Bank efficiency and non-performing loans in Brazil: causality tests

Abstract

In general, periods of financial instability are associated with low bank efficiency and high default rates in banks' credit portfolios. Hence, this article intends to analyze the relationship between non-performing loans and bank efficiency in order to identify the primary instigator of bank failures. We use a Data Envelopment Analysis (DEA) model to estimate technical, allocative and economic efficiencies. Furthermore, both a dynamic panel and a panel VAR are applied to a recent semi-annual data ranging from june 2000 to june 2007, indicating that variations on bank efficiency precede loan quality deterioration. Therefore, bank efficiency measures may be important early warning indicators of financial instability, in particular, DEA efficiency measures should be used as macroeconomic tools to evaluate the financial system.

Key words: Bank efficiency, Non-performing loans, Causality, Financial stability

1 Introduction

In the last years, financial stability has been the center of many studies, specially after the burst of the subprime crisis in August 2007. In the current financial system, characterized as more liberalized, competitive and integrated [Koutsomanoli-Filippaki and Mamatzakis, 2009], preventing financial turmoils, such as bank failures, is of utmost relevance. To this end, an important issue is to try to identify which factors promote financial crises in order to obtain early warning indicators as an attempt to guarantee the soundness of the banking system. In this sense, literature has shown the important role that bank-level fundamentals play in explaining bank failures [Gonzalez-Hermosillo, 1999].

In particular, empirical works have verified that not only bank inefficiency contributes to the risk of failure [Wheelock and Wilson, 1995, Podpiera and Podpiera, 2005], but that asset quality deteriorates rapidly before the bankruptcy [Gonzalez-Hermosillo, 1999], as well. Hence, a fundamental question is whether there exists a causality relationship between bank efficiency and loan quality. As Podpiera and Weill [2008] note, if one event precedes the other, then it might be the core instigator of bank failures.

The 'bad luck' hypothesis [Berger and DeYoung, 1997] states that an increase in non-performing loans is caused by an unexpected exogenous event (bad luck), such as economic slowdowns or firms' breakdown. Banks will, consequently, incur higher costs in order to monitor these problem loans, decreasing efficiency. The extra expenses include diverse factors, among which: the extra costs of dealing with delinquent borrowers, the occasional reduction in the respective collateral value and the price of renegotiating debts. Thus, under this hypothesis, soaring problem loans lead to low levels of bank efficiency, i.e., there is a negative relationship between the two variables.

On the other hand, the 'bad management' hypothesis [Berger and DeYoung, 1997] considers that the low efficiency is related to subpar management, which is also reflected on a poor credit monitoring, affecting loan quality and resulting in high default (endogenous, in this case). Once again, bank efficiency and non-performing loans are negatively related, though in a different temporal ordering.

Two alternative hypotheses, by contrast, state a positive inter-temporal relationship from bank efficiency to credit risk. Firstly, under the 'skimping' hypothesis [Berger and DeYoung, 1997], firms maximizing profits in the long run may appear more efficient in the short run, when reducing costs by devoting less resources to monitoring and underwriting loans. Nonetheless, the financial institution will consequently have to cope with delinquent borrowers and the respective costs of dealing with these clients in the future. Conversely, the 'risk-averse management' hypothesis [Koutsomanoli-Filippaki and Mamatzakis, 2009] assumes bank managers to be risk-averse, so that they could increase operational expenses allocated to loan monitoring and screening, reducing bank efficiency, with the purpose of compressing default in credit portfolios. The fear of facing financial crises and information asymmetry are possible explanations for the positive relationship between bank efficiency and problem loans, in this case [1] .

Surprisingly, there seems to be only a limited number of studies testing if non-performing loans are caused by exogenous or endogenous events, what results in an inconclusive relationship between the credit risk variable and bank efficiency. The seminal work on this subject, Berger and DeYoung [1997], shows that for the industry of North American commercial banks as a whole, there is a negative relationship between problem loans and cost efficiency in both directions for the period from 1985 to 1994, although the data favor the 'skimping' hypothesis for a subset of banks persistently efficient across time. Accordingly, Rossi et al. [2009] also provide evidences for both the 'bad luck' and the 'bad management' hypotheses in the case of large Austrian commercial banks between 1997 and 2003. By contrast, Williams [2004], with a sample of European savings banks over the years 1990-1998, supports the 'bad management' hypothesis, indicating that low levels of cost efficiency deteriorate loan quality, consequently augmenting the probability of bank failure, in line with Fiordelisi et al. [2010] that use a data set of European commercial banks with respect to the period 1995-2007. Koutsomanoli-Filippaki and Mamatzakis [2009], alternatively, support the 'bad luck' hypothesis when analyzing 27 European Union members from 1998 to 2006. However, the 'bad management' hypothesis is not rejected, specially in the case of low financial development countries, albeit it performs a lower impact.

Nonetheless, there is still a poor focus on emerging countries. Firstly, Rossi et al. [2005] investigate nine European transition countries during the period 1995-2002 noting an exogenous origin for credit risk, consistent with the 'bad luck' hypothesis. Subsequently, Podpiera and Weill [2008] conclude that the main cause of bank failures in the Czech Republic between 1994 and 2005 is bad management.

In emerging countries, financial crises become particularly important as far as they reduce economic growth given that a great portion of the investments relies on banks' credit offer. Brazil is a meaningful case, since besides constituting one of the largest banking systems in Latin America, it also represents a complex financial market with roughly one third of banking assets of private

[1] A more detailed description of the hypotheses can be found in the related literature.

banks, one third of state-owned and one third of foreign participation. Hence, the evidences to be presented here are of multiple nationality and multiple bank type interest.

Our main contribution points in the direction of identifying relevant factors that may anticipate a systemic instability, in order to help bank regulators and supervisors to assure a healthy financial system. For this matter, by indicating that increases/decreases in bank efficiency precede variations on non-performing loans, our results provide evidence that DEA efficiency measures may be helpful in assessing potential future risks in the banking system. Therefore, they may be included in the regulator's toolkit to evaluate financial stability.

The remainder of this paper is structured as follows. Section 2 describes the methodology employed. Section 3 presents the data. The empirical results are available in section 4, whereas section 5 displays our final considerations.

2 Methodology

2.1 Efficiency Measure

The NPL ratio is directly available from the data, but as cost functions are not straightly observable, efficiency must be estimated. To estimate efficiency, we use the Data Envelopment Analysis (DEA) approach, in accordance to Staub et al. [2010]. We estimate three types of bank efficiency. Technical efficiency is associated with the efficient use of inputs within the bank technology, evaluating the amount of input used to produce certain output. On the other hand, allocative efficiency is related to how the mix of inputs affects the production process, since changes in the mix of inputs used can increase/decrease output. Overall efficiency is measured by economic efficiency.

Two important advantages of the non-parametric DEA methodology justify its use here. Firstly, the DEA demands very few assumptions about the technological production function, avoiding arbitrary suppositions about the efficient frontier shape. Moreover, the DEA is a widely employed technique and easy to work with complex relations between multiple outputs. In this sense, as Banker and Natarajan [2004] point out, technical, allocative and economic efficiencies can be computed through the use of total expenditures data only, without direct knowledge of particular input prices.

In studies of contextual variables [2], the DEA relies on statistical procedures of

[2] Factors likely to affect efficiency.

two stages. On the first stage, the efficiencies are estimated. On the second, the regression is modeled with efficiency as the dependent variable and contextual variables as independent. Technical problems may arise from this procedure due to possible correlation among efficiency measures. However, if the contextual variables are exogenous to the production process, Simar and Wilson [2007], Souza and Staub [2007], and Banker and Natarajan [2008] indicate that the two stage analysis is viable and, under certain error circumstances, can even capture stochastic and non-parametric efficiency results [3]. Motivated by these recent findings in DEA, we consider an extension for panel data in this work.

Consider a production process with n production units (banks). Each unit employs varying quantities of p distinct inputs x to produce varying quantities of s distinct products y. Denote by $Y = (y_1, \ldots, y_n)$ the production matrix $s \times n$ of the n banks and by $X = (x_1, \ldots, x_n)$ the input matrix $p \times n$. The element $y_r \geq 0$ is the production vector $s \times 1$ of bank r and x_r is the input vector $p \times 1$ used by bank r to produce y_r (the condition $l \geq 0$ means that at least one l component is strictly positive). The matrices $Y = (y_{ir})$ and $X = (x_{ir})$ must satisfy: $\sum_i l_{ir} > 0$ and $\sum_r l_{ir} > 0$, where l is x or y. In our application, $p = 3$ and $s = 3$. Furthermore, it will be necessary that $x_r, y_r > 0$ (which means that all components of the input and the output vectors are strictly positive).

In line with Banker and Natarajan [2004], we deal with technical, allocative and economic efficiencies employing aggregate cost variables. In this context, $C = (c_1, \ldots, c_n)$ is the total cost vector, where c_r is the total production cost of bank r and $V = (v_1, \ldots, v_n)$ is the input cost matrix. Here, v_{ir} is the expenditure of bank r with the input i (the ith component of vector v_r). The knowledge of a vector of input prices $g = (g_1, \ldots, g_p)$ results in $v_{ir} = g_i x_{ir}$ and $c_r = \sum_{i=1}^{p} v_{ir}$.

The procedure can be formulated as a linear programming. Thus, economic efficiency is computed as

$$\theta_r^e = \text{argmin}\,\{\theta;\ Y\lambda \geq y_r,\ C\lambda \leq \theta c_r,\ \lambda 1 = 1,\ \lambda \geq 0\}.$$

Technical efficiency is calculated as

$$\theta_r^{tec} = \text{argmin}\,\{\theta;\ Y\lambda \geq y_r,\ V\lambda \leq \theta v_r,\ \lambda 1 = 1,\ \lambda \geq 0\}.$$

[3] See Banker and Natarajan [2008].

Finally, allocative efficiency equals the ratio

$$\theta_r^a = \frac{\theta_r^e}{\theta_r^{tec}},$$

where θ is the measured bank efficiency and λ represents a vector of constants that form linear combinations of all inputs and outputs of the observed banks. The efficiency measures are computed for each bank for each of the T years to generate a panel of observations $(\theta_{it}^e, \theta_{it}^{tec}, \theta_{it}^a)$ with $t = 1, \ldots, T$ and $i = 1, \ldots, n$. The restriction, $\lambda 1 = 1$, results in a model with Variable Returns to Scale , i.e., for a given output level, minimizing the amount of input used still produce the output level predetermined.

We adopt statistical models to evaluate the significance of covariables (factors) in the efficiency measures, assuming independence between factors and errors. The models we use adjust to the category of dynamic panel data analysis and consider the serial correlation in the bank population. Cross-correlations among banks during the periods induced by DEA calculations seem to be despicable and, pursuant to Souza and Staub [2007] and Banker and Natarajan [2008], are not patterned.

2.2 Arellano Bond Dynamic Panel

The presence of the lagged dependent variable among the explanatory variables characterizes a dynamic model. Therefore, since more efficient banks in a specific year are inclined to remain efficient in the following year, we consider a dynamic model in the DEA response.

Our dynamic panel of Arellano and Bond [1991] considers

$$y_{i,t} = \alpha y_{i,t-1} + \beta x_{i,t-1} + u_{i,t} \qquad i = 1, \ldots, N \qquad t = 1, \ldots, T_i, \tag{1}$$

where α and β are the parameters to be estimated, y is EFF, the efficiency measured by DEA and x represents NPL, the ratio of non-performing loans divided by total loans [4]. We consider as non-performing loans the ones that are 90 days past due or more. We also estimate the model with $y_{i,t} = NPL$ in order to reproduce the panel VAR methodology and test the robustness of our results, as well as, to identify the optimal lag order for the panel VAR approach.

[4] We regress all our variables in ln. This applies to both the Arellano-Bond dynamic panel and the panel VAR.

Since the variables NPL and EFF are strongly affected by transitory conditions [Berger and DeYoung, 1997], one must control for these factors. We eliminate the aggregate bank-specific macro shocks by subtracting the respective mean of each variable calculated for each bank type-year. For this reason, we do not include time dummies in our Arellano-Bond dynamic panel estimation.

The error term is $u_{i,t} = v_i + \epsilon_{i,t}$, with $\epsilon_{i,t}$ representing the idiosyncratic disturbances. The bank-specific effects, v_i are assumed to be stochastic and not-correlated with $\epsilon_{i,t}$. Furthermore, we consider both the bank efficiency (EFF) and the rate of non-performing loans (NPL) as endogenous variables, i.e., correlated with the non-observed error term $\epsilon_{i,t}$.

To eliminate the unobserved individual fixed effects, the model takes first differences of equation (1):

$$y_{i,t} - y_{i,t-1} = \alpha(y_{i,t-1} - y_{i,t-2}) + \beta(x_{i,t-1} - x_{i,t-2}) + (\epsilon_{i,t} - \epsilon_{i,t-1}). \qquad (2)$$

In this work, we employ the System-GMM [5] and following Roodman [2006] we report all our specification choices. To begin with, the Arellano-Bond and Blundell-Bond estimators have one- and two-step variants. The two-step is asymptotically more efficient; however the reported two-step standard errors tend to be severely downward biased [Arellano and Bond, 1991, Blundell and Bond, 1998]. Therefore, we use a finite-sample correction to the two-step covariance matrix derived by Windmeijer [2005], which makes two-step robust estimations more efficient than one-step robust, especially for System-GMM.

An additional problem is that when instruments are many, they tend to overfit the instrumented variables and bias the results toward those of OLS/GLS. Hence, we reduce the width of the instrument matrix and employ fewer instruments than the number of groups (99 banks). The constant is not included in our dynamic panel specification to keep consistency with the VAR model replication [Love and Zicchino, 2006], which does not consider the constant term in its estimation since the fixed effects and the macro shocks are eliminated.

[5] The System-GMM [Arellano and Bover, 1995, Blundell and Bond, 1998] is an extension of the Difference-GMM [Arellano and Bond, 1991] estimator. Mainly, the former makes an additional assumption that first differences of instrumenting variables are uncorrelated with the fixed effects. In contrast with the Arellano and Bond [1991] estimator - which uses only lagged levels of the endogenous variables as instruments for equation (2), the System-GMM estimator consists of two equations - the original equation, or the level equation (1), which employs lagged differences as instruments; and the transformed or in-difference equation (2), which, as stated above, adopts lagged levels of the variables as instruments. Thus, the Arellano-Bond is extended, improving its efficiency [Roodman, 2006].

As a method based on instrumental variables, the GMM relies on the validity of the moment conditions. For this purpose, we test for over-identifying restrictions using the Hansen statistic [6]. The null hypothesis is that the instruments are valid, that is, not correlated with the error term and that the excluded instruments were correctly excluded from the estimated equation. Moreover, it is important to notice that the test has low power if the model includes a very large set of excluded instruments. Therefore, we also investigate the Difference-in-Hansen tests of exogeneity of instrument subsets [7].

In addition to the test of joint validity of instruments, we test for autocorrelation in the idiosyncratic disturbance term ϵ_{it} employing the Arellano and Bond autocorrelation test under the null of zero serial correlation in order to verify if there are lags which are invalid instruments. Since negative first-order serial correlation is expected in differences due to the relation between $\Delta\epsilon_{it}$ and $\Delta\epsilon_{i,t-1}$ via the term $\epsilon_{i,t-1}$, we actually look for r-order serial correlation with $r = 2, \dots, T$. While the test is applied to the residuals in differences, first-order serial correlation in levels would be indicated by second-order correlation in differences (AR(2) test). In this case, the $y_{i,t-2}$ variable, for example, would be endogenous to the $\epsilon_{i,t-1}$ term in the error component in differences $(\epsilon_{i,t} - \epsilon_{i,t-1})$, what would make it an invalid instrument. Thus, it should be required to instrument with deeper lags. We therefore employ the Arellano-Bond test to determine the optimal lag order to be later used in the panel VAR model.

2.3 VAR Model

To appraise the relationship between bank efficiency and non-performing loans, we also adopt the traditional VAR approach, which considers the system variables as endogenous, combined with the panel data analysis [Love and Zicchino, 2006]. Herewith, it is permitted individual non-observed heterogeneity as follows:

$$z_{i,t} = \Gamma_0 + \Gamma_1 z_{i,t-1} + f_i + d_{b,t} + e_t, \tag{3}$$

where $z_{i,t}$ is a vector of two variables {NPL,EFF}. Γ_0 and Γ_1 are parameters to be estimated and f_i are fixed effects that allow for individual heterogeneity.

[6] The Sargan statistic is a special case of Hansen's J under the assumption of homoscedasticity.

[7] The test is defined as the difference between the Hansen statistic with a lower number of instruments, excluding the subset of possible invalid instruments, and the equation with all instruments, including those of suspicious validity. The null hypothesis is that both the reduced number of instruments and the additional suspicious instruments are valid.

$d_{b,t}$ is a dummy that captures internal macroeconomic shocks which affect all banks. This dummy was eliminated when we subtracted the respective average of each variable calculated for each bank type-year.

The model is estimated through GMM and calculated after the fixed effects have been removed by the use of the forward mean-differencing, or so called 'Helmert' transformation [8]. Once the VAR variables are obtained, impulse response functions are engendered with their respective confidence intervals at a significance level of 5% being estimated by applying Monte Carlo simulation. The analysis of these functions makes it possible to draw conclusions about the response of each efficiency to the rate of non-performing loans and vice-versa, since they show the response of a specific variable to a one standard error shock in another variable in the system, keeping all other disturbances equal to zero.

However, to isolate shocks it is required to decompose the residuals so that they are orthogonal. A common way to deal with this problem is to choose a specific ordering with earlier variables being more exogenous and later ones more endogenous. To fit this purpose, we assume bank efficiency to be relatively more exogenous than non-performing loans. We believe this is a plausible assumption given that our data favor the hypotheses in line with the idea of bank efficiency preceding variations in the rate of problem loans. Therefore, non-performing loans is considered an endogenous risk variable.

To complete the interpretation of our findings, we also expose the matrix with variance decompositions, which describe the percent of one variable explained by innovations accumulated over time in another variable.

We expect the bank efficiency response to a shock in NPL to be negative, once an increase in the rate of problem loans should generate low efficiency levels as the bank could not be able to retrace completely to the borrowers all the costs proceeding from the increase in default ('bad luck' hypothesis).

On the other hand, the rate of non-performing loans can either react positively or negatively to a variation in bank efficiency. In the first case, there are two different views: a reduction in bank efficiency caused by higher expenses with loan monitoring should, similarly, result in less problem loans ('risk-averse management' hypothesis); or, by contrast, an increase in bank efficiency triggered by less expenses with loan monitoring may boost default in the future ('skimping' hypothesis). In the second case, however, a less efficient management can magnify the rate of default since inefficiency might be verified not only in high operational expenses, but also in failure to monitor and appropriately control credit portfolios ('bad management' hypothesis).

[8] See Love and Zicchino [2006] for a more detailed discussion.

3 Data Sampling

We use a semi-annual data set from COSIF (National Plan of Banking Accounts) [9], with respect to the period from june 2000 to june 2007. We eliminate from our sample all banks which included less than 3 observations. In particular, this procedure led to a total of 99 banks, resulting in a panel of 1316 observations. However, there are 1217 observations when we include 1 semi-annual lag in the model.

Table 1 shows the descriptive statistics for the variables employed in our models. Overall, there seems to be a great variation across banks as verified by the minimum and maximum of all variables. This fact illustrates that our banks are considerably heterogeneous. NPL ranges from 0% to approximately 46%, whereas the values for the bank efficiencies are located between the following intervals: technical efficiency ranges from 0.019 to 1, allocative efficiency from 0.033 to 2.652 and economic efficiency from 0.002 to 1. The mean of our bank efficiencies is in line with previous findings [Staub et al., 2010], showing that Brazilian banks operate far from the efficient frontier if compared to other countries.

For the calculus of the DEA efficiency measures, we must define inputs and outputs. In this sense, we follow the intermediation approach [10]. Here, banks can be seen as intermediators between investors and savers. In other words, it collects deposits to convert them into loans by the use of capital (operational expenses net of personnel expenses) [11] and labor (personnel expenses) as inputs. Moreover, borrowed funds and interest expenses generated by the institutions are also considered as inputs.

As noted by Berger and Humphrey [1991], deposits can be included as outputs once they are assumed to be proportional to the service outputs offered to depositors. Besides, loans and investments are important outputs in the Brazilian case as they constitute about two thirds of banking assets.

Under the DEA methodology, one compares the product creation of each individual bank with similar production units. Higher interest expenses imply greater relative usage of borrowed funds. Thus, an efficient bank is capable of using less inputs, such as interest, operational and personnel expenses, and produce more output, such as deposits, loans and investments.

[9] Account report that all financial institutions must state to the Central Bank of Brazil on monthly basis.
[10] The great majority of the DEA studies adopt an intermediation approach [Sathye, 2003].
[11] Proxy for capital expenses.

4 Empirical Results

When adjusting the data for the Arellano-Bond dynamic panel (Table 2), the Hansen test suggests that the models are correctly specified, considering that there are no evidences of correlation between instruments and errors. The Difference-in-Hansen tests do not reject the null of exogeneity of a subset of instruments, reinforcing the absence of correlation between errors and instruments.

Regarding the tests for autocorrelation in the residuals in differences, the AR(1) test rejects the null hypothesis of no existence of first-order serial autocorrelation to the bank efficiencies and NPL equations, as expected, indicating a negative first-order serial autocorrelation. Contrarily, the AR(2) test does not detect evidences of second-order serial correlation in all equations, implying that all instruments are valid. We opt for a specification with a lower number of instruments following Roodman [2006]. We test several specifications and choose the most parsimonious one.

The estimated Arellano-Bond coefficients show that the positive NPL response to a variation on technical efficiency is expressive and significant at a 1% level, suggesting that the causality would run from bank efficiency to non-performing loans. Furthermore, as the coefficients of the allocative and economic efficiencies on the NPL equation are not significant, the most plausible hypotheses seem to be the 'skimping' and the 'risk-averse management' from the positive technical efficiency coefficient. Accordingly, when technical, allocative and economic efficiencies are the dependent variables, the supposition of bank efficiency preceding changes in credit default and not otherwise is confirmed, for there is a close to zero and statistically not significant effect of problem loans on all three types of bank efficiency.

Looking at the VAR results, Table 3 indicates that the impact of NPL on technical efficiency is close to zero and not significant. On the other hand, the effect of technical efficiency on problem loans is found to be positive and significant at a 5% level, once more indicating that the causality would run from bank efficiency to non-performing loans. Turning to the allocative efficiency case, it performs a poor and not significant response to a variation in NPL, whereas non-performing loans react negatively to changes in this bank efficiency, with a significance level of 10%. Consequently, the panel VAR statistical inference is considerably similar to the Arellano-Bond dynamic panel outcomes, supporting the idea of bank efficiency explaining variations in the rate of non-performing loans in credit portfolios. Finally, looking at the economic efficiency, it appears to have a negative impact on problem loans and to be despicably affected by the credit risk variable. Nonetheless, none of the coefficients are significant.

To a full comprehension of the findings, however, one must analyze the impulse response functions generated along with the variance decomposition matrix. Figures 1 to 3 display the graphics of the orthogonalized impulse response functions for the VAR model. It is important to note that the impulse response functions reflect the same relationships estimated by the VAR parameters.

To begin with, the first row of the technical efficiency outcomes in Figure 1 confirms our previous results considering that there seems to be a close to zero effect of NPL on technical efficiency both in the estimated coefficients and in the impulse response functions. Moreover, the confidence interval is wide in the pick response, which is low in magnitude. The evidence for a reverse causation appears to be much stronger as the response of NPL to a one standard deviation shock in technical efficiency is negative and large in magnitude with a low confidence interval in the short run, whereas it becomes positive before the first semester and converges to the equilibrium in the long run. Hence, we might expect bank efficiency to be the origin of higher rates of non-performing loans.

The allocative efficiency outcomes are presented in Figure 2. It can be seen from the first row that a shock to NPL has little impact on bank efficiency. The pick response of the allocative efficiency to innovations of NPL occurs in the first semester with a large confidence interval, while it converges to zero thereafter. On the other hand, the reaction of the allocative efficiency to a one-standard deviation shock in non-performing loans is initially positive and large in magnitude, with a smaller confidence interval, becoming negative and converting to the equilibrium afterwards. Therefore, since the reaction of NPL to a variation in the allocative efficiency is much more expressive, the idea of bank efficiency preceding problem loans is reinforced.

In addition, turning to the economic efficiency in Figure 3, we find little evidence of a causality in either direction as there are small responses and large confidence intervals for shocks in the studied variables [12] .

We view our results as supportive of an idea where it is of great relevance to analyze bank efficiency as the macroprudential indicator (NPL) appears to be mainly originated endogenously. By the technical efficiency outcomes, higher default rates seem to be originated in internal problems, consistent with the 'bad management' hypothesis. In other words, supbar managers may not only incur high expenses, but poorly monitor loans, as well, ultimately leading to a greater rate of problem loans. The positive relationship afterwards indicates a risk-averse management, i.e., when verifying an increase in the credit risk

[12] It is important to note that we reject the 'bad luck' hypothesis, according to which non-performing loans would be caused by the exogenous component 'bad luck', given that the effect of NPL on bank efficiency is considerably close to zero and not significant on all the estimated models.

variable, managers tend to, consequently, incur additional costs so as to monitor loan quality and control the rise in problem loans. Hereupon, the response of NPL to a shock in bank efficiency converges to the equilibrium. Thus, managers might not only spend excessively, but also inefficiently monitor loans, therefore, generating a high default rate, according to the 'bad management' hypothesis [Berger and DeYoung, 1997]. The uncertainty brought up by the increase in non-performing loans can signal the beginning of a period of financial instability, leading (risk-averse) managers to allocate more resources to loan monitoring, worsening bank efficiency, but restraining, by contrast, the rate of problem loans, along with the 'risk-averse management' hypothesis.

Although less significant [13], the allocative efficiency presents an alternative interpretation. When maximizing profits, firms may reduce their fund destined to loan monitoring and control, in order to create the appearance of efficiency in the short run, since less resources can initially support a constant amount of loans. Nonetheless, when the rate of default goes high in the long-run, bank inefficiency emerges, showing that the covered inefficiency is the primary cause of non-performing loans. Hence, as an effort to overcome its inefficiency, banks might try to increase bank efficiency, but bear the consequences of rising problem loans later on. By not being able to maintain this strategy, it becomes most likely to verify that the subpar management is the primary explanation for an increase in default, in line with our main conclusion [14].

The variance decomposition will enrich our discussion. It exhibits the percent of variation in the row variable explained by the column variable. The total effect accumulated over time is in line with our prior evidences, favoring the hypothesis of bank efficiency explaining NPL's innovation for the technical and allocative cases [15]. Mainly, bank efficiency seems to explain more of the variations in non-performing loans, than otherwise, showing that the 'bad management', 'risk-averse' and 'skimping' hypotheses are the most likely assumptions in this paper.

[13] From the results, we expect technical efficiency to be the main cause of inefficiency triggering high default rates in Brazil.
[14] Although we have evidence for a positive association between bank efficiency and NPL ('skimping' and risk-averse management' hypotheses), it seems, however, that 'bad management' is the primary and initial instigator of problem loans.
[15] We view the percent variation of the economic efficiency explained by NPL and vice-versa as considerably similar.

5 Final Considerations

We use the Granger-causality tests to identify the intertemporal relationship between non-performing loans and bank efficiency for Brazilian banks over the years 2000-2007. With all the caveat concerning our investigation [16] , the Arellano-Bond dynamic model and the panel-VAR yield, in general, [17] a statistical significant relationship ranging from bank efficiency to non-performing loans. I appears, overall, that inefficiency is the primary trigger of soaring problem loans. Hence, it is fundamental to study bank efficiency since it precedes a macroprudential indicator: non-performing loans. In other words, bank efficiency measures are important for the banking system regulation for providing informational content. Thus, we suggest that bank efficiency measures, in particular DEA efficiency measures, should be used as an early warning indicator of financial stability.

Our empirical observation embodies other important policy implications as a matter of reducing the probability of bank failures. At first, the principals for assuring a healthy banking system, such as, supervisors, regulators, owners and managers, should bare in mind that the main risks faced by banking institutions are most likely originated in internal errors, principally, in subpar management [Berger and DeYoung, 1997]. Thereby, the improvement of bank efficiency must be the focus of regulatory polices in order to restrain credit risk. Rossi et al. [2009] recall the relevance of the Basel II regulatory framework, which emphasizes a supervisory review process as an important step for a more risk-sensitive system. At this rate, a diminishing probability of bank failures would be conditioned to better managerial practices, what demands incentives to improve human capital in order to ameliorate performance.

We conclude that there are strong evidences that a low cost efficiency may result in greater vulnerability, given that such bad performance should be related to an increase in the system credit risk. This would also demand special effort of banking supervisors to reduce the exposure to risk practices [Koutsomanoli-Filippaki and Mamatzakis, 2009]. Another suggested aspect to improve stability is linked to a more rigorous financial regulation as a matter of constricting risk, in particular, systemic risk. The aforementioned regulation is in vigor in Brazil and explains, partially, why the recent financial crisis had little effect on our domestic banking system. Lastly, one should also emphasize the importance of transparency, like the disclosure of off balance sheet items and of risky management activities in order to improve banking stability and reduce the incidence of crises [Nier, 2005, Breuer, 2006].

[16] We acknowledge that the causality in question follows the Granger-causality technique, i.e., is in line with inter-temporal precedence.
[17] The results are robust for both models.

References

M. Arellano and S. Bond. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58:277–297, 1991.

M. Arellano and O. Bover. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1):29–51, 1995.

R.D. Banker and R. Natarajan. Statistical tests based on DEA efficiency scores. *In W.W. Cooper, L. Seiford and J. Zhu, eds., Handbook of Data Envelopment Analysis, Boston: Kluwer Academic Publishers*, 2004.

R.D. Banker and R. Natarajan. Evaluating contextual variables affecting productivity using Data Envelopment Analysis. *Operations Research*, 56: 48–58, 2008.

A.N. Berger and R. DeYoung. Problem loans and cost efficiency in commercial banks. 21(6):849–870, 1997.

A.N. Berger and D.B. Humphrey. The dominance of inefficiencies over scale and product mix economies in banking. *Journal of Monetary Economics*, 28(2):117–148, 1991.

R. Blundell and S. Bond. Initial conditions and moment restrictions in dynamic panel data models. *Jounal of Econometrics*, 87(1):115–143, 1998.

J.B. Breuer. Problem bank loans, conflicts of interest, and institutions. *Journal of Financial Stability*, 2:266–285, 2006.

F. Fiordelisi, D. Marques-Ibanez, and P. Molyneux. Efficiency and risk in european banking. Working Paper Series 1211, European Central Bank, 2010.

B. Gonzalez-Hermosillo. Determinants of ex-ante banking system distress: A macro-micro empirical exploration of some recent episodes. IMF Working Papers 99/33, International Monetary Fund, 1999.

A. Koutsomanoli-Filippaki and E.C. Mamatzakis. Performance and Merton-type default risk of listed banks in the EU: A panel VAR approach. *Journal of Banking & Finance*, 33(11):2050–2061, 2009.

I. Love and L. Zicchino. Financial development and dynamic investment behavior: evidence from panel VAR. *The Quarterly Review of Economics and Finance*, 46:190–210, 2006.

E.W. Nier. Bank stability and transparency. *Journal of Financial Stability*, 1:342–354, 2005.

A. Podpiera and J. Podpiera. Deteriorating cost efficiency in commercial banks signals an increasing risk of failure. Working Papers 2005/06, Czech National Bank, Research Department, 2005.

J. Podpiera and L. Weill. Bad luck or bad management? Emerging banking market experience. *Journal of Financial Stability*, 4:135–148, 2008.

D.M. Roodman. How to do xtabond2: An introduction to "Difference" and "System" GMM in Stata. Working Papers 103, Center for Global Development, 2006.

S. Rossi, M. Schwaiger, and G. Winkler. Managerial behavior and cost/profit efficiency in the banking sectors of central and eastern european countries. Working Papers 96, Oesterreichische Nationalbank (Austrian Central Bank), March 2005.

S.P.S. Rossi, M.S. Schwaiger, and G. Winkler. How loan portfolio diversification affects risk, efficiency and capitalization: A managerial behavior model for Austrian banks. *Journal of Banking & Finance*, 33(12):2218–2226, 2009.

M. Sathye. Efficiency of banks in a developing economy: The case of India. *European Journal of Operational Research*, 148(3):662–671, August 2003.

L. Simar and P.W. Wilson. Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1):31–64, 2007.

G.S. Souza and R.B. Staub. Two stage inference using Data Envelopment Analysis efficiency measurements in univariate production models. *International Transactions in Operational Research*, 14:245–258, 2007.

R.B. Staub, G. Souza, and B.M. Tabak. Evolution of bank efficiency in Brazil: A DEA approach. *European Journal of Operational Research*, 202:204–213, 2010.

D.C. Wheelock and P.W. Wilson. Why do banks disappear? the determinants of u.s. bank failures and acquisitions. Working Papers 1995-013, Federal Reserve Bank of St. Louis, 1995.

J. Williams. Determining management behaviour in european banking. *Journal of Banking & Finance*, 28(10):2427–2460, 2004.

F. Windmeijer. A finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of Econometrics*, 126(1):25–51, 2005.

Table 1. Descriptive statistics for variables in the Arellano-Bond and the VAR Model.

Variable	Mean	Std. Dev.	Min	Max
NPL	0.0201708	0.0336288	0	0.4581309
Technical	0.6558859	0.2522241	0.0191443	1
Allocative	0.6935725	0.2393757	0.0335856	2.652515
Economic	0.4727922	0.2606144	0.0019998	1

Statistics are for N = 1316.

Table 2. Arellano-Bond Dynamic Panel to NPL and technical, allocative and economic
efficiencies.

VARIABLES	(1) Technical	(2) NPL	(3) Allocative	(4) NPL	(5) Economic	(6) NPL
Technical_{t-1}	0.409***	1.150***				
	(0.0690)	(0.267)				
Allocative_{t-1}			0.505***	-0.591		
			(0.0941)	(0.555)		
Economic_{t-1}					0.693***	-0.374
					(0.0767)	(0.513)
NPL_{t-1}	0.00246	0.384***	0.00338	0.631***	-0.00555	0.713***
	(0.00887)	(0.0837)	(0.00670)	(0.137)	(0.00847)	(0.153)
Number of observations	1,217	1,217	1,217	1,217	1,217	1,217
Number of banks	99	99	99	99	99	99
Number of instruments	10	28	28	48	10	26
F Test	20.46***	29.09***	14.42***	10.81***	43.15***	13.81***
AR(1)	-5.074***	-3.794***	-3.681***	-3.579***	-4.447***	-3.715***
AR(2)	0.145	1.098	1.468	1.260	1.403	1.295
Hansen Test	8.962	28.14	29.31	56.47	8.037	29.33
Hansen Test Excluding Group	8.28	27.07	25.50	26.70	6.02	28.18
Difference-in-Hansen	0.68	1.07	3.80	29.77	2.02	1.15

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3. VAR model to NPL and technical, allocative and economic efficiencies.

	Technical	NPL	Allocative	NPL	Economic	NPL
$\text{Technical}_{(t-1)}$	0.3689***	1.0088**				
	(0.069)	(0.446)				
$\text{Allocative}_{(t-1)}$			0.4674***	-1.0963*		
			(0.075)	(0.571)		
$\text{Economic}_{(t-1)}$					0.5051***	-0.0340
					(0.088)	(0.590)
$\text{NPL}_{(t-1)}$	0.0051	0.4676***	0.0055	0.4874***	0.0099	0.4691***
	(0.008)	(0.062)	(0.007)	(0.061)	(0.009)	(0.062)

Coefficients of VAR regression to non-performing loans and technical, allocative and economic efficiencies. The symbols ***,**,* represent statistical significance at 1%, 5% and 10% levels, respectively. Standard errors in parentheses.

Figure 1. Technical Efficiency - NPL

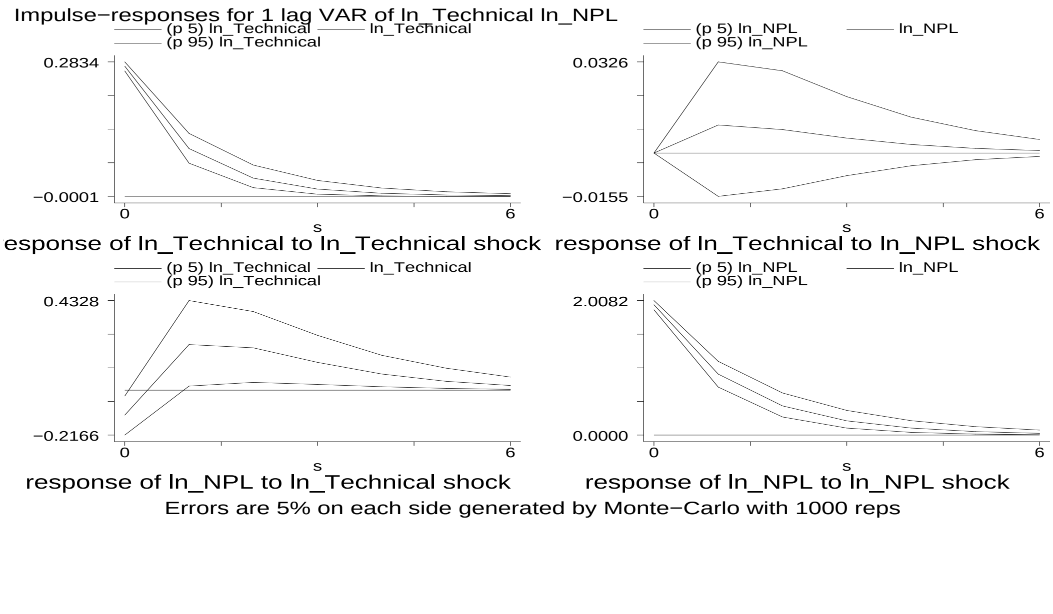

Figure 2. Allocative Efficiency - NPL

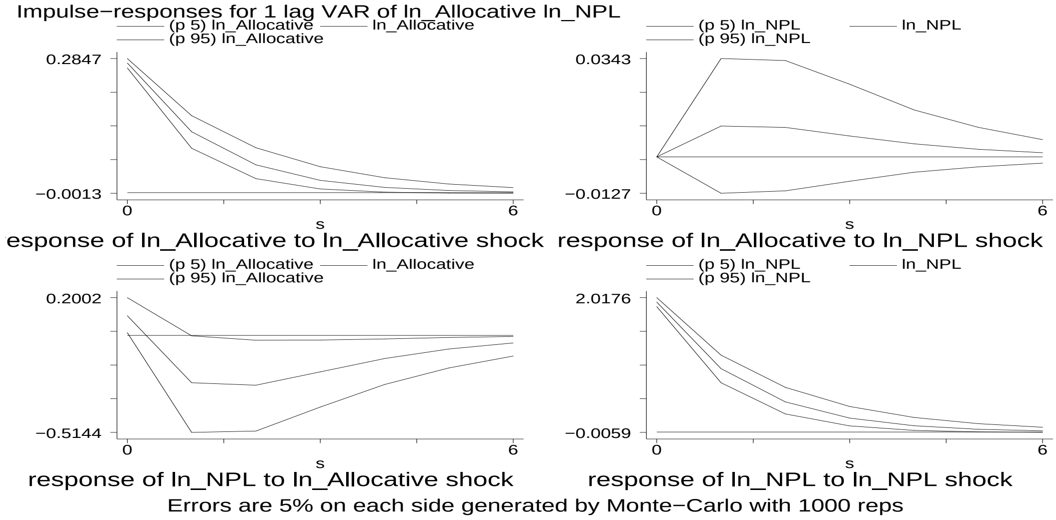

Figure 3. Economic Efficiency - NPL

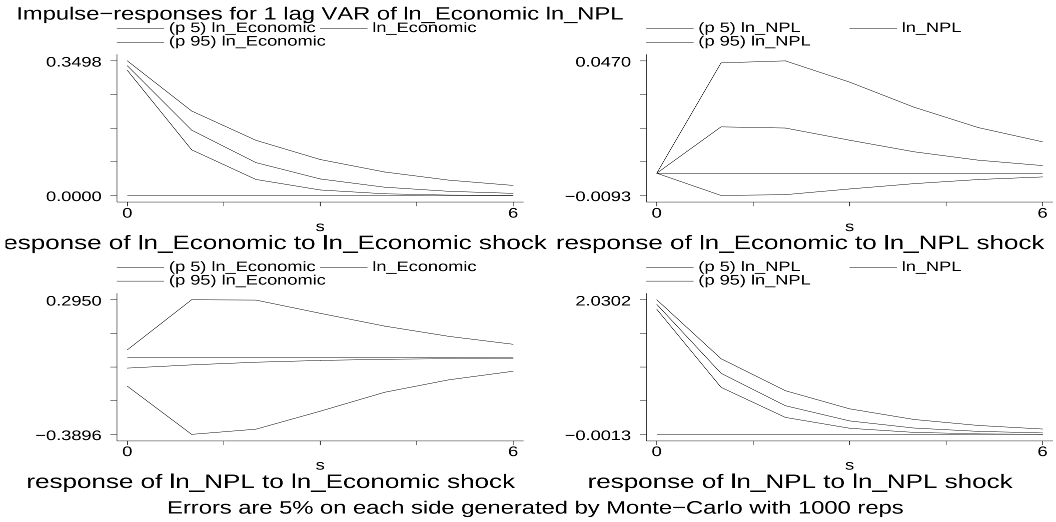

Table 4. Variance decompositions for NPL and bank efficiency.

	s	Technical	NPL		Allocative	NPL		Economic	NPL
Technical	10	.99756452	.00243548	Allocative	.99685667	.00314333	Economic	.99313895	.00686105
NPL	10	.02632684	.97367316	NPL	.03934097	.96065903	NPL	.00098696	.99901304
Technical	20	.99756446	.00243554	Allocative	.99685655	.00314345	Economic	.99313814	.00686186
NPL	20	.0263275	.9736725	NPL	.03934275	.96065725	NPL	.00098697	.99901303

Percent of variation in the row variable explained by column variable.