit-efficiency measurements. Estrada and Osorio believe that these findings reflect the existence of some market power in the Colombian banking sector.

VII. Final Remarks

The studies on economies of scale and scope for the Colombian banking sector were mostly carried out in the 1980s and early 1990s. In general, their findings favored the existence of scale economies in banks during the period. Nevertheless, the lack of studies for the second half of the 1990s, especially after the financial crisis of 1999, rules out a definite conclusion about the persistence of scale economies.

On the other hand, the results of studies about economic efficiency are rather more heterogeneous than those on scale economies. In fact, the different efficiency estimators found in the Colombian literature fell within a relatively broad range (between 28% and 73%). Such vast differences among the studies reviewed are problematic, since they show that the measurements of bank efficiency in Colombia are quite sensitive to the functional forms of costs, the variables chosen as inputs and outputs, the sample used, and estimation methodologies. Even more, they leave unanswered questions about the most adequate assumptions for the Colombian banking case.

However, there is one common conclusion suggested by most X-efficiency related studies. They find that there is still potential for the costs of Colombian banks to be more efficiently managed. This potential is actually greater than in developed countries, where banks are closer to their efficient frontier⁶. Colombian banks have, in fact, begun in recent years to exploit this capacity for improving their cost management, as evidenced by the studies that have measured temporal variations in cost efficiency.

Finally, there is some evidence that the efficiency of Colombian banks is influenced both by idiosyncratic factors of institutions (type of ownership, levels of financial capital, type of business, size of branch network), and by environmental factors jointly affecting the whole sector (regulatory burden, economic cycle, market concentration, financial crises). Nevertheless, the differences between the determinants found in some studies (Table 4) will require further research in order to explain them.

⁶ Berger and Humphrey (1997) report that studies on the United States estimate bank X-efficiency level to range between 61% and 95%.

References

Badel, A. (2002), "Sistema Bancario Colombiano: ¿Somos Eficientes a Nivel Internacional?", *Archivos de Economía*, DNP, Document 190.

Banco de la República (2003), *Reporte de Estabilidad Financiera*, December.

Berger, A. and D. Humphrey (1997), "Efficiency of Financial Institutions: International Survey and Directions for Future Research", *European Journal of Operational Research* 98.

_____ and L. Mester (1997), "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?", *Journal of Banking and Finance* 21.

Bernal, O. and S. Herrera (1983), "Producción, Costos y Economías de Escala en el Sistema Bancario Colombiano", *Ensayos Sobre Política Económica*, Banco de la Republica, No. 3.

Castro, C. (2001), "Eficiencia-X en el sector bancario colombiano", *Desarrollo y Sociedad*, Universidad de los Andes, No. 48, September.

Estrada, D. and P. Osorio (2004), "Efectos del capital Financiero en la Eficiencia del Sistema Bancario Colombiano", forthcoming.

Ferrufino, A. (1991), "Reestimación y Ampliación de la Evidencia Sobre las Economías de Escala en el Sistema Financiero Colombiano", *Ensayos Sobre Política Económica*, Banco de la Republica, No. 19.

Janna, M. (2003), "Eficiencia en Costos, Cambios en las Condiciones Generales del Mercado, y Crisis en la Banca Colombiana: 1992-2002", *Borradores Semanales de Economía*, Banco de la República, No. 260.

Suescún, R. (1987), "Nueva Evidencia Sobre Economías de Escala en la Banca Colombiana", *Ensayos Sobre Política Económica*, Banco de la Republica, No. 12.

_____ and M. Misas (1996), "Cambio tecnológico, ineficiencia de escala e ineficiencia-X en la banca colombiana", *Borradores Semanales de Economía*, Banco de la República, No. 59.