European Journal of Operational Research 240 (2015) 562–571

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Dynamic effects in inefficiency: Evidence from the Colombian banking sector

Jorge E. Galán [a,*], Helena Veiga [a,b,c], Michael P. Wiper [a,b]

[a] *Department of Statistics, Universidad Carlos III de Madrid, Spain*
[b] *Instituto Flores de Lemus, Universidad Carlos III de Madrid, Spain*
[c] *BRU/UNIDE, Avenida das Forças Armadas, 1600-083 Lisboa, Portugal*

ARTICLE INFO

Article history:
Received 23 January 2014
Accepted 3 July 2014
Available online 10 July 2014

Keywords:
Productivity and competitiveness
Banks efficiency
Bayesian inference
Dynamic effects
Heterogeneity

ABSTRACT

Firms face a continuous process of technological and environmental changes that requires them to make managerial decisions in a dynamic context. However, costs and constraints prevent firms from making instant adjustments towards optimal conditions and may cause inefficiency to persist in time. We propose a dynamic inefficiency specification that captures differences in the adjustment costs among firms and non-persistent effects of inefficiency heterogeneity. The model is fitted to a ten year sample of Colombian banks. The new specification improves model fit and have effects on efficiency estimations. Overall, Colombian banks present high inefficiency persistence but important differences between institutions are found. In particular, merged banks present low adjustment costs that allow them to recover rapidly efficiency losses derived from merging processes.

1. Introduction

The decision making process followed by producers is dynamic in nature. Technology and environment change continuously and variations with respect to their current production conditions have to be considered by firms. However, firms face restrictions and costs in the adjustment process. Regulation, quasi-fixed or indivisible inputs, and transaction, information and other adjustment costs are important factors preventing firms from making free and instant adjustments towards optimal conditions. In this context, firms may not only be inefficient at some point, but this inefficiency may persist from one period to the next, and firms may find it optimal to remain partly inefficient in the short-run.

This issue has been little studied in the efficiency measurement literature but has recently become an important concern. In stochastic frontier models, first introduced by Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977), we can find two alternative approaches to deal with time dependent inefficiencies. The first approach defines deterministic time specifications for the evolution of efficiency. As examples we find the proposals by Kumbhakar (1990) and Battese and Coelli (1992) where a time invariant inefficiency measure is multiplied by a parametric function of time, the model by Cornwell, Schmidt, and Sickles (1990) that defines producer specific parameters, and the proposal by Lee and Schmidt (1993) where time dummies are used. These models have the problem of imposing arbitrary restrictions on the short-run efficiency and are not able to model firm-level dynamic behaviour. A more recent approach involves the dynamic behaviour of inefficiency by considering models that estimate long-run efficiency. These models recognize a persistence effect of firms' inefficiency over time and specify its evolution as an autoregressive process. In this context, Ahn, Good, and Sickles (2000) defined an error structure intended to capture the relationship between the short and long-run dynamics. This pioneer proposal has been criticized for its economic foundations and for modeling autoregressive processes on nonnegative variables.

An alternative proposal that avoids this problem and argues that improvements in efficiency depend on adjustment costs was introduced by Tsionas (2006). Under this framework, factors affecting short-run efficiency may not be rapidly adjusted when adjustment costs are high, implying that inefficiency persists from one period to the next. This model was applied to a sample of US banks and very high inefficiency persistence was found out, suggesting the presence of high adjustment costs in the banking sector. Previous studies have also found evidence of inefficiency persistence in financial institutions. Tortosa-Ausina (2002), in an analysis of transition probabilities of efficiency, found that most of Spanish banks remain in the same state of relative inefficiency in consecutive periods.

* Corresponding author. Address: Department of Statistics, Universidad Carlos III de Madrid, C/Madrid 126, 28903 Getafe, Spain. Tel.: +34 916249674; fax: +34 916249849.

E-mail addresses: jegalan@est-econ.uc3m.es (J.E. Galán), mhveiga@est-econ.uc3m.es (H. Veiga), mwiper@est-econ.uc3m.es (M.P. Wiper).

http://dx.doi.org/10.1016/j.ejor.2014.07.005
0377-2217/

In this context, the model proposed by Tsionas (2006) becomes very relevant in accounting for inefficiency persistence. This model presents two main characteristics. The first one is that it assumes a constant persistence parameter for all banks in the sector. However, it is possible that banks with different characteristics face different costs of adjustment. Large institutions may benefit from scale economies to adapt faster new processes while foreign institutions may face lower costs due to cheaper access to multiple sources of funding or more diversification (see Chen & Liao, 2011). In a recent application to French banks, Huang and Chen (2009) extends the model in Ahn et al. (2000) to consider rational expectations and test the homogeneous persistence restriction. Their findings suggest that this restriction lead to biased estimators and underestimation of technical efficiency.

The second characteristic of the model in Tsionas (2006) is that it allows the inclusion of observed heterogeneity in the inefficiency. The effects of heterogeneity on bank efficiency have been studied by Bos, Koetter, Kolari, and Kool (2009). These authors find not only, that including observed heterogeneity in the frontier and in the inefficiency distribution leads to significant changes in the efficiency levels and rankings, but also that how these variables are included is relevant. This can be particularly important in a dynamic framework, since including covariates as inefficiency drivers in an autoregressive specification implies that they have persistent effects over time. In this context, it would mean that the effects that these firm characteristics and environmental conditions produce on inefficiency can not be easily adjusted by firms. This would be the case of a public bank that is less efficient because of attending rural customers in remote places. This characteristic may not be easily altered and may consequently induce high adjustment costs. However, changing other conditions such as some managerial practices or the risk exposure of short-run investment portfolios may be easier to adjust. In these cases, heterogeneity sources should be allowed to be inefficiency drivers but modeled out of the dynamic specification.

In this work we propose a model that accounts for firm specific persistent effects in the inefficiency and is able to model observed heterogeneity in and out of the inefficiency dynamics. We apply the new specification to a sample of Colombian banks during the last decade. The Colombian banking sector is of interest since it has been characterized by the arrival of foreign institutions and several mergers and acquisition (M&A) processes that have increased the differences in terms of size among banks during this period.

Certainly, the effects of foreign ownership, size and M&A on banks efficiency have been studied previously under static formulations. Regarding foreign ownership and size, divergent results have been obtained previously. On the one hand, using Bayesian stochastic frontiers, Tecles and Tabak (2010) and Assaf, Matousek, and Tsionas (2013) found foreign and large banks to be more cost and profit efficient in Brazil and Turkey, respectively. Using nonparametric methods, Ray and Das (2010) and Sathye (2003) also found positive effects of foreign ownership on efficiency in Indian banks. On the other hand, negative effects of foreign ownership have been found by Lensik, Meesters, and Naaborg (2008) in a study including a sample of 105 countries; while, a negative impact of size was found by Hartman and Storbeck (1996) for banks in Sweden, and no size effects were concluded for the case of Brazil by Staub, da Silva e Souza, and Tabak (2010). In the dynamic context, Tsionas (2006) identified size to have persistent effects on cost efficiency of US banks. Concerning M&A, previous studies have found none or very little improvement on input-oriented technical efficiency or cost efficiency (see Amel, Barnes, Panetta, & Salleo, 2004, for a review. However, merged banks have been found to present different time patterns than non-merged institutions and their efficiency to be highly dependent on time (see Cuesta & Orea, 2002)). In this context, introducing time dependency into a dynamic structure and allowing merged banks to follow their own dynamics may lead to different conclusions.

In particular we have two main aims: firstly, to evaluate the impact of adding more flexibility to the persistence parameter and separating heterogeneity from the dynamics on efficiency estimations; secondly, to identify the importance of adjustment costs in explaining input-oriented technical efficiency of Colombian banks and the effects of size, foreign ownership and M&A as inefficiency drivers.

The rest of the paper contains five additional sections. In Section 2, we present the dynamic specification proposed for the inefficiency. In Section 3, we show how to carry Bayesian inference for this model and some model comparison criteria. In Section 4, we present the derivation of the stochastic input distance function. In Section 5, we present the empirical application including a brief description of the Colombian banking sector, the data, and the main results in terms of the model specification and their implications for Colombian banks. Section 6 concludes the paper.

2. Heterogeneity and dynamic inefficiency

The inclusion of variables that capture firm characteristics in the inefficiency component is important to distinguish properly heterogeneity from inefficiency but they do not necessarily capture inefficiency adjustment processes or have persistent effects. Moreover, the evolution of the inefficiency over time can be attributed to the heterogeneity of adjustment costs and then be directly modeled by firm-specific persistence parameters.

We propose a specification for the inefficiency in the context of dynamic stochastic frontier models that is flexible in terms of the treatment of heterogeneity. The first characteristic is the separation of two components within a log-linear specification for the inefficiency. One component, which is unobserved, will follow an autoregressive process that captures the portion of inefficiency that is transmitted from one period to the next. The second component is a vector of observed covariates driving the inefficiency level. In this sense this could be seen as an extension of the model by Galán, Veiga, and Wiper (2013), where an unobserved random parameter is modeled in the inefficiency distribution along with observed covariates, but where the unobserved part follows a first order autoregressive process. This component can also include observed variables that will capture persistent effects of heterogeneity in the inefficiency as in the specification followed by Tsionas (2006) for the whole inefficiency.

The second characteristic that we allow for is the modeling of a firm specific persistence parameter in the dynamic component that captures heterogeneity among banks in terms of the proportion of the inefficiency that is transmitted to the next periods. Different specific persistence parameters may suggest heterogeneity in the adjustment costs of banks.

The model is given by:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it} - u_{it}, \quad v_{it} \sim N(0, \sigma_v^2) \tag{1}$$

$$\log u_{it} = \theta_{it} + \mathbf{z}_{it}\boldsymbol{\gamma} + \xi_{it}, \quad \xi_{it} \sim N(0, \sigma_\xi^2) \tag{2}$$

$$\theta_{it} = \omega + \mathbf{h}_{it}\boldsymbol{\psi} + \rho_i \theta_{i,t-1} + \eta_{it}, \quad \eta_{it} \sim N(0, \sigma_\eta^2), \quad t = 2 \dots T \tag{3}$$

$$\theta_{i1} = \frac{\omega + \mathbf{h}_{i1}\boldsymbol{\psi}}{1 - \rho_i} + \eta_{i1}, \quad \eta_{i1} \sim N\left(0, \frac{\sigma_\eta^2}{1 - \rho_i^2}\right), t = 1. \tag{4}$$

The stochastic frontier is represented by (1) where y_{it} represents the output for firm i at time t, $\mathbf{x}_{it}$ is a row vector that contains the input quantities, $\boldsymbol{\beta}$ is a vector of parameters, v_{it} is an idiosyncratic error assumed to follow a normal distribution, and u_{it} is the inefficiency component. Eq. (2) is the log-linear specification for the inefficiency where θ_{it} represents the dynamic component, $\mathbf{z}_{it}$ is a row vector of firm specific heterogeneity variables, γ is a vector of parameters and ξ_{it} is a white noise process with constant variance σ^2_ξ. The unobserved dynamic component θ_{it} follows an autoregressive process represented by (3) where ω is a constant, $\mathbf{h}_{it}$ is a row vector of observed covariates, ψ is a vector of parameters, ρ_i is the firm specific persistence parameter measuring the proportion of the dynamic part of the inefficiency that is transmitted from one period to the next for every firm, and η_{it} represents unobserved random shocks in the dynamic component and follows a normal distribution with variance σ^2_η.[1] The dynamic process is assumed to be stationary and (4) initializes it.

Stationarity ensures that the dynamics of the log-inefficiency do not diverge to negative or positive infinity. If this condition is not imposed, efficiency scores could be equal to one or zero in the long-run. The first case would contradict the adjustment costs theory that motivates the dynamic formulation and the second case would imply that totally inefficient firms do not exit the market. Therefore, the persistence parameters are required to satisfy $|\rho_i| < 1$. A value close to 1 for this parameter means high persistence of the inefficiency dynamic component and slow adjustment of firms towards optimal conditions.

Modeling firm specific persistence parameters imply that even when no covariates are included, firms may present differences in their adjustment of common factors and therefore different long-run inefficiencies. However, as we present in Section 3 we model these parameters using a hierarchical structure common in Bayesian statistics that links them to a common parameter for the whole sector. Moreover, firms in the sector share also a common long-run dynamic component ω and common elasticities for the covariates given by the vectors ψ and γ.

The proposed dynamic specification given by Eqs. (2)–(4) encompasses other models in the literature and permits us to compare some assumptions by including restrictions. For instance, homogeneous costs of adjustments for all banks can be studied by imposing $\rho_i = \rho$. If $\rho = 0$ the model is reduced to a static formulation with no adjustment costs but where an unobserved component θ_{it} captures latent inefficiency heterogeneity as in Galán et al. (2013). Additionally, if η_{it} is also equal to 0, the model takes the form of the Battese and Coelli (1995) static formulation. Finally, if $\rho_i = \rho$ and $\xi_{it} = \gamma = 0$ the model reduces to the dynamic model in Tsionas (2006).

3. Bayesian inference

In order to fit our models, we use Bayesian inference as in Tsionas (2006) and Emvalomatis (2012). However, maximum likelihood techniques can also be implemented for estimating dynamic specifications (see Emvalomatis, Stefanou, & Lansink, 2011).

Bayesian inference for stochastic frontier models was introduced by van den Broeck, Koop, Osiewalski, and Steel (1994). Among its main advantages are the formal incorporation of parameter uncertainty and the derivation of posterior densities of efficiencies for every individual firm.

We assume proper but relatively disperse prior distributions throughout. In particular, the distributions assumed for the parameters in the distance function are: $\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\beta)$ where $\boldsymbol{\Sigma}^{-1}_\beta$ is a diagonal matrix with precision priors set to 0.001 for all coefficients. The variance of the idiosyncratic error term is inverse Gamma distributed, that is equivalent to $\sigma^{-2}_v \sim G(a, b)$ where the priors for shape and rate parameters are set to 0.01.

The specification in (2) implies that the inefficiency follows a log-normal distribution. Then $u_{it}|\theta_{it}, \mathbf{z}_{it}, \gamma, \sigma^2_\xi \sim LN(\theta_{it} + \mathbf{z}_{it}\gamma, \sigma^2_\xi)$, where the location component is composed of the unobserved dynamic parameter and the observed heterogeneity component. The distribution for the unobserved parameter modeling the dynamics is $\theta_{it}|\theta_{i,t-1}, \omega, \mathbf{h}_{it}, \psi, \rho_i, \sigma^2_\eta \sim N(\omega + \mathbf{h}_{it}\psi + \rho_i\theta_{i,t-1}, \sigma^2_\eta)$ for $t = 2 \ldots T$. Given stationarity we have $\theta_{i1}|\omega, \mathbf{h}_{i1}, \psi, \rho_i, \sigma^2_\eta \sim N\left(\frac{\omega + \mathbf{h}_{i1}\psi}{1-\rho_i}, \frac{\sigma^2_\eta}{1-\rho_i^2}\right)$. The distribution for the common constant ω is normal with priors set to -1.5 and 1 for the mean and precision.[2] The distribution of the parameter vector of the observed covariates in the dynamic component is: $\psi \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\psi)$ where $\boldsymbol{\Sigma}^{-1}_\psi$ is a diagonal matrix with precision priors set to 0.1. Finally, the distribution for the firm characteristic parameters in the inefficiency are: $\gamma \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\gamma)$ where $\boldsymbol{\Sigma}^{-1}_\gamma$ is a diagonal with priors set to 0.1 for every precision coefficient.

Regarding the persistence parameters, we assume $|\rho_i| < 1$ to assure stationarity. Since the persistence parameters are allowed to vary across firms, we define a hierarchical structure where $\rho_i = 2k_i - 1$ and $k_i \sim \beta(k, 1-k)$ with $k \sim \beta(r, s)$ and priors for these parameters set to 0.5. In the case that the homogeneous persistence restriction $\rho_i = \rho$ is imposed, we assume $\rho = 2k - 1$ with k defined as previously.

The variances are assumed to follow inverse gamma distributions where $\sigma^{-2}_\eta, \sigma^{-2}_\xi \sim G(n, d)$ with priors set to $n = 10, d = 0.01$ and $n = 0.5, d = 0.005$, respectively.[3]

Markov Chain Monte Carlo (MCMC) methods and in particular the Gibbs Sampling algorithm with data augmentation can be used here (see Koop, Steel, & Osiewalski, 1995, for its introduction to Bayesian inference of stochastic frontier models). The implementation of our models is carried out using the WinBUGS package (see Griffin & Steel, 2007, for a general procedure). The MCMC algorithm involves 50,000 iterations where the first 20,000 are discarded and a thinning equal to 6 is used to remove autocorrelations. Therefore, 5000 iterations are used for the posterior inference.

Sensitivity analysis was performed by allowing changes in the priors of the parameters in the inefficiency component. In particular, different priors for ω imply different priors on the efficiency but in our experiments, no important differences were obtained in the posterior distributions. For the persistence parameter ρ we studied the sensitivity to the use of a truncated normal distribution and posterior results were also found to be robust to the use of this alternative. Small changes in the values of n and d in the priors of σ^{-2}_η and σ^{-2}_ξ were also examined with no evidence of posterior sensitivity. Finally, we also checked the posterior distribution of the idiosyncratic errors $\mathbf{v}$ to check the normality and non autocorrelation assumptions. We found no evidence of non-normality or of autocorrelation in this case. Note however that in cases where the idiosyncratic errors do not appear to be normal, one possibility is to model using a heavy tailed distribution such as student-*t* (see Griffin & Steel, 2007, for the implementation of this assumption under a Bayesian framework). In the case of autocorrelation, it is also possible to think of an autoregressive structure for this component.

3.1. Comparison criteria

As model selection criteria, we use a version of the Deviance Information Criterion (DIC) called DIC_3 and the Log Predictive Score (LPS), which compares the models in terms of their predictive performance. Both criteria are calculated using the MCMC output.

[1] It can be noticed that if unobserved random shocks are only allowed to affect the inefficiency via the scale parameter of its distribution, then $\eta_{it} = 0$ and the inefficiency can be modeled following other nonnegative distributions with firm specific and time-varying mean. However, other distributions can change the interpretation of the inefficiency parameters.

[2] These values center the efficiency prior distributions at 0.8 similar to other Bayesian empirical applications in banking.

[3] The first is the same prior used by Tsionas (2006) for the random shocks variance in the inefficiency equation and the second is that suggested by West and Harrison (1997) for the state equation of Bayesian dynamic linear models.

The DIC_3 is a stable variant of the within sample measure of fit introduced by Spiegelhalter, Best, Carlin, and van der Linde (2002) commonly used in Bayesian analysis. Defining the deviance of a model with parameters θ as $D(\theta) = -2\log f(\mathbf{y}|\theta)$, where $\mathbf{y}$ are the data, then $DIC = 2\overline{D(\theta)} - D(\bar{\theta})$. However, using an estimator of the density $f(\mathbf{y}|\theta)$ instead of the posterior mean $\bar{\theta}$ is more stable. This alternative specification was first proposed by Richardson (2002) and presented by Celeux, Forbes, Robert, and Titterington (2006) to overcome problems when the original DIC is implemented to random effects and mixture models. Li, Zeng, and Yu (2012) also remark on the lack of robustness of the original DIC in models with data augmentation. The formulation for this criterion is:

$$DIC_3 = -4E_\theta[\log f(\mathbf{y}|\theta)|\mathbf{y}] + 2\log\widehat{f}(\mathbf{y}). \tag{5}$$

On the other hand, LPS is a proper scoring rule introduced by Good (1952) that examines how well a model performs when its implied predictive distribution is compared with observations not used in the inference sample. In this sense, it evaluates the out-of-sample behaviour of different models by mean of their divergence between the actual sampling density and the predictive density.[4] In general, the procedure consists of partitioning the sample into two sets. The first, is a training data set used to fit the model and the second is a prediction set used to evaluate the predictive performance of the first set. In particular, our implementation defines as the training data set one that contains the observations up to the penultimate time period at which data are observed for each firm. Therefore, the prediction data set will consist of the last observed time period for each firm. That is, if t_i represents the index of the last time point when data are observed for firm i, the predictive set contains the set of observations y_{1,t_1} to y_{k,t_k} for the k firms in the sample. The average of the log predictive density functions evaluated at observed out-of-sample values are then calculated. The formula for the LPS is the following:

$$LPS = -\frac{1}{k}\sum_{i=1}^{k}\log f(y_{i,t_i}|\text{previous data}) \tag{6}$$

4. Stochastic input distance function

We represent the technology in (1) with an input distance function. This allows us to consider input quantities while accounting for multiple products, which avoids the problem of using firm specific prices that may not fulfill the competitive input prices assumption (see Berger & Mester, 2003).

We assume that banks employ an $N \times 1$ vector of inputs $\mathbf{x} = (x_1, x_2, \ldots, x_N)'$ to provide an $M \times 1$ vector of service outputs $\mathbf{q} = (q_1, q_2, \ldots, q_M)'$. We also define the following input set:

$$L_g(q) = \mathbf{x} : \mathbf{x} \text{ and technology } g \text{ can produce } \mathbf{q}, \tag{7}$$

where the technology g satisfies the axioms of closeness, boundedness, strong disposability and convexity as described by Färe and Primont (1995). Based on the input set, it is possible to define the following input distance function that describes the regular technology g:

$$D_I(\mathbf{x}, \mathbf{q}, g) = \sup_\lambda\{\lambda : \mathbf{x}/\lambda \in L_g(\mathbf{q}) \geqslant 1\}, \tag{8}$$

where λ denotes the maximum amount by which an input vector can be radially contracted while the output vector remains constant. We assume that in each period all banks use the best available technology. Therefore, for bank i in period t, the Debreu–Farrell input-oriented measure of technical efficiency (TE) is:

$$TE(x_{it}, q_{it}, t) \equiv 1/D_I(x_{it}, q_{it}, t). \tag{9}$$

[4] See Griffin and Steel (2004) and Ferreira and Steel (2007) for previous applications of LPS in stochastic frontier models.

The input distance function is homogeneous of degree one, a non-decreasing concave function of inputs and a non-increasing quasi-concave function of outputs (see Färe & Primont, 1995). Linear homogeneity implies that it is possible to normalize all the inputs in the distance function by an arbitrarily chosen input $x_{N_{it}}$:

$$1/x_{N_{it}} = D_I(x_{it}/x_{N_{it}}, q_{it}, t)\exp(-u_{it}), \tag{10}$$

where $u_{it} \equiv \ln D_I(x_{it}/x_{N_{it}}, q_{it}, t) \geqslant 0$. Therefore, a bank is technically efficient if and only if $u_{it} = 0$ or similarly, $TE(x_{it}, q_{it}, t) = 1$.

Regarding the technology representation, we use a translog functional form to parameterize the distance function. So, we define $v_{it} \equiv \ln D_I(x_{it}/x_{N_{it}}, q_{it}, t) - TL(x_{it}/x_{N_{it}}, q_{it}, t)$, where $TL(.)$ is the translog function. Then, (10) becomes:

$$y_{it} = TL(x_{it}/x_{N_{it}}, q_{it}, t) + v_{it} - u_{it}, \tag{11}$$

where $y_{it} \equiv -\ln x_{N_{it}}$. If any outputs or normalized inputs are stochastic then v_{it} is stochastic and (11) becomes a standard translog stochastic frontier model. For the estimation, the inefficiency term u_{it} is assumed to follow a nonnegative distribution and the random noise component v_{it} is assumed to follow a normal distribution.

From (11), the technical efficiency of individual firms in each period is calculated as:

$$TE_{it} = \exp(-u_{it}). \tag{12}$$

Returns to scale (RTS) can be derived as the sum of output elasticities as follows:

$$RTS = -\left(\sum_{m=1}^{M}\frac{\partial \ln D_I(\mathbf{x}, \mathbf{q}, t)}{\partial \ln q_m}\right)^{-1}, \tag{13}$$

where a RTS measure less than 1 indicates that the production technology present decreasing returns to scale. On the other hand, increasing returns to scale are observed if the RTS measure is larger than 1, while if it is equal to 1 it indicates constant returns to scale.

Finally, technical change (TC) measuring common shifts in the input distance function is given by:

$$TC = \left(\frac{\partial \ln D_I(\mathbf{x}, \mathbf{q}, t)}{\partial t}\right). \tag{14}$$

5. Empirical application

In this section we describe briefly the Colombian banking sector, give details about the model specification and data, and present the estimation results by comparing different specifications derived from the general model and analyzing the empirical implications.

5.1. The Colombian banking sector

The Colombian banking sector has experienced major changes in the last thirty years. It passed from a high regulated and low competitive system in the eighties to a more flexible and foreign capital open system in the nineties. From 1998 to 2002 the country suffered a deep financial crisis that lead to a rearrangement of the banking sector. This implied a reduction of the number of banks and a concentration of commercial and mortgage activities under the same institution. This reorganization process occurred during the period 2002 till 2009, which was characterized by an environment of economic recovery, high foreign capital flows and an increase of the services provided by banks. During these years, several M&A processes took place leading to a reduction of the number of financial institutions, which passed from more than forty mortgage and commercial banks in the mid 90s to less than twenty in 2009. Foreign capital banks had also played an important role in Colombia during the period of study and they accounted for almost 40% of the banking entities in 2009.

Previous efficiency studies of Colombian banking system have mainly evaluated costs and profit efficiency and have focused on the crisis and immediate post-crisis periods (see Janna, 2003, for a review on applications to the Colombian banking sector). All of these studies have shown similar results in terms of an increase in the efficiency of the sector during the mid-nineties, decreases in efficiency during the crisis period and a recovery on these indicators in the following years. The effect of bank mergers in Colombia has been studied by Estrada (2005) for the 1994–2004 period who found gains in cost efficiency specially for relatively inefficient pre-merger banks. Clavijo, Rojas, Salamanca, Montoya, and Rizo (2006) also studied M&A from 1990 to 2005 finding decreases in efficiency in the subsequent periods to the processes. However, most of these occurred during the crisis period.

5.2. Data and model specification

The data set contains information from 31 commercial banks, which represents 87% of the total assets in the Colombian banking sector. This is an unbalanced panel data set of quarterly data from 2000 to 2009 from the local central bank and the supervisory agency. During the period, nineteen M&A processes were carried out involving banks in the sample and here, we shall consider post-merged institutions as different banks. Regarding ownership, nine banks in the sample are foreign-owned and only one is public-owned.

In this work, we use the intermediation approach where banks are assumed to collect deposits and other liabilities in order to transform them into earning assets. In particular, we follow the assets approach proposed by Sealey and Lindley (1977) where banks use labor, capital and deposits to produce loans, securities and other investments.[5]

We select three inputs and two outputs. Inputs are quantities of deposits (x_1), labor (x_2) and physical capital (x_3), including premises and other fixed assets. As outputs we consider the total loans (y_1), including consumer, industrial, commercial and real state loans; and the total investments and other securities (y_2). All monetary variables are expressed in thousand of millions of pesos and are in real terms of 2009 by deflating by the consumer price index. Regarding the inefficiency heterogeneity variables, they are included either inside the dynamic inefficiency component or out of it but the same variable is not simultaneously included in both parts. These variables are the log of total banks assets (z_1/h_1), its square (z_2/h_2) and foreign ownership (z_3/h_3). Public ownership is not considered since the sample contains only one bank with public capital. Table 1 reports summary statistics of these variables.

Input x_3 is used as a numeraire to accomplish linear homogeneity in inputs and a translog input distance function derived from (11) is used. The estimated model including the dynamic specification in (2)–(4) for the inefficiency distribution is the following:

$$-\ln x_{3_{it}} = \beta_0 + \sum_{m=1}^{2}\beta_m \ln y_{m_{it}} + \sum_{r=1}^{2}\delta_r \ln\left(\frac{x_{r_{it}}}{x_{3_{it}}}\right) + \frac{1}{2}\sum_{m=1}^{2}\sum_{n=1}^{2}\beta_{mn}\ln y_{m_{it}}\ln y_{n_{it}} + \frac{1}{2}\sum_{r=1}^{2}\sum_{s=1}^{2}\delta_{rs}\ln\left(\frac{x_{r_{it}}}{x_{3_{it}}}\right)\ln\left(\frac{x_{s_{it}}}{x_{3_{it}}}\right) + \sum_{m=1}^{2}\sum_{r=1}^{2}\eta_{mr}\ln y_{m_{it}}\ln\left(\frac{x_{r_{it}}}{x_{3_{it}}}\right) + \kappa_1 t + \frac{1}{2}\kappa_2 t^2 + \sum_{m=1}^{2}\phi_m t \ln y_{m_{it}} + \sum_{r=1}^{2}\varphi_r t \ln\left(\frac{x_{r_{it}}}{x_{3_{it}}}\right) - u_{it} + v_{it} \quad (15)$$

$$\log u_{it} = \theta_{it} + \sum_{p=1}^{3}\gamma_p z_{p_{it}} + \xi_{it};\ \xi_{it} \sim N(0, \sigma_\xi^2) \quad (16)$$

$$\theta_{it} = \omega + \rho_i\theta_{i,t-1} + \sum_{p=1}^{3}\psi_p h_{p_{it}} + \eta_{it};\ \eta_{it} \sim N(0, \sigma_\eta^2);\quad t = 2\ldots T \quad (17)$$

$$\theta_{i1} = \frac{\omega + \sum_{p=1}^{3}\psi_p h_{p_{i1}}}{1-\rho_i} + \eta_{i1};\ \eta_{i1} \sim N\left(0, \frac{\sigma_\eta^2}{1-\rho_i^2}\right);\quad t = 1. \quad (18)$$

[5] Different approaches have been found to lead to similar conclusions (see Mester, 1993).

Table 1
Summary statistics.

Variable	Mean	Std. deviation	Minimum	Maximum
Deposits	3 886 117.8	4 661 207.2	146 005.1	29 600 000
Labor	2 984.5	2 676.6	79	20 780
Physical capital	93 036.7	101 070.5	5 359.075	710 837.1
Loans	3 305 469.4	4 195 981.4	132 508.6	27 900 000
Investments	1 357 952.7	1 472 720.2	32 466.74	8 277 268
Assets	5 643 177.9	6 576 929.2	319 757.3	41 700 000

In addition to linear homogeneity in inputs, we impose cross-effects symmetry by requiring $\beta_{mn} = \beta_{nm}$ and $\delta_{rs} = \delta_{sr}$.

5.3. Estimation results

Using the specification in (15)–(18) we estimate four different models by adding some restrictions. Model I follows the same inefficiency specification in Tsionas (2006) by including all three heterogeneity variables in the inefficiency dynamics. Therefore, $\rho_i = \rho$ and $\gamma_1, \gamma_2, \gamma_3$ and ξ_{it} are all equal to 0. Model II consider all heterogeneity variables out of the dynamic component but still restricts persistence to be common for all banks. Thus, $\rho_i = \rho$ and ψ_1, ψ_2 and ψ_3 are equal to zero. Model III combines heterogeneity variables in and out the dynamic component. In particular, following results in Models I and II, we set γ_3, ψ_1, and ψ_2 equal to zero and we keep $\rho_i = \rho$. Finally, Model IV uses the same combination of heterogeneity variables in Model III but allows for bank specific persistence parameters ($\rho_i \neq \rho$).

Table 2 presents the estimation results for all models. If we compare Model I to Model II we observe two main relevant results. First, Model II exhibits lower values for DIC_3 and LPS suggesting better fit and prediction performance when heterogeneity is modeled out of the inefficiency dynamics. Second, variables regarding size become relevant as technical inefficiency drivers and seem to present negative but decreasing effects. This would suggest that size affects the inefficiency level at every period but that its effects can be rapidly adjusted. On the other hand, foreign ownership presents relevant negative effects in technical inefficiency under both models. Consequences of these differences in the technical efficiency estimations are explored by selecting banks with different characteristics in terms of size and ownership.

Fig. 1 compares the posterior technical efficiency distributions for two banks with different sizes obtained from both models. One bank from the first quartile (Bank A) and one bank from the fourth quartile (Bank B) of the sample are selected in terms of assets level. We observe that while in Model I the posterior distributions of the technical efficiencies of both banks are almost undistinguishable, in Model II Bank B seems to have a high probability of being more efficient than Bank A. This shows that size becomes important for differentiating banks in terms of their technical efficiency only when it is modeled out of the dynamic component. This warns us of the possibility of biases in efficiency estimations and wrong conclusions about the effects of heterogeneity variables in dynamic inefficiency models when their effect is only considered as part of the dynamics.

Foreign ownership is found to be a relevant inefficiency driver when it is included both in and out of the dynamic component. Thus, we explore the consequences for the efficiency estimations of a foreign bank under both models and in two different periods. Fig. 2 shows that posterior efficiency distributions are quite similar in both models in the first period but that they become different at the end of the sample period. Since the value of the variable for this bank is the same during all the sample, this may suggest that the effect of including a covariate in the inefficiency dynamics could be cumulative after many periods despite of the lower estimate for the coefficient. A possible reason is that given that persistence

Table 2
Posterior mean and standard deviation of parameter distributions.

Parameter	Model I $\gamma_1, \gamma_2, \gamma_3 = 0$ $\rho_i = \rho$		Model II $\psi_1, \psi_2, \psi_3 = 0$ $\rho_i = \rho$		Model III $\psi_1, \psi_2, \gamma_3 = 0$ $\rho_i = \rho$		Model IV $\psi_1, \psi_2, \gamma_3 = 0$ $\rho_i \neq \rho$	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Distance function								
β_0	−5.3801	0.3807	−5.3883	0.3813	−5.4324	0.2145	−5.2458	0.3438
$\beta_1(\ln y_i)$	−0.0216	0.0010	−0.0443	0.0021	−0.0344	0.0019	−0.0167	0.0013
$\beta_2(\ln y_c)$	−0.0065	0.0027	−0.0029	0.0012	−0.0061	0.0005	−0.0080	0.0007
$\beta_3(\ln dep)$	0.1290	0.0169	0.2284	0.0299	0.2240	0.0560	0.1773	0.0260
$\beta_4(\ln l)$	0.1029	0.0243	0.1172	0.0277	0.1198	0.0341	0.0947	0.0272
$\beta_5(t)$	0.0206	0.0097	0.0283	0.0133	0.0153	0.0070	0.0193	0.0031
$\beta_6(t^2)$	−0.0004	0.0002	−0.0003	0.0002	−0.0002	0.0001	−0.0003	0.0001
$\phi_1(1/2 \ln y_i^2)$	−0.0808	0.0232	−0.0844	0.0243	−0.0857	0.0200	−0.0873	0.0234
$\phi_2(\ln y_i \ln y_c)$	0.0734	0.0225	0.0713	0.0218	0.0692	0.0179	0.0644	0.0153
$\phi_3(1/2 \ln y_c^2)$	−0.1429	0.0223	−0.1458	0.0227	−0.1346	0.0219	−0.1343	0.0153
$\phi_4(1/2 \ln dep^2)$	−0.2878	0.0988	−0.3098	0.1064	−0.3230	0.0491	−0.4601	0.0433
$\phi_5(\ln dep \ln l)$	0.1136	0.0491	0.1334	0.0577	0.1293	0.0379	0.1572	0.0569
$\phi_6(1/2 \ln l^2)$	−0.0699	0.0437	−0.1004	0.0628	−0.0753	0.0299	−0.0380	0.0158
$\delta_1(\ln y_i \ln dep)$	−0.0332	0.0367	−0.0296	0.0327	−0.0122	0.0164	−0.0163	0.0175
$\delta_2(\ln y_i \ln l)$	−0.0088	0.0123	−0.0187	0.0260	−0.0251	0.0157	−0.0360	0.0127
$\delta_3(\ln y_c \ln dep)$	0.1530	0.0306	0.1707	0.0342	0.1420	0.0165	0.1486	0.0233
$\delta_4(\ln y_c \ln l)$	−0.0171	0.0244	−0.0170	0.0244	−0.0118	0.0123	−0.0460	0.0130
$\kappa_1(\ln y_i t)$	0.0007	0.0012	0.0011	0.0018	0.0004	0.0011	0.0028	0.0047
$\kappa_2(\ln y_c t)$	−0.0050	0.0022	−0.0058	0.0025	−0.0043	0.0010	−0.0054	0.0018
$\kappa_3(\ln dept)$	0.0096	0.0054	0.0092	0.0051	0.0097	0.0028	0.0117	0.0074
$\kappa_4(\ln lt)$	−0.0068	0.0037	−0.0062	0.0034	−0.0068	0.0027	−0.0016	0.0114
Inefficiency								
ω	0.0022	0.0003	0.0024	0.0002	0.0015	0.0002	0.0012	0.0002
$\psi_1(\ln assets)$	−0.0001	0.0003						
$\psi_2(\ln assets^2)$	0.0000	0.0001						
$\psi_3(foreign)$	−0.0033	0.0008			−0.0041	0.0003	−0.0032	0.0003
ρ	0.9954	0.0172	0.9749	0.0168	0.9670	0.0183	0.8867	0.0452
$\gamma_1(\ln assets)$			−0.0245	0.0011	−0.0189	0.0010	−0.0458	0.0017
$\gamma_2(\ln assets^2)$			0.0002	0.0001	0.0009	0.0002	0.0023	0.0007
$\gamma_3(foreign)$			−0.2061	0.0162				
σ_v^2	0.0012	0.0008	0.0011	0.0007	0.0013	0.0009	0.0017	0.0009
σ_η^2	0.0006	0.0002	0.0003	0.0001	0.0007	0.0002	0.0010	0.0002
σ_ξ^2			0.0005	0.0002	0.0002	0.0001	0.0002	0.0001
DIC_3	−2789.60		−2949.20		−2961.52		−3094.10	
LPS	−213.72		−253.48		−261.79		−295.63	

Note: In Model IV the values reported for ρ correspond to the posterior means of the individual ρ_i's.

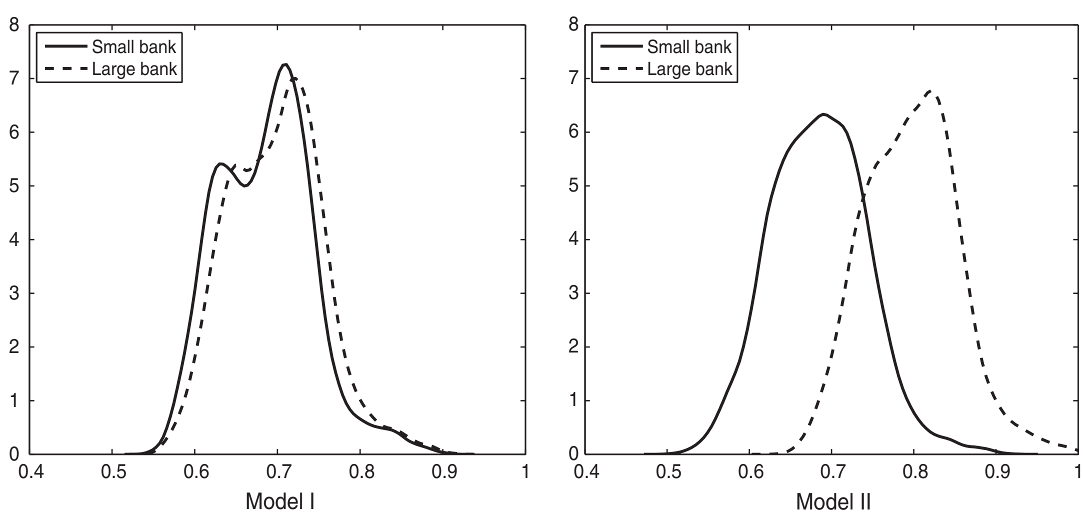

Fig. 1. Posterior technical efficiency distribution for a small and large bank.

is very high, most of the effect of the covariate is transmitted to the next period, where once more it affects the inefficiency.

Since the location of foreign ownership may lead to different efficiency estimates, we estimate a third model. Model III includes foreign ownership in the inefficiency dynamics while keeps the assets variables out of this component. Results in terms of fitting performance and prediction improve compared to those of the previous models and the coefficient for the variable remains relevant

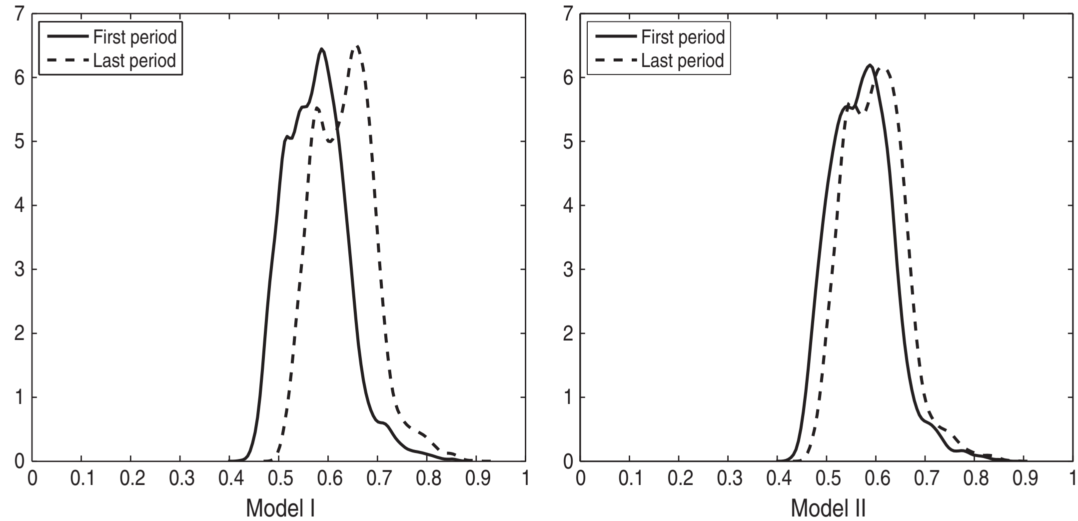

Fig. 2. Posterior technical efficiency distribution of a foreign bank.

as inefficiency driver. This suggests that foreign-owned banks present lower technical inefficiency and that the effects derived from this type of ownership are persistent over time.

In general, inefficiency persistence is found to be very high in all models. This result is consistent to that obtained by Tsionas (2006) for US banks. This suggests that adjustment costs are very important in the banking sector and force these institutions to remain inefficient in the short run. However, banks may also present differences in these adjustment costs depending on their own characteristics. We explore these differences among banks by including a firm-specific persistence parameter in Model IV. We observe that this model exhibits lower DIC_3 and *LPS* values compared to Model III, which suggests that recognizing heterogeneous costs of adjustment may improve the fit and predictive performance of the model and have effects on the evolution of efficiency if the estimated persistence parameters are very different between banks.

In order to identify these differences, in Fig. 3 we plot the posterior mean and 95% probability intervals for ρ_i. We classify banks in three main categories: foreign and domestic banks, large and small banks, and merged and non-merged institutions.[6] We observe differences in the average posterior mean between all complementary groups. This suggests that size, foreign ownership, and being involved in M&A processes are possible sources of inefficiency persistence. In particular, foreign, large and merged banks are more likely to present lower adjustment costs than their counterparts. However, the persistence parameters between merged and non-merged institutions are those with the highest probability of being different as suggested from the very small overlapping between both probability intervals. The average posterior mean for ρ_i among merged banks (0.71) is the lowest compared to those of the other groups and it is not only far from that estimated for non-merged banks (0.94) but also from that estimated for ρ in Model III (0.97) when the parameter is assumed to be common to all banks.

These differences may have important effects in the dynamic behavior of inefficiency over time between both groups of banks. To illustrate these effects we plot in Fig. 4 the evolution of the mean posterior technical efficiency estimated from models III and IV for merged and non-merged banks. It is observed that efficiency of merged banks decreases immediately in both models after these processes are carried out. However, Model IV identifies a rapid recovery of the efficiency of merged banks that starts around three years after the merging process and reaches the non-merged efficiency levels after five years. This pattern is totally different from that identified in Model III, where technical efficiency of merged institutions seems to remain lower than that of non-merged banks.

These results contrast with those of previous studies that measure the effect of banks M&A whether on cost or input-oriented technical efficiency (see Amel et al., 2004, for a review on studies in developed countries). However, the pattern on the evolution of input-oriented technical efficiency that we see for Colombian merged banks in Model IV is similar to that identified by Cuesta and Orea (2002) in a study of output-oriented technical efficiency of Spanish merged banks. In that study, technical efficiency was found to exhibit a concave pattern with negative but decreasing effects during the first six years after mergers, and positive increasing effects after that point. Although the model estimated by Cuesta and Orea (2002) is not dynamic in nature, it allows merged banks to follow a different temporal pattern to that of non-merged institutions. This may suggest that mergers lead to different evolution processes of the inefficiency and that models recognizing these differences are more appropriate.

With respect to the inefficiency drivers in the Colombian banking sector, foreign ownership and size are found to have positive effects on technical efficiency. However, the impact is decreasing for size. Moreover, we identify that the effects of size on inefficiency can be rapidly adjusted by Colombian banks, while the advantages presented by foreign banks are difficult and costly to reach or adjust.

Finally, we compare TE, TC and RTS by groups of banks following the results from Model IV. Table 3 summarizes these findings. We observe that foreign banks in Colombia present higher technical efficiency than domestic institutions, as well as higher technical change during the last decade. These findings coincide with those reported in other recent studies for developing countries using non-dynamic models (see Claessens & Horen, 2012). Beyond managerial practices, the reasons could be related to more diversification, parents expertise, or access to cheaper and multiple sources of financial resources (see Chen & Liao, 2011). In contrast to domestic institutions, foreign banks also present increasing returns to scale, suggesting that these institutions have more room to raise their production scale and possibly to take M&A decisions. Foreign

[6] We define small and large banks as those below and above the median of assets level, respectively.

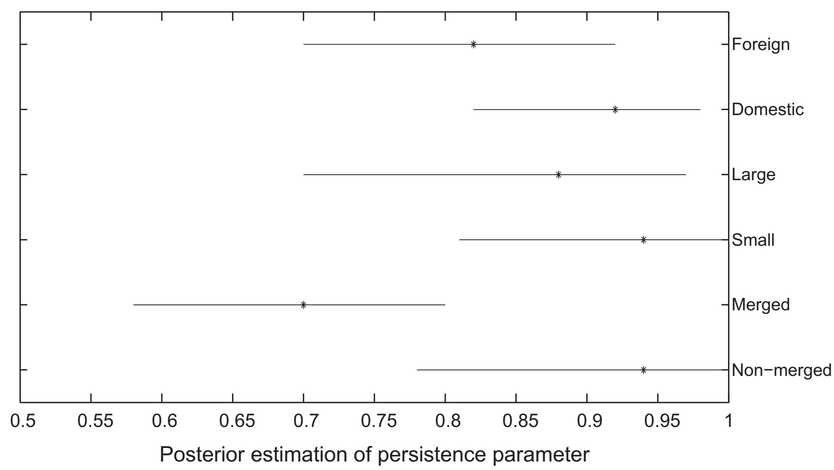

Fig. 3. Posterior median and 95% probability intervals for firm specific persistence parameters by type of bank.

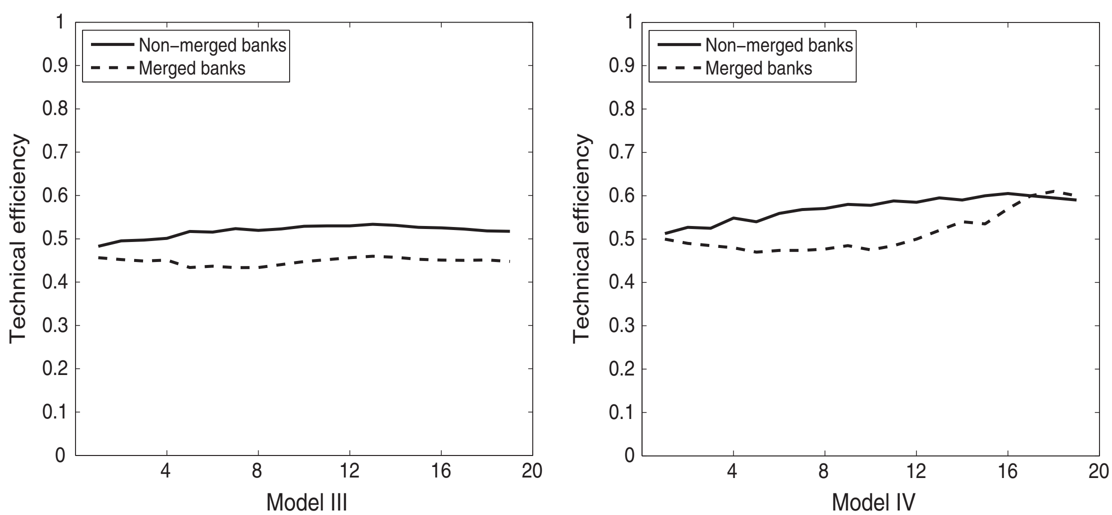

Fig. 4. Evolution of mean posterior efficiencies for merged and non-merged banks.

banks in Colombia are characterized for being specialized in corporate clients, offer complex products and have few branches with low operations. In a recent study, Das and Kumbhakar (2011) also found similar scale economies for foreign banks in India.

In terms of size, we find that large banks present higher TE and TC than small institutions during the period. However, large institutions are found to operate at decreasing returns to scale in contrast to small banks. Higher efficiency of large banks and potential scale gains for small banks were also recently found by Tabak and Tecles (2010) and Tecles and Tabak (2010) in India and Brazil, respectively. In particular, in Colombia most large domestic banks are those involved in merger processes. Since merged institutions also present decreasing returns to scale, this may suggest that these processes led banks to be oversized. Also, on average, merged banks exhibit lower technical efficiency than non-merged institutions. However, they are found to present lower adjustment costs that allow them to adjust quicker towards optimal conditions. Thus, they would be able to present higher efficiency after some periods. Finally, technical change is also found to be higher for merged than for non-merged banks and it can be a consequence of the reorganization processes implied by mergers.

Table 3
TE, TC and RTS by type of bank.

Bank type	TE	TC	RTS
Foreign	0.6011	0.0307	1.0986
Domestic	0.5476	0.0251	0.9180
Large	0.5853	0.0285	0.9202
Small	0.5304	0.0278	1.0316
Merged	0.5076	0.0326	0.9021
Non-merged	0.5512	0.0267	1.0633

6. Concluding remarks

In the presence of adjustment costs, firms do not find it optimal to adapt their processes towards efficiency in the short-run. This behaviour can be captured through a dynamic specification for the inefficiency term. One of the most relevant contributions in this context is that by Tsionas (2006) where the inefficiency is allowed to have persistent effects over time and to be driven by inefficiency covariates. In this work we have extended this idea in order to recognize heterogeneity in the adjustment costs among firms and non-persistent effects of observed heterogeneity.

Our findings suggest that modeling covariates in and out of the inefficiency dynamics have implications on the identification of inefficiency determinants and on the efficiency estimations. In particular, we find that foreign ownership has negative and persistent effects in technical inefficiency of Colombian banks. This may suggest that some characteristics of foreign banks such as country diversification and access to cheaper funding sources can be difficult and costly to obtain or change. On the other hand, the effects of bank size on technical inefficiency are found to be rapidly adjusted.

Colombian banks are also found to present very high inefficiency persistence, coinciding with previous findings in the US and Spanish banking sectors (see Tsionas, 2006, respectively; Tortosa-Ausina, 2002). However, important differences are observed among banks with different characteristics when firm specific persistence parameters are modeled. Foreign, large and merged institutions are found to present lower adjustment costs than their counterparts. This suggests that these institutions may benefit from diversification or economies of scale when carrying out adjustments in their short-run operations as these are costly for domestic, small and non-merged banks. This finding is particularly important for merged banks since this characteristic allows them to recover rapidly from efficiency losses observed after merger processes.

These results are of interest not only for financial institutions, but also for regulators given the importance that M&A have had in the sector in recent years and the role of foreign banks in developing countries. In particular, although, our findings reveal important decreases in efficiency of merged institutions during the initial years after these processes are carried out, the lower inefficiency persistence of banks involved in M&A and the non-persistent effects of size on inefficiency may validate these processes in the mid-term. However, the Colombian regulator should be aware of the results on economies of scale, which leave little margin for non-merged institutions to increase their size and reveal decreasing returns for merged and large banks. Exploring market power considerations would be also of interest for future policy decisions in the sector. In general, bank efficiency may be a useful indicator for financial stability considerations given that banks with low efficiency have been found to be more prone to future defaults (see Berger & DeYoung, 1997). In this regard, those banks with high inefficiency persistence should be drawn to the attention of the regulator.

Overall, the proposed specification encompasses other models in the literature and adds more flexibility in terms of considering inefficiency heterogeneity in a dynamic context. This improves the fit and predictive performance of the model and allows us to capture effects that have not been previously identified with other models. As possible extensions of interest we can think of modeling potential unobserved technological heterogeneity sources as in Galán and Pollitt (2014) in the context of dynamic SFA models, and the inclusion of other bank inefficiency drivers such as loans and investment attributes, (see Tecles & Tabak, 2010). Work is currently in progress on including risk measures and recognizing bank specific effects of risk exposure as inefficiency drivers.

Acknowledgements

The authors would like to thank Christopher O'Donnell for his discussion of the paper and his valuable suggestions as well as Mike Tsionas and the participants of the 13th European Workshop on Efficiency and Productivity for the comments received. Financial support from the Spanish Ministry of Education and Science, research Projects ECO2012-3401, MTM2010-17323, SEJ2007-64500 and ECO2012-38442 is also gratefully acknowledged.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ejor.2014.07.005.

References

Ahn, S., Good, D., & Sickles, R. (2000). Estimation of long-run inefficiency levels: A dynamic frontier approach. *Econometrics Review, 19*, 461–492.
Aigner, D., Lovell, C., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics, 6*, 21–37.
Amel, D., Barnes, C., Panetta, F., & Salleo, C. (2004). Consolidations and efficiency in the financial sector: A review of the international evidence. *Journal of Banking and Finance, 28*, 2493–2519.
Assaf, A., Matousek, R., & Tsionas, E. (2013). Turkish bank efficiency: Bayesian estimation with undesirable outputs. *Journal of Banking and Finance, 37*, 506–517.
Battese, G., & Coelli, T. (1992). Frontier production, function, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis, 3*, 153–169.
Battese, G., & Coelli, T. (1995). A model for technical inefficiency effects in a stochastic frontier production model for panel data. *Empirical Economics, 20*, 325–332.
Berger, A., & DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks. *Journal of Banking and Finance, 21*(6), 849–870.
Berger, A., & Mester, L. (2003). Explaining the dramatic changes in the performance of U.S. banks: Technological change, deregulation, and dynamic changes in competition. *Journal of Financial Intermediation, 12*, 57–95.
Bos, J., Koetter, M., Kolari, J., & Kool, C. (2009). Effects of heterogeneity on bank efficiency scores. *European Journal of Operational Research, 195*, 251–261.
Celeux, G., Forbes, F., Robert, C., & Titterington, D. (2006). Deviance information criteria for missing data models. *Bayesian Analysis, 4*, 651–674.
Chen, S., & Liao, C. (2011). Are foreign banks more profitable than domestic banks? Home- and host-country effects of banking market structure, governance, and supervision. *Journal of Banking and Finance, 35*, 819–839.
Claessens, S., Horen, N., 2012. *Foreign banks: Trends, impact and financial stability*. Working Paper, International Monetary Fund.
Clavijo, S., Rojas, C., Salamanca, C., Montoya, G., & Rizo, C. (2006). Mergers and acquisitions in the colombian financial sector: Impact on efficiency (1990–2005). *ICFAI Journal of Mergers and Acquisitions, 3*, 38–61.
Cornwell, C., Schmidt, P., & Sickles, R. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics, 46*, 185–200.
Cuesta, R., & Orea, L. (2002). Mergers and technical efficiency in spanish savings banks: A stochastic distance function approach. *Journal of Banking and Finance, 26*, 2231–2247.
Das, A., & Kumbhakar, S. (2011). Productivity and efficiency dynamics in indian banking: An input distance function approach incorporating quality of inputs and outputs. *Journal of Applied Econometrics, 93*, 205–234.
Emvalomatis, G. (2012). Adjustment and unobserved heterogeneity in dynamic stochastic frontier models. *Journal of Productivity Analysis, 37*, 7–16.
Emvalomatis, G., Stefanou, S., & Lansink, A. (2011). A reduced-form model for dynamic efficiency measurement: Application to dairy farms in Germany and The Netherlands. *American Journal of Agricultural Economics, 93*, 131–174.
Estrada, D., 2005. *An analysis of banking mergers in Colombia*. Financial Stability Report July, 69–79, Banco de la República's Monetary and Reserves Division.
Färe, R., & Primont, D. (1995). *Multi-output production and duality: Theory and applications*. Netherlands: Kluwer Academic Publishers.
Ferreira, J., & Steel, M. (2007). Model comparison of coordinate-free multivariate skewed distributions with an application to stochastic frontiers. *Journal of Econometrics, 137*, 641–673.
Galán, J., & Pollitt, M. (2014). *Inefficiency persistence and heterogeneity in Colombian electricity distribution utilities*. Cambridge Working Paper in Economics, EPRG Working Paper 1403, University of Cambridge.
Galán, J., Veiga, H., & Wiper, M. (2013). Bayesian estimation of inefficiency heterogeneity in stochastic frontier models. *Journal of Productivity Analysis*. http://dx.doi.org/10.1007/s11123-013-0377-4.
Good, I. (1952). Rational decisions. *Journal of the Royal Statistical Society B, 14*, 107–114.
Griffin, J., & Steel, M. (2004). Semiparametric Bayesian inference for stochastic frontier models. *Journal of Econometrics, 123*, 121–152.
Griffin, J., & Steel, M. (2007). Bayesian stochastic frontier analysis using WinBUGS. *Journal of Productivity Analysis, 27*, 163–176.
Hartman, T., & Storbeck, J. (1996). Input congestion in loan operations. *International Journal of Production Economics, 46-47*, 413–421.
Huang, T., & Chen, Y. (2009). A study on long-run inefficiency levels of a panel dynamic cost frontier under the framework of forward-looking rational expectations. *Journal of Banking and Finance, 33*, 842–849.
Janna, M., 2003. *Banking efficiency in Colombia: A review of the literature*. Financial Stability Report December, 78-85, Banco de la República's Monetary and Reserves Division.
Koop, G., Steel, M., & Osiewalski, J. (1995). Posterior analysis of stochastic frontier models using Gibbs sampling. *Computational Statistics, 10*, 353–373.
Kumbhakar, S. (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics, 46*, 201–212.
Lee, Y., & Schmidt, P. (1993). *The measurement of productive efficiency: Techniques and applications. Chapter production frontier model with flexible temporal variation in technical inefficiency*. Oxford University Press, Inc.
Lensik, R., Meesters, A., & Naaborg, I. (2008). Bank efficiency and foreign ownership: Do good institutions matter? *Journal of Banking and Finance, 32*, 834–844.

Li, Y., Zeng, T., & Yu, J. (2012). *Robust deviance information criterion for latent variable models*. Singapore Management University, Research Collection School of Economics (Open Access). Paper 1403.

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb–Douglas production functions with composed errors. *International Economic Review, 8*, 435–444.

Mester, L. (1993). Efficiency in the savings and loans industry. *Journal of Banking and Finance, 17*, 267–286.

Ray, S., & Das, A. (2010). Distribution of cost and profit efficiency: Eevidence from Indian banking. *European Journal of Operational Research, 201*, 297–307.

Richardson, S. (2002). Discussion of Spiegelhalter et al. *Journal of the Royal Statistical Society, Series B, 64*, 626–627.

Sathye, M. (2003). Efficiency of banks in a developing economy: The case of India. *European Journal of Operational Research, 148*, 662–671.

Sealey, C., & Lindley, J. (1977). Inputs, outputs and a theory of production and cost at depository financial institutions. *Journal of Finance, 32*, 1252–1266.

Spiegelhalter, D., Best, N., Carlin, B., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, 64*(4), 583–616.

Staub, R., da Silva e Souza, G., & Tabak, B. (2010). Evolution of bank efficiency in Brazil: A DEA approach. *European Journal of Operational Research, 202*, 204–213.

Tabak, B., & Tecles, P. (2010). Estimating a Bayesian stochastic frontier for the Indian banking system. *International Journal of Production Economics, 125*, 96–110.

Tecles, P., & Tabak, B. (2010). Determinants of bank efficiency: The case of Brazil. *European Journal of Operational Research, 207*, 1587–1598.

Tortosa-Ausina, E. (2002). Exploring efficiency differences over time in the Spanish banking industry. *European Journal of Operational Research, 139*, 643–664.

Tsionas, E. (2006). Inference in dynamic stochastic frontier models. *Journal of Applied Econometrics, 21*, 669–676.

van den Broeck, J., Koop, G., Osiewalski, J., & Steel, M. (1994). Stochastic frontier models: A Bayesian perspective. *Journal of Econometrics, 61*, 273–303.

West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models. Springer series in statistics*. New York: Springer-Verlag.