

What Does Google Say about Credit Developments in Brazil?

Abstract

In this paper multivariate State Space (SS) models are used to evaluate and forecast household loans in Brazil, taking into account two Google search terms in order to identify credit demand: *financiamento* (type of loan used to finance goods) and *empréstimo* (a more general type of loan). Our framework is coupled with nonlinear features, such as Markov-switching and threshold point. We explore these nonlinearities to build identification strategies to disentangle the supply and demand forces which drive the credit market to equilibrium over time. We also show that the underlying nonlinearities significantly improves the performance of SS models on forecasting the household loans in Brazil, particularly in short-term horizons.

Keywords: Google trends; credit market; household loans; state space models; Markov switching; threshold models.

JEL codes: E50; C32; C53.

1 Introduction

An in-depth understanding about how credit market works is of great concern for both academics and practitioners. According to Bernanke (1993), previously left aside by conventional theory, the credit became an important element of contemporary macroeconomics. Within this context, Bernanke and Blinder (1988) were one of the first to place the relationship between bank loans and aggregate demand under the spotlight, as part of a last decades growing literature focusing on the role of credit at monetary policy that complemented the traditional money view.¹

Following this literature, when assessing the response of loans to an interest rate shock, Bernanke and Blinder (1992) draw attention to a similar timing pattern between unemployment and the bank assets dynamics, an evidence that does not allow to rule out that this observed behavior reflects a passive demand effect and not the supply linkage required by the credit view. The underlined controversy enhances one of the main problems for researchers when evaluating the credit developments: the difficult task to identify the supply and demand forces that determines the equilibrium on loans market over time. As an additional example, after identifying one cointegration relationship where real loans are positively related to real GDP and negatively to real interest rates at Euro Area, Calza et al. (2003) assume that these results are linked to a long-run demand equation and estimate a Vector Error Correction (VEC) model in order to evaluate the short term dynamics of the credit demand. However, they recognize that cannot be excluded the hypothesis that supply factors were also in place, which could diminish the reliability of their results.

In order to overcome this kind of problem, researchers have been resorting to a variety of identification strategies. For instance, assuming that the quantity of loans observed in the banking system is the balance given by the intersection of the demand and supply curves at each point in time, Suzuki (2004) observes the price of loans in order to identify the shifts of the supply and demand curves behind the change in the observable quantity. Estimating a Vector Autoregression (VAR) with Japanese data, after an interest rate hike the author verifies a loan price increase and a loan quantity decrease, results that are claimed to be coherent with a supply shift and supported by the credit view of monetary policy.²

As alternative approaches, Del Giovane et al. (2011) rely on bank's surveys responses about loan supply/demand conditions combined with micro-data on loan quantities and prices to disentangle the forces shaping the lending to non-financial firms in Italy. Focusing on evaluating the balance sheet credit channel at Spain, Jimenez et al. (2012) separate bank loan supply from demand (i) studying loan applications, (ii) exploiting theoretically motivated interactions between economic and monetary conditions and (iii) observing how strong balance-sheet variables are. More recently, using bank-level data on lending in Italy,

¹The so-called credit view is evaluated through two main transmission channels that act at the loans supply: the balance sheet channel (Gertler and Gilchrist 1993) and the lending channel (Kashyap and Stein 1994). Bernanke and Gertler (1995) summarize this view of the credit channel and its role within the monetary policy transmission.

²The author focus on solving the supply versus demand puzzle, first noted by Bernanke and Blinder (1992), that arises from the similar timing pattern between the loans demand and the loans supply responses after a monetary tightening, linked respectively to the money and the credit view of monetary policy. As such, after a monetary tightening, the resulting reduction of bank loans does not, of itself, indicate whether the supply schedule of bank loans shifts left or the demand schedule for bank loans shifts left.

Del Giovane et al. (2017) identify shifts in demand and supply based on the responses of Italian banks to the Eurosystem’s Bank Lending Survey. In turn, Cipollini and Parla (2018) use Structural VAR analysis to disentangle credit demand and supply shocks and their effect on Italian economy during the 2008 to 2014 crisis period.

Considering studies applied to Brazilian economy, Coelho et al. (2017) propose a method based on the identification through heteroskedasticity approach (Rigobon 2003) to deal with the simultaneity bias that could arise on the interaction between supply and demand when determining the market equilibrium. The authors identification strategy relies on the need that the conditional variances of credit demand and supply change according to the monetary policy regime. Basically, in periods of high-variance (low-variance) of the policy rate, supply shifts more (less) than demand. Then, equilibrium observations from these periods will trace out demand (supply) more than supply (demand).

As Einav and Levin (2013) emphasize, the growth of internet triggered a data revolution that made available newly-rich informational sets to economics analyses. In the wake of this phenomenon, known as Big Data, one of the growing research streams is to explore web search engine data in order to nowcast or forecast economic and financial variables. In this environment, the Google Trends stands out as an useful tool, providing data based on Google Search that shows how often a particular search term is entered in comparison with all other search terms in different regions and languages (Jun et al. 2018).

There is a growing literature exploring Google Trends data in order to forecast financial and economic variables. Since Choi and Varian (2009b) and Choi and Varian (2009a) showed how Google Trends can be used to nowcast a set of time series (such as auto sales, retail sales, home sales, travel behavior and labor market variables), the number of studies applying this tool has increased. We can find works employing this type of data all around the world for a variety of problems: auto sales in Chile (Carriere-Swallow and Labbé 2011); private consumption in US (Vosen and Schmidt 2011); labour and housing markets both at UK (McLaren and Shanbhogue 2011) and US (Dietzel 2016); cinema admissions in UK (Hand and Judge 2012); stock market moves on US (Preis et al. (2013) and Hu et al. (2018)); business cycle turning points at US (Chen et al. 2015); energy data at US (Hassani and Silva 2016); hotel registrations in Puerto Rico (Rivera 2016); exchange rates at OECDs countries (Bulut 2017); building uncertainty indices for US and Australia (Castelnuovo and Tran 2017); tourism demand at Belgium and Austria (Onder 2017); and global oil consumption (Yu et al. 2018), *inter alia*.³

Regarding the credit market, Zeybek and Ugurlu (2015) use weekly Google Trends data to forecast the loan demand in Turkey. The authors show that Google search query data is successful at nowcasting loan demand. In turn, employing machine learning methods, Burdeau and Kintzler (2017) make an assessment of the value-added of Google Trends and Google Correlate to forecast French credit flows for house purchases. The authors find evidences that this data is useful to predict this variable several months in advance.

In other interesting use of Google Trends data, focusing regulatory initiatives regarding the change in the coverage and backing of deposit insurances in Europe, Fecht et al. (2019) show that Google searches for the term ”deposit insurance” and related strings can serve as early warning indicators for deposit shifts from local private banks to fully guaran-

³For a more complete discussion about research developments using Google Trends, see Jun et al. (2018).

teed public banks in Germany, reflecting depositors' fears about bank defaults. Accordingly to the authors, these findings show that the resilience of banks to investors' fears of bank failures is severely impaired by asymmetries in the deposit insurance coverage. When banks' liquidity is questioned, these asymmetries can lead to a reallocation of funds leaving some banks in a liquidity shortage while others are awash with liquidity.

Following this literature, we use Google Trends data to evaluate Brazilian credit market developments in two ways: (i) exploring nonlinearities in the relationship between the search terms linked to credit demand and the observable loan quantity, we propose identification strategies to disentangle credit supply and demand forces shaping household loans market; (ii) with the same settings, we perform in-sample and out-of-sample analysis, assessing how close we get to loans' market data generating process.

The Google search terms that we associate to credit demand boil down to two strings: *financiamento* and *empréstimo*. An important feature of these words is that, although sometimes they can be used as synonymous on Portuguese language, they are often linked to different types of loans in Brazil. The word *financiamento* usually refers to a longer term and collateralized credit used for acquiring goods such as houses and cars. In turn, the word *empréstimo* usually applies to a more expensive, shorter term and riskier loan. Also, while web searching for the former tends to take place when people have positive prospects over the economy, the same is not necessarily true for the latter. The search for the word *empréstimo* is also suitable for times of uncertainty, when people need to borrow money to supplement their income. The same hypothesis is not reasonable for the *financiamento* web searches.

Taking into account the abovementioned, our identification strategies rely on the thesis that nonlinearities in the way how these variables are related to the observable loan quantities over time can be useful to identify shifts in the credit market supply and demand forces. As far as we know, we are the first to explore Google Trends data to address this concern. An advantage of our identification strategies is that, while the procedures at previous literature usually explore data that is not fully available to the public or that exist just at some countries, our method is based on a freely available dataset and could be reproduced in all regions covered by Google Trends.

With the aim of studying the potential nonlinearities, we employ multivariate State Space (SS) models that allows for Markov-switching and threshold parameters.⁴ As the main contributions, besides proposing new identification strategies to disentangle credit supply and demand forces, we also obtained evidence showing that some of the underlying nonlinearities improve the performance of SS models on forecasting the household loans path and are useful to get closer to the data generating process linked to credit market in Brazil. We also show the potential role of web search engine data as complementary to evaluate the credit market condition, a task that is usually based on surveys held by policy makers or banking sectors.

The rest of the paper is structured as follows. Section 2 provides the data description. Section 3 focuses on the identifications strategies description. Section 4 presents our models. Besides presenting and discussing our results, Section 5 provides in-sample and out-of-sample analysis, comparing the performance of the built SS models against a set of naive ones. Then, section 6 brings the final remarks.

⁴A statistical theory for threshold estimation in the regression context is developed by Hansen (2000). Tong (2010) presents a selective review about the threshold model evolution in time series analysis.

2 Data description

Our data was collected for Brazilian economy from January/2004 to March/2019, totalling 183 observations. As Ubiergo (2012) highlights, the adoption of an inflation-targeting regime and better economic fundamentals helped Brazil to sustain significantly lower real interest rates in the last decade compared with the previous one. In the aftermath of the 2008 financial crisis, prices pressures that stemmed from overstimulus policies triggered a monetary tightening cycle that lasted almost four years. Latter, the combination of 2015/2016 severe recession and inflation deceleration opened room for Central Bank to begin a policy rate reduction cycle. As recession was left behind, a lower interest rate and inflation environment, joint to a better prospect for the economy, set up the basis for a loan market recovery. All these events characterized the general economic background of credit market developments in Brazil along our sample.

The following time series belong to our database: (i) non-earmarked new operations⁵ of personal credit, a type of loan granted to individuals not bound to any specific destination; (ii) monthly average interest rate from the same credit line; (iii) monthly Google Trends Indexes for *financiamento* and *empréstimo* search terms. The first two time series are provided by Central Bank of Brazil (CBB). In turn, the last two time series were obtained with the `gtrends` R package. All time series are log-linearized and normalized.

Table 1 presents information on the full history (2004-Present) of top searched queries related to the *empréstimo* and *financiamento* terms at Google.⁶ Exploring this data, we can denote some important differences on web searching these words. First, while there is no clear purpose when people search by *empréstimo*, the word *financiamento* is often associated with financing goods, as cars/vehicles (*carro/veículos*) and houses (*casa*). Second, we identified the word *negativado* among the ones associated with the term *empréstimo*, an expression that is often used in Brazil to refer to people with bad credit record. This result illustrates a characteristic of the type of loan linked to the word *empréstimo*, that is usually riskier (and consequently, more expensive) than other alternatives.

Therefore, an important feature of these words is that, although sometimes they can be used as synonymous on Portuguese language, they are often linked to different types of loans in Brazil. Overall, while web searching for *financiamento* tends to take place when people have positive prospects over the economy, the same is not necessarily true for the word *empréstimo*. Considering these aspects, we can explore shifts in the relationship among these variables over time to evaluate nonlinearities⁷ linked to credit market developments in Brazil.

⁵In Brazil, non-earmarked operations are financing and loans which rates are freely negotiated between financial institutions and borrowers, i.e., market rates. In non-earmarked operations, financial institutions have autonomy to decide loans destination. These operations contrast with earmarked operations, which rates are regulated and funds are mandatory or provided by the government. Then, once our interest relies on evaluating supply and demand credit market forces, we focused on the market linked non-earmarked operations.

⁶According to Google Trends website, the relative frequency allow us to rank the most searched terms. As an example, the value “100” indicates the most searched term, while “50” for other queries means that this search has half of the most searched term frequency.

⁷In the Appendix section, we performed various linearity tests on our observed variables. As we can see at Table 8, most of them rejected the linearity null hypothesis. In this sense, our identification strategies can be built to explore these data patterns.

Table 1: Related queries: *empréstimo* and *financiamento* search terms.

Rank	<i>empréstimo</i>		<i>financiamento</i>	
	Related queries	Relative freq.	Related queries	Relative freq.
1	empréstimo	100	caixa financiamento	100
2	empréstimo pessoal	83	simulador financiamento	52
3	consignado	73	financiamento veículos	48
4	empréstimo consignado	73	financiamento de veículos	42
5	caixa empréstimo	69	simulador de financiamento	39
6	fazer empréstimo	53	simulação financiamento	37
7	crefisa	49	simulação	37
8	crefisa empréstimo	48	financiamento santander	33
9	empréstimo online	44	financiamento carro	30
10	empréstimo negativado	39	financiamento veículo	28
11	empréstimo para negativado	32	financiamento casa	27
12	empréstimo pessoal	26	financiamento carros	25
13	bradesco empréstimo	25	simulação de financiamento	24
14	banco do brasil empréstimo	24	bradesco financiamento	23
15	empréstimo do banco do brasil	23	juros financiamento	23

Source: Google

3 Identification strategies

First of all, it is important to highlight that our framework differs from structural settings as Rigobon (2003) and Lanne et al. (2010). The identification strategies built by these authors are indicated in situations where we intend to identify unobserved supply and demand shocks from observed information about equilibrium (prices and quantities), in order to recover parameters from a theoretical model with endogenous variables (e.g., the slope of supply and demand curves). Instead, our identification strategies are empirical and aim to provide information about which unobserved supply and demand forces shifted the equilibrium in credit market each point in time. The information we are interested in is useful by both banks and policy makers, as they could use it to evaluate if an additional supply of loans would match demand or if an additional credit stimulus would be effective.⁸

Basically, assuming that web search engine data are good proxies for credit demand, we can identify some of these curves shifts by cross-referencing this data on information about the observed equilibrium. In particular, as we observe credit equilibrium⁹ and also proxies for loans' demand with different risk levels, we claim that our framework is able to identify two kind of curve shifts in households loan markets: (i) through the Google indexes we presented in last sections, we identify credit demand shifts accordingly to different risk

⁸For instance, central banks usually conduct credit market condition surveys, showing interest to evaluate supply and demand conditions individually. As Annibal and Koyama (2011) emphasize, such surveys are held in countries as Brazil (Pesquisa Trimestral de Condições de Crédito), United States (Senior Loan Officer Opinion Survey on Bank Lending Practices), England (Credit Conditions Survey), Japan (Senior Loan Officer Opinion Survey on Bank Lending Practices), Chile (Estándares de Aprobación en el Mercado del Crédito Bancario) and Euro Area (Bank Lending Survey).

⁹We highlight that similar to Suzuki (2004), our identification strategies assume that the observable quantity of bank loans is the balance given by interactions of non-observed demand and supply forces in credit market. In this sense, as time goes by, a change in the observed quantity could be associated with a shift of the demand curve, a shift of the supply curve, or both.

levels; (ii) exploring this information about risk on empirical models that address the co-movements among our observed variables (credit equilibrium and the demand proxies), we identify specific supply shifts linked to changes in the level of risk in credit market.

Considering the last curve shift mentioned, it is important to highlight that there is an extensive literature that relates credit supply shifts to changes in borrower (demand) risk. For instance, as emphasized by Freixas and Rochet (2008), the usual graphical analysis of supply and demand does not work in the context of the credit market. The reason is that the credit supply function may be backward-bending, what gives rise to a situation of credit rationing (the demand for credit exceeds supply at the prevailing interest rate). The theoretical background usually emphasize the role of asymmetric information to explain this phenomenon. As examples, Stiglitz and Weiss (1981) explore the concept of adverse selection to develop a model that enhances this supply curve shape. In turn, moral hazard supports the models developed by Jaffee and Russell (1976) and Bester and Hellwig (1987). Last but not least, Williamson (1987) relies on monitoring costs due to default risk.

Inside this view, departing from Williamson (1987), Kick et al. (2015) present a framework that is particularly illustrative for our identification strategy. As the authors note, assuming that a bank's loan supply function (L) is an increasing function of the expected (conditional) profit (Π) of a representative standard debt contract, it depends not only on the payment obligation to borrowers but also on the general risk level in the credit market. Moreover, defining $F(\chi/\theta)$ as the cumulative probability distribution function of the outcome χ of a representative borrower's investment project conditional on the general risk θ in the credit market, where θ shifts $F(\chi/\theta)$ in the sense of first-order stochastic dominance (FSD), Kick et al. (2015) highlight that a higher risk level θ makes low realizations of the project outcome χ more likely. Formally:

$$\frac{\partial F(\chi/\theta)}{\partial \theta} > 0 \quad \forall \chi \quad (3.1)$$

As a result, a higher level of risk θ causes a bank reduce its loan supply for any nominal payment obligation R , i.e.

$$\frac{\partial L(\mathbb{E}(\Pi(\chi/\theta)))}{\partial \theta} = L'(\cdot) \frac{\partial \mathbb{E}(\Pi(\chi/\theta))}{\partial \theta} < 0 \quad \forall R \quad (3.2)$$

Taking into account all the information abovementioned, at our framework, the identification strategy for the non-observed demand forces is straightforward. We assume that an increase (decrease) in web searches by the terms *finanziamento* and *empréstimo* would work as proxies for an increase (decrease) in loans demand. The greater the interest by these queries, the greater the strength of demand forces acting in credit market. Moreover, as enhanced before, each search term provides information about different types of loans demand. The *finanziamento* term is often associated with less risky and collateralized loans that households usually apply to finance goods, whereas the *empréstimo* term can be linked to riskier non-collateralized loans with no specific purpose.

In turn, the identification strategy for the non-observed supply forces explores the co-movement among our observed variables. More specifically, on one hand, in periods where the demand indexes are not moving in tandem with the observed loan measure, the equilibrium observations from this period will trace out supply more than demand. On the

other hand, in periods where the demand indexes are moving in tandem with the observed loan measure, we cannot rule out the possibility that demand or both demand/supply forces are acting on credit market balance.

Particularly, taking into account the potential relationships among the observed variables and the distinct risk degrees associated with our demand proxies, we could expect at least three reasonable situations where an increase (decrease) in our demand indexes should and should not coincide with an increase (decrease) in the observed loan measure. At our framework, we assume that eventual changes across these particular situations (or, regimes) would provide information on specific supply shifts linked to lenders risk aversion.

As a first case, consider the situation when all observable variables dynamics do coincide. Given that all observable variables are now moving in the same direction, this situation should be characterized by a common positive correlation among the loan quantity measure and both demand indexes. In other words, an increase (decrease) in both lower and higher risk demand proxies are now matched with an increase (decrease) in credit market observed equilibrium, an event that would become most likely as an economy moves up (slows down) and lenders are less (more) risk averse. In this situation, we cannot rule out that both demand and supply forces are driving the equilibrium to the same direction.

A second case is given by a situation where the variables dynamics do not coincide. In this regime, both demand indexes are not moving in the same direction of the observed loan quantity and the equilibrium from this period will trace out supply more than demand. In the opposite direction to the first case, an increase (decrease) in both lower and higher risk demand proxies are now matched with a decrease (increase) in the credit market observed equilibrium, an event that would show a stronger dissociation between demand and supply forces at credit market. For instance, this result could be expected in a scenario where lenders would be extremely risk averse and unwilling to meet a higher demand for any type of loans (even the collateralized ones), driving the observed equilibrium downward.

In the third possible situation, the variables dynamics also do not coincide. But now, the riskier demand index is not moving in tandem with both observed loan measure and less risky demand index. To illustrate this result, as an economy slows down (moves up) and the unemployment rate rises (falls), we could expect a reduction (an increase) in the observed loan measure and we cannot rule out that supply forces are acting in this new equilibrium due to lenders higher (lower) risk aversion. At the same time, the fears of job loss rise (decrease) and households would be unwilling (willing) to apply for a home or an auto loan, a decision that results in a demand reduction (increase) for the less risky and usually collateralized type of credit, that we linked to the *finanziamento* term. Nevertheless, in these conditions, would not be a surprise if we observe an increase (a decrease) in the household demand for loans in order to supplement their income, a riskier type of credit that we linked to the *empréstimo* term.

Departing from these elements, a built-in piece of our identification strategies is a nonlinear co-movement among the observed variables. In other words, our framework must be able to evaluate if the observed loan quantity is or is not moving in tandem with the demand indexes as time goes by. In next section, we describe a baseline model designed to implement this framework- Also, we propose an alternative setting that explores information about loans price to perform a robustness check on our main approach.

4 Models

In order to implement the identification strategies discussed at previous section, we applied SS models. Exploring this class of models, the co-movements are analyzed taking a state variable as reference, introduced at our setting as a common trend extracted from the following observed variables: (i) non-earmarked new operations of personal credit (l_t); (ii) Google Trends Index for the *financiamento* search term (f_t); (iii) and Google Trends Index for the *empréstimo* search term (e_t).

Moreover, to evaluate if the observed variables dynamics do or do not coincide over time, the framework is coupled with regime-switching features. Basically, these features were introduced in two ways: (i) as a first approach, the regime shifts evolve according to a Markov Chain; (ii) as a second approach, the regime shifts are triggered by loans' price volatility in relation to an estimated threshold point.¹⁰ While the former approach provides our baseline models, the latter will be explored for robustness check and comprises the group of benchmark models at both in-sample and out-of-sample validations.

We estimated the parameters of these models minimizing the negative of a log-likelihood function recursively, exploring a Kalman Filter algorithm. Following Hamilton (1994), the maximization of the log-likelihood function is started by making an initial guess to the numerical values of the unknown parameters. In the first round, the likelihood function value results from this initial parameter values. Then, under the assumption that the initial state and the errors (at both state and measure equations) are normal, numeric optimization methods are successively applied to update the parameter values until the negative of the log-likelihood is minimized. In the Markov-Switch models, the regimes are changing due to a latent Markov process, where we estimate the parameters minimizing the negative of an conditional log-likelihood derived from a mixture of normal densities. In this conditional log-likelihood, each normal density is weighted by a predicted probability associated to an initially unknown regime. In turn, the threshold model is similar to the Markov Switch model discussed. However, instead of random parameters and regime-changes due to a latent Markov process, now the estimation of the parameters is influenced by an observed exogenous process given by a stochastic volatility model.

4.1 Dynamic SS Models with Markov Switching (DSSMS)

The first class of models are the Dynamic SS models with Markov-switching (DSSMS) parameters, where we allow for a regime change to deal with our potential nonlinearity thesis among our search terms and the observable loan quantity. Basically, this specification is characterized by two principles. First, the state s_t is assumed to be Markovian. That is, the system dynamics at time t is generated by one of m possible regimes evolving according to a Markov chain, with the particular regime being initially unknown. Second, the measure

¹⁰Following Piger (2007), there is two broad categories of regime-switching models: "Threshold" models and "Markov-switching" models. As the author emphasize, the primary difference between these approaches is in how the evolution of the state process is modeled. Threshold models, introduced by Tong (1983), assume that regime shifts are triggered by the level of observed variables in relation to an unobserved threshold. Markov-switching models, introduced by Goldfeld and Quandt (1973), Cosslett and Lee (1985), and Hamilton (1989), assume that the regime shifts evolve according to a Markov chain.

equation dependent variables are independent given s_t . The general idea is that the value of the state at time t (s_t) specifies the distribution of the observations at time t .

Formally, the state s_t is a Markov chain taking values in a finite state $1, \dots, m$ conditional to a distribution

$$\pi_j = Pr(s_t = j) \quad (4.1)$$

and stationary transition probabilities

$$\pi_{ij} = Pr(s_{t+1} = j / s_t = i) \quad (4.2)$$

for $t = 0, 1, 2, \dots$ and $i, j = 1, \dots, m$. Since the second principle of the model is that the observations are conditionally independent, we must specify the distributions. The general form of them is denoted bellow.

$$\pi_j(y_t) = p(y_t | s_t = j) \quad (4.3)$$

In this setting, y_t gives the measure equation vector of dependent variables. On our models, these variables are linked through measure equations, supposing they have a common trend (x_t). As a special feature, this class of models is coupled with a Hidden Markov process that determines whether the coefficients linked to the search terms change over time. Thus, as time goes by, the measure equations could assume two structures ($m = 2$):

$$\begin{bmatrix} l_t \\ f_t \\ e_t \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2^i \\ \beta_3^i \end{bmatrix} [x_t] + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{t,l} \\ \epsilon_{t,f} \\ \epsilon_{t,e} \end{bmatrix} \quad (4.4)$$

and

$$\begin{bmatrix} l_t \\ f_t \\ e_t \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2^j \\ \beta_3^j \end{bmatrix} [x_t] + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{t,l} \\ \epsilon_{t,f} \\ \epsilon_{t,e} \end{bmatrix}, \quad (4.5)$$

where

$$R^i = var \begin{pmatrix} \epsilon_{t,l} \\ \epsilon_{t,f} \\ \epsilon_{t,e} \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12}^i & r_{13}^i \\ r_{21}^i & r_{22}^i & r_{23}^i \\ r_{31}^i & r_{32}^i & r_{33}^i \end{pmatrix} \text{ and } R^j = var \begin{pmatrix} \epsilon_{t,l} \\ \epsilon_{t,f} \\ \epsilon_{t,e} \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12}^j & r_{13}^j \\ r_{21}^j & r_{22}^j & r_{23}^j \\ r_{31}^j & r_{32}^j & r_{33}^j \end{pmatrix} \quad (4.6)$$

The noises ($\epsilon_{t,k}$, where k denotes the dependent variables at the measure equation) are normally distributed and characterized by an observation covariance matrix given by R . The matrix R is symmetric ($r_{21} = r_{12}$, $r_{31} = r_{13}$ and $r_{32} = r_{23}$) and positive semi-definite across regimes. As we denoted, the coefficients β_2 and β_3 and the variance/covariance parameters linked to both search terms (r_{22} , $r_{21} = r_{12}$, r_{33} , $r_{31} = r_{13}$ and $r_{32} = r_{23}$) are allowed to switch between two initially unknown regimes ($i, j = 1, 2$).

In turn, the transition equation gives the law of motion for x_t . For simplicity, this variable is modeled as a first order autoregressive process:

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \epsilon_{t,x} \quad (4.7)$$

where the state variance is given by $Q^x = \text{var}(\epsilon_{t,x})$ with a normally distributed noise ($\epsilon_{t,x} \sim \mathcal{N}(0, Q^x)$).

Adding the assumptions that the noises at transition and measure equations are uncorrelated and the initial state is normally distributed ($x_0 \sim \mathcal{N}(\mu_{x,0}, \Sigma_{x,0})$), we estimate the parameters by Maximum Likelihood. The log-likelihood may be written as

$$\begin{aligned} \ln L_Y(\Theta) = & \pi_i(t) \frac{1}{2} \left(\sum_{t=1}^n \ln |\Sigma_t(\Theta^i)| + \sum_{t=1}^n \epsilon_t(\Theta^i)' \Sigma_t(\Theta^i)^{-1} \epsilon_t(\Theta^i) \right) \\ & + \pi_j(t) \frac{1}{2} \left(\sum_{t=1}^n \ln |\Sigma_t(\Theta^j)| + \sum_{t=1}^n \epsilon_t(\Theta^j)' \Sigma_t(\Theta^j)^{-1} \epsilon_t(\Theta^j) \right) \end{aligned} \quad (4.8)$$

where the conditional probability (π) and the parameter vector (Θ) - both evaluated at regimes $i, j = 1, 2$ - are estimated observing the measure equation innovations (ϵ_t) and the innovations covariance matrix (Σ_t).

Our thesis is that these nonlinearities can be useful to get a better representation of the data generating process, helping to identify changes in credit market condition and to disentangle supply/demand forces acting in equilibrium.

We consider four specifications following this general structure. The first two are DSSMS models allowing regime changes just for the *empréstimo* term related parameters, with (DSSMS/E/DR) and without (DSSMS/E/NDR) diagonal observation covariance matrix (R). The remaining models are extensions allowing for regime changes on all parameters related to both *empréstimo* and *financiamento* search terms, with (DSSMS/E&F/DR) and without (DSSMS/E&F/NDR) diagonal observation covariance matrix (R).

Combining the identification strategies proposed for demand and supply, we can evaluate how these forces are acting on credit market equilibrium. For instance, if web searches for the *empréstimo* term are increasing, showing a growing demand for riskier loans, while β_3 is negative (positive) and the remaining measure equation parameters β_1 and β_2 are positive, we have a situation where the riskier demand index is not moving (is moving) in the same direction that the common factor, the observed loan quantity and the less risky demand index. This is an example where the supply and demand forces would be more desynchronized (synchronized), with a lower (higher) probability that lenders meet the total market needs for riskier loans.

4.2 Dynamic SS Models with a Threshold Parameter (DSST)

Our second class of models are the Dynamic SS Models with a Threshold Parameter (DSST), an extension where we introduce a threshold coefficient (γ) to deal with our underlying nonlinearity thesis about the data generating process. The time series indicator used as reference for the threshold parameter is the personal credit interest rate volatility. Departing from the log-linearized square of the demeaned underlying interest rate (r_t), this

variable is modelled as a stochastic volatility process given by a new state (h_t) introduced at the framework. Then, the measurement equation becomes

$$\begin{bmatrix} l_t \\ f_t \\ e_t \\ w_t \end{bmatrix} = \begin{cases} \begin{bmatrix} \beta_1 & 0 \\ \beta_2^i & 0 \\ \beta_3^i & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ h_t \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{t,l} \\ \epsilon_{t,f} \\ \epsilon_{t,e} \\ \epsilon_{t,w} \end{bmatrix} & \text{if } h_{t-1} > \gamma \\ \begin{bmatrix} \beta_1 & 0 \\ \beta_2^j & 0 \\ \beta_3^j & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ h_t \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{t,l} \\ \epsilon_{t,f} \\ \epsilon_{t,e} \\ \epsilon_{t,w} \end{bmatrix} & \text{if } h_{t-1} \leq \gamma \end{cases} \quad (4.9)$$

where $w_t = \log r_t^2$ and $\epsilon_{t,w} = \log \epsilon_{t,r}^2$.¹¹ In turn, now we have state equations denoted by

$$\begin{bmatrix} x_t \\ h_t \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \delta_0 \end{bmatrix} + \begin{bmatrix} \alpha_1 & 0 \\ 0 & \delta_1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ h_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{t,x} \\ \epsilon_{t,h} \end{bmatrix} \quad (4.10)$$

Analogously to x_t , the law of motion for the state h_t is given by an AR(1) process and the variance is equal to $Q^h = \text{var}(\epsilon_{t,h})$, with a normally distributed noise ($\epsilon_{t,h} \sim \mathcal{N}(0, Q^h)$).

Once more, adding the assumptions that the noises at transition and measure equations are uncorrelated and the initial states are normally distributed ($x_0 \sim \mathcal{N}(\mu_{x,0}, \Sigma_{x,0})$ and $h_0 \sim \mathcal{N}(\mu_{h,0}, \Sigma_{h,0})$), we estimate the parameters by Maximum Likelihood. The log-likelihood function is usual, denoted as

$$\ln L_Y(\Theta) = \frac{1}{2} \sum_{t=1}^n \ln |\Sigma_t(\Theta)| + \frac{1}{2} \sum_{t=1}^n \epsilon_t(\Theta)' \Sigma_t(\Theta)^{-1} \epsilon_t(\Theta) \quad (4.11)$$

where the parameter vector is Θ (which includes the threshold coefficient, γ), the measure equation innovations are represented by ϵ_t and Σ_t gives the innovations covariance matrix.

Similar to the previous class of models, we consider four specifications following this general structure. The first two are DSST models allowing a regime change just for the *empréstimo* coefficient parameter (β_3), with (DSST/E/DR) and without (DSST/E/NDR) diagonal observation covariance matrix (R). The remaining models are extensions allowing for regime changes at the coefficients related to both *empréstimo* and *financiamento* search terms (β_3 and β_2), with (DSST/E&F/DR) and without (DSST/E&F/NDR) diagonal observation covariance matrix (R).

Thus, the identification strategy for the credit supply assumes that banks are able to determine the cost of households loans and that changes in these costs could be associated with changes in credit market risk degree. More precisely, if the price volatility triggers changes at the relationship among our demand indexes and the observed loan quantity, this could be viewed as a signal that lenders are responding to risk environment changing the cost of the loans (therefore, a supply side feature).¹²

¹¹Thus, we apply a traditional stochastic volatility model where $r_t = \epsilon_{t,r} \exp(\frac{h_t}{2})$

¹²As Tabak et al. (2015) show, the Brazilian banking industry functions under conditions of monopolistic competition and bank's market power fluctuates throughout time. This means that, at least on some moments, financial institutions are able to lead the market by changing the interest charged on loans.

On this framework, the indicator time series acts as the switching mechanism, allowing for parameters β_2 and β_3 to change conditionally on loan’s cost volatility.¹³ For instance, if a high-volatility regime is matched by an increase (a decrease) of our riskier demand measure and at this regime our riskier demand measure is not moving in tandem with the observed equilibrium,¹⁴ we cannot rule out the possibility that supply is shifting in response to the higher (lower) risk in credit market. We take this result as a signal that, while the demand for riskier loans are increasing (decreasing) the equilibrium observations from these periods would be tracing out a specific kind of supply force linked to lenders risk appetite. Once again, this specific supply shift is associated with an eventual change in the sign of parameter β_3 , a movement that would be informative on the odds that the riskier loan demand is being met by supply. But now, instead of a Markov switch, the changes on β_3 sign depend on a threshold linked to the volatility dynamics of loans price.

5 Results and Discussion

In this section, besides evaluating the estimated parameters and providing a robustness check for our baseline setting, we make in-sample and out-of-sample analysis, comparing the performance of the nonlinear SS models against a set of naive ones. In these last exercises, we include among the benchmarks a standard linear Gaussian SS model, known as Dynamic SS Linear (DSSL) model. This framework rules out regime changes on parameters estimation and becomes a reference to evaluate whether the nonlinear settings provide gains compared with the linear one. The DSSL is detailed at Appendix section.

5.1 Estimated parameters and nonlinearities evaluation

The parameters estimated in the state equations are presented at Table 2. In turn, we show the estimated parameters from the measurement equations at Table 3.

As a general result, we observe throughout the transition equations parameters that the state variables exhibited a high degree of persistence in all models (all specifications present α_1 positive and close to one). Observing the parameters linked to the common factor (x_t), at some specifications we cannot statistically reject the hypothesis that this state variable follows a random walk.

Also, compared with our linear benchmark (DSSL), we observe an improvement of the models when we introduce the underlying nonlinearities, once we identify a greater number of parameters statistically different from zero at 10% on these specifications. We interpret this result as a clue that, allowing for the underlying nonlinearities, we are closer to the credit market data generating process than we were with the DSSL framework.

Therefore, we assume that changes in the loan’s cost volatility are linked with supply conditions, in a way that is possible to use this data variation to disentangle supply from demand.

¹³The idea of using the loan’s interest rate volatility as reference for the threshold estimation came from Coelho et al. (2017). Based on the assumption that credit supply reacts more quickly to shocks of monetary policy than credit demand, these authors explore the policy rate volatility dynamics within an identification strategy to estimate supply and demand parameters at credit market. Herein, instead of policy rate, we rely on the interest rate charged over the household loans applied at our models.

¹⁴That is, β_3 is negative while β_1 is positive.

Table 2: State equations estimated parameters

	DSSL/ DR	DSSL/ NDR	DSSMS/ E/DR	DSSMS/ E/NDR	DSSMS/ E&F/DR	DSSMS/ E&F/NDR	DSST/ E/DR	DSST/ E/NDR	DSST/ E&F/DR	DSST/ E&F/NDR
$\ln(Q^x)$	-3.547*** (0.718)	-3.591*** (0.719)	-3.306*** (0.504)	-2.978*** (0.267)	-3.546*** (0.734)	-1.993*** (0.241)	-2.910*** (0.323)	-2.783*** (0.000)	-2.556*** (0.281)	-1.978*** (0.000)
α_1	0.988*** (0.006)	0.985*** (0.006)	0.989*** (0.005)	0.985*** (0.005)	0.988*** (0.006)	0.976*** (0.013)	0.986** (0.005)	0.994*** (0.000)	0.987*** (0.005)	0.985*** (0.000)
α_0	0.761 (0.667)	0.576 (0.482)	1.078 (0.771)	1.002*** (0.413)	0.799 (0.71)	0.757 (0.609)	1.065** (0.463)	3.792*** (0.000)	1.71** (0.85)	1.372*** (0.000)
$\ln(Q^h)$	—	—	—	—	—	—	-0.225*** (0.000)	-0.054*** (0.000)	-0.533*** (0.000)	0.002*** (0.000)
δ_1	—	—	—	—	—	—	0.989*** (0.000)	0.904*** (0.000)	0.987*** (0.000)	0.961*** (0.000)
δ_0	—	—	—	—	—	—	3.730*** (0.000)	-0.882*** (0.000)	1.147*** (0.000)	1.739*** (0.000)
γ	—	—	—	—	—	—	-0.426*** (0.000)	0.299*** (0.000)	-0.577*** (0.000)	0.452*** (0.000)
$\ln(\pi_{ii}/(1 - \pi_{ii}))$	—	—	4.824*** (0.926)	2.426*** (0.602)	3.951*** (0.682)	3.517*** (0.973)	—	—	—	—
$\ln(\pi_{ji}/(1 - \pi_{ji}))$	—	—	-3.346*** (1.222)	-4.103*** (0.674)	-3.142*** (0.758)	-4.454*** (0.794)	—	—	—	—

* $pvalue < 0.10$, ** $pvalue < 0.05$, *** $pvalue < 0.01$. Standard errors between ().

Table 3: Measurement equations estimated parameters

	DSSL/ DR	DSSL/ NDR	DSSMS/ E/DR	DSSMS/ E/NDR	DSSMS/ E&F/DR	DSSMS/ E&F/NDR	DSST/ E/DR	DSST/ E/NDR	DSST/ E&F/DR	DSST/ E&F/NDR
β_1	2.386 (1.687)	2.427 (1.715)	1.819** (0.891)	1.415*** (0.324)	2.375 (1.719)	1.227*** (0.288)	1.274*** (0.371)	1.114*** (0.000)	0.938*** (0.231)	0.640*** (0.000)
β_2	1.672 (1.189)	1.749 (1.245)	1.279** (0.635)	0.962*** (0.231)	—	—	0.830*** (0.235)	0.773*** (0.000)	—	—
β_2^i	—	—	—	—	1.931 (1.404)	0.971*** (0.242)	—	—	0.766*** (0.1971)	0.485 (0.000)
β_2^j	—	—	—	—	0.768 (0.645)	0.512*** (0.208)	—	—	0.459 (0.1437)	0.209 (0.000)
β_3	0.151 (0.218)	-0.070 (0.242)	—	—	—	—	—	—	—	—
β_3^i	—	—	-0.659* (0.339)	-0.556*** (0.152)	-0.956 (0.705)	-0.452*** (0.126)	-0.556*** (0.203)	-0.235*** (0.000)	-0.377*** (0.121)	-0.238*** (0.000)
β_3^j	—	—	3.082** (1.512)	2.349*** (0.545)	3.794 (2.753)	2.07*** (0.492)	1.084*** (0.359)	0.461*** (0.000)	0.799*** (0.22)	0.626*** (0.000)
$\ln(r_{11})$	-4.487*** (0.149)	-1.969*** (0.220)	-4.431*** (0.143)	-2.237*** (0.171)	-4.46*** (0.15)	-7.869* (4.691)	-4.479*** (0.150)	-2.231*** (0.000)	-4.529*** (0.153)	-1.434*** (0.000)
$\ln(r_{22})$	-0.647*** (0.106)	-0.350*** (0.058)	-0.652*** (0.106)	-0.326*** (0.017)	—	—	-0.646*** (0.106)	-0.337*** (0.000)	-0.696*** (0.11)	-0.373*** (0.000)
$\ln(r_{22}^i)$	—	—	—	—	-0.92*** (0.13)	-1.954** (1.019)	—	—	—	—
$\ln(r_{22}^j)$	—	—	—	—	-0.336 (0.218)	-0.394*** (0.082)	—	—	—	—
$\ln(r_{33})$	-0.002 (0.048)	-0.275 (0.203)	—	—	—	—	-0.372*** (0.107)	-0.090*** (0.000)	-0.436*** (0.11)	-0.806*** (0.000)
$\ln(r_{33}^i)$	—	—	-0.874*** (0.114)	-0.513*** (0.037)	-0.95*** (0.12)	-1.096*** (0.139)	—	—	—	—
$\ln(r_{33}^j)$	—	—	-3.868*** (0.429)	-1.560*** (0.383)	-1.649*** (0.233)	-2.888 (2.688)	—	—	—	—
$\ln(r_{44})$	—	—	—	—	—	—	-0.707*** (0.000)	-0.619*** (0.000)	-0.232*** (0.000)	-0.537*** (0.000)

* $pvalue < 0.10$, ** $pvalue < 0.05$, *** $pvalue < 0.01$. Standard errors between ().

continued on next page

Table 3: Measurement equations estimated parameters

	DSSL/ DR	DSSL/ NDR	DSSMS/ E/DR	DSSMS/ E/NDR	DSSMS/ E&F/DR	DSSMS/ E&F/NDR	DSST/ E/DR	DSST/ E/NDR	DSST/ E&F/DR	DSST/ E&F/NDR
r_{21}	-	-0.074 (0.076)	-	-0.123 (0.098)	-	-	-	0.088*** (0.000)	-	0.554*** (0.000)
r_{21}^i	-	-	-	-	-	0.552*** (0.134)	-	-	-	-
r_{21}^j	-	-	-	-	-	0.096 (1.108)	-	-	-	-
r_{31}	-	0.650*** (0.191)	-	-	-	-	-	-0.133*** (0.000)	-	-0.742*** (0.000)
r_{31}^i	-	-	-	-0.114 (0.101)	-	0.131 (0.187)	-	-	-	-
r_{31}^j	-	-	-	0.109* (0.061)	-	0.199 (0.459)	-	-	-	-
r_{32}	-	-0.052 (0.115)	-	-	-	-	-	-0.162*** (0.000)	-	0.056*** (0.000)
r_{32}^i	-	-	-	-0.021 (0.057)	-	0.167* (0.099)	-	-	-	-
r_{32}^j	-	-	-	-0.178*** (0.059)	-	-0.28 (0.322)	-	-	-	-
r_{41}	-	-	-	-	-	-	-	0.077*** (0.000)	-	-0.015*** (0.000)
r_{42}	-	-	-	-	-	-	-	-0.040*** (0.000)	-	0.125*** (0.000)
r_{43}	-	-	-	-	-	-	-	-0.274*** (0.000)	-	0.408*** (0.000)

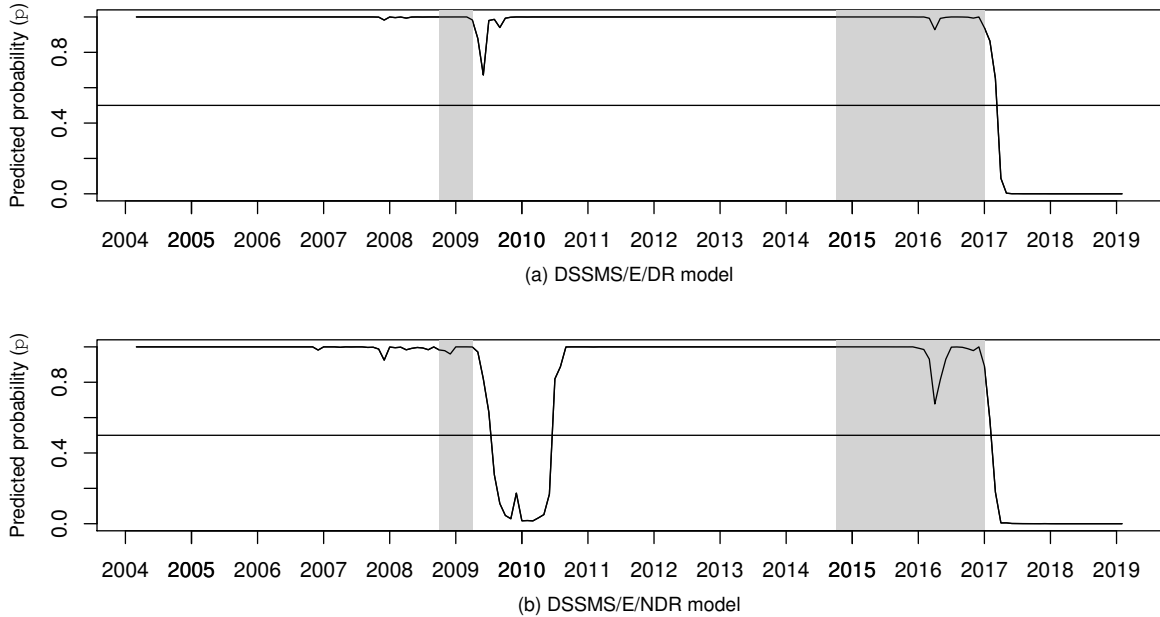
* $pvalue < 0.10$, ** $pvalue < 0.05$, *** $pvalue < 0.01$. Standard errors between ().

Focusing on β_2 , we highlight that all specifications provided a positive signal for this parameter. Notwithstanding, when we allowed for nonlinearities at this coefficient, we could not reject the hypothesis that there is no difference in the value of the parameter between the two regimes. The same is true for the observation matrix variance term r_{22} .

Then, evaluating our main concern, we verify that all nonlinear models (DSSMS and DSST) confirmed our initial expectation, with β_3 holding different signs across regimes. Furthermore, excluding the DSSMS/E&F/DR specification, all of them presented β_3 statistically different from zero at both regimes. Also, we highlight that the specifications DSSMS/E/DR and DSSMS/E/NDR registered an observation matrix variance term r_{33} that is statistically different across regimes. More precisely, at the regime i β_3 is negative (e.g., the parameter is -0.659 on the DSSMS/E/DR specification) and the variance is higher (e.g., $\exp(-0.874) > \exp(-3.868)$ on the DSSMS/E/DR specification).

Regarding the DSSMS models, we highlight that the transition probabilities (π_{ij} , $i, j = 1, 2$) are statistically different from zero at a 10% level on all specifications, another evidence that favors our thesis about the existence of nonlinearities on the data generating process. Also, we can see in Figures 1 and 2 the dynamics of the predicted probabilities. As a rule of thumb, if this probability gets above 0.5, we observe the regime i (linked to a negative β_3). Otherwise, the model points to regime j (linked to a positive β_3). Still on these charts, excluding the DSSMS/E/DR specification, the longer moments when regime j prevailed were in the aftermath of the two recessions observed at our sample: (i) between 2009 and 2010; and (ii) starting at the end 2016 and lasting till the end of our sample.

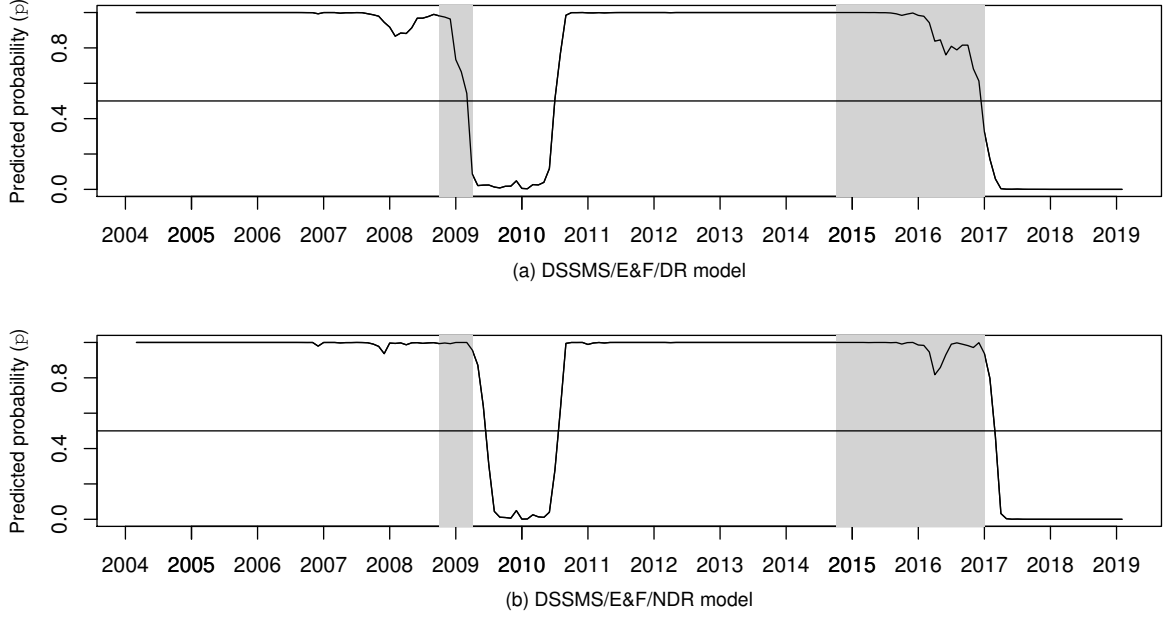
Figure 1: Predicted probability - (a) DSSMS/E/DR and (b) DSSMS/E/NDR



Note: Gray-shaded areas show periods of recession in Brazil according to CODACE/IBRE (2017).

Evaluating this result, as the economy faces a contraction phase, a positive β_1 and a negative β_3 show that as the observed loan measure decreases with the slow down, the index linked to riskier loans do not. We take this result as an evidence that at least part

Figure 2: Predicted probability - (a) DSSMS/E&F/DR and (b) DSSMS/E&F/NDR



Note: Gray-shaded areas show periods of recession in Brazil accordingly to CODACE/IBRE (2017).

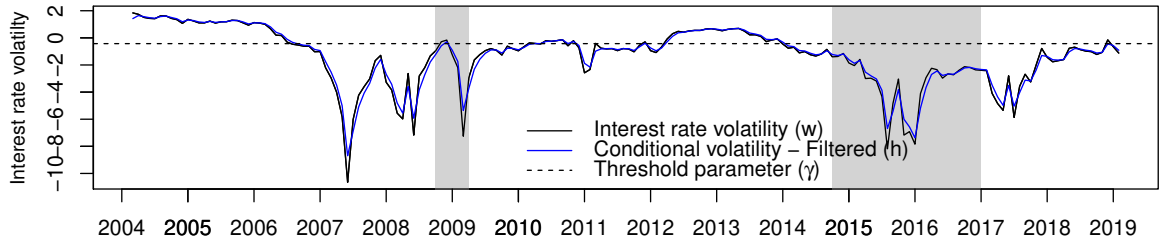
of this riskier demand is not being met by supply and the level of risk aversion is higher among lenders. In turn, as the economy changes to an expansion phase with both β_1 and β_3 positive, the demand for riskier loans is increasing with the observed loan measure and we cannot rule out that at least part of this demand is attained by supply, an indication that the level of risk aversion is lower among lenders. In other words, given some conditions about the business cycle, the changes in β_3 sign (the coefficient linked to the riskier type of loan) across regimes would inform a specific supply shift related to lenders risk-appetite over time. Following our thesis, this means that these recovery moments were marked by a synchronized increase in both supply and demand of credit, with a higher bank sector risk appetite. But, in most part of the sample, the regime i prevailed, showing that our sample period was dominated by a more risk averse position of banks and desynchronized supply/demand forces acting on credit market equilibrium.

Taking into account the DSST, in a general way, the Figures 3 and 4 present two more well defined moments of higher interest rate volatility: (i) from the beginning of the sample to the end of 2006; (ii) from the beginning of 2012 to the beginning of 2014. In turn, moments of lower volatility can be identified both preceding and in the aftermath of the recessions comprising our sample (2009 and 2015/2016).

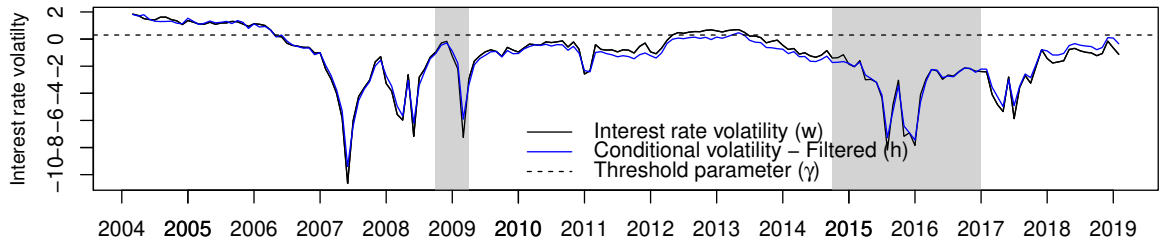
Considering the way we defined the threshold, at moments of higher volatility ($h_{t-1} > \gamma$, a condition that we associate to supply shifts) we identify the regime i and there is a negative correlation between the *empréstimo* search term and the common factor ($\beta_3 < 0$). Otherwise, the model points to the regime j (linked to $\beta_3 > 0$, a condition that we associate to lower supply shifts). This result goes at the same direction of the coefficients and variances related to the same web search term in the Markov-switching models.

As a robustness check, Table 4 details the nonlinear models' regime classification. In

Figure 3: Interest rate volatility vs Threshold - (a) DSST/E/DR and (b) DSST/E/NDR



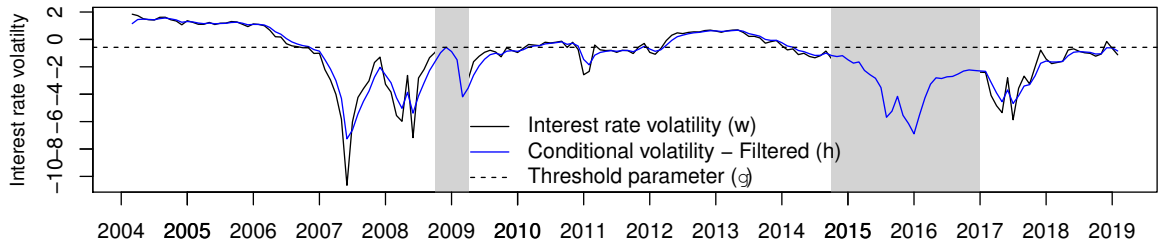
(a) DSST/E/DR



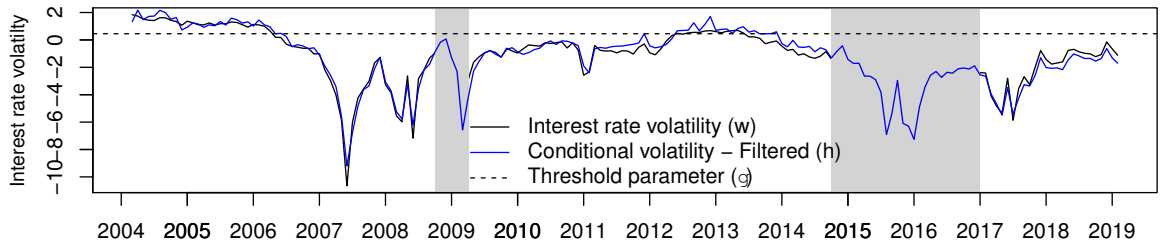
(b) DSST/E/NDR

Note: Gray-shaded areas show periods of recession in Brazil accordingly to CODACE/IBRE (2017).

Figure 4: Interest rate volatility vs Threshold - (a) DSST/E&F/DR and (b) DSST/E&F/NDR



(a) DSST/E&F/DR



(b) DSST/E&F/NDR

Note: Gray-shaded areas show periods of recession in Brazil accordingly to CODACE/IBRE (2017).

turn, Table 5 shows how the regimes overlapped at both nonlinear models. Here, our goal is to evaluate how often the baseline (DSSMS) and alternative (DSST) approaches pointed to the same regime. The higher success rate was reached when comparing the DSST/E&F/NDR with the DSSMS/E&F/DR (66,1%). Then, despite the high dispersion, some DSST models were able to get the same regime that the DSSMS models on the most part of our sample.

Finally, we provide a possible interpretation and discuss the evolution of our state variable, that is the common trend among three variables: (i) nonearmarked new operations of personal credit, capturing information about supply and demand; (ii) Google Trends Index for the *financiamiento* search term, capturing information about the demand of less risky loans; (iii) and Google Trends Index for the *empréstimo* search term, capturing information about the demand of riskier loans. Considering our results, while the first two variables presented a positive correlation with the state variable during all the sample, the correlation sign between the *empréstimo* search term and the state variable changed across two regimes. Taking the prevailing dynamics of the DSSMS models as reference, at periods of normality and recession, the correlation between the state variable and the riskier demand index is negative; once recessions are over and an expansion moment begins, the correlation between the state variable and the riskier demand index becomes positive. In this sense, when supplemented with information about the actual regime, the state variable could be viewed as a credit market condition index that carries information about loans' demand quality.

Table 4: Regime classification.

Model	Regime i ($\beta_3 < 0$)			Regime j ($\beta_3 > 0$)		
	Period	Months	% of sample	Period	Months	% of sample
DSSMS/E/DR	Jan/04 to Mar/17	159	86.9%	Apr/17 to Mar/19	24	13.1%
DSSMS/E/NDR	Jan/04 to Jul/09 & Jul/10 to Feb/17	147	80.3%	Aug/09 to Jun/10 & Mar/17 to Mar/19	36	19.7%
DSSMS/E&F/DR	Jan/04 to Mar/09 & Jul/10 to Dec/16	141	77.0%	Apr/09 to Jun/10 & Jan/17 to Mar/19	42	23.0%
DSSMS/E&F/NDR	Jan/04 to Jun/09 & Aug/10 to Feb/17	145	79.2%	Jul/09 to Jul/10 & Mar/17 to Mar/19	38	20.8%
DSST/E/DR	Jan/04 to Aug/06 & Dec/08 & Apr/10 to Nov/10 & Dec/11 & Apr/12 to Jan/14	64	35.0%	Sep/06 to Nov/08 & Jan/09 to Mar/10 & Dec/10 to Nov/11 & Jan/12 to Mar/12 & Feb/14 to Mar/19	119	65.0%
DSST/E/NDR	Jan/04 to Apr/06 & Apr/13 to May/13	30	16.4%	May/06 to Mar/13 & Jun/13 to Mar/19	153	83.6%
DSST/E&F/DR	Jan/04 to Nov/06 & Mar/10 to Dec/10 & Dec/11 & Apr/12 to Mar/14	70	38.3%	Dec/06 to Feb/10 & Jan/11 to Nov/11 & Jan/12 to Mar/12 & Apr/14 to Mar/19	113	61.7%
DSST/E&F/NDR	Jan/04 to Oct/06 & Dec/06 & Nov/08 to Dec/08 & Dec/09 & Jun/10 to Dec/10 & Apr/11 & Jun/11 to Jan/12 & Mar/12 to Jun/14 & Aug/14 & Dec/14	84	45.9%	Nov/06 & Jan/07 to Oct/08 & Jan/09 to Nov/09 & Jan/10 to May/10 & Jan/11 to Mar/11 & May/11 & Feb/12 & Jul/14 & Sep/14 to Nov/14 & Jan/15 to Mar/19	99	54.1%

Table 5: Overlapping regimes (% of total sample)

Model	DSSMS/ E/DR	DSSMS/ E/NDR	DSSMS/ E&F/DR	DSSMS/ E&F/NDR	DSST/ E/DR	DSST/ E/NDR	DSST/ E&F/DR	DSST/ E&F/NDR
DSSMS/ E/DR	100.0%	93.3%	90.0%	92.2%	47.2%	28.3%	50.6%	58.3%
DSSMS/ E/NDR	93.3%	100.0%	96.7%	98.9%	50.6%	35.0%	52.8%	62.8%
DSSMS/ E&F/DR	90.0%	96.7%	100.0%	96.7%	53.9%	38.3%	56.1%	66.1%
DSSMS/ E&F/NDR	92.2%	98.9%	96.7%	100.0%	50.6%	36.1%	52.8%	62.8%
DSST/ E/DR	47.2%	50.6%	53.9%	50.6%	100.0%	81.1%	95.6%	86.7%
DSST/ E/NDR	28.3%	35.0%	38.3%	36.1%	81.1%	100.0%	77.8%	70.0%
DSST/ E&F/DR	50.6%	52.8%	56.1%	52.8%	95.6%	77.8%	100.0%	87.8%
DSST/ E&F/NDR	58.3%	62.8%	66.1%	62.8%	86.7%	70.0%	87.8%	100.0%

5.2 In-sample and out-of-sample evaluation

In this subsection, we evaluate and compare the models performance at both in-sample and out-of-sample perspectives. On the former perspective, our goal is to understand if the nonlinearities extensions help to get closer to the credit market data generating process. As a complementary test, in the latter perspective we assess the models' forecast accuracy.

Once the models are non-nested, for the in-sample analysis we perform the Vuong test (Vuong 1989).¹⁵ On the null hypothesis of this test, the competing models are equally close to the true data generating process. All the tests are made on pairs, where we compare the nonlinear (first model) specification against the linear one (second model). In a simpler way, we can interpret the test as follows: (i) when the test statistic is positive and the null hypothesis is rejected, the first model is closer to the data generating process; (ii) when the test statistic is negative and the null hypothesis is rejected, the second model is the best one; (iii) if we do not reject the null hypothesis, there is no statistical difference between the models. Table 6 provides the Vuong test results.

We observe that all tests rejected the null hypothesis. But, while the statistics were positive for the nonlinear Markov-switching models, the same is not true for the threshold specifications. Then, the Vuong test shows that, comparing with the linear specification, the DSSMS models are the closest to the data generating process. This evidence favors our assumption that the nonlinearities linked to the Google search data could be useful to evaluate the credit market developments in Brazil.

An additional advantage of the SS models is that they allow us to forecast the observed variables. Therefore, we evaluate the SS models performance among them and also against two additional naive models: a random walk and an AR(1) equation.

The Table 7 presents the root mean square error (RMSE) for each model from the first to the twelfth month ahead. The out-of-sample forecasts comprise the last 48 monthly observations, evaluated on an expanding window basis (from April 2015 to March 2019). The significance probability values (*pvalues*) for the Diebold-Mariano (DM) test (Diebold and Mariano 1995), which addresses the null hypothesis of no difference between two models, are also provided. Besides the naive ones, this test is presented against the DSSL/DR model.

The best model for each one of the forecasting horizons were marked in bold. The DSSMS models were the ones that stand out as the best on out-of-sample accuracy in most of forecast horizons. Anyway, applying the DM test, the differences against the naive models were statistically significant at a 10% level just for the one month ahead forecast, where the DSSMS/E&F/DR was the one with lower forecast error. Evaluating these results, the DSSMS models seem to be the best ones, specially on shorter forecast horizons. In brief, these findings highlight the benefits obtained with the introduction of nonlinearities on models.

Once more, we consider the improvement reached at the out-of-sample forecasts when allowed for the underlying nonlinearities as another evidence favoring the thesis that these nonlinear features could be useful to get closer to the data generating process associated with credit market. We must highlight that all this was accomplished with a parsimonious model, exploring a set of three variables in case of DSSMS model.

¹⁵More formally, according to the author, the test allows to evaluate if the distributions in the competing models closest to the data generating process are observationally identical.

Table 6: In-sample fit comparisons: Vuong test statistics.

		Linear model (DSSL)	
		Diagonal covariance matrix (DR)	Non-diagonal covariance matrix (NDR)
Markov-Switch - <i>empréstimo</i> term	Diagonal	(6.476)***	(5.831)***
	Covariance	[6.219]***	[5.769]***
	Matrix (DR)	<5.809>***	<5.671>***
(DSSMS/E)	Non-diagonal	(7.060)***	(6.419)***
	Covariance	[6.431]***	[6.013]***
	Matrix (NDR)	<5.426>***	<5.364>***
Markov-Switch - <i>empréstimo</i> and <i>financiamento</i> terms	Diagonal	(6.984)***	(6.347)***
	Covariance	[6.553]***	[6.138]***
	Matrix (DR)	<5.865>***	<5.804>***
(DSSMS/E&F)	Non-diagonal	(4.577)***	(4.110)***
	Covariance	[3.830]***	[3.560]***
	Matrix (NDR)	<2.636>***	<2.682>***
Threshold model - <i>empréstimo</i> term	Diagonal	(-12.029)***	(-12.233)***
	Covariance	[-12.310]***	[-12.372]***
	Matrix (DR)	<-12.759>***	<-12.595>***
(DSST/E)	Non-diagonal	(-12.865)***	(-13.190)***
	Covariance	[-13.433]***	[-13.617]***
	Matrix (NDR)	<-14.340>***	<-14.298>***
Threshold model - <i>empréstimo</i> and <i>financiamento</i> terms	Diagonal	(-11.174)***	(-11.427)***
	Covariance	[-11.481]***	[-11.602]***
	Matrix (DR)	<-11.970>***	<-11.881>***
(DSST/E&F)	Non-diagonal	(-13.657)***	(-13.950)***
	Covariance	[-14.267]***	[-14.420]***
	Matrix (NDR)	<-15.241>***	<-15.169>***

Notes: Unadjusted statistic between (); AIC-adjusted statistic between []; and BIC-adjusted statistic between < >. * *pvalue* < 0.10, ** *pvalue* < 0.05, *** *pvalue* < 0.01

Table 7: Out-of-sample forecasting comparisons: RMSEs* and DM test**

Model	Forecast horizons					
	1 month ahead	2 months ahead	3 months ahead	4 months ahead	5 months ahead	6 months ahead
RW	0.176	0.166	0.167	0.177	0.166	0.167
AR(1)	0.177	0.176	0.175	0.162	0.156	0.157
DSSL/DR	0.148 (0.056) [0.043]	0.161 (0.801) [0.601]	0.184 (0.093) [0.762]	0.146 (0.199) [0.117]	0.165 (0.958) [0.400]	0.187 (0.210) [0.095]
DSSL/NDR	0.281 (0.0002) [0.0002] ⟨5.3e-7⟩	0.30 (0.004) [0.016] ⟨0.001⟩	0.324 (0.011) [0.037] ⟨0.014⟩	0.275 (0.114) [0.063] ⟨0.025⟩	0.294 (0.089) [0.094] ⟨0.073⟩	0.318 (0.088) [0.106] ⟨0.123⟩
DSSMS/E/DR	0.145 (0.034) [0.026] ⟨0.465⟩	0.148 (0.288) [0.355] ⟨0.020⟩	0.166 (0.936) [0.756] ⟨0.827⟩	0.147 (0.167) [0.041] ⟨0.838⟩	0.148 (0.316) [0.618] ⟨0.013⟩	0.167 (0.997) [0.496] ⟨0.036⟩
DSSMS/E/NDR	0.157 (0.280) [0.250] ⟨0.220⟩	0.173 (0.708) [0.933] ⟨0.327⟩	0.193 (0.215) [0.584] ⟨0.845⟩	0.153 (0.361) [0.709] ⟨0.429⟩	0.169 (0.868) [0.398] ⟨0.807⟩	0.189 (0.286) [0.145] ⟨0.905⟩
DSSMS/E&F/DR	0.143 (0.026) [0.019] ⟨0.047⟩	0.153 (0.391) [0.421] ⟨0.020⟩	0.173 (0.539) [0.933] ⟨0.706⟩	0.144 (0.133) [0.034] ⟨0.793⟩	0.154 (0.409) [0.976] ⟨0.055⟩	0.173 (0.637) [0.235] ⟨0.071⟩
DSSMS/E&F/NDR	0.168 (0.653) [0.62] ⟨0.236⟩	0.17 (0.779) [0.869] ⟨0.508⟩	0.189 (0.182) [0.631] ⟨0.279⟩	0.166 (0.675) [0.673] ⟨0.309⟩	0.171 (0.798) [0.4309] ⟨0.760⟩	0.189 (0.263) [0.130] ⟨0.843⟩
DSST/E/DR	0.146 (0.036) [0.027] ⟨0.338⟩	0.154 (0.473) [0.453] ⟨0.006⟩	0.175 (0.336) [0.986] ⟨0.016⟩	0.147 (0.158) [0.181] ⟨0.788⟩	0.156 (0.472) [0.851] ⟨0.013⟩	0.175 (0.413) [0.246] ⟨0.110⟩
DSST/E/NDR	0.183 (0.720) [0.728] ⟨0.036⟩	0.20 (0.131) [0.461] ⟨0.079⟩	0.223 (0.049) [0.239] ⟨0.154⟩	0.176 (0.984) [0.398] ⟨0.251⟩	0.193 (0.269) [0.145] ⟨0.341⟩	0.218 (0.122) [0.115] ⟨0.420⟩
DSST/E&F/DR	0.146 (0.038) [0.029] ⟨0.350⟩	0.153 (0.440) [0.445] ⟨0.01⟩	0.178 (0.281) [0.917] ⟨0.150⟩	0.148 (0.183) [0.241] ⟨0.587⟩	0.154 (0.479) [0.893] ⟨0.010⟩	0.180 (0.315) [0.165] ⟨0.102⟩
DSST/E&F/NDR	0.185 (0.690) [0.697] ⟨0.031⟩	0.196 (0.194) [0.538] ⟨0.133⟩	0.220 (0.062) [0.270] ⟨0.217⟩	0.177 (0.978) [0.396] ⟨0.257⟩	0.189 (0.450) [0.287] ⟨0.523⟩	0.213 (0.184) [0.137] ⟨0.525⟩

Notes: *Calculated with expanding window for last 48 months of the sample.

**DM test pvalues: RW between (); AR(1) between []; and DSSL/DR between ⟨⟩.

continued on next page

Table 7: Out-of-sample forecasting comparisons: RMSEs* and DM test**

Model	Forecast horizons					
	7 months ahead	8 months ahead	9 months ahead	10 months ahead	11 months ahead	12 months ahead
RW	0.167	0.167	0.177	0.166	0.167	0.167
AR(1)	0.156	0.201	0.272	0.253	0.255	0.279
DSSL/DR	0.188 (0.253) [0.082]	0.190 (0.241) [0.643]	0.151 (0.108) [0.006]	0.178 (0.517) [0.198]	0.198 (0.331) [0.286]	0.202 (0.344) [0.175]
DSSL/NDR	0.316 (0.118) [0.114] ⟨0.168⟩	0.315 (0.146) [0.300] ⟨0.213⟩	0.266 (0.338) [0.372] ⟨0.201⟩	0.287 (0.251) [0.83] ⟨0.296⟩	0.311 (0.217) [0.941] ⟨0.344⟩	0.310 (0.237) [0.522] ⟨0.386⟩
DSSMS/E/DR	0.168 (0.978) [0.558] ⟨0.007⟩	0.169 (0.953) [0.003] ⟨1.3e-8⟩	0.153 (0.037) [0.008] ⟨0.827⟩	0.152 (0.408) [0.073] ⟨ 2e-16 ⟩	0.172 (0.894) [0.156] ⟨0.000⟩	0.173 (0.879) [0.096] ⟨7e-6⟩
DSSMS/E/NDR	0.188 (0.266) [0.091] ⟨0.995⟩	0.187 (0.269) [0.409] ⟨0.887⟩	0.150 (0.084) [0.003] ⟨0.845⟩	0.166 (0.975) [0.062] ⟨0.455⟩	0.185 (0.362) [0.069] ⟨0.596⟩	0.185 (0.395) [0.007] ⟨0.538⟩
DSSMS/E&F/DR	0.174 (0.668) [0.280] ⟨0.034⟩	0.175 (0.674) [0.003] ⟨0.010⟩	0.149 (0.001) [0.009] ⟨ 0.705 ⟩	0.158 (0.563) [0.076] ⟨0.001⟩	0.177 (0.721) [0.145] ⟨0.041⟩	0.178 (0.728) [0.074] ⟨0.024⟩
DSSMS/E&F/NDR	0.191 (0.285) [0.220] ⟨0.882⟩	0.192 (0.311) [0.72] ⟨0.908⟩	0.173 (0.839) [0.019] ⟨0.278⟩	0.182 (0.297) [0.056] ⟨0.865⟩	0.200 (0.399) [0.383] ⟨0.950⟩	0.203 (0.395) [0.266] ⟨0.947⟩
DSST/E/DR	0.175 (0.487) [0.161] ⟨0.107⟩	0.175 (0.515) [9e-5] ⟨0.111⟩	0.150 (0.040) [0.008] ⟨0.817⟩	0.160 (0.682) [0.067] ⟨0.032⟩	0.176 (0.610) [0.048] ⟨0.218⟩	0.177 (0.622) [0.012] ⟨0.227⟩
DSST/E/NDR	0.217 (0.120) [0.085] ⟨0.490⟩	0.216 (0.130) [0.652] ⟨0.561⟩	0.172 (0.890) [0.001] ⟨0.478⟩	0.191 (0.161) [0.007] ⟨0.686⟩	0.215 (0.069) [0.233] ⟨0.726⟩	0.215 (0.044) [0.079] ⟨0.779⟩
DSST/E&F/DR	0.181 (0.359) [0.162] ⟨2.5e-06⟩	0.182 (0.377) [0.002] ⟨0.221⟩	0.154 (0.072) [0.011] ⟨0.722⟩	0.159 (0.646) [0.068] ⟨6e-6⟩	0.186 (0.474) [0.142] ⟨0.261⟩	0.187 (0.490) [0.070] ⟨0.270⟩
DSST/E&F/NDR	0.212 (0.213) [0.137] ⟨0.608⟩	0.210 (0.218) [0.781] ⟨0.680⟩	0.171 (0.876) [0.004] ⟨0.536⟩	0.182 (0.600) [0.010] ⟨0.936⟩	0.206 (0.183) [0.182] ⟨0.880⟩	0.206 (0.158) [0.056] ⟨0.944⟩

Notes: *Calculated with expanding window for last 48 months of the sample.

**DM test pvalues: RW between (); AR(1) between []; and DSSL/DR between ⟨⟩.

6 Final Remarks

In this paper, aiming to evaluate potential nonlinearities linked to credit developments in Brazil, we applied SS models to combine household loans time series with Google Trends Indexes collected for the strings *empréstimo* and *financiamento*. Basically, our set of models comprises extensions of the traditional linear SS specification to deal with the underlying nonlinearities: Markov-switching and Threshold models. All specifications were evaluated with diagonal and non-diagonal observation covariance matrix.

As main contributions, we show that nonlinear features on the relationship among these variables and the household loans, such as Markov-switching and threshold point, can be explored in two ways. First, we propose identification strategies to disentangle the supply and demand forces which drive the credit market to equilibrium over time. An advantage of our identification strategies is that, while the procedures at previous literature usually explore data that is not fully available to public or that exist just at some countries, our method is based on a freely available data and could be reproduced in all regions covered by Google. Second, we show that the underlying nonlinearities improve the performance of SS models on both in-sample and out-of-sample perspectives, with the Markov-switching SS representation standing out against the other specifications on both aspects.

We interpret our findings as evidences that the Google web searches could be useful to identify the data generating process linked to credit market, exploring the information content of terms linked to loan dynamics at a country. And the use goes beyond the usual nowcasting and forecasting, once we can apply this data to help on other complexes tasks, as disentangling the credit supply and demand or identifying more precisely the credit market condition in the economy throughout time.

References

- Annibal, C. and S. Koyama (2011). Pesquisa Trimestral de Condições de Crédito no Brasil. Working Papers Series 245, Central Bank of Brazil, Research Department.
- Bernanke, B. (1993). Credit in the macroeconomy. *Quarterly Review* (Spr), 50–70.
- Bernanke, B. and A. Blinder (1988). Credit, money, and aggregate demand. *American Economic Review* 78(2), 435–39.
- Bernanke, B. and A. Blinder (1992). The federal funds rate and the channels of monetary transmission. *American Economic Review* 82(4), 901–21.
- Bernanke, B. S. and M. Gertler (1995). Inside the black box: The credit channel of monetary policy transmission. *The Journal of Economic Perspectives* 9(4), 27–48.
- Bester, H. and M. Hellwig (1987). *Moral Hazard and Equilibrium Credit Rationing: An Overview of the Issues*, pp. 135–166. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Bulut, L. (2017, 11). Google trends and the forecasting performance of exchange rate models. *Journal of Forecasting* 37.
- Burdeau, E. and E. Kintzler (2017, 11). Assessing the use of google trends to predict credit developments. *61th ISI World Statistics Congress*.

- Calza, A., C. Gartner, and J. Sousa (2003). Modelling the demand for loans to the private sector in the euro area. *Applied Economics* 35(1), 107–117.
- Carriere-Swallow, Y. and F. Labbé (2011). Nowcasting with google trends in an emerging market. *Journal of Forecasting*.
- Castelnuovo, E. and T. D. Tran (2017). Google it up! a google trends-based uncertainty index for the united states and australia. *Economics Letters* 161(C), 149–153.
- Chan, K. S. (1991). Percentage points of likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)* 53(3), 691–696.
- Chen, T., E. P. K. So, L. Wu, and I. K. M. Yan (2015). The 2007–2008 u.s. recession: What did the real-time google trends data tell the united states? *Contemporary Economic Policy* 33(2), 395–403.
- Choi, H. and H. Varian (2009a). Predicting initial claims for unemployment benefits. *Google technical report*.
- Choi, H. and H. Varian (2009b). Predicting the present with google trends. *Google technical report*.
- Cipollini, A. and F. Parla (2018). Credit demand and supply shocks in italy during the great recession. *Applied Economics* 50(53), 5795–5813.
- CODACE/IBRE (2017). *Comunicado de Datação de Ciclos Mensais Brasileiros*. Comitê de Datação de Ciclos Econômicos - Instituto Brasileiro de Economia - Fundação Getulio Vargas.
- Coelho, C., J. M. P. Mello, M. Garcia, and R. Rigobon (2017). A method for identifying aggregate credit supply and demand parameters using heteroskedascity an application for brazil.
- Cosslett, S. and L.-F. Lee (1985, 01). Serial correlation in latent variable models. *Journal of Econometrics* 27, 79–97.
- Del Giovane, P., G. Eramo, and A. Nobili (2011). Disentangling demand and supply in credit developments: A survey-based analysis for italy. *Journal of Banking and Finance* 35(10), 2719–2732.
- Del Giovane, P., A. Nobili, and F. M. Signoretti (2017, June). Assessing the Sources of Credit Supply Tightening: Was the Sovereign Debt Crisis Different from Lehman? *International Journal of Central Banking* 13(2), 197–234.
- Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13(3), 253–63.
- Dietzel, M. A. (2016, 03). Sentiment-based predictions of housing market turning points with google trends. *International Journal of Housing Markets and Analysis* 9, 108–136.
- Einav, L. and J. Levin (2013). The data revolution and economic analysis. In *Innovation Policy and the Economy, Volume 14*, pp. 1–24. National Bureau of Economic Research.
- Fecht, F., S. Thum, and P. Weber (2019, 05). Fear, deposit insurance schemes, and deposit reallocation in the german banking system. *Journal of Banking & Finance* 105.

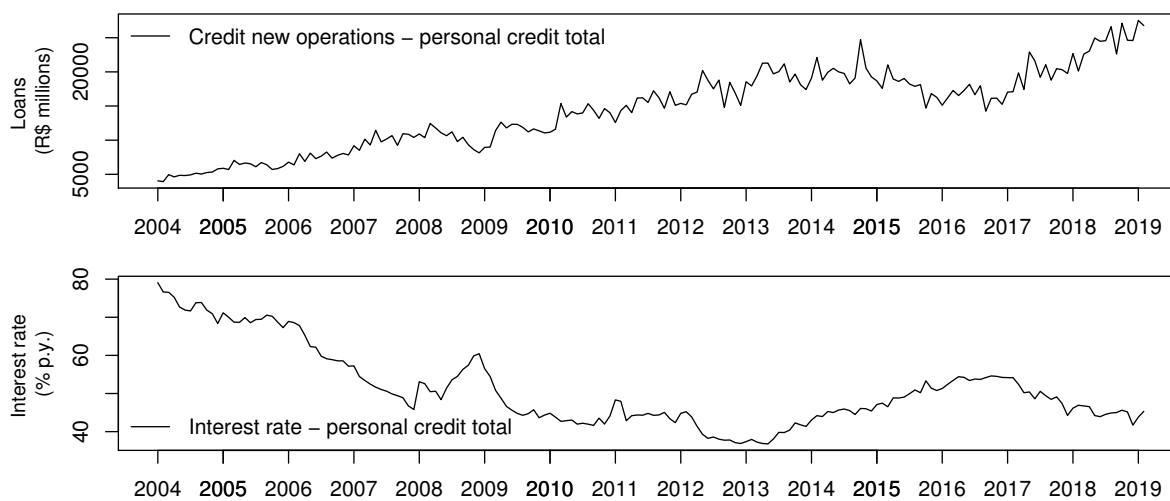
- Freixas, X. and J. Rochet (2008). *Microeconomics of Banking, 2nd Edition* (2 ed.), Volume 1. The MIT Press.
- Gertler, M. and S. Gilchrist (1993). The role of credit market imperfections in the monetary transmission mechanism: Arguments and evidence. *Scandinavian Journal of Economics* 95(1), 43–64.
- Goldfeld, S. and R. Quandt (1973, 02). A markov model for switching regressions. *Journal of Econometrics* 1, 3–15.
- Hamilton, J. (1989, 02). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–84.
- Hamilton, J. (1994). *Time Series Analysis* (1 ed.). Princeton University Press.
- Hand, C. and G. Judge (2012, 07). Searching for the picture: forecasting uk cinema admissions using google trends data. *Applied Economics Letters* 19, 1051–1055.
- Hansen, B. (2000). Sample splitting and threshold estimation. *Econometrica* 68(3), 575–604.
- Hassani, H. and E. Silva (2016, 12). Forecasting energy data with a time lag into the future and google trends. *International Journal of Energy and Statistics* 4, 1650020.
- Hu, H., L. Tang, S. Zhang, and H. Wang (2018, 04). Predicting the direction of stock markets using optimized neural networks with google trends. *Neurocomputing* 285.
- Jaffee, D. M. and T. Russell (1976). Imperfect information, uncertainty, and credit rationing. *The Quarterly Journal of Economics* 90(4), 651–666.
- Jimenez, G., S. Ongena, J.-L. Peydro, and J. Saurina (2012, August). Credit Supply and Monetary Policy: Identifying the Bank Balance-Sheet Channel with Loan Applications. *American Economic Review* 102(5), 2301–2326.
- Jun, S.-P., H. S. Yoo, and S. Choi (2018). Ten years of research change using google trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change* 130, 69 – 87.
- Kashyap, A. and J. Stein (1994). Monetary policy and bank lending. In *Monetary Policy*, pp. 221–261. National Bureau of Economic Research, Inc.
- Keenan, D. M. (1985). A tukey nonadditivity-type test for time series nonlinearity. *Biometrika* 72(1), 39–44.
- Kick, T., T. Pausch, and B. Ruprecht (2015). The winner’s curse: Evidence on the danger of aggressive credit growth in banking. Technical report.
- Koller, W. and M. Fischer (2002, 09). Testing for non-linear dependence in univariate time series: An empirical investigation of the austrian unemployment rate. *Networks and Spatial Economics* 2, 191–209.
- Lanne, M., H. Lütkepohl, and K. Maciejowska (2010). Structural vector autoregressions with markov switching. *Journal of Economic Dynamics and Control* 34(2), 121 – 131.
- Lo, M. C. and E. Zivot (2001). Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics* 5(4), 533–576.

- McLaren, N. and R. Shanbhogue (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin* 51(2), 134–140.
- McLeod, I. and W. Li (1983, 07). Diagnostic checking arma time series models using squared-residual autocorrelation. *Journal of Time Series Analysis* 4, 269–273.
- Onder, I. (2017, 08). Forecasting tourism demand with google trends: Accuracy comparison of countries versus cities. *International Journal of Tourism Research*.
- Piger, J. (2007, 08). Econometrics: Models of regime changes. *Complex Systems in Finance and Econometrics*.
- Preis, T., H. S. Moat, and H. E. Stanley (2013). Quantifying trading behavior in financial markets using google trends. *Scientific Reports*.
- Rigobon, R. (2003). Identification through heteroskedasticity. *The Review of Economics and Statistics* 85(4), 777–792.
- Rivera, R. (2016, 12). A dynamic linear model to forecast hotel registrations in puerto rico using google trends data. *Tourism Management* 57.
- Stiglitz, J. and A. Weiss (1981, 02). Credit rationing in markets with imperfect information. *American Economic Review* 71, 393–410.
- Suzuki, T. (2004, 02). Credit channel of monetary policy in japan: Resolving the supply versus demand puzzle. *Applied Economics* 36, 2385–2396.
- Tabak, B. M., G. M. Gomes, and M. da Silva Medeiros (2015). The impact of market power at bank level in risk-taking: The brazilian case. *International Review of Financial Analysis* 40, 154 – 165.
- Teräsvirta, T., C.-F. Lin, and C. W. J. Granger (1993). Power of the neural network linearity test. *Journal of Time Series Analysis* 14(2), 209–220.
- Tong, H. (1983, 01). *Threshold Model in Non-Linear Time Series Analysis*, Volume 21.
- Tong, H. (2010, 04). Threshold models in time series analysis—30 years on. *Statistics and Its Interface* 2.
- Tsay, R. S. (1986, 08). Nonlinearity tests for time series. *Biometrika* 73(2), 461–466.
- Ubiergo, A. S. (2012). The puzzle of brazil’s high interest rates. IMF Working Papers 12/62, International Monetary Fund.
- Vosen, S. and T. Schmidt (2011). Forecasting private consumption: survey-based indicators vs. google trends. *Journal of Forecasting* 30(6), 565–578.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57(2), 307–333.
- Williamson, S. (1987). Costly monitoring, loan contracts, and equilibrium credit rationing. *The Quarterly Journal of Economics* 102(1), 135–145.
- Yu, L., Y. Zhao, L. Tang, and Z. Yang (2018, 01). Online big data-driven oil consumption forecasting with google trends. *International Journal of Forecasting*.
- Zeybek, O. and E. Ugurlu (2015, 03). Nowcasting credit demand in turkey with google trends data. *Journal of Applied Economic Sciences* 10, 293–300.

7 Appendix

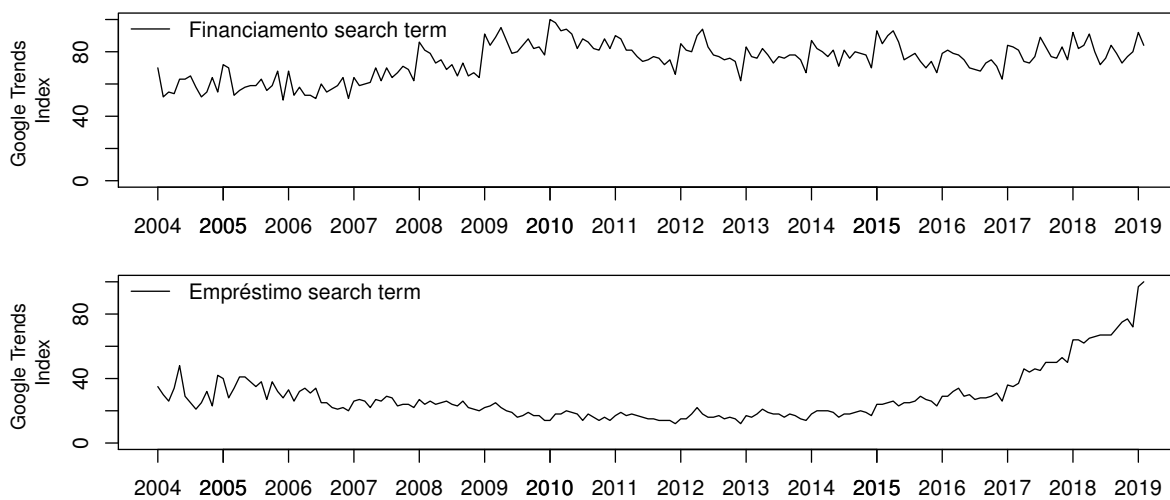
7.1 Data

Figure 5: Household loans and interest rate (personal credit).



Source: Central Bank of Brazil.

Figure 6: Google Trends Indexes - Financiamento and Empréstimo.



Source: Google.

7.2 Estimated common factors (x_t)

Figure 7: Common factor - (a) DSSL/DR and (b) DSSL/NDR.

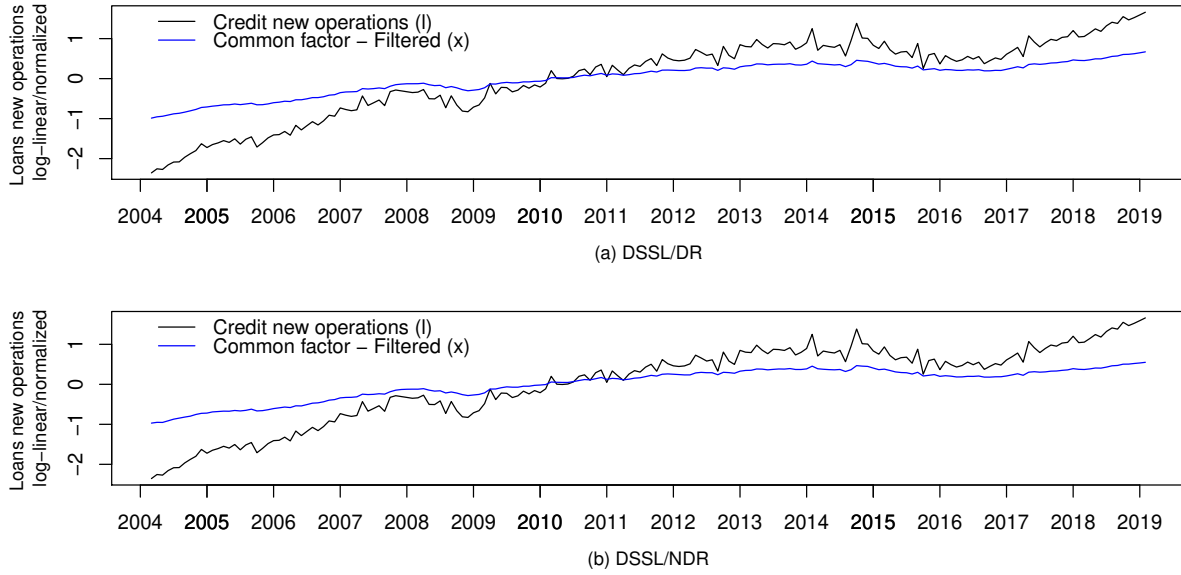


Figure 8: Common factor - (a) DSSMS/E/DR and (b) DSSMS/E/NDR.

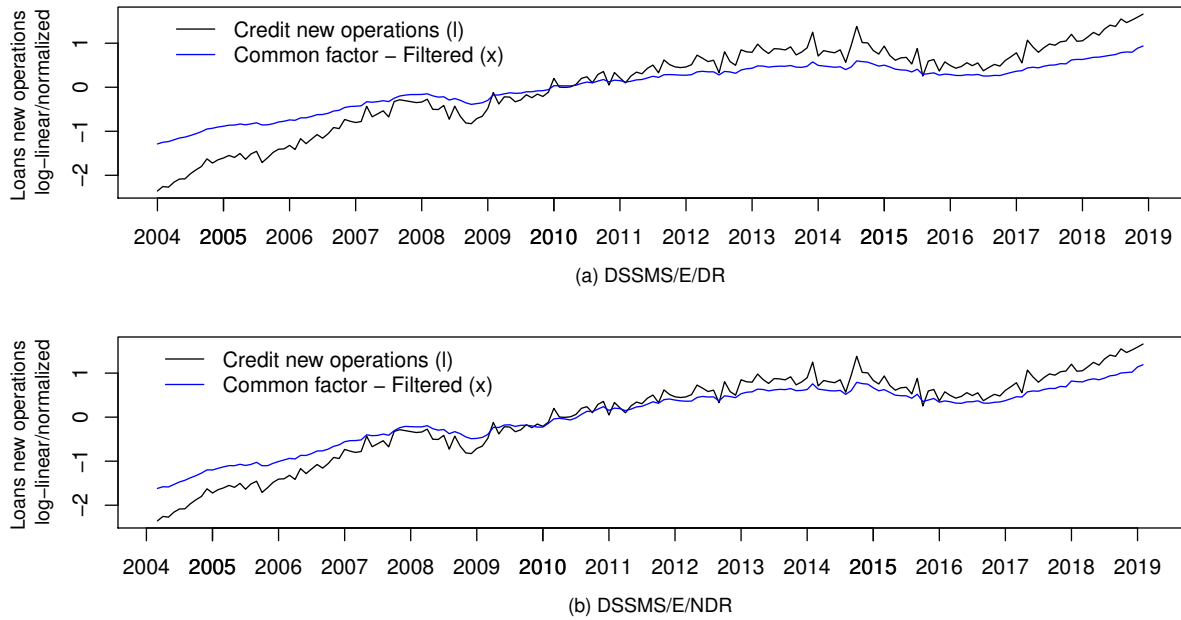
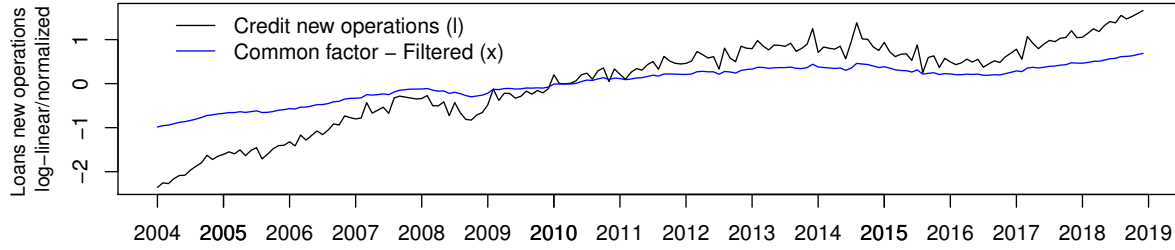
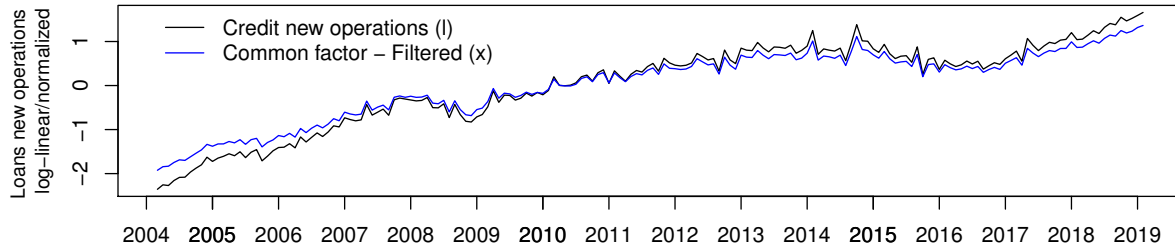


Figure 9: Common factor - (a) DSSMS/E&F/DR and (b) DSSMS/E&F/NDR.

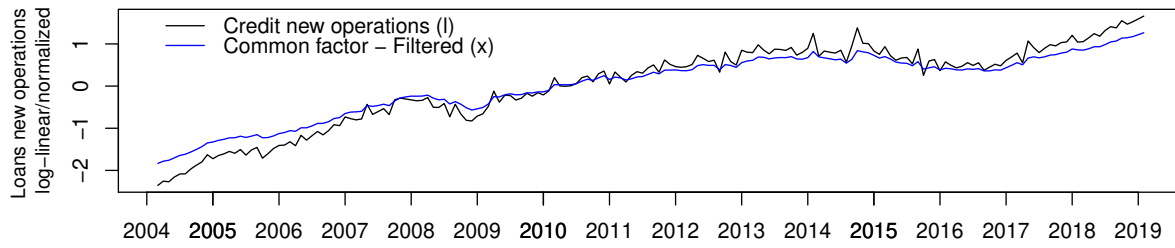


(a) DSSMS/E&F/DR

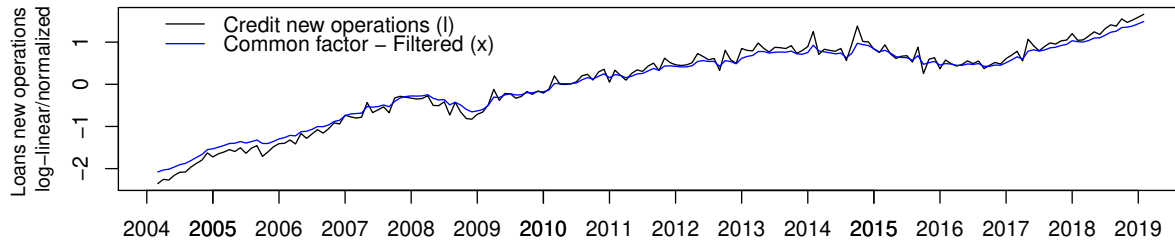


(b) DSSMS/E&F/NDR

Figure 10: Common factor - (a) DSST/E/DR and (b) DSST/E/NDR.

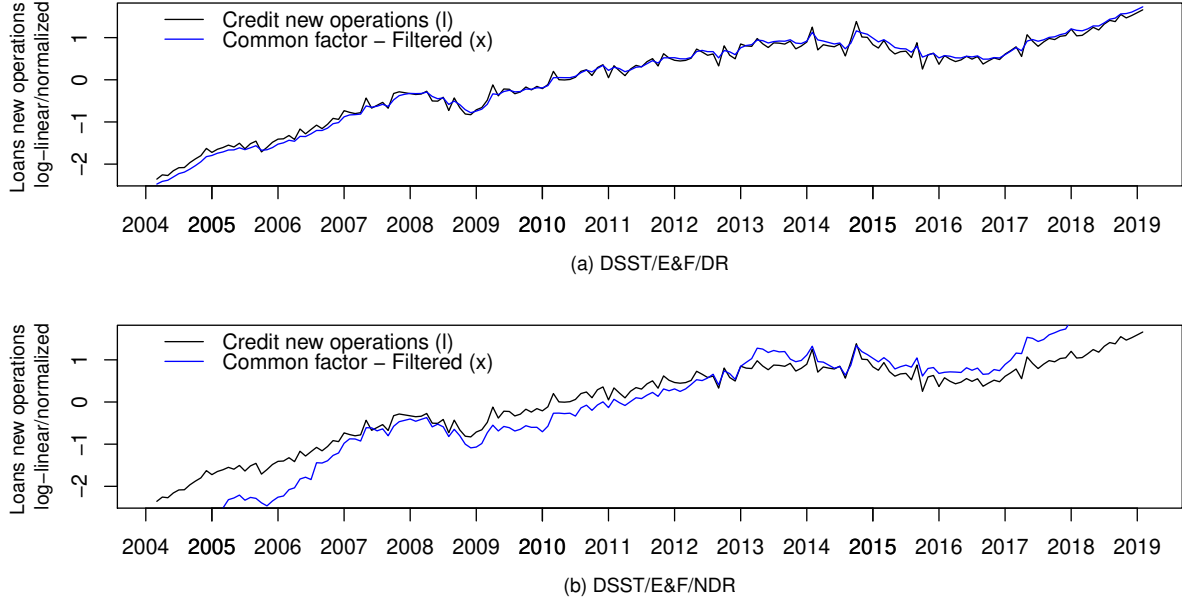


(a) DSST/E/DR



(b) DSST/E/NDR

Figure 11: Common factor - (a) DSST/E&F/DR and (b) DSST/E&F/NDR.



7.3 Linearity tests

In Table 8, we present the results of some linearity tests. The Keenan's one-degree test for nonlinearity against the null hypothesis that the time series follows some autoregressive (AR) linear process. The McLeod-Li test checks for the presence of conditional heteroscedascity by computing the Ljung-Box (portmanteau) test with the squared data. Accordingly to Koller and Fischer (2002), the McLeod-Li test is an indirect test and based on the fact that by fitting a linear model to the data, the inherent non-linearity has been swept into the residuals. Tsay's test evaluates quadratic nonlinearity in a time series. We also carry out the likelihood ratio test for threshold nonlinearity, with the null hypothesis being a normal AR process and the alternative hypothesis being a TAR model with homogeneous, normally distributed errors. The Teraesvirta's neural network test for neglected nonlinearity, with the null hypotheses of linearity in "mean". The Lo & Zivot test has three tests available: (i) linear vs 1 threshold TVAR; (ii) linear vs 2 thresholds TVAR; (iii) 1 threshold TVAR vs 2 thresholds TVAR. The tests (i) and (ii) can be seen as linearity tests, whereas the third can be seen as a specification test. In general, the most of the tests rejected linearity. Considering the Lo & Zivot test, we reject linearity in tests (i) and (ii), while the specification test (iii) favors just one regime change (1 threshold TVAR).

Table 8: Linearity tests

Tests	<i>pvalue</i>		
	l_t	f_t	e_t
Keenan's one-degree test for nonlinearity (Keenan 1985)			
H0: time series follows some AR process	0.045	0.643	0.025
McLeod-Li test (McLeod and Li 1983)			
H0: time series follows some ARIMA process	0.000	0.000	0.000
Tsay's Test for nonlinearity (Tsay 1986)			
H0: time series follows some AR process	0.977	0.029	0.000
Likelihood ratio test for threshold nonlinearity (Chan 1991)			
H0: time series follows AR process / H1: time series follows TAR process	0.273	0.000	0.028
Teraesvirta's neural network test* (Teräsvirta et al. 1993)			
H0: Linearity in "mean"	0.000	0.000	0.000
Lo & Zivot multivariate linearity test (Lo and Zivot 2001)			
H0: Linear VAR / H1: 1 threshold TVAR		0.001	
H0: Linear VAR / H1: 2 thresholds TVAR		0.003	
H0: 1 threshold TVAR / H1: 2 thresholds TVAR		0.332	

Notes: *Calculated for regression, where $l_t = f(f_t, e_t)$, $f_t = f(l_t, e_t)$ and $e_t = f(l_t, f_t)$.

7.4 Dynamic SS Linear models (DSSL)

Among the benchmarks in the in-sample and out-of-sample evaluations, we evaluated the standard linear Gaussian SS model, known as Dynamic SS Linear (DSSL) model. In this framework we rule out regime changes on parameters estimation, resulting at the following state-space representation:

$$\begin{bmatrix} l_t \\ f_t \\ e_t \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} [x_t] + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{t,l} \\ \epsilon_{t,f} \\ \epsilon_{t,e} \end{bmatrix} \quad (7.1)$$

and

$$R = var \begin{pmatrix} \epsilon_{t,l} \\ \epsilon_{t,f} \\ \epsilon_{t,e} \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \quad (7.2)$$

with $\epsilon_{t,l} \sim \mathcal{N}(0, r_{11})$, $\epsilon_{t,f} \sim \mathcal{N}(0, r_{22})$ and $\epsilon_{t,e} \sim \mathcal{N}(0, r_{33})$. Also, R is symmetric ($r_{21} = r_{12}$, $r_{31} = r_{13}$ and $r_{32} = r_{23}$) and positive semi-definite.

Once again, the transition equation gives the law of motion for x_t and follows a first order autoregressive process:

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \epsilon_{t,x} \quad (7.3)$$

where the state variance is given by $Q^x = var(\epsilon_{t,x})$ with a normally distributed noise ($\epsilon_{t,x} \sim \mathcal{N}(0, Q^x)$). Adding the assumptions that the noises at transition and measure equations are uncorrelated and the initial state is normally distributed ($x_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$), we estimate the parameters by Maximum Likelihood.

We consider two specifications following this basic structure. The first is a DSSL with diagonal R observation covariance matrix (DSSL/DR). The second one is a DSSL with non-diagonal R observation covariance matrix (DSSL/NDR).