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