Macroeconomic indicators explain and predict default? A study using Brazilian data.*

Emerson Fernandes Marçal† Pedro L. Valls Pereira‡

January 14, 2016

Abstract

The objective of this study is to examine the role of macroeconomic indicators in explaining the evolution of default rates of the Brazilian economy. Time series techniques are used to evaluate the performance of macroeconomic indicators in explaining in sample and out-of-sample behavior of the default series. The work uses default indicators from January 2000 to September 2015. The study evaluates whether and how the relationship between industrial production. Short-term interest rate (Selic) and inflation are related to the default indicators. The economic cycle represented by the cyclical component of industrial production has a direct effect on the indicator of default. The effect of interest rates occurs indirectly. Changes in interest rates generate impacts on the economic cycle and these affect the default. There is evidence of a mechanism of feedback between default and macroeconomic indicators. i.e. there is Granger causality in both directions. Finally, a real time prediction exercise is done using the estimated models. The best results are obtained for the series of total default and the default for national private banks for horizons from six to twelve months. In these cases, models that contain these indicators can generate better predictions for default variable. Finally, we discuss ways to improve the forecasting exercise.

JEL codes: E37. C51. C53.

Keywords: Forecasting, Model Selection, Default.

Contents

*Financial support from FEBRABAN is acknowledge. The second author also acknowlegde the suppport from CNPq and FAPESP.

†Coordinator of CEMAP- EESP-FGV and Associate Professor at EESP-FGV and Mackenzie. E-mail: emerson.marcal@fgv.br

‡Coordinator of CEQEF-EESP-FGV and Professor at EESP-FGV. E-mail: pedro.valls@fgv.br

1 Introduction

The analysis of the default levels of savings is important because an unexpected and sharp increase in defaults can generate repercussions for the financial system as a whole, affecting the transmission mechanism of monetary policy and have negative effects for the economy. Policy makers and managers of banks' portfolios are interested in understanding the relationship that exist between the macroeconomic environment and the evolution of default rates. The development of default prediction models is extremely useful in the development of such agents' likely scenarios over a longer horizon.

The objective of this study is to evaluate the influence of macroeconomic indicators in default indicators in Brazil. Two types of analysis will be made.

The first evaluates whether and how macroeconomic indicators - interest rates. economic cycle and inflation - contribute in explaining default rates for the Brazilian economy. The second type of analysis evaluates whether there are and what is the size of the prediction gain for the default series from models that use macroeconomic information

The work is divided into seven sections including this introduction. Section two, presents a brief review of the literature. In section three, the methodology of this work is discussed in detail. In section four, describes the database. In section five, we present the result of the estimation and forecast exercise. Section six limitations and possible extensions of work are proposed. Finally. in section seven presents the concluding remarks.

2 Motivation and Literature Review

There is an extensive literature in economics linking financial environment and macro-economic environment suggesting a strong and feedback relationship between them. Bernanke et al (1999) give a classic work on this line. In this paper, the authors develop a model in which there is a strong interaction between economic cycle and credit cycle.

Empirical studies confirm the relationship between economic cycle and default. Beck et al (2013) use panel data to a group of countries show that variables other than just the economic cycle such as exchange rate and stock prices can influence the level of defaults on countries. Claessens et al (2012) analyses the relationship between the business cycle and financial cycle for a large sample of countries. The authors' results suggest that recessions are associated with stronger default crisis and the effect of the asymmetrical economic cycle.

In Brazil there is work exploiting microdata. This work is difficult to replicate, as it requires no public access database. Correa et al (2014) use micro data available from Brazil's Central Bank information to assess the relationship between credit and economic cycle. The authors conclude that a recession increases the likelihood of rising defaults and losses in loan portfolios. The work focuses on consumer credit analysis and acquisition of vehicles covering the period 2008 to 2013. In this work, the authors explore both the temporal and cross-section dimensions because they have access to data at the individual level. Tabak et al (2008) study the relationship between bank concentration and defaults controlled by macroeconomic factors. The results suggest that greater bank concentration can lead to lower levels of default.

The work Linardi & Ferreira (2008) explores a public database as well as the present work. The authors estimate an autoregressive model to evaluate the relationship between default rates and macroeconomic factors. The authors' results suggest the existence of relationship between the variables. The authors also propose a methodology to evaluate scenarios for default, but the predictive power of the models is not evaluated.

3 Methodology

This section discusses in more detail the methodology to be used. Two types of analysis are done. The first type of analysis focuses on the estimation of the model in sample. The goal in this case is to assess the extent to which macroeconomic indicators help explain the behavior of default rates and how these relationships is over time. The technique used to assess this relationship is a VAR model. The second type of analysis evaluates if the macroeconomic indicators are able to anticipate and in what forecast horizons can predict default. It is an out-of-sample prediction exercise. The evaluation of the predictive quality is made used recent selection techniques of forecasting models, such as Hansen et al (2011).

3.1 Analysis in sample

In the in sample analysis, we use a linear multivariate model such as Vector Autoregressive Model (VAR). We evaluate to what extent this model is a good representation of the data generating process. This is done using recent advances in model selection area given by Autometrics algorithm developed by Hendry & Doornik (2014). The algorithm uses model selection techniques that follows LSE Econometrics developed by David Hendry ([Hendry (1995)).

3.1.1 Vector Auto Regressive - VAR

We chose to model the data using a Vector Autoregressive (VAR). This model can be viewed as a reduced form of a macro-economic model. A VAR estimates the relationship between the variables and the lags of them. There is no contemporary relationship between the variables:

$$Y_t = \mu + \Gamma_1 Y_{t-1} + \cdots + \Gamma_p Y_{t-p} + \varepsilon_t \tag{1}$$

where ε_t is a vector of disturbances normally distributed and not serially correlated, Ω the matrix of variance and covariance of the errors which is the same for all periods of time and $\theta = [\mu, \Gamma_1, \cdots, \Gamma_p, \Omega]$ contains the parameters of the model. The vector Y_t contains indicators of default and macroeconomic variables.

3.1.2 Variables relevant tests

Gourieroux & Jasiak (2001) argue that restrictions may be imposed on the VAR so that the process can be seen as uncorrelated for any desired time period and is a strong indicator of the absence of any causal relationship between the variables.

If all matrices Γ_i with i ranging from 1 to p (VAR order) are diagonal and the matrix of variance and covariance Ω is also diagonal, then the temporal correlation of the process is zero for any point in time.

3.1.3 Granger Causality

The diagonal restriction of the matrices Γ_i is equivalent to the absence of Granger causality in either direction. The Granger causality concept [Granger (1969)] can be defined as follows :

Definition 1 *Granger Causality (Hendry (1995) page 176) Let* y *and* x *be two variables whose joint distribution is given by* $D_Z(y_t, x_t|Z_{t-1})$ *where* $Z = [y, x]$ *and* Z_{t-1} *represents all past history up to time* $t-1$*. Considering the factorization of the joint distribution given by* $D_Z(y_t, x_t|Z_{t-1}) = D_Z(y_t|x_t, Z_{t-1})D_Z(x_t|Z_{t-1})$*. If* $D_Z(x_t|Z_{t-1}) = D_Z(x_t|X_{t-1})$ *then the variable* y_{t-1} *does not cause in Granger sense* x_t*.*

It is noteworthy that the Granger causality concept is associated with temporal precedence and not as a relationship of cause and effect. Especially in models in which agents are forward looking to make their decisions, the Granger causality may occur in the opposite direction to the one implied by cause and effect relationship.

3.2 Analysis out of the sample

A good fit in the sample does not necessarily guarantee that the same model will be able to perform well in forecasting future values. The field of forecasting in economics is one of the most challenging. The economic variables such as part of a complex system is difficult to anticipate. Changes of all kinds such as crises. structural changes. measurement errors. instability of the data generating process can make a well formulated model has no predictive power. There is also the fact that many forecast variables in Economics are from the financial sector. which makes predictability extremely profitable. In an efficient market in the information sense the opportunities of economic profits should be low or non-existent based on public information. Therefore predictability should be low or non-existent in this context.

3.2.1 Evaluation predictive power of the model

In making an economic forecast the analyst must build a model and using this model to forecast. The accuracy of forecast can be evaluated using several criteria. A first criterion relates to bias. If the forecasts are submitted to systematic errors in either direction, they should not be used in principle. Another criterion relates to deviations from the observed value. Two sets of prediction have the same average value of the variable is chosen the one with less dispersion.

In order to evaluate the forecast errors some metrics have been established in the literature. Two very popular examples are the mean squared error (MSE) and mean absolute error (MAE) defined below:

$$MSE = \sum_{\tau-1}^{N} \left(\widetilde{y}_{t+\tau+h}^{t+\tau-1} - y_{t+\tau+h} \right)^2 \tag{2}$$

$$MAE = \sum_{\tau-1}^{N} |\widetilde{y}_{t+\tau+h}^{t+\tau-1} - y_{t+\tau+h}| \tag{3}$$

where $\widetilde{y}_{t+\tau+h}^{t+\tau-1}$ represents the forecast for the variable y at time $t+\tau+h$ conditional on the information available at time $t+\tau-1$ and h represents the number of step ahead in the forecasting.

One way of selecting among several forecasting models is to calculate such measures and to rank the model from them, opting for the model with smaller MSE and or MAE.

3.2.2 Exercise of pseudo real time forecast

An analyst who wants to build a prediction model can simulate a real-time exercise to evaluate a set of models have good predictive power. The analyst must choose an evaluation window and an estimation window. In the estimation window models are estimated, the variables chosen and forecasts generated for the data in the evaluation window. The choice of models may not contain any information that was not available to the analyst at the time he would have generated the forecast for the exercise to be valid. Models built from the estimation window data can then be tested for predictive power in the evaluation window. If there is a good performance it is possible that a real exercise will have good performance as well. The selection of predictive models is a computational intensive process, to the extent that the forecast has to be done all the time in order to try to simulate the forecast would have been done in real time.

3.2.3 Choice of benchmark

The choice of benchmark is something important. One possible criterion concerns the construction of naive models that would easily overcome by more sophisticated models. One example is the random walk in which it is assumed that the best forecast for h periods ahead is exactly the value of the last observation. The random walk model is widely used as a benchmark in Finance and has been often a "tough" opponent to beat. When forecasting exchange rate for short horizons is very difficult to beat the random walk as shown in the seminal work of Meese & Rogoff (1983).

Another benchmark used for data with some persistence is an autoregressive model of order 1. Castle et al (2014) show that it is very difficult to beat him in inflation forecasting exercises.

3.3 The model confidence set - MCS

The model confidence set (MCS) is a forecasting model selection technique developed by Hansen et al (2011). It consists of an algorithm that ranks a set of predictions from a set of models. M^* contains the set of best type selected from an initial set of models. The M^0 set is the set that contains the best models defined from a predictive quality criteria.

Definition 2 *The set that contains the best models is defined by* $M^* = \{i \in M^0 : E(d_\tau^{i,j} \leq 0 \text{ for all } j \in M^0\}$

Let $M^\dagger$ be the complementary set, i.e. $M^\dagger = \{i \in M^0 : E(d_\tau^{i,j} > 0$ for all $j \in M^0\}$ in which $g(e_\tau^i)$ is some loss function and

$$d_\tau^{i,j} = g(e_\tau^i) - g(e_\tau^j) \tag{4}$$

$$e_\tau^i = \widetilde{y}_{t+\tau+h}^{t+\tau-i} - y_{t+\tau+h} \tag{5}$$

MCS selects models using an equivalency test, δ_M, and an elimination rule, ρ_M. The equivalence rule is applied to the set $M = M^0$. If the equivalence rule is rejected at a selected confidence level, then there is, with high probability, a group of bad models in terms of predictive power that must be eliminated from the set of good models. In this case an elimination rule, ρ_M is used to remove models with low predict power form the set of good models. Having done this, we use the equivalence rule again. The procedure is repeated until the equivalency predictive hypothesis in the analyzed set, δ_M is not rejected. The set of models of the last step $(\widetilde{M}_F)$ is selected and must contains the best models to a certain level of significance.

The null hypothesis of equivalence test is given by:

$$H_M^0 : E(d_\tau^{i,j}) = 0 \text{ for all } i, j \in M \tag{6}$$

where $M \subset M^0$.

The alternative hypothesis is given by:

$$H_M^0 : E(d_\tau^{i,j}) \neq 0 \text{ for all } i, j \in M \tag{7}$$

An important point to emphasize is that there may be better models out of the initial set of models "candidates" M^0. The goal is to rank a particular set of models to obtain M^*.

The null hypothesis can be tested from the following statistics[1]

$$T_D = \sum_{i \in M} t_i^2 \tag{8}$$

where $t_i = \frac{\overline{d}_i}{\sqrt{\widehat{VAR}(\overline{d}_i)}}$ and $\overline{d}_i = \frac{1}{M} \sum_{j \in M} \overline{d}_{ij}$

[1] There are other possible choices

The test statistic given by (8) has a non-standard statistical distribution but can be simulated using bootstrap techniques. The delete rule is given by:

$$\rho_M = \arg\max_i (t_i) \tag{9}$$

3.3.1 The algorithm

The MCS algorithm has the following steps:

(i) Initializes the procedure by setting the initial set of model to be analyzed $M = M^0$;

(ii) Tests H_M^0 using δ_M and a significance level α;

(iii) If H_M^0 is not rejected the procedure ends and the final set is $\widehat{M}^*_{1-\alpha} = M$, otherwise we use the elimination rule ρ_M to delete an object from M set and back to step (i).

The authors suggest that the MCS have the following statistical properties:

(i) $\lim_{n\to\infty} (M^* \subset \widehat{M}^*_{1-\alpha}) > 1 - \alpha$ and;

(ii) $\lim_{n\to\infty} (i^\dagger \in \widehat{M}^*_{1-\alpha}) = 0$ for all $i^\dagger \in M^\dagger$

3.3.2 Ranking the models: p-values for MCS

The elimination rule, ρ_M defines a sequence of random sets, $M_0 = M_1 \supset M_2 \supset \cdots \supset M_{m_0}$, where $M_i = \{\rho_i, \cdots, \rho_{m_0}\}$ and m_0 is the number of elements in M_0, ρ_{m_0} is the first element to be eliminated, ρ_{m_1} is the second to be eliminated, and so on. In the end, only one model survives. Set up the p-value of the final model to one. It is stored in the p-values of the deleted models if they are larger than the p-value of the model previously eliminated. If the p-value of the current removal is stored. The p-values of MCS are important as make it easier for the analyst to assess a given set $\widehat{M}^*_{1-\alpha}$.

3.4 The pseudo exercise in real time

The data collected allow us to create several models to predict default. The sample is divided into two parts. The first part is used to estimate models and using them to forecast, while the second part is used to evaluate the predictive power over various horizons. In this exercise seeks to simulate a real-time forecast. Used as close as possible to the set of information available to agents at the time the prediction as made. In other words, the models are estimated and reviewed at each point in time in order to incorporate the information gain that arises from the time spans. Each model generates forecasts up to one year ahead. There are so twelve groups' forecasts because we are using monthly data.

One problem with the simulation of a real-time exercise has to do with the data sources. Many of the series are discontinued or have revised methodology or undergo review by new information unprocessed. Thus, the information set is not exactly the one available to agents, but very close. The exercise uses a slightly higher set of information available to the agents at the time that made real time prediction. This may create a small bias toward constructing better predictive models. Because of the original database can be extremely costly in terms of time and resources, a possible refinement of the exercise is left for future extension.

Recently Cusinato et al (2013) analyses the effect of data revisions by the Brazilian Bureau of Statistics (IBGE) for quarterly GDP when estimating the Brazilian potential output. The authors suggest that between the estimates made from the initial disclosure and revise data the difference was 0.7% in annualized terms. which is something quite substantial. To our knowledge, we are the first to carry on the estimation of potential output for Brazilian data using industrial productions, which is observed at a monthly basis.

3.5 Building Output gap

The output gap was estimated using the Hodrick and Prescott (1997). This filter allows the analyst to separate trend and cycle. The trend is interpreted as potential output as the cyclic term is interpreted as a deviation of full employment. There are other possibility in the literature as Baxter & King (1999) or the state space model discussed in Harvey (1990) and Durbin & Koopman (2001). An application of extracting the output gap using state space model to Brazilian data is given in Valls Pereira (1986).

4 Description of the data base

The default data was obtained in the Central Bank of Brazil website. Recently the institution has promoted a methodology to review the default series discontinuing most of the existing series and not performing the reconstruction of the new series for earlier periods. Thus, most of the indicators of default in major breakdown levels so as individuals and companies, credit arrangements, among others are available only from 2010, which limits the analysis. It would be interesting, if technically possible; the Central Bank continued to disclose the default series in the old methodology.

Given the data constraint, we chose to use the aggregate default series that has been available since January 2000 and the only breakdown available since 2000 given by private (domestic and foreign bank) and public.

The collected macroeconomic indicators were the base interest rate Selic practiced in the market, the total industrial production, the broad consumer price index (IPCA)

<TABLE 1 HERE>

Figure 3 shows the evolution of the cycle component extracted using the full sample. The first chart shows the estimated cycle using seasonal unadjusted data and it is difficult to visualize the cycle. In the second graph shows, the cycle component using seasonal adjusted data and now the cycle is clearer.

<FIGURE 1 HERE>
<FIGURE 2 HERE>
<FIGURE 3 HERE>

5 Results

In this section, we present the results of the estimated models default. Two types of exercises are presented. In the first, the analysis is done in sample and try to demonstrate the link between default and macroeconomic variables, in particular the monetary cycle and the business cycle. In the second part, we report the results of a real-time simulation exercise to assess the predictive ability of macroeconomic indicators of default.

5.1 Tests in the sample

First unit root tests are carry out to determine the order of integration of variables. This is done so that all variables entering the VAR estimation have the same order of integration. Next a VAR is estimated for various models using the default series and macroeconomic indicators and Autometrics model selection is used to simplify the models. Next Causality tests are presented.

5.1.1 Unit Root Tests

Initially we determine the order of integration for the variables as this is important for the analysis that follows. The presence or absence of non-stationary processes changes the way to handle time series analysis. A detailed discussion is made for example in Hamilton (1994).

Table 2 shows the Augment Dickey & Fuller test for all the series in level. The results suggest that the null hypothesis of a unit root can be rejected for most of the series except the Selic rate, industrial production and private foreign bank default.

<TABLE 2 HERE>

5.1.2 Analysis of estimated models

VAR model were estimated using the default variables and the macroeconomic indicators. The estimated models are described below. The starting point was a VAR with twelve lags, seasonal dummies and correction for outliers. The final models are reported below and the simplification strategy followed a general to specific approach along the lines proposed by the LSE Econometrics popularized by the work of David Hendry. The model selection uses Autometrics developed by Hendry & Doornik (2014). The full sample is used that covers the period from January 2000 to September 2015. Based on this strategy it was possible to obtain models that show good results in terms of specification and can be viewed with a good approximation data generator process.

<TABLE 3 HERE>

In all specifications assessed there is a strong relationship between economic cycle and default rates. The output gap is significant in the default equation in all specifications. The interest rate, whether in the form of $\Delta ln(Selic)$ or Deflated Selic (real interest) does not seem to cause direct influence on default. However this does not imply that it is irrelevant to explain what happens to default. There is an important dynamic between nominal and real interest rate and the business cycle. Such that factors that influence the interest end up affecting the business cycle which in turn ends up generating effect on default. The evolution of the level of default also influence the macroeconomic variables.

It is also worth noting that it impulse dummies are used to control additives outliers and innovation outliers.[2] The Autometrics selection algorithm suggests that the estimated structure has points of instability: 2002 (11), 2002 (12), 2003 (2), 2003 (10), 2003 (11), 2004 (2), 2005 (4), 2005 (5), 2005 (6), 2008 (12) and 2009 (4). The 2002 points are associated with instability caused by the elections of 2002. The point in 2008 is probably related to effects of the international crisis on Brazil. The 2013 (11) may be associated with the release of the payroll loans Law. The other points have no clear interpretation but statistical significance.

<TABLE 4 HERE>
<TABLE 5 HERE>
<TABLE 6 HERE>

5.1.3 Granger Causality

The results of Granger causality tests between the variables in the different models are presented in the following tables. For all models, we obtain evidence of Granger causality from the default variable to the macroeconomic indicators and in the opposite direction in the joint tests. This does not necessarily imply that all macroeconomic indicators causes or is caused in the sense of Granger

[2] Nielsen (2004) for a discussion.

by default variable, but by at least one indicator causes or is caused in the sense of Granger by the default variable.

In all models the variable Output Gap Granger cause defaults and is Granger caused at a 5% level of significance, i.e. there is a strong relationship between business cycle and default on all models.

The effect of interest rates in nominal terms is assessed through the Selic rate or in real terms is assessed by Selic rate deflated by the IPCA does not seem to have a direct causal effect. The evidence of a direct relationship between interest rates and default is not strong. The null hypothesis is not rejected for traditional significance levels. This does not mean that interest is an irrelevant variable in explaining default. The effect of interest has on the output gap and vice versa is pronounced and statistically robust. Thus, the channel exists, but the action occurs indirectly mediated by the business cycle.

<TABLE 7 HERE>
<TABLE 8 HERE>
<TABLE 9 HERE>
<FIGUE 4 HERE>

5.2 Evaluation of Predictive Power

A good fit in sample of the model does not necessarily becomes good model for prediction out of the sample. Unexpected changes, structural instabilities, uncertainty in the estimation of parameters and other factors could lead models whose variables have a good fit in sample, generate a low predictive performance. In this section we wish to assess whether macroeconomic variables help to improve the predictive power of the models and in what time horizons. The forecast window covers the period from January 2008 to the end of the sample.

5.2.1 Looking at the timeline

The primary source of data for the default variables is the Central Bank of Brazil. The institution informs that the disclosure of these indicators is made in the publication "Note to press" with a lag of at least four weeks (one month) with respect to the last business day of the month. The first disclosure of the industrial production index is done with a two months lags. The inflation rate measured by the IPCA is made about a week after the closing of the month. The Selic rate is known at the end of the business day. Thus, industrial production is the series that most restricts the development of the forecast. All data is at best one month ahead of the data industrial production.

For example, on the last day of November, we have available the October default data, the Selic rate and IPCA of November and industrial production in September. With this real-time analyst is able to generate forecasts for October, November, December, January next year and so on. Note that the October forecast for default is only feasible when this data is already available. Thus the prediction step forward is not helpful.

In the forecast construction is done in two steps forward - the end of December as an example - would be possible to use the inflation data, interest and default available until November in real time, but the choose in this work to use forecast values. The reason is due to a conservative approach for model evaluation. Since we are working with the revised data rather than the first disclosure of production data, we chose to discard the available information from other indicators to try to control the bias of better performance of our procedure related to the feasible real time approach. But the execution in real time, there is no reason to discard such information. As the industrial production data are lagged two periods, the forecasts made in real time with models for two steps ahead, in practice, amount to exercise Nowcast, while the forecast 3 steps ahead is equivalent to one month ahead and so on. This nomenclature will be used for here onward. A Nowcast exercise is to use a model to predict a variable in the current moment, but not yet released from available indicators that bring information about it.

The evaluation of a forecast model, some questions need to be evaluated. One way of evaluating is to choose a well-naive benchmark that any reasonable model should be able to beat. We will choose as the benchmark random walk. In this model, the best prediction for all time horizons is given by the variable at the last observable date.

5.2.2 Simulation exercise results

In this part of the article, we present the ranking of the estimated models. As a general conclusion, the predictive performance of macroeconomic indicators is best for private sector default. The predictive performance of the models in the public sector is much lower.

Total Default Private Sector:

Table below shows the results of the models classified by the MCS algorithm. We report only the finalist's models that could not be excluded from the level of 10%. The model with p-value of 1 is the best model in terms of mean square error or mean absolute error, but you cannot say that the best performance across the rest who are in final assembly happened by chance.

Nowcast the horizon and one month ahead is hard to beat the random walk model. This result is expected in that changes in macroeconomic variables affect with a lag on the default indicators. In the six months ahead analysis, models with macroeconomic indicators seem to have the best performance by placing the random walk model at a level close to being eliminated from the final set. Models with two macroeconomic indicators (1 and 2) are better model than the model with 3 indicators (model 3) on the horizon of 12 months. Models that assume stationarity of the series of default performs better than models that do not make this assumption. This result is in line with the unit root test, suggesting that determine the order of integration may be important to forecast.

The following nomenclature is as defined in Table 2. The models M1-b, M2-b and M3-b contain the same variables as models M1, M2 and M3, with the difference that the default series is not stationary and in models with Mi-b (i=1,2,3) it is model in first difference and in level in the other cases.

Default Private Domestic Banks:

The table below shows the performance of macroeconomic indicators for the components of the total default: the series of default of domestic banks. In this case the performance of models with macroeconomic indicators is far superior to the horizon of six months and one year. In none of these cases, not the random walk or AR-12 model are in the final set. In the twelve-month period the M1 model is alone. In the case of a year, only one model is finalist, eliminating all other competitors. In the short term, the performance is still lower than the random walk.

<TABLE 11 HERE>

Default Private Foreign Banks

The other private sector default component is given by the portfolio of foreign banks. In this case the Brazilian macroeconomic indicators do not have a good performance. The reason may be the stochastic characteristic of the series. The number of default shows clear signs of non-stationary, which makes the random walk a difficult opponent to beat. Models that assume that the number of defaults is not stationary are those with greater predictive power, although none have outperformed the random walk, but some of them are in the final set for the horizons of one, six and twelve months.

<TABLE 12 HERE>

Default Public Banks

Finally the number of defaults of public banks has clearly signs of structural change and its forecast from macroeconomic indicators is clearly ineffective. In all cases the random walk model was superior all model and is the one that is present in the final set of models.

<TABLE 13 HERE>

6 Limitations of Article and Possible Extensions

This work can be improved on several fronts. The first point concerns a better measurement of real and monetary business cycle. We used only the Selic rate and industrial production as a proxy of the economic environment. Still got a strong evidence of the importance of these to explain the dynamics of private default in Brazil. A wide number of indicators can be employed like, employment indicators, production, survey of economic activity, many types of interest rates and monetary indicators that could be used for better understanding of the macroeconomic environment.

Regarding the model in sample, there is the possibility to advance in the identification of economic shocks direction and seek a better understanding of the reaction dynamic of default to different types of shock. A classic example of identification strategy can be found in Blanchard and Quah (1988).

There is also the possibility of using multivariate state space models developed by Harvey (1990), and Durbin & Koopman (2001). This model can also be an attractive option in constructing forecasts.

Regarding the forecast there is a broad research agenda that can be used. A comprehensive review is made in Elliot & Timmermann (2008). There are important advances both in the area of generation forecasts and in model selection techniques.

A first refinement is in line with the search for better measurement of the economic cycle. Factorials models in the spirit proposed by Stock & Watson (2002a-b) are options to be tried.

One of the big reasons why the economic forecasts fail has to do with unanticipated changes in conditional mean of the process see Clements & Hendry (1998) and Clements & Hendry (2001). Statistical models with flexible structures can help improve the predictive power to the extent that structural changes can be addressed and incorporated into the preparation of forecasts.

In this spirit, there is a bias correction literature. In the case of a forecast introduce a systematic bias error in a certain direction; there are techniques that allows trying to correct the bias by adding corrective terms the forecast held (Issler & Lima (2009)).

Currently, there is a possibility of making a wide range of models for forecasting. In a seminal text, Granger & Ramanathan (1984) proposes that a combination of model predictions may have a predictive power better than each of the individual forecasts. Since then a vast literature discusses forecasting combination of techniques and the potential gains of the same see Timmermann (2006).

There is also the possibility of improving this work when it comes to model selection techniques. The method proposed by Hansen et al (2011) and used in this work is a clear advance in this area, but more can be done. For example, it is possible to use other metrics with loss of functions that take into account asymmetry. Both mean square error, the average absolute error metrics are symmetry loss function. Often, analysts give different values to positive and negative errors. For example, a central bank using inflation targets might be more concerned to underestimate than overestimate inflation in the near future. A commercial bank might be more concerned about the underestimation of the level of default.

Finally, a possible refinement is done in Linardi & Ferreira (2008). Is to use the estimated VAR to develop a range of possible forecasts and assign the same probability of occurrence. This makes it possible to construct scenarios for maximum and minimum values for evolution of default and not just the forecast average values. This way you can build boundaries from which there is low probability of being exceeded by the level of default. Such a strategy is important to establish values for stress testing of the portfolio and maximum loss simulations.

7 Concluding Remarks

This study aimed to assess the extent to which macroeconomic indicators help to explain the dynamics of the Brazilian default. Using private default data was possible to show that macroeconomic indicators show a robust relationship with the business cycle and the monetary cycle. The data generating process was a Vector Autoregressive (VAR) and the best model was selected using the general to specific approach proposed by the LSE. This yields models that meet the conditions to be a good approximation of the data in terms of specification tests such as normality, absence of correlation and homoscedasticity in the residuals, and no evidence of structural change.

The models suggest that there is a strong and direct relationship between economic cycle and default. The monetary cycle is also important, but their influence is given via impact that generates the economic cycle. The effect exists but is essentially indirect.

A real-time forecasting simulation exercise was also done and suggests that models that incorporate macroeconomic variables has a better predictive performance over longer horizons for the private default series. For public default, the predictive performance of the models with macroeconomic variables is very disappointing.

Finally, we discuss a number of limitations and possible extensions of this work, especially regarding the development of models with good predictive power for default. The possibility of accurate predictions construction is an important research topic in economics and recent advances have been made.

8 References

Baxter, Marianne & King, Robert G. (1999) Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics*, **81**: 575–593.

Beck, Roland; Jakubik, Petr & Piloiu, Anamaria (2013) Non-performing loans: What matters in addition to the economic cycle. European Central Bank working paper no 1515.

Bernanke, Ben S.; Gertler, Mark & Gilchrist, Simon (1999) The financial accelerator in a quantitative business cycle framework. Em *Handbook of Macroeconomics*, 1:1341–1393, Editors: J. B. Taylor & M. Woodford, Elsevier Science: Amsterdam.

Blanchard, Olivier J. & Quah, Danny (1988) The dynamic effects of aggregate demand and supply disturbances. *American Economic Review,* **79**: 655-673.

Castle, Jennifer; Hendry, David & Clements, Michael P. (2014) Robust Approaches to Forecasting. Economics Series Working Papers 697, University of

Oxford, Department of Economics.

Claessens, Stijn ; Kose, M. Ayhan & Terrones, Marco E. (2012) How do business and financial cycles interact? *Journal of International Economics*, **87**(1):178 – 190.

Clements, Michael P & Hendry, David F. (2001) *Forecasting non-stationary economic time series*. MIT Press: Boston.

Clements, Michael P & Hendry, David F. (1998) *Forecasting economic time series*. Cambridge University Press: Cambridge.

Correa, Arnildo da S.; Marins, Jaqueline T. M.; das Neves, Myrian B. E. & da Silva, Antonio C. M. (2014) Credit default and business cycles: An empirical investigation of brazilian retail loans. *Revista Brasileira de Economia*, 68 (3): 337-362.

Cusinato, Rafael T.; Minella, André & Pôrto Júnior, Sabino da S. (2013) Output gap and gdp in brazil: a real-time data analysis. *Empirical Economics*, **44** (3): 1113-1127.

Durbin, James & Koopman, Siem J. (2001) *Time Series Analysis by State Space Methods*. Oxford University Press: Oxford.

Elliott, Graham & Timmermann, Allan (2008) Economic forecasting. *Journal of Economic Literature*, 46 (1): 3–56.

Gourieroux, Christian & Jasiak, Joann (2001) *Financial Econometrics*. Princeton University Press: Princeton.

Granger, Clive W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3): 424–438.

Granger, Clive, W. & Ramanathan, Ramu (1984) Improved methods of combining forecasts. *Journal of Forecasting*, 3 (2): 197-204.

Hamilton, James D. (1994) *Time Series Analysis*. Princeton University Press: Princeton.

Hansen, Peter R.; Lunde, Asger & Nason, James M. (2011) The model confidence set. *Econometrica*, **79** (2): 453–497.

Harvey, Andrew C. (1990) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press: Cambridge.

Hendry, David F. & Doornik, Jurgen A. (2014) *Empirical Model Discovery and Theory Evaluation: Automatic Selection Methods in Econometrics*. MIT Press: Boston.

Hendry, David F. (1995) *Dynamic Econometrics*. Oxford University Press: Oxford.

Hodrick, Robert J. & Prescott, Edward C. (1997) Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking,* **29** (1): 1–16.

Issler, João V. & Lima, Luiz R. (2009) A panel data approach to economic forecasting: The bias-corrected average forecast. *Journal of Econometrics*, **152** (2): 153–164.

Linardi, Fernando de M. & Ferreira, Mauro. S. (2008) Avaliação dos determinantes macroeconômicos da inadimplência bancária no brasil. XXXVI Encontro Nacional de Economia, Salvador, Bahia.

Meese, Richard A. & Rogoff, Kenneth (1983) Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, **14** (1): 3–24.

Nielsen, Heino B. (2004) Cointegration analysis in the presence of outliers. *The Econometrics Journal*, **7** (1): 249–271.

Stock. James H. & Matson, Mark W. (2002-a) Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, **97** (460): 1167-1179.

Stock. James H. & Matson, Mark W. (2002-b) Macroeconomic forecasting using diffusion indexex. *Journal of Business & Economic Statistics*, **20** (2): 147-162

Tabak, Benjamin, M.; Guerra, Solange, M.; Lima, Eduardo, J. A. & Chang, Eui J. (2008) The stability–concentration relationship in the brazilian banking system. *Journal of International Financial Markets, Institutions and Money*, **18** (4):388–397.

Timmermann, Alan (2006) Forecast combinations, Em G. Elliot, C. Granger & A. Timmermann (ed.) *Handbook of Economic Forecasting*, vol 1, n0 1: 135-196. Elsevier|: Amsterdam.

Valls Pereira, Pedro L. (1986) Estimação do hiato do produto via componentes não observados. *Brazilian Review of Econometrics*, **6** (2): 81–95.

9 Tables and Figures

Table 1: Description of Variables

Series	Source	Sample	Frequency	Transformation
Total Private Default	Central Bank of Brazil	2000(1) to 2015(9)	Monthly	None
Private Domestic Bank Default	Central Bank of Brazil	2000(1) to 2015(9)	Monthly	None
Private Foreign Bank Default	Central Bank of Brazil	2000(1) to 2015(9)	Monthly	None
Public Bank Default	Central Bank of Brazil	2000(1) to 2015(9)	Monthly	None
Industrial Production	IBGE	1975(1) to 2015(7)	Monthly	Concatenation
Selic	Central Bank of Brazil	1980(1) to 2015(7)	Monthly	None
Consumer Price Index (IPCA)	IBGE	1980(1) to 2015(7)	Monthly	None

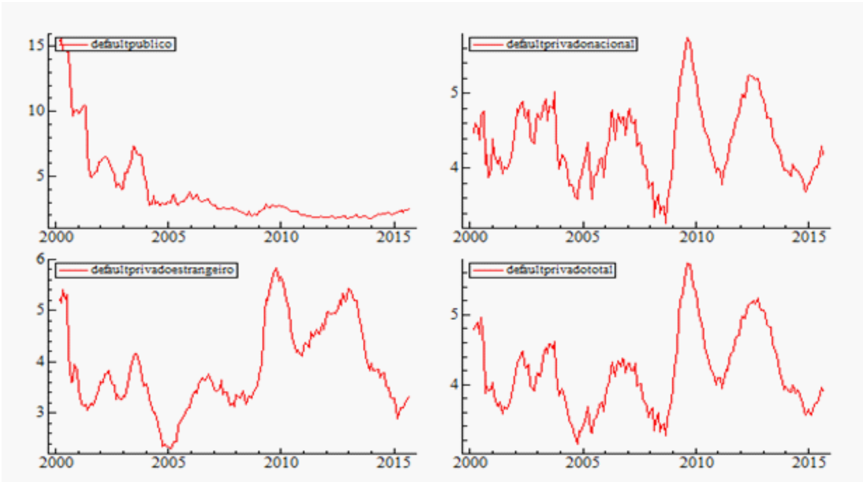
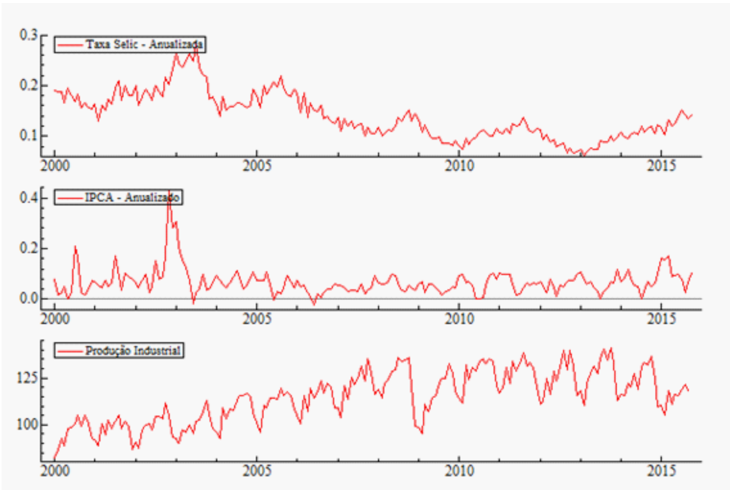


Figure 1: Default Series



Figue 2: Macroeconomic Indicators

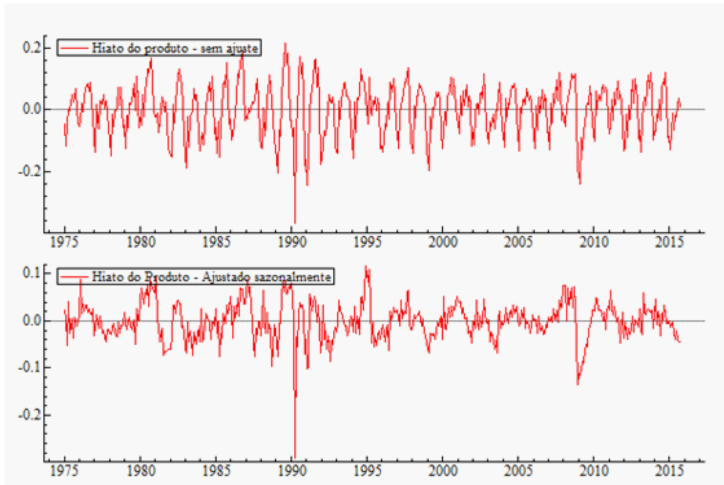


Figure 3: Output Gap using the full sample

Table 2: Unit Root Tests				
Variables Included	Seasonality	Lags	t-Statistics	Sample
Total Private Default	Yes	3	-3.493	2000(1) to 2015(9)
Private Domestic Bank Default	Yes	3	-3.408	2000(1) to 2015(9)
Private Foreign Bank Default	Yes	2	-2.376	2000(1) to 2015(9)
Public Bank Default	Yes	2	-5.037	2000(1) to 2015(9)
Selic	No	3	-1.882	2000(1) to 2015(9)
Industrial Production	Yes	12	-1.693	1975(1) to 2015(9)
Real Selic Deflated By IPCA	Yes	0	-5.675	2000(1) to 2015(9)
Note: Critical Values: 5% ≥ -2.882 and 1% ≥ -3.47				

Table 3: List of Models estimated for Total Default				
Model	Variable Included	Lags	Seasonality	Outliers Corrections
1	Default $\Delta \ln(Selic)$ Output Gap	12	Yes	Yes
2	Default Output Gap lnReal Selic	12	Yes	Yes
3	Default $\Delta \ln(Selic)$ Output Gap lnReal Selic	12	Yes	Yes

Table 4: Estimation of Model 1: Total Private Default, $\Delta \ln Selic$ and Output Gap

$X_{1t} =$	$\underset{(0.02)}{1.069 X_{1t-1}}$	$\underset{(0.0176)}{-0.1136 X_{1t-7}}$	$\underset{(7.03)}{+9.36 X_{2t-1}}$	$\underset{(9.21)}{+5.733 X_{2t-3}}$
$\underset{(0.17)}{-1.276 X_{3t-1}}$	$\underset{(0.297)}{-0.1618 X_{3t-3}}$	$\underset{(0.23)}{+0.8544 X_{3t-4}}$	$\underset{(0.23)}{+0.4008 X_{3t-10}}$	$\underset{(0.0314)}{-0.07617 \mathrm{CS}_{t-1}}$
$\underset{(0.051)}{-0.2017 \mathrm{CS}_{t-2}}$	$\underset{(0.0395)}{+0.1521 \mathrm{CS}_{t-3}}$	$\underset{(0.0395)}{+0.1046 \mathrm{CS}_{t-4}}$	$\underset{(0.0323)}{+0.05777 \mathrm{CS}_{t-6}}$	$\underset{(0.0288)}{+0.08823 \mathrm{CS}_{t-7}}$
$\underset{(0.034)}{+0.01186 \mathrm{CS}_{t-8}}$	$\underset{(0.0289)}{+0.1696 \mathrm{CS}_{t-9}}$	$\underset{(0.031)}{+0.1346 \mathrm{CS}_{t-10}}$	$\underset{(0.0594)}{+0.1562 \mathrm{DI{:}2003(2)}_t}$	$\underset{(0.0682)}{-0.3799 \mathrm{DI{:}2005(5)}_t}$
$\underset{(0.0685)}{-0.235\ \mathrm{DI{:}2005(6)}_t}$	$\underset{(0.0827)}{+0.3445 \mathrm{I{:}2002(12)}_t}$	$\underset{(0.0865)}{-0.3845 \mathrm{I{:}2003(11)}_t}$	$\underset{(0.084)}{+0.1471 \mathrm{I{:}2008(12)}_t}$	$\underset{(0.0588)}{+0.1835}$
$X_{2t} =$	$\underset{(0.0000173)}{-0.0002078 X_{1t-1}}$	$\underset{(0.000152)}{+4.86 * 10^{-5} X_{1t-7}}$	$\underset{(0.0606)}{-0.454\ X_{2t-1}}$	$\underset{(0.0794)}{+0.3455 X_{2t-3}}$
$\underset{(0.0146)}{-0.0006022 X_{3t-1}}$	$\underset{(0.0256)}{+0.005569 X_{3t-3}}$	$\underset{(0.00198)}{-0.0004785 X_{3t-4}}$	$\underset{(0.0137)}{-0.04591 X_{3t-10}}$	$\underset{(0.0000271)}{-0.001473 \mathrm{CS}_{t-1}}$
$\underset{(0.000439)}{-0.001238 \mathrm{CS}_{t-2}}$	$\underset{(0.00034)}{-0.0005494 \mathrm{CS}_{t-3}}$	$\underset{(0.0000271)}{+0.01752\ \mathrm{CS}_{t-4}}$	$\underset{(0.000278)}{+0.001594 \mathrm{CS}_{t-6}}$	$\underset{(0.0000248)}{+0.0005082 \mathrm{CS}_{t-7}}$
$\underset{(0.0000293)}{-0.0003908 \mathrm{CS}_{t-8}}$	$\underset{(0.0000249)}{-0.0006096 \mathrm{CS}_{t-9}}$	$\underset{(0.000267)}{-0.001042 \mathrm{CS}_{t-10}}$	$\underset{(0.000512)}{+0.01951\ \mathrm{DI{:}2003(2)}_t}$	$\underset{(0.000588)}{+0.000397 \mathrm{DI{:}2005(5)}_t}$
$\underset{(0.000591)}{+0.0009989 \mathrm{DI{:}2005(6)}_t}$	$\underset{(0.0000713)}{+0.001167 \mathrm{I{:}2002(12)}_t}$	$\underset{(0.000746)}{-0.001061 \mathrm{I{:}2003(11)}_t}$	$\underset{(0.000724)}{-0.0006703 \mathrm{I{:}2008(12)}_t}$	$\underset{(0.000506)}{+0.000669}$
$X_{3t} =$	$\underset{(0.00552)}{-0.01415 X_{1t-1}}$	$\underset{(0.00485)}{+0.01502 X_{1t-7}}$	$\underset{(0.0818)}{-5.156\ X_{2t-1}}$	$\underset{(2.54)}{-0.8133 X_{2t-3}}$
$\underset{(0.0468)}{+0.5078 X_{3t-1}}$	$\underset{(0.0818)}{+0.5054 X_{3t-3}}$	$\underset{(0.0633)}{-0.4699 X_{3t-4}}$	$\underset{(0.0437)}{+0.01304 X_{3t-10}}$	$\underset{(0.00865)}{+0.01333 \mathrm{CS}_{t-1}}$
$\underset{(0.014)}{+0.1764 \mathrm{CS}_{t-2}}$	$\underset{(0.0109)}{+0.05007 \mathrm{CS}_{t-3}}$	$\underset{(0.0113)}{+0.1264 \mathrm{CS}_{t-4}}$	$\underset{(0.00889)}{+0.1338\ \mathrm{CS}_{t-6}}$	$\underset{(0.00792)}{+0.09278 \mathrm{CS}_{t-7}}$
$\underset{(0.00936)}{+0.0866\ \mathrm{CS}_{t-8}}$	$\underset{(0.00796)}{+0.1057\ \mathrm{CS}_{t-9}}$	$\underset{(0.0189)}{+0.04848 \mathrm{CS}_{t-10}}$	$\underset{(0.0164)}{+0.03067 \mathrm{DI{:}2003(2)}_t}$	$\underset{(0.0188)}{+0.01799 \mathrm{DI{:}2005(5)}_t}$
$\underset{(0.0189)}{+0.04848 \mathrm{DI{:}2005(6)}_t}$	$\underset{(0.0228)}{-0.02954 \mathrm{I{:}2002(12)}_t}$	$\underset{(0.0238)}{-0.003409 \mathrm{I{:}2003(11)}_t}$	$\underset{(0.0231)}{-0.1404 \mathrm{I{:}2008(12)}_t}$	$\underset{(0.0162)}{-0.003359}$
Vector AR1 - 7 test	$F(63, 382) = 0.96766$	[0.5493]		
Vector Normality test	$\aleph^2(6) = 12.164$	[0.0584]		
Vector Hetero test	$F(180, 810) = 0.96215$	[0.6197]		
Vector RESET test	$F(18, 404) = 1.5980$	[0.0590]		

Note: X_1 is Total Private Default; X_2 is $\Delta \ln Selic$; X_3 is Output Gap and CS are Monthly centered seasonal dummies

Note: I:$i(j)$ is a Impulse Indicator Saturation (IIS) in year i month j and DI:$i(j)$ is a Difference IIS in year i month j

Table 5: Estimation of Model 2: Total Private Default, Output Gap and lnReal_Selic

$X_{1t} =$	$\underset{(0.0426)}{1.047}\ X_{1t-1}$	$\underset{(0.0811)}{+0.2944} X_{1t-3}$	$\underset{(7.03)}{-0.3988} X_{1t-4}$	$\underset{(0.15)}{-1.044} X_{2t-1}$
$\underset{(0.238)}{+0.1556} X_{2t-3}$	$\underset{(0.136)}{+0.2855} X_{2t-4}$	$\underset{(1.81)}{-1.277} X_{3t-1}$	$\underset{(0.0374)}{-0.1047}\mathrm{CS}_{t-2}$	$\underset{(0.0369)}{-0.1077}\mathrm{CS}_{t-3}$
$\underset{(0.0408)}{+0.1199}\mathrm{CS}_{t-4}$	$\underset{(0.0312)}{+0.09736}\mathrm{CS}_{t-6}$	$\underset{(0.0267)}{+0.1068}\mathrm{CS}_{t-7}$	$\underset{(0.0303)}{+0.08698}\mathrm{CS}_{t-8}$	$\underset{(0.0273)}{+0.1082}\mathrm{CS}_{t-9}$
$\underset{(0.0319)}{+0.09729}\mathrm{CS}_{t-10}$	$\underset{(0.0892)}{-0.3451}\mathrm{DI{:}2002(11)}_t$	$\underset{(0.125)}{+0.3417}\mathrm{I{:}2002(11)}_t$	$\underset{(0.0866)}{-0.3653}\mathrm{I{:}2003(11)}_t$	$\underset{(0.0843)}{-0.4442}\mathrm{I{:}2005(5)}_t$
$\underset{(0.0851)}{+0.1841}\mathrm{I{:}2008(12)}_t$	$\underset{(0.0877)}{+0.2995}\mathrm{I{:}2009(4)}_t$	$\underset{(0.0587)}{+0.2454}$		
$X_{2t} =$	$\underset{(0.0117)}{-0.01412} X_{1t-1}$	$\underset{(0.0222)}{+0.05262} X_{1t-3}$	$\underset{(0.017)}{+0.06312} X_{1t-4}$	$\underset{(0.041)}{+0.4384} X_{2t-1}$
$\underset{(0.0651)}{+\ 0.404}\ X_{2t-3}$	$\underset{(0.0373)}{-\ 0.443}\ X_{2t-4}$	$\underset{(0.496)}{-0.8498} X_{3t-1}$	$\underset{(0.0102)}{+0.1593}\mathrm{CS}_{t-2}$	$\underset{(0.0101)}{+0.0424}\mathrm{CS}_{t-3}$
$\underset{(0.0112)}{+0.1169}\mathrm{CS}_{t-4}$	$\underset{(0.00853)}{+\ 0.1313}\ \mathrm{CS}_{t-6}$	$\underset{(0.00732)}{+0.09563}\mathrm{CS}_{t-7}$	$\underset{(0.0083)}{+0.08367}\mathrm{CS}_{t-8}$	$\underset{(0.00748)}{+\ 0.1173}\ \mathrm{CS}_{t-9}$
$\underset{(0.00874)}{+0.05605}\mathrm{CS}_{t-10}$	$\underset{(0.0244)}{+0.0525}\mathrm{DI{:}2002(11)}_t$	$\underset{(0.0343)}{+0.04197}\mathrm{I{:}2002(11)}_t$	$\underset{(0.0237)}{+0.002756}\mathrm{I{:}2003(11)}_t$	$\underset{(0.0231)}{+0.01108}\mathrm{I{:}2005(5)}_t$
$\underset{(0.0233)}{-0.1382}\mathrm{I{:}2008(12)}_t$	$\underset{(0.024)}{-0.05134}\mathrm{I{:}2009(4)}_t$	$\underset{(0.0233)}{-0.1382}$		
$X_{3t} =$	$\underset{(0.00013)}{-0.0001633} X_{1t-1}$	$\underset{(0.000145)}{+9.827 * 10^{-5}} X_{1t-7}$	$\underset{(0.0633)}{-0.4908} X_{2t-4}$	$\underset{(0.00429)}{+0.001012} X_{2t-1}$
$\underset{(0.00681)}{+\ 0.0109}\ X_{2t-3}$	$\underset{(0.0039)}{-0.01586} X_{2t-4}$	$\underset{(0.0519)}{+\ 0.739}\ X_{3t-1}$	$\underset{(0.00107)}{+0.002961}\mathrm{CS}_{t-2}$	$\underset{(0.00106)}{-6.25 * 10^{-5}}\mathrm{CS}_{t-3}$
$\underset{(0.00117)}{+0.001984}\mathrm{CS}_{t-4}$	$\underset{(0.000893)}{+0.001252}\mathrm{CS}_{t-6}$	$\underset{(0.000766)}{+0.001014}\mathrm{CS}_{t-7}$	$\underset{(0.000869)}{-0.0005899}\mathrm{CS}_{t-8}$	$\underset{(0.00796)}{-0.0005167}\mathrm{CS}_{t-9}$
$\underset{(0.000915)}{-0.0006461}\mathrm{CS}_{t-10}$	$\underset{(0.00256)}{-0.006234}\mathrm{DI{:}2002(11)}_t$	$\underset{(0.00359)}{-0.01008}\mathrm{I{:}2002(11)}_t$	$\underset{(0.00248)}{-0.0001184}\mathrm{I{:}2003(11)}_t$	$\underset{(0.00241)}{+0.002545}\mathrm{I{:}2005(5)}_t$
$\underset{(0.00244)}{+0.001506}\mathrm{I{:}2008(12)}_t$	$\underset{(0.00251)}{-0.004782}\mathrm{I{:}2009(4)}_t$	$\underset{(0.00168)}{+0.00477}$		
Vector AR1 - 7 test	$F(63, 388) = 1.0608$	[0.3613]		
Vector Normality test	$\aleph^2(6) = 9.686$	[0.1385]		
Vector Hetero test	$F(132, 827) = 0.84135$	[0.8931]		
Vector RESET test	$F(18, 410) = 1.5411$	[0.0725]		

Note: X_1 is Total Private Default; X_2 is Output Gap; X_3 is lnReal_Selic and CS are Monthly centered seasonal dummies

Note: I:$i(j)$ is a Impulse Indicator Saturation (IIS) in year i month j and DI:$i(j)$ is a Difference IIS in year i month j

Table 6: Estimation of Model 3: Total Private Default, $\Delta \ln Selic$, Output Gap and lnReal_Selic

$X_{1t} =$	$\underset{(0.0161)}{1.072}\ X_{1t-1}$	$\underset{(0.018)}{-0.1201} X_{1t-7}$	$\underset{(7.86)}{+8.201} X_{2t-1}$	$\underset{(7.38)}{+1.43}\ X_{2t-3}$
$\underset{(7.51)}{-4.438} X_{2t-5}$	$\underset{(8.63)}{+9.399} X_{2t-9}$	$\underset{(6.01)}{-8.328} X_{2t-11}$	$\underset{(0.152)}{-0.663}\ X_{3t-1}$	$\underset{(0.222)}{+0.6191} X_{3t-8}$
$\underset{(0.245)}{-0.004911} X_{3t-9}$	$\underset{(0.157)}{+0.4221} X_{3t-10}$	$\underset{(1.93)}{-0.5077} X_{4t-1}$	$\underset{(1.85)}{-0.8028} X_{4t-12}$	$\underset{(0.0321)}{-\ 0.149}\ \mathrm{CS}_{t-2}$
$\underset{(0.0335)}{+0.001232} \mathrm{CS}_{t-4}$	$\underset{(0.0351)}{-0.2028} \mathrm{CS}_{t-5}$	$\underset{(0.0335)}{-0.002004} \mathrm{CS}_{t-6}$	$\underset{(0.0443)}{+0.03732} \mathrm{CS}_{t-7}$	$\underset{(0.0329)}{+0.1574} \mathrm{CS}_{t-9}$
$\underset{(0.0403)}{+0.08252} \mathrm{CS}_{t-10}$	$\underset{(0.091)}{-0.3017} \mathrm{DI{:}2002(11)}_t$	$\underset{(0.0843)}{+0.02391} \mathrm{DI{:}2003(10)}_t$	$\underset{(0.0588)}{+0.03472} \mathrm{DI{:}2004(2)}_t$	$\underset{(0.0836)}{-0.02286} \mathrm{DI{:}2005(4)}_t$
$\underset{(0.123)}{+0.326} \mathrm{I{:}2002(11)}_t$	$\underset{(0.126)}{-0.3854} \mathrm{I{:}2003(11)}_t$	$\underset{(0.12)}{-0.5115} \mathrm{I{:}2005(5)}_t$	$\underset{(0.0821)}{+0.1621} \mathrm{I{:}2008(12)}_t$	$\underset{(0.0723)}{+0.2092}$
$X_{2t} =$	$\underset{(0.00013)}{-0.0001633} X_{1t-1}$	$\underset{(0.000145)}{+9.827 * 10^{-5}} X_{1t-7}$	$\underset{(0.0633)}{-0.4908} X_{2t-1}$	$\underset{(0.0595)}{+0.3694} X_{2t-3}$
$\underset{(0.0605)}{+0.08857} X_{2t-5}$	$\underset{(0.0696)}{+0.2806} X_{2t-9}$	$\underset{(0.0485)}{-0.3311} X_{2t-11}$	$\underset{(0.00122)}{+0.002636} X_{3t-1}$	$\underset{(0.00179)}{+0.005414} X_{3t-8}$
$\underset{(0.00198)}{-0.00388} X_{3t-9}$	$\underset{(0.00127)}{-0.004081} X_{3t-10}$	$\underset{(0.0155)}{-0.05664} X_{4t-1}$	$\underset{(0.0149)}{-0.001074} X_{4t-12}$	$\underset{(0.000259)}{-0.0005029} \mathrm{CS}_{t-2}$
$\underset{(0.00027)}{+0.001289} \mathrm{CS}_{t-4}$	$\underset{(0.000196)}{+0.0003597} \mathrm{CS}_{t-5}$	$\underset{(0.00027)}{+0.001749} \mathrm{CS}_{t-6}$	$\underset{(0.000358)}{+0.001123} \mathrm{CS}_{t-7}$	$\underset{(0.000265)}{-0.000196} \mathrm{CS}_{t-9}$
$\underset{(0.000325)}{-0.000612} \mathrm{CS}_{t-10}$	$\underset{(0.000733)}{-0.0001302} \mathrm{DI{:}2002(11)}_t$	$\underset{(0.00068)}{-0.001993} \mathrm{DI{:}2003(10)}_t$	$\underset{(0.000474)}{-0.001692} \mathrm{DI{:}2004(2)}_t$	$\underset{(0.000674)}{+0.0009124} \mathrm{DI{:}2005(4)}_t$
$\underset{(0.000993)}{+0.001428} \mathrm{I{:}2002(11)}_t$	$\underset{(0.00102)}{-0.002141} \mathrm{I{:}2003(11)}_t$	$\underset{(0.000967)}{+0.0006891} \mathrm{I{:}2005(5)}_t$	$\underset{(0.000663)}{-2.61 * 10^{-5}} \mathrm{I{:}2008(12)}_t$	$\underset{(0.000583)}{+\ 0.0014}$

Table 6 (cont): Estimation of Model 3: Total Private Default, $\Delta \ln Selic$, Output Gap and lnReal_Selic

$X_{3t} =$	$\underset{(0.00471)}{-0.0090017X_{1t-1}}$	$\underset{(0.00525)}{+0.01626X_{1t-7}}$	$\underset{(2.29)}{-9.71\,X_{2t-1}}$	$\underset{(2.15)}{+9.043X_{2t-3}}$
$\underset{(2.19)}{-8.32\,X_{2t-5}}$	$\underset{(2.52)}{+11.54X_{2t-9}}$	$\underset{(1.76)}{-1.468X_{2t-11}}$	$\underset{(0.0443)}{+0.6624X_{3t-1}}$	$\underset{(0.0647)}{+0.1757X_{3t-8}}$
$\underset{(0.0716)}{-0.3892X_{3t-9}}$	$\underset{(0.0459)}{+0.3194X_{3t-10}}$	$\underset{(0.563)}{-1.614X_{4t-1}}$	$\underset{(0.539)}{+1.99\,X_{4t-12}}$	$\underset{(0.00937)}{+0.08269\text{CS}_{t-2}}$
$\underset{(0.00978)}{+0.1063\,\text{CS}_{t-4}}$	$\underset{(0.0103)}{+0.006228\text{CS}_{t-5}}$	$\underset{(0.00979)}{+0.0893\,\text{CS}_{t-6}}$	$\underset{(0.013)}{+0.1074\text{CS}_{t-7}}$	$\underset{(0.00961)}{+0.09792\text{CS}_{t-9}}$
$\underset{(0.0118)}{+0.02443\text{CS}_{t-10}}$	$\underset{(0.0266)}{+0.07214\text{DI:}2002(11)_t}$	$\underset{(0.0246)}{-0.02053\text{DI:}2003(10)_t}$	$\underset{(0.0172)}{-0.02398\text{DI:}2004(2)_t}$	$\underset{(0.0244)}{+0.08005\text{DI:}2005(4)_t}$
$\underset{(0.036)}{-0.04949\text{I:}2002(11)_t}$	$\underset{(0.0368)}{+0.03894\text{I:}2003(11)_t}$	$\underset{(0.035)}{+0.08558\text{I:}2005(5)_t}$	$\underset{(0.024)}{-0.1381\text{I:}2008(12)_t}$	$\underset{(0.0211)}{-0.03319}$
$X_{4t} =$	$\underset{(0.000458)}{-0.0001049X_{1t-1}}$	$\underset{(0.00051)}{-0.00006258X_{1t-7}}$	$\underset{(0.223)}{+0.4594X_{2t-1}}$	$\underset{(0.209)}{+0.1404X_{2t-3}}$
$\underset{(0.213)}{+0.6271X_{2t-5}}$	$\underset{(0.245)}{+0.5024X_{2t-9}}$	$\underset{(0.171)}{-0.1429X_{2t-11}}$	$\underset{(0.0043)}{-0.003477X_{3t-1}}$	$\underset{(0.00629)}{+3.898*10^{-5}X_{3t-8}}$
$\underset{(0.00696)}{-0.001828X_{3t-9}}$	$\underset{(0.00446)}{-0.002642X_{3t-10}}$	$\underset{(0.0547)}{+0.7139X_{4t-1}}$	$\underset{(0.0524)}{+0.1108X_{4t-12}}$	$\underset{(0.00951)}{+0.002271\text{CS}_{t-2}}$
$\underset{(0.000951)}{+0.002574\text{CS}_{t-4}}$	$\underset{(0.000997)}{+0.002591\text{CS}_{t-5}}$	$\underset{(0.000951)}{+0.002113\text{CS}_{t-6}}$	$\underset{(0.00126)}{+0.001354\text{CS}_{t-7}}$	$\underset{(0.000934)}{-0.0006061\text{CS}_{t-9}}$
$\underset{(0.00114)}{+0.0006846\text{CS}_{t-10}}$	$\underset{(0.00258)}{-0.004886\text{DI:}2002(11)_t}$	$\underset{(0.00239)}{+0.005454\text{DI:}2003(10)_t}$	$\underset{(0.00167)}{-0.001284\text{DI:}2004(2)_t}$	$\underset{(0.00237)}{+0.004227\text{DI:}2005(4)_t}$
$\underset{(0.00349)}{-0.01236\text{I:}2002(11)_t}$	$\underset{(0.00358)}{+0.009088\text{I:}2003(11)_t}$	$\underset{(0.00341)}{-0.002535\text{I:}2005(5)_t}$	$\underset{(0.00233)}{-0.002818\text{I:}2008(12)_t}$	$\underset{(0.00205)}{+0.003968}$
Vector AR1 - 7 test	$F(112, 459) = 1.3439$	[0.0192]		
Vector Normality test	$\aleph^2(8) = 14.948$	[0.0602]		
Vector Hetero test	$F(140, 516) = 0.9569$	[0.6177]		
Vector RESET test	$F(32, 499) = 1.0563$	[0.3862]		

Note: X_1 is Total Private Default; X_2 is $\Delta \ln Selic$; X_3 is Output Gap; X_4 is LnReal_Selic and CS are Monthly centered seasonal dummies

Note: I:$i(j)$ is a Impulse Indicator Saturation (IIS) in year i month j and DI:$i(j)$ is a Difference IIS in year i month j

Table 7: Granger Causality - Model 1					
Variables	Cause in Granger sense	Variables	Statistics	Distribution	p-value
X_2, X_3	$\longrightarrow$	X_1	103.02	$\aleph^2(6)$	[0.0000]
X_2	$\longrightarrow$	X_1	2.2567	$\aleph^2(2)$	[0.3226]
X_3	$\longrightarrow$	X_1	102.50	$\aleph^2(4)$	[0.0000]
X_1	$\longrightarrow$	X_2, X_3	13.438	$\aleph^2(4)$	[0.0093]
X_1	$\longrightarrow$	X_2	2.1304	$\aleph^2(2)$	[0.3447]
X_1	$\longrightarrow$	X_3	9.7748	$\aleph^2(2)$	[0.0075]
X_2	$\longrightarrow$	X_3	7.3027	$\aleph^2(2)$	[0.0260]
X_3	$\longrightarrow$	X_2	29.217	$\aleph^2(4)$	[0.0000]
Note: X_1 is Total Private Default; X_2 is $\Delta \ln Selic$; X_3 is Output Gap					

Table 8: Granger Causality - Model 2					
Variables	Cause in Granger sense	Variables	Statistics	Distribution	p-value
X_2, X_3	$\longrightarrow$	X_1	72.74	$\aleph^2(4)$	[0.0000]
X_2	$\longrightarrow$	X_1	68.931	$\aleph^2(3)$	[0.0000]
X_3	$\longrightarrow$	X_1	0.4975	$\aleph^2(1)$	[0.4806]
X_1	$\longrightarrow$	X_2, X_3	31.462	$\aleph^2(6)$	[0.0000]
X_1	$\longrightarrow$	X_2	19.956	$\aleph^2(3)$	[0.0002]
X_1	$\longrightarrow$	X_3	11.644	$\aleph^2(3)$	[0.0087]
X_2	$\longrightarrow$	X_3	18.306	$\aleph^2(3)$	[0.0004]
X_3	$\longrightarrow$	X_2	2.9412	$\aleph^2(1)$	[0.0863]
Note: X_1 is Total Private Default; X_2 is Output Gap; X_3 is lnReal_Selic					

Table 9: Granger Causality - Model 3					
Variables	Cause in Granger sense	Variables	Statistics	Distribution	p-value
X_2, X_3, X_4	$\longrightarrow$	X_1	98.957	$\aleph^2(11)$	[0.0000]
X_2	$\longrightarrow$	X_1	7.3321	$\aleph^2(5)$	[0.1971]
X_3	$\longrightarrow$	X_1	66.290	$\aleph^2(4)$	[0.0000]
X_4	$\longrightarrow$	X_1	0.35472	$\aleph^2(2)$	[0.8375]
X_1	$\longrightarrow$	X_2, X_3, X_4	23.935	$\aleph^2(6)$	[0.0005]
X_1	$\longrightarrow$	X_2	4.6208	$\aleph^2(2)$	[0.0992]
X_1	$\longrightarrow$	X_3	9.6149	$\aleph^2(2)$	[0.0082]
X_1	$\longrightarrow$	X_4	2.8766	$\aleph^2(2)$	[0.2373]
X_2	$\longrightarrow$	X_3	101.12	$\aleph^2(5)$	[0.0000]
X_2	$\longrightarrow$	X_4	13.257	$\aleph^2(5)$	[0.0211]
X_3	$\longrightarrow$	X_2	19.2866	$\aleph^2(4)$	[0.0006]
X_3	$\longrightarrow$	X_4	1.2866	$\aleph^2(4)$	[0.8636]
X_4	$\longrightarrow$	X_2	14.698	$\aleph^2(2)$	[0.0000]
X_4	$\longrightarrow$	X_3	17.141	$\aleph^2(2)$	[0.0002]

Note: X_1 is Total Private Default; X_2 is $\Delta \ln Selic$; X_3 is Output Gap; X_4 is lnReal_Selic

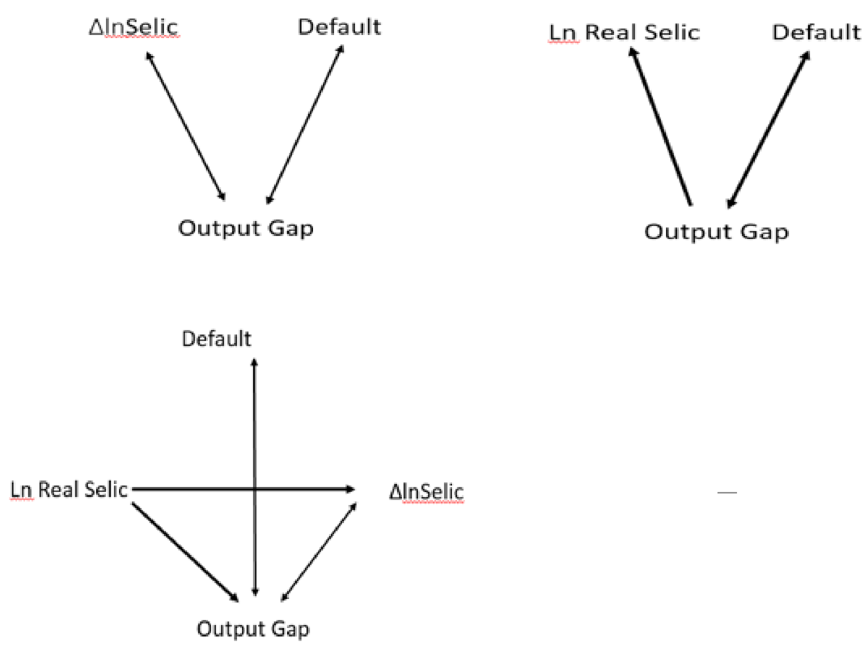


Figure 4: Resume of Granger Causality Test; 5 % of significance

Table 10: Quality assessment of the forecast of estimated models - Total Default

Nowcast			1 month ahead			6 months ahead			12 months ahead		
Models	MSE(1)	p-value	Models	MSE(1)	p-value	Models	MSE(1)	p-value	Models	MSE(1)	p-value
RW1	0.01779	1.0000	AR-12	0.04951	1.0000	M2	0.35973	1.0000	M1	0.81443	1.0000
AR-12	0.02529	0.1611	RW-2	0.05120	0.9868	M1	0.37714	0.9526	AR-12	0.83128	0.9196
M1-b	0.03088	0.1611	M1-b	0.05329	0.9868	AR-12	0.38475	0.9526	RW-13	**1.0774**	0.2002
M1	0.03212	0.1611	M1	0.05385	0.9868	M3	0.39321	0.4889	M2	0.93265	0.1525
			M2	0.05620	0.9656	RW-7	0.45866	0.2418			
			M2-b	0.06074	0.8250						
			M3	0.06298	0.8250						
			M3-b	0.07142	0.3331						
Nowcast			**1 month ahead**			**6 months ahead**			**12 months ahead**		
Models	MAD(2)	p-value	Models	MAD(2)	p-value	Models	MAD(2)	p-value	Models	MAD(2)	p-value
RW-1	0.10674	1.0000	M1	0.16690	1.0000	M2	0.43601	1.0000	M1	0.68310	1.0000
AR-12	0.11736	0.4865	AR-12	0.16747	0.9787	M1	0.44491	0.7978	AR-12	0.80300	0.1583
M1	0.12373	0.4865	M1-b	0.17069	0.9177	AR-12	0.47868	0.7629	M3	**0.78097**	0.1209
M1-b	0.1269	0.4222	M2	0.18048	0.8974	M3	**0.47153**	0.2319	RW-13	0.86075	0.1209
M2	0.13275	0.2986	RW-2	0.18132	0.8974	RW-7	0.53105	0.1290			
			M3	0.18712	0.8974						
			M2-b	0.19032	0.6885						
			M3-b	0.19674	0.6885						

Note: (1) MSE*1000; (2) MAD*1000 and Models are the ones in the Final Set using 10% level of significance

Table 11: Quality assessment of the forecast of estimated models - Default Private Domestic Banks

Nowcast			1 month ahead			6 months ahead			12 months ahead		
Models	MSE$^{(1)}$	p-value	Models	MSE$^{(1)}$	p-value	Models	MSE$^{(1)}$	p-value	Models	MSE$^{(1)}$	p-value
RW1	0.01973	1.0000	RW-2	0.05389	1.0000	M1	0.22354	1.0000	M1	0.38421	1.0000
			M1-b	0.0601	0.7279	M2	0.2962	0.1866			
			M1	0.06218	0.7279						
			AR-12	0.06722	0.5135						
			M2	0.0764	0.2788						
			M3	0.08758	0.1225						
Nowcast			1 month ahead			6 months ahead			12 months ahead		
Models	MAD$^{(2)}$	p-value	Models	MAD$^{(2)}$	p-value	Models	MAD$^{(2)}$	p-value	Models	MAD$^{(2)}$	p-value
RW-1	0.10674	1.0000	M1-b	0.18986	1.0000	M1	0.35894	1.0000	M1	0.45515	1.0000
			RW-2	0.19022	0.9858	M2	0.39947	0.7978			
			M1	0.19272	0.9858	M3	0.44812	0.7629			
			AR-12	0.20221	0.9095						
			M2	0.21316	0.7612						
			M3	0.2275	0.4269						

Note: $^{(1)}$ MSE*1000; $^{(2)}$ MAD*1000 and Models are the ones in the Final Set using 10% level of significance

Table 12: Quality assessment of the forecast of estimated models - Default Private Foreign Banks

Nowcast			1 month ahead			6 months ahead			12 months ahead		
Models	MSE$^{(1)}$	p-value	Models	MSE$^{(1)}$	p-value	Models	MSE$^{(1)}$	p-value	Models	MSE$^{(1)}$	p-value
RW1	0.03075	1.0000	RW-2	0.0781	1.0000	RW-7	0.58817	1.0000	RW-13	1.33957	1.0000
			AR-12	0.10004	0.5460	M2-b	0.66678	0.6450	AR-12	1.51307	0.5804
			M2-b	0.10436	0.5460	AR-12	0.66702	0.6450	M2-b	1.91159	0.5804
			M2	0.11875	0.5460	M2	0.72851	0.5918	M2	1.92637	0.5804
			M1-b	0.13893	0.5460	M1-b	0.8914	0.2185	M1	2.0181	0.5804
			M1	0.15667	0.5350	M1	0.93073	0.2185	M3	2.05606	0.5804
			M3-b	0.15147	0.2764	M3	0.97248	0.2185	M1-b	2.11029	0.5804
			M3	0.1587	0.1374	M3-b	1.00175	0.2185	M3-b	2.70316	0.3542
Nowcast			**1 month ahead**			**6 months ahead**			**12 months ahead**		
Models	MAD$^{(2)}$	p-value	Models	MAD$^{(2)}$	p-value	Models	MAD$^{(2)}$	p-value	Models	MAD$^{(2)}$	p-value
RW-1	0.12652	1.0000	RW-2	0.19670	1.0000	RW-7	0.57512	1.0000	RW-13	0.98225	1.0000
						M2-b	0.61694	0.7016	M2-b	1.02884	0.7793
						AR-12	0.62743	0.7016	AR-12	1.06999	0.7793
						M2	0.64774	0.6474	M2	1.08069	0.7793
						M1-b	0.68224	0.4714	M3	1.09629	0.7793
						M1	0.7018	0.4714	M1	1.10647	0.7793
						M3-b	0.70681	0.4714	M1-b	**1.10084**	0.6563
						M3	0.71866	0.4714	M3-b	1.1894	0.2540

Note: (1) MSE*1000; (2) MAD*1000 and Models are the ones in the Final Set using 10% level of significance

Table 13: Quality assessment of the forecast of estimated models - Default Public Banks

Nowcast			1 month ahead			6 months ahead			12 months ahead		
Models	MSE$^{(1)}$	p-value	Models	MSE$^{(1)}$	p-value	Models	MSE$^{(1)}$	p-value	Models	MSE$^{(1)}$	p-value
RW1	0.00974	1.0000	RW–2	0.01756	1.0000	RW-7	0.06682	1.0000	RW-13	0.13804	1.0000
Nowcast			1 month ahead			6 months ahead			12 months ahead		
Models	MAD$^{(2)}$	p-value	Models	MAD$^{(2)}$	p-value	Models	MAD$^{(2)}$	p-value	Models	MAD$^{(2)}$	p-value
RW-1	0.07511	1.0000	RW-2	0.10670	1.0000	RW-7	0.19849	1.0000	RW-13	0.29350	1.0000

Note: $^{(1)}$ MSE*1000; $^{(2)}$ MAD*1000 and Models are the ones in the Final Set using 10% level of significance