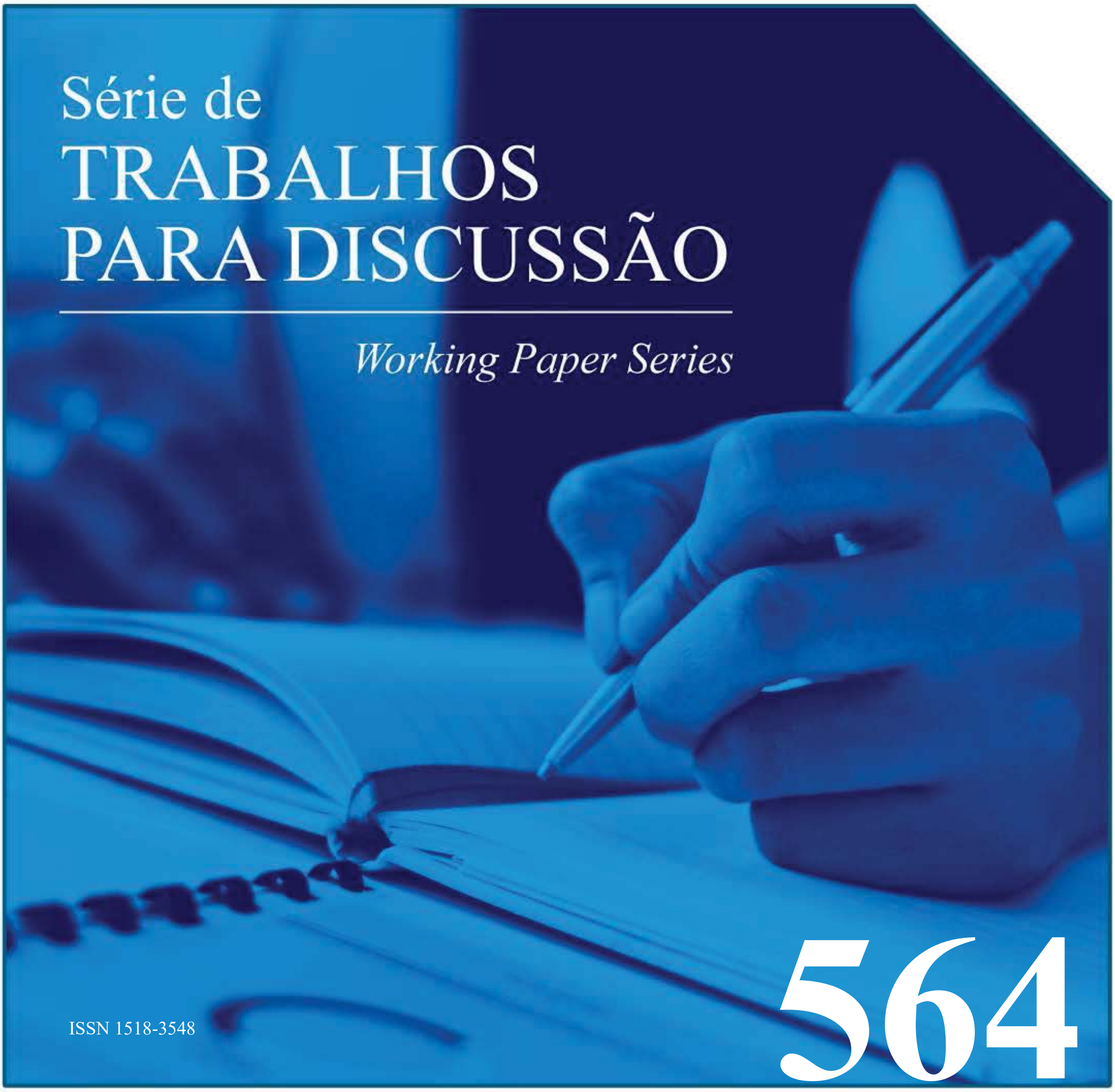


August 2022

Nowcasting Brazilian GDP with Eletronic Payments Data

Raquel Nadal Cesar Gonçalves

BANCO CENTRAL DO BRASIL

ISSN 1518-3548
CGC 00.038.166/0001-05

Working Paper Series	Brasília	no. 564	Agosto	2022	p. 3-36

Working Paper Series

Edited by the Research Department (Depep) – E-mail: workingpaper@bcb.gov.br

Editor: Rodrigo Barbone Gonzalez

Co-editor: Eurilton Alves Araujo Jr

Head of the Research Department: André Minella

Deputy Governor for Economic Policy: Diogo Abry Guillen

The Banco Central do Brasil Working Papers are evaluated in double-blind referee process.

Citizen Service Division

Banco Central do Brasil

Deati/Diate

SBS – Quadra 3 – Bloco B – Edifício-Sede – 2º subsolo

70074-900 Brasília – DF – Brazil

Toll Free: 0800 9792345

Fax: +55 (61) 3414-2553

Internet: http//www.bcb.gov.br/?CONTACTUS

Non-technical Summary

GDP nowcasting models based on coincident indicators are very useful to monitor how shocks and policies impact economic conditions in the short term. In the Brazilian case, GDP series are released only about 2 months after the closure of the quarter, reinforcing the importance of using coincident economic indicators in nowcasting models to infer about the pace of activity.

Electronic payments data can be classified as coincident indicators since they allow to track economic transactions, working as a proxy to the firms' production of goods and services; and are released in a timely basis, long before other relevant activity indicators. Another advantage lies in the possibility of data disaggregation until the level of economic division, by payment instrument, either in terms of number of transactions or deflated value, generating large numbers of series that can help to anticipate economic fluctuations.

This paper seeks to measure how data from different electronic payment instruments contribute to improving the nowcasting accuracy of GDP and of its sectoral components. To do so, the GDP nowcasting accuracy obtained by factor models using payments data and other economic indicators is compared with the accuracy of models without payments data. This comparison is performed at two distinct horizons: right after the closure of the quarter to be predicted, when payments data are already available but data from other relevant economic indicators are still unknown; and about 15 days before the GDP release, when data from other coincident indicators are also available. The results show that payments data contribute significantly to improving GDP nowcast accuracy in both horizons, but mainly just after the closure of the quarter, which suggests, additionally, that these data are also able to anticipate some of the information that is contained in coincident indicators.

The use of retail electronic payment data from different instruments in the nowcasting models of Brazilian GDP is innovative and enhances the Central Bank' efforts to anticipate the pace of economic activity.

Sumário Não Técnico

Modelos de previsão do PIB baseados em indicadores antecedentes são muito úteis para monitorar como choques e políticas impactam as condições econômicas no curto prazo. No caso do Brasil, onde as séries do PIB são divulgadas apenas cerca de 2 meses após o fechamento do trimestre, torna-se ainda mais relevante a utilização de indicadores econômicos antecedentes em modelos de previsão para inferir sobre o ritmo de atividade.

Dados de pagamento podem ser classificados como indicadores antecedentes por permitirem rastrear transações econômicas, funcionando como *proxy* para a produção de bens e serviços pelas firmas, e por estarem disponíveis sem defasagem significativa, antes de outros indicadores antecedentes relevantes para previsão da atividade. Outra vantagem está na possibilidade de desagregar esses dados até o nível de divisão econômica, por instrumento de pagamento, tanto em termos de número de transações como em termos de valor deflacionado, gerando uma grande quantidade de séries que podem auxiliar a antecipar flutuações.

Este artigo busca mensurar como transações realizadas por meio de diferentes instrumentos de pagamento contribuem para melhorar a acurácia preditiva do PIB e de seus componentes setoriais. Para isto, a acurácia preditiva do PIB obtida com modelos de fatores dinâmicos usando dados de pagamento e outros indicadores econômicos é comparada à acurácia preditiva de modelos sem dados de pagamentos. A comparação é realizada em dois períodos distintos: logo após o fechamento do trimestre, quando já se conhecem os dados de pagamento, mas os outros indicadores antecedentes ainda não são foram divulgados; e cerca de 15 dias antes da divulgação do PIB, quando outros indicadores que antecipam a atividade também já foram divulgados. Os resultados mostram que dados de pagamentos contribuem de maneira significativa para melhorar a acurácia preditiva do PIB em ambos os horizontes, ainda que os ganhos sejam mais relevantes no primeiro período. Essa conclusão sugere, em particular, que dados de pagamento antecipam inclusive parte da informação de indicadores antecedentes.

A utilização de dados de pagamento eletrônicos de diferentes instrumentos de pagamento em modelos de previsão do PIB brasileiro é inovadora, contribuindo para antecipar com mais acurácia o ritmo da atividade econômica.

Nowcasting Brazilian GDP with Electronic Payments Data

Raquel Nadal Cesar Gonçalves*†

Abstract

Electronic payments data are usually available on a more timely basis than other coincident economic indicators and can be disaggregated into the level of economic divisions, by number of transactions and value, being potentially useful to anticipate the pace of economic activity. This paper seeks to measure how data from electronic payment instruments contribute to improving the nowcasting accuracy of GDP and its sectoral components. To do so, the nowcasting accuracy of complete models, with economic indicators and payments data, is compared with the accuracy of base models, without payments data, in two horizons: right after the closure of the quarter to be predicted, when payments data are already available; and about 15 days before the GDP release, when data from other coincident economic indicators are also known. The results show payments data contribute significantly to improving GDP nowcast accuracy in both horizons, but mainly just after the closure of the quarter.

Keywords: payments data, credit transactions, dynamic factor model, nowcasting.

JEL Codes: C53, E17, E32, E42

* Banco Central do Brasil. email: raquel.nadal@bcb.gov.br. I would like to thank Saulo Benchimol Bastos, Leonardo Nogueira Ferreira, Rodrigo Barbone Gonzales, Gustavo Silva Araujo, Rogerio Antonio Lucca, Fabio Martins Trajano de Arruda, Cyntia Freitas Azevedo, Renata Modesto Barreto, Mario Rubem do Coutto Bastos, and Franz Mainhardt Carpes for the opportunity, reviews and sugestions throughout this study.

1. Introduction

In Brazil, GDP is released with a lag of about 60 days. During this interval, GDP nowcasts based on coincident economic variables are very useful to monitor how shocks, policies, and production decisions affect economic conditions in the short term.

This paper seeks to evaluate if data from different payment instruments are useful to improve the nowcasting accuracy of quarterly growth rates of Brazilian GDP and its sectoral components (Agricultural, Industrial, and Services Sector). To do so, it is assumed that the number and the deflated value of credited payment transactions to firms work as a proxy to the overall production of goods and services by the companies, potentially allowing to anticipate the pace of economic activity as other coincident and current economic indicators.

Electronic payments data have the benefit of being compiled electronically and therefore available on a very timely basis, virtually free of sample errors. Another advantage lies in the fact that payment instruments can be matched by a tax identifier, allowing to track economic activity even at a very disaggregated level. The single disadvantage relies on the short-time series in terms of monthly observations for some payment instruments (despite the millions of transactions per day).

To nowcast GDP counting only with a few monthly observations for some of the payments data series, this study uses a dynamic factor model (DFM) estimated by the expectation-maximization (EM) algorithm, which can deal with arbitrary patterns of missing values and allows for a mixed frequency database. Payment series are used in nowcast models in an aggregated manner, considering the total number of transactions, total number of transactions above BRL 5000.00, and total deflated value of transactions made with all payment instruments; and in a disaggregated manner, by payment instrument, by sector, and by CNAE division[1], also considering the number of transactions, number of transactions above BRL 5000.00, and deflated value.

[1] CNAE – Portuguese abbreviation for National Classification of Economic Activities. At all, there are 87 CNAE divisions.

To avoid noises that may come from the use of this large-scale data set in the estimation of the common factors that compose the DFMs, a series selection based on Lasso[2] is adopted.

To measure the contribution of payments data to nowcast accuracy of GDP, the difference between out-of-sample errors obtained from models with all selected variables (complete models) and from models without payments data (base models) is calculated in two horizons: right after the closure of the quarter that corresponds to the GDP to be nowcasted, when payments data are already available; and about 15 days before the release of GDP (or 45 days after the closure of the quarter), when many other relevant economic indicators are also known.

As the base models are nested within the complete models, the statistical significance of the difference between out-of-sample errors is calculated by using the hypothesis tests developed by McCracken (2007). The methodology is like the one described by Diebold and Mariano (1995), but the critical values are simulated for nested models.

The results show electronic payments data significantly contribute to improving the nowcast accuracy of GDP and its main components (GDP of Industry Sector and Services Sector) right after the closure of the quarter to be predicted. Fifteen days before the release of the GDP, the contributions of payments data become smaller but are still statistically significant. Besides, since the difference between the errors from complete and base models becomes smaller in the second horizon, it is plausible to infer as well that payments data can anticipate not only GDP series but also other current and coincident economic indicators that are used in the nowcasting models.

The findings reveal the importance of monitoring the evolution of payments data to track economic activity and reinforces their use in models of central banks, seeking to optimize monetary policy decisions. This importance becomes even more substantial with the results obtained in additional tests, which reveal that models using payments data keep generating accuracy gains also considering two-step-ahead forecasts.

[2] Least Absolute Shrinkage and Selection Operator.

The method used in this study is related to the nowcasting literature, as it takes advantage of both mixed frequency and factors models developments. This literature has started with Mitchell and Burns (1938), but only evolved after the construction of simple indexes based on coincident indicators by Stock and Watson (1991). Nunes (2005) and Camacho and Perez-Quiros (2008) developed models for very short-term forecasting, next refined by Ghysels, Santa-Clara, and Valkanov (2004) with the incorporation of variables measured at different frequencies. Clements and Galvão (2008) and Kuzin, Marcellino, Andreou, and Ghysels (2010) and Schumacher (2011) provide some evidence of improvements in quarterly forecasts using monthly data.

While a strand of literature evolved in the development of mixed-frequency models, a related strand worked on the construction of factors for nowcasting (Forni et al., 2005b; D’Agostino and Giannone, 2006) and on dealing with data set properties, like arbitrary patterns of missing values (Banbura et al., 2014) and ragged-edge, in which data components are released at different dates (Giannone, Reichlin and Small, 2008; Banbura, Giannone and Reichlin, 2010).

This paper makes profit from all these literature advances as the quarterly GDP nowcasts are estimated through mixed frequency factors models, using ragged edge data.

The use of payment data in nowcasting models has only recently been exploited, becoming more plausible as the technology for data compilation, storage, and treatment evolved (Aprigliano, Ardizzi, and Monteforte, 2019). First works indeed find that fluctuations in transactions made by an isolated payment instrument can improve the GDP and household consumption forecasts in the medium and short term (Galbraith and Tkacz, 2007, 2009; Esteves, 2009; Carlsen and Storgaard, 2010; Duarte, Rodrigues, and Rua, 2017). Nonetheless, more recent studies prefer to use a set of payment instruments to forecast economic activity since this procedure allows to endogenize individual choices of payment instrument or the arrival of new technologies (Barnett et al., 2016; Galbraith and Tkacz, 2018; Aprigliano, Ardizzi, and Monteforte, 2019).

Although other studies have already estimated factor models to forecast Brazilian activity indicators (Bragoli et al., 2015; Dahlhaus et al., 2017; Gomes, 2018), the use of retail electronic payment data from different payment instruments in DFMs to make Brazilian GDP nowcasting more accurate is innovative, as far as known.

The rest of the paper is organized as follows. Section 2 gives a brief overview of the payment system data, providing some evidence on the relationship between payment transactions and activity. Section 3 exhibits the identification strategy, whereas Section 4 presents the results. Section 5 concludes.

2. Overview of the Payments System in Brazil and correlations with economic activity

2.1. Payment System Data in Brazil

A payment system is composed of a set of payment instruments and rules, procedures, and technologies used to settle money transfers between economic agents (Aprigliano, Ardizzi, and Monteforte, 2019).

There are wholesale payments and retail payments. While wholesale payments involve intrabank transactions, refinancing operations with national central banks, and transactions between financial markets, retail payments are closely related to economic activity, involving money transactions between individuals and firms (Padoa-Shioppa, 2004). In this paper, only retail payments that occur inside the Brazilian Payment System (SPB)[3] are considered in the nowcasting models. Other remark is that regional payments data are not used as the main objective of the article is to nowcast total and sectoral GDPs.[4]

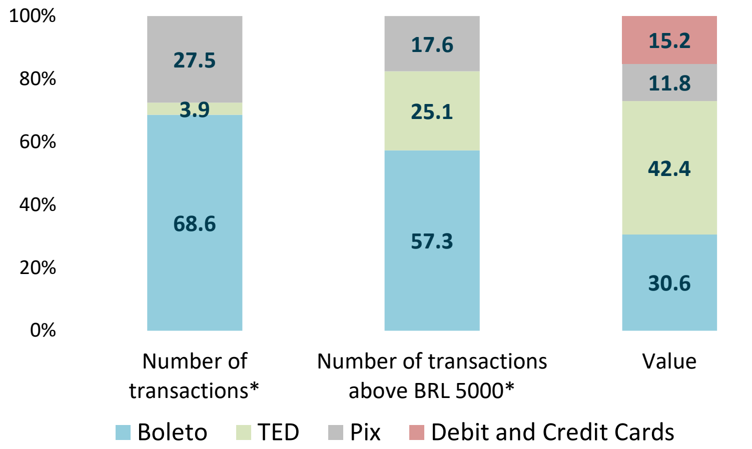
[3] SPB includes only electronic transactions. It excludes payments made with cash, highly significant in terms of the number of transactions, but with lower value, being, therefore, less correlated with activity fluctuations.

[4] Regional payments data, nonetheless, is also available and could be used both to nowcast regional GDPs, which tend to be disclosed with even more lags than the total GDP, and to rank leading indicators in terms of nowcasting relevance for each Brazilian region (by using the same methodology of this article).

Nowadays, the SPB has as payment instruments Doc[5], TED[6], Boleto[7], Pix[8], covenant, GRU[9] and FGTS[10], automatic debit, lottery, and debit and credit card. This study uses data from TED, Pix, Boleto, and debit and credit card, which represented together more than 65% of the total number of transactions in the SPB in 2020, being therefore quite representative of the electronic payments in Brazil.

Figure 1 illustrates the share of each instrument in the total payments considered in this study, in terms of number of transactions, number of transactions above BRL 5000.00[11] and deflated value[12] (data from September 2021).

Figure 1 – Share of payment instruments in the SPB



*The total number of transactions with debit and credit cards are not available for the BCB; only an aggregation of the transactions.

The payments in Figure 1 are only interbank credited payments to firms. So far, there is still no traceable and identifiable data for intrabank payments by tax identifier.[13] The

[5] Doc: Payment instrument that is like a Credit Order Document. It consists of a payment instrument to execute electronic transfers, liquidated with some delay.
[6] TED: Portuguese abbreviation for Electronic Available Transfer, the first electronic payment instrument that allowed real-time transfers in business days and commercial hours in Brazil.
[7] Boleto is similar to a bill, exhibiting bar codes.
[8] Pix: Instrument that allows for instant payments, being liquidated in real-time 24 hours every day.
[9] GRU: Portuguese abbreviation for Federal Tax Liability Payment Form.
[10] FGTS: Portuguese abbreviation for Time-of-Service Guarantee Fund (FGTS) payment form.
[11] As transactions with higher values tend to be more correlated with economic activity.
[12] The values are deflated using the IPCA – the main Brazilian CPI.
[13] The lack of data for intrabank transactions could be a major problem if significant changes in the proportion between interbank and intrabank transactions had been observed during the sample horizon or if intrabank transactions represented the biggest portion between total transactions. Nonetheless, estimates of the intrabank transaction do not show significant changes that seem to harm the results presented in this

choice of working only with credited payments to firms reflects the strategy of capturing fluctuations in the Brazilian economic activity from a production perspective.

Payments made with TED have been introduced slowly in the SPB. As of mid-2003, TED's transactions could only be made for values above BRL 5000.00, being firstly reduced to BRL 3000.00 in May 2010, and afterwards to BRL 2000.00 and BRL 1000.00. Transactions of any value using TED started to be allowed only at the beginning of 2016. Therefore, although series with the number of transactions above BRL 5000.00 made with this payment instrument are available since mid-2003, series without structural breaks for the total number of transactions and deflated value start only in 2016.

Before the introduction of Pix, transactions made with TED were the only type of electronic transaction with real-time liquidation in business day and commercial hours in Brazil, being, therefore, less risky than other available instruments. Due to this historical feature, payments with TED still represent a little more than 40% of the total value transacted (Figure 1). However, for being a paid transaction, the total number of payments with TED historically had a much smaller market share compared to other instruments, such as Boleto. Recently, with the introduction of Pix, TED's share in the total number of transactions dropped even more, representing less than 4% when only TED, Boleto, and Pix are considered (Figure 1).

In the case of Boleto, there has never been a superior limit for the value of transactions nor costs for the users. Therefore, in opposition to what is observed to TED, Boleto payments have a large share in the total number of transactions considered in Figure 1, but only a little more than 30% in terms of the total value of transactions. Payments with Boleto only started to be tracked in November 2018. Thus, before this date, there is no available data for Boleto transactions in the BCB.

Pix transactions were only introduced in the SPB in November 2020. This instrument can be considered a substitute for TED and Boleto, since transactions made using Pix have instantaneous liquidation, even on weekends and holidays, 24 hours a day, are not paid

paper. For the agricultural sector, the portion of intrabank transactions may be more representative than the interbank. In this sector, firms' transactions tend to be concentrated in Banco do Brasil, the main bank that provides rural credit in Brazil.

(when issued by a natural person) and allow for charges through QR codes. Gradually, therefore, the share of TED and Boleto tends to drop in favor of Pix's share, both in total number of transactions and deflated value.

In 2020, transactions with debit and credit cards, together, represented about 74% of the total number of transactions considered in this study[14]. Nonetheless, in the database of cards payments that is available for the BCB nowadays[15], it is not possible to verify the total number of transactions in a month, but only an aggregation of transactions by value and firm. It is worth noting as well that for the BCB even these aggregated data are only available since November 2017. Concerning the deflated value, the share of debit and credit cards in the total value transacted was about 15% in September 2021.

2.2. Payments data and economic activity

As Irving Fisher wrote in his quantitative theory of money, "the money paid in any transaction is the equivalent of the goods bought at the price of sale" (Fisher, 1912). In this sense, payments data represent coincident indicators since they allow to track economic conditions and are released in a timely basis, before other activity indicators, potentially representing an important source of information for short-term forecasting.

Figures 2.1 and 2.2 show the evolution of economic activity in parallel with the fluctuations in the number of transactions and number of transactions above BRL 5000.00 for different payment instruments, while Figure 2.3 illustrates credited payments to firms in terms of deflated value for each payment instrument, also simultaneously to the evolution of the activity economic indicator.

The figures exhibit data from December 2018 to September 2021. The activity indicator is represented by the monthly GDP series, after seasonal adjustment. This series, available on the BCB website, is calculated by simple extrapolation of quarterly GDP, released by

[14] TED, Pix, Boleto and debit and credit cards.
[15] Provided by Eletronic Funds Transfer (known as Sitraf), from the Interbank Payment Clearing House (known as CIP).

the IBGE[16]. In the specific case of cards, the number of transactions reflects an aggregated of credited payments to firms.

As Pix substituted mainly TED's transactions[17], payments made by TED and Pix were added and are shown together in the figures. The procedure allows to compensate both for the short-time series of Pix and for the structural breaks in TED's series since Pix's entry. It is worth noticing, nevertheless, that Pix also replaced transactions made with Boleto, cards, and cash, which can be easily seen by the climb in the total number of transactions made with TED and Pix together after October 2020 (Figure 1). Such an expressive increase cannot be seen in the number of transactions above BRL 5000.00 nor in the deflated value of payments made by Pix and TED, which reveals that the substitution between these instruments is stronger for transactions with smaller values.

Figure 2.1 – Monthly GDP and number of transactions
Seasonally adjusted index (02/2020=100)

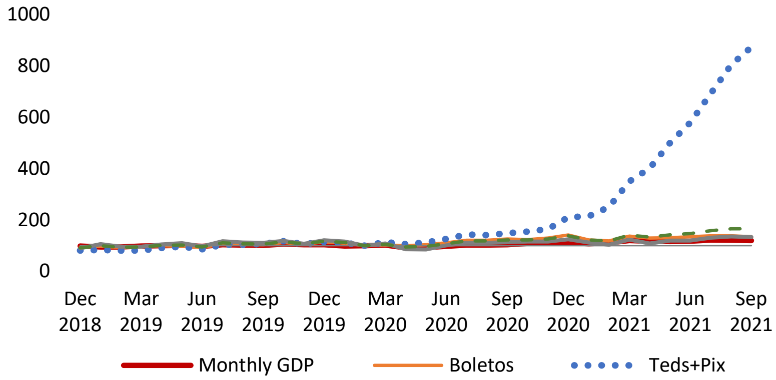
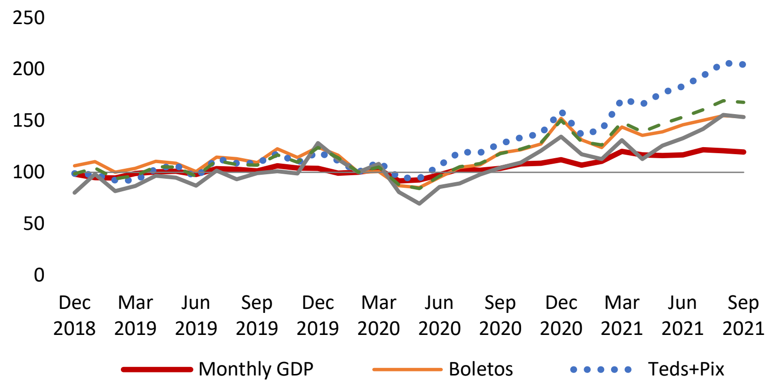
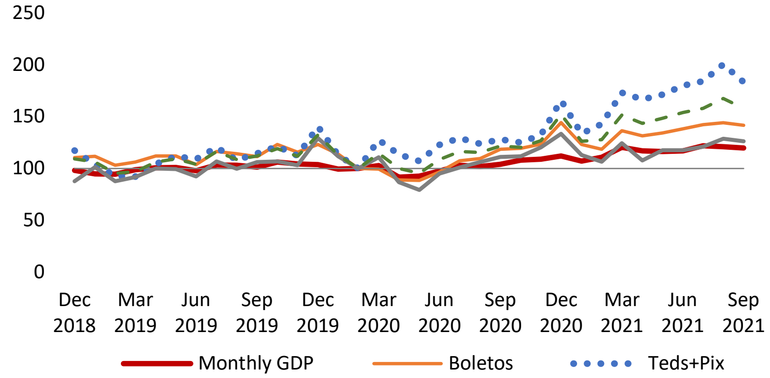


Figure 2.2 – Monthly GDP and number of transactions above BRL 5000.00
Seasonally adjusted index (02/2020=100)



[16] IBGE – Portuguese abbreviation for "Instituto Brasileiro de Geografia e Estatística" or Brazilian Institute of Geography and Statistics.
[17] As said before, these instruments have high similarities between them.

Figure 2.3 – Monthly GDP and deflated value of transactions
Seasonally adjusted index (02/2020=100)



In the figures, it can be noticed that fluctuations in payment transactions and in monthly GDP show similar patterns. From December 2018 to September 2021, the correlation between the total number of transactions and GDP was about 0.94. This correlation is higher than the correlations obtained considering the total number of transactions of each payment instrument individually.

The correlation of payments data with activity index when only the number of transactions above BRL 5000.00 is considered reaches 0.96, being the highest between all the series shown in Figures 2.1 to 2.3. Finally, considering the deflated value of transactions, the correlation between monthly GDP and transactions with all payment instruments reaches 0.95, also higher than the correlation achieved when the deflated value of each payment instrument is considered.

The visual inspection and the correlations of payment transactions with monthly GDP suggest that payment series can be a useful source of information to nowcast activity, remaining the need of assessing the information content of these variables relative to other economic indicators that are already regularly monitored.

This study uses data from 2,073 series, from which 1,998 are payments data, composed of aggregated and disaggregated series at the level of payment instrument, economic sector, and CNAE division. The series' values represent the number of transactions, the number of transactions above BRL 5000.00, and the deflated value. As said before, TED and Pix are added. Data from cards, nonetheless, are subdivided into 3 groups consisting of anticipation, debit, and credit data. Besides these, there are still the Boleto series and the series that aggregates the values of all these instruments. Sectors are represented by Agricultural, Industrial, and Services sectors and there are 87 CNAE divisions.

The data set of other indicators is composed of 75 series available on the BCB or IBGE websites, involving mainly total and sectoral activity series, besides other monetary and labor market series. Annex I exhibits all the current and coincident economic indicators that are tested to compose the nowcasting models, together with their approximated lags of release concerning the closure of the GDP quarter, in days.

From the whole dataset used in this study, only a few are selected to compose the DFM for total GDP and its sectoral components. The selection of payments series and other economic indicators are based on their capacity of anticipating activity growth rates, as identification strategy better explains, next.

3. Identification Strategy

This study aims at measuring the contribution of payments data to the nowcasting accuracy of the quarterly growth rates of Brazilian GDP and its sectoral components (released by the IBGE) on two distinct horizons: right after the closure of the quarter corresponding to GDP and 15 days before the release of the activity indicator (about 2.5 months after the closing date of GDP).

To do so, DFMs to forecast GDP using payments data and other economic indicators are firstly estimated. The out-of-sample nowcasting accuracy of these models is next compared to the accuracy of DFMs without payments data (using only other economic indicators).

To nowcast GDP and its sectoral components, all the 2,073 series, including payments data and other economic indicators, could be included as explanatory variables for the DFMs. However, this large-scale data set might add more noise than information into the estimation of the factors in the models. As highlighted by Boivin and Ng (2006), although factor models provide a parsimonious way of handling such a large data set, some pre-selection schema may improve the nowcasting performance.

How to select time series? In Amstad, Potter, and Rich (2017), for example, the selection of series for the DFMs is made based on judgment and practical experience, while da Gama Machado, Nadal, and Kawaoka (2020) use Granger causality tests. In this paper, as in Bai and Ng (2008) and Aprigliano, Ardizzi, and Monteforte (2019), the screen of the variables is based on a Lasso algorithm (Zou and Hastie, 2005)[18], performed for each model and for each out-of-sample period. Selection depends therefore on the variables' ability to anticipate the target, represented here by quarterly growth rates of total and sectoral GDP.

For each out-of-sample quarter to be predicted, monthly explanatory series are firstly seasonally adjusted[19] and then transformed into quarterly growth rate variables. Lasso regression is then performed having as dependent variables the quarterly growth rate of total GDP or its components and as explanatory variables quarterly percentage rate of coincident indicators[20].

In each out-of-sample period, Lasso selects the regressors for each GDP series based on the solution of the following minimization problem:

$$min\left(\sum_{T}^{t=1}(y_t - \sum_{j=1}^{N}\beta_j X_{jt})^2\right) s.t. \sum_{j=1}^{N}\left|\beta_j\right| \leq \tau, \tag{1}$$

with τ representing the tuning parameter, set equal to the value that minimizes the cross-validated error across the sample. The model selects n_L series among all N collected in the data set:

[18] Lasso algorithm is a special case of the elastic net model, with α being set to 0 (see Zou and Hastie, 2005). By adopting this procedure, it is possible to choose the regressors that best anticipate activity between the whole data set. Adaptative lasso could have been used instead (Zou, 2006). This last procedure is more appropriated for non-orthogonal predictors, presenting the "oracle properties" as it re-weights the coefficients that are first estimated through elastic net. Nonetheless, for this paper's purpose, the "inconsistency" of the coefficients is less important as it does not change selection.

[19] Seasonal adjustment is done using X-13 Arima wherever possible. To use this seasonal adjustment method, 36 monthly observations are needed, at least. For Boletos and total payments data, only available since December 2018, an alternative method based on an STL decomposition with correction for working days is used. STL: "Seasonal and Trend decomposition using Loess".

[20] As all the explanatory series are used in the same scale, expressing quarterly seasonal adjusted growth rates, there is no need to re-standardize them (through Z-score or standard deviation methods) for the Lasso selection.

$$\hat{L}_{n_L} = \left\{ j \in \{1, 2, \dots, N\} : \left| \hat{\beta}_{L_j} \right| > 0 \right\}. \quad (2)$$

In the Lasso models for total GDP growth rates, the whole data set is considered for selection (all payments series and other economic indicators), whilst in the case of sectoral GDP growth rate series, only sectoral variables are available as candidates, avoiding eventual meaningless selection. By doing so, a specific subset of the data is screened to compose the factors in nowcasting models of total and sectoral GDP series in each out-of-sample period of each investigated horizon.

As a robustness check, three alternative Lasso models are estimated for each GDP series in each out-of-sample period: one first with all data set in the pool of candidates; another one having only payments data; and a last one, excluding payments series, having as candidates only other economic indicators.[21] The picked variables in these three Lasso models are considered in the complete model, while only the other economic indicators, excluding payments data, are employed in the base model.

It is important to note that the same economic indicators, selected through Lasso models, are employed in the base case (without payments data) and in the complete models of each GDP series. Therefore, the base models are nested within the complete models.

The n_L series selected by Lasso for each of the GDP components in each out-of-sample period are subsequently used in DFMs as follows.

Let x_t be the vector representing the monthly growth rate of the n_L time series selected before, already seasonally adjusted. The general specifications of the DFM are represented by:

$$x_t = \mu + \Lambda f_t + \varepsilon_t \quad (3)$$

$$f_t = \sum_{i=1}^{p} \Lambda_i f_{t-i} + \beta u_t, \; u_t \sim i.i.d. N\left(0, I_q\right). \quad (4)$$

[21] Similar procedure is done by Aprigliano, Ardizzi and Monteforte (2019).

In equation (3), x_t is described as a function of an intercept μ and r unobserved common factors denoted by f_t. The factors, in this case, capture the co-movements of the selected series, working as a proxy for fundamental shocks that conduct the activity performance. The variables in x_t are loaded into the unobserved factors f_t by means of Λ. The parameter μ represents a vector of constants, while the error can be described as an $AR(1)$ to allow for serial correlation:

$$\varepsilon_{i,t} = \alpha_i \varepsilon_{i,t-1} + e_{i,t},\ e_{i,t} \sim i.i.d.N\ (0, \sigma^2) \tag{5}$$

where $E\left[e_{i,t}e_{j,s}\right] = 0\ for\ i,j = 1, \dots, n$ and $i \neq j$.

Equation (4) imposes to unobserved factors f_t the structure of a $VAR(p)$. In this study, $p = 4$ since the data to be predicted have a quarterly frequency. The finding results, though, are robust to different lags.

The EM method is used to estimate (3). First presented by Banbura, Giannone, and Reichlin (2010), this method can deal with mixed-frequency databases, as is the case of the data in this study, and with arbitrary patterns of missing values (Banbura and Modugno, 2014).

The EM method also allows to define factors for different subgroups of the data set (blocks) but, in opposition to the two stages method, that builds on an exactly DFM[22], the number of shocks to the factors q must be set equal to the number of factors r. In this study, the models are estimated with only one global block, representing the series picked by Lasso; but with two factors, aiming at capturing distinct patterns in data. Therefore, $q = r = 2$.[23]

To predict a quarterly series with monthly data, the procedure used is similar to the one described in Mariano and Murasawa (2003): the quarterly series is assumed to be observable only in the third month of the quarter, and as unobservable in the other two months. It is also assumed that the quarterly series, represented in this study by the GDP

[22] A case where the error components from equation (1) are assumed to be mutually uncorrelated at all lags.
[23] The finding results are robust for changes in the parameters used for estimation.

series, is equal to the sum of its monthly values. Thus, by approximation, it is possible to express its quarterly growth rate through a function of its monthly unobservable differences (as Annex II shows).

The EM algorithm estimates the missing values, the parameters, and the unobservable common factors using a recursive structure in two steps. As described by Valk, de Mattos, and Ferreira (2019), in the E-step, the algorithm calculates the conditional expectation of the likelihood function using the estimates of the static parameter θ from the previous iteration, θ_j. Next, in the M-step, new parameters θ_{j+1} are estimated by maximization of the likelihood function with respect to θ. Convergence is achieved whenever the absolute change in the likelihood function is less than 10^{-4} (tolerance level employed by the algorithm). The recursive process starts with estimates based on Principal Components Analysis (PCA).[24]

As the base models are nested within the complete models, the asymptotic variance of the difference between predictive accuracies (measured in terms of out-of-sample errors) is zero under the null hypothesis that the additional variables have no predictive ability. In this case, asymptotic normality of standard test statistics fails (Galbraith and Tkacz, 2018), requiring the use of McCracken (2007) test statistics and critical values, designed specifically for the cases of nested models.

The predictive accuracy of each estimated model is measured by the mean absolute error (MAE) of out-of-sample nowcasts, from the first quarter of 2019 to the third quarter of 2021. Real-time *vintages* from total GDP and payment series are used to measure MAEs, a procedure of particular interest when the objective is to nowcast GDP in the short term. Specifically for other economic indicators, *pseudo-real time data* are employed. [25]

[24] See Giannone, Reichlin and Small (2008).

[25] *Real-time vintages*: seasonally adjusted series in each period, without considering non-available information at the respective moment or *ex-post* revisions. In the case of *pseudo-real times vintages*, vintages reflect only changes in seasonal adjustment, not revisions.

4. Results

Annex III exhibits payments data and other economic indicators selected by Lasso for each GDP series in all predicted quarters. Although Lasso was estimated individually for the first and the second horizon of analysis in each quarter[26], the selected series in both horizons were used together in each DFM for the quarter. This procedure allows for a better comparison between horizons, avoiding higher errors in the last horizon because of the inclusion or exclusion of a series.

The sample available for the Lasso selection started only in January 2018 to avoid a biased selection towards TED together with Pix series, much longer than the other instruments series. Over time, the sample window grows.

For total GDP, Lasso selected 68 different series considering both horizons and 9 quarters, being 43 of payments data and 25 of other economic indicators. For the Agricultural sector, 43 series were selected, with 32 of them represented by payment data. In the case of the industrial and service sector, 102 and 74 series were selected, respectively, being 88 and 53 payment series.

Variables of all types of instruments were screened in at least one horizon for all GDP series. Nonetheless, as the word cloud[27] in Figure 3 reveals, the number of transactions above BRL 5000.00 made with debit cards in CNAE division 38, followed by the number of transactions above BRL 5000.00 made with Boleto in CNAE divisions 30 and 79 were among the most selected series considering all quarters. [28]

[26] First only using data available after the closure of GDP, and afterwards using data disclosed until 15 days before the release of the activity indicator.

[27] The words clouds presented on figures 3 and 4 were constructed considering the frequency of each variable' selection in Lasso models: the higher the frequency of selection, the higher the size of the words. An alternative way of presenting the words cloud consists of considering the loads that each variable received in Lasso to determine its frequency. In this case, however, the most appropriated Lasso would be the Adaptative Lasso, which corrects loads' bias that is verified in the case of multicollinear predictors.

[28] Division 38 refers to the economic activity of "collection, treatment, and disposal of waste and material recovery"; division 30 refers to "manufacturing of transport equipment other than motor vehicles"; and division 79 is related to "travel agency, tour operation and other reservation service and related activities".

Figure 3 – Payment series selected by Lasso for GDP series*

*Figure 3 considers the selection made by Lasso for Total, Agricultural, Industrial, and Services GDP series in all quarters and both horizons. Bigger letters denote more frequent selection.

In the case of other economic indicators, the most selected series were the transport index from the Monthly Survey of Services, together with the capital and extractive index from the Monthly Industrial Production (Figure 4). Nonetheless, many other series representing different economic activities were also picked by Lasso. The oil and energy consumption, the steel production, the truck sales, the services expectation confidence index, and the expectations for the agricultural production of many products are among the selected series.

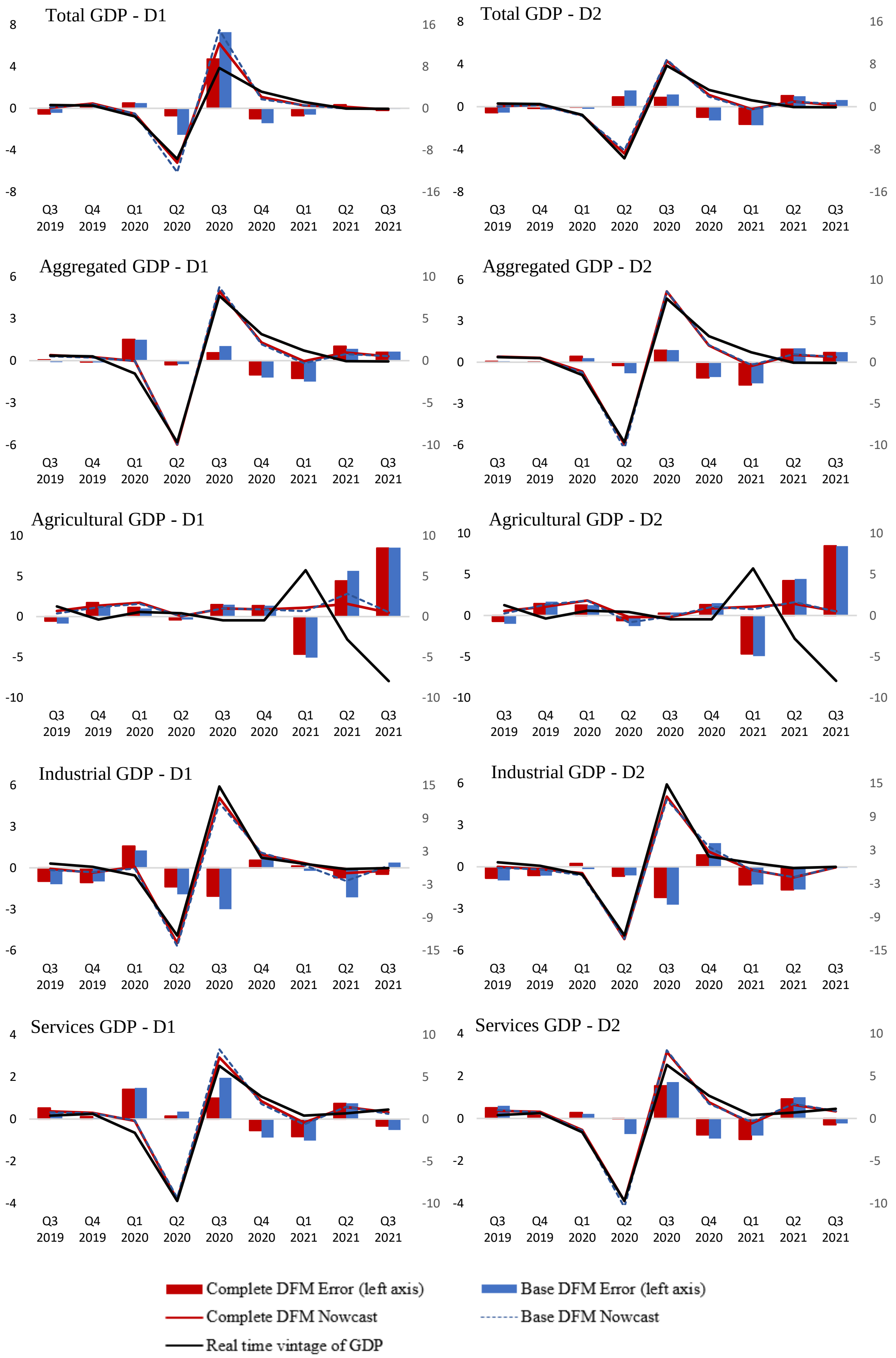
Figure 4 – Other economic series selected by Lasso for GDP series*

*Figure 4 considers the selection made by Lasso for Total, Agricultural, Industrial, and Services GDP series in all quarters and both horizons. Bigger letters denote more frequent selection. See annex I to translate the codes.

After Lasso selection, DFMs were estimated from the first quarter of 2009 until the quarter before the one to be predicted (growing window sample). Nowcasts were generated for the closed interval between the third quarter of 2019 and the third quarter of 2021.

Figure 5 exhibits the one-step-ahead out-of-sample nowcasts and errors for the Total GDP; for the Aggregated GDP calculated based on the sectoral GDPs nowcasts and their weights (known *a priori*); and for Agricultural, Industry, and Services GDP. The GDP series in the figures show the real-time vintages of the activity indicators. In each figure, the first row presents the nowcasts and the errors for the first horizon, just after the end of the quarter (D1), while the second shows the nowcasts and the errors for the second horizon, 15 days before the release of the quarterly GDP data (D2).

Figure 5 – Out-of-sample nowcasts and errors for GDP series



In D1, the improvement in the nowcast accuracy achieved by the complete models (with payments data) is noticeable, while in D2 it is less obvious. It is also visible that the out-of-sample nowcasting errors for the Agricultural GDP in both horizons are higher than the errors obtained for the other activity series. This result is probably due to the fewer Agricultural coincident indicators on monthly basis and to the lack of intrabank payments in the available data set. In the Agricultural sector, diversely from the others, the payments tend to be strongly concentrated in only one public bank[29], which may reduce the forecasting ability of these data.

Table 1 exhibits the MAEs obtained with the complete and the base models right after the GDP closure (D1), when payments data are already known for the whole quarter but most of the other coincident economic indicators are not. The table also shows the ratio between out-of-sample errors, with below one values indicating higher nowcast accuracy of the models with payments data; and the *t* statistic for the difference between out-of-sample errors (FA-t). To measure the statistical significance of FA-t, McCracken's critical values to nested models and recursive windows are employed (McCracken, 2007). Under the null hypothesis, the MAEs from base models are equal to the ones of complete models, and under the alternative hypothesis, the MAEs from complete models are smaller.

Table 1 – MAE of complete and base models in D1, ratios between MAEs, and statistical significance of the difference between errors

	Complete DFM	Base DFM	Ratio Complete/Base	FA-t
Total GDP	1.01	1.46	**0.69**	1.34**
Aggregated GDP	0.71	0.80	**0.89**	1.43**
Agricultural Sector	2.68	2.87	0.93	1.27**
Industrial Sector	0.99	1.34	**0.74**	1.88***
Services Sector	0.62	0.83	**0.75**	1.89***

***Denotes rejection of the null hypothesis at the 1%; ** denotes rejection of the null hypothesis at the 5%; and * denotes rejection of the null hypothesis at the 10%.

Right after the closure of GDP (D1), Table 1 reveals that the use of payments data improves the Total GDP nowcast accuracy by 31% when it is directly predicted using DFMs, and by about 11% when it is calculated based on sectoral GDP nowcasts. In both

[29] Due to governmental programs and credit policies.

cases, the difference between MAEs from base and complete models are statistically significant at 5% level.

The use of payments data also improves the accuracy of the GDP from the Industrial Sector, by 26%, and of the GDP from Services Sector, by 25%. For these sectoral GDPs, the difference between MAEs in D1 is also statistically significant, this time at 1% level. In the case of Agricultural GDP, the accuracy improvement with the use of payments data is smaller – of only 7%, but the difference between MAEs is still significant at 5%.

Table 2 replicates the results from Table 1 considering the available data 15 days before the GDP release (D2). On this horizon, only the nowcast accuracy of Agricultural GDP seems to improve as much as in D1 (8%). For the other activity series, the improvements in nowcast accuracy are smaller than in D1, but still statistically significant at the 1% level in the case of Total GDP, at 5% in the case of Industrial and Services GDP, and at the 10% level for Aggregated GDP.

Table 2 – MAE of complete and base models in D2, ratios between MAEs, and statistical significance of the difference between errors

	Complete DFM	Base DFM	Ratio Complete/Base	FA-t
Total GDP	0.72	0.92	**0.78**	2.96***
Aggregated GDP	0.67	0.72	**0.93**	0.75*
Agricultural Sector	2.55	2.77	0.92	3.11***
Industrial Sector	0.93	1.09	**0.85**	1.48**
Services Sector	0.60	0.70	**0.86**	1.09**

***Denotes rejection of the null hypothesis at the 1%; ** denotes rejection of the null hypothesis at the 5%; and * denotes rejection of the null hypothesis at the 10%.

These results suggest payments data can improve the nowcast accuracy of activity series mainly just after the closure of the GDP´s quarter, when other coincident economic indicators are still unknown. There are still gains of including payments data in the nowcast models when relevant coincident indicators are already known, but in this case, the accuracy improvement is smaller, which may indicate that part of the information contained in other economic indicators can be partially anticipated by payments data.

Additional tests show, moreover, that the accuracy gains from using payments data to forecast GDP are still observed for two-step-ahead forecasts[30]. These gains are statistically significant at the 1% and 5% levels (Table 3). For the next quarters (three and four-steps ahead), the gains are less significant[31].

Table 3 – Statistical significance of the difference between errors from base and complete models considering two, three and four steps ahead

	2 steps ahead	3 steps ahead	4 steps ahead
Total GDP	1.91***	1.01*	0.40
Aggregated GDP	1.41**	1.02*	1.01*
Agricultural Sector	1.43**	0.35	0.70
Industrial Sector	1.36**	1.39**	0.85*
Services Sector	1.52**	0.91*	1.01*

***Denotes rejection of the null hypothesis at the 1%; ** denotes rejection of the null hypothesis at the 5%; and * denotes rejection of the null hypothesis at the 10%.

Payment data still have few observations on a monthly time basis, especially regarding the number and value of transactions with Boleto and Pix. The results presented in this paper probably tend to improve as the available sample becomes larger. However, it is already possible to obtain nowcast accuracy gains with the use of payments data in models to nowcast GDP.

5. Conclusions

This study aims at assessing how data from payment instruments can contribute to improving the predictive accuracy of GDP and its sectoral components. The accuracy gain is measured by the difference between out-of-sample nowcasts errors from DFMs using payments data and other economic indicators (complete model) and DFMs using only other economic indicators (base model).

[30] To compute the statistical significance of the difference between errors from complete and base models in two, three and four steps ahead, the model uses data available in D2, when the whole set of payments data and other economic indicators are already known.
[31] As there is still no GDP vintage available for the first and second quarter of 2022, the accuracy gains computed for three and four-step-ahead forecasts count, respectively, with only 8 and 7 observations. For the one and two-step-ahead forecasts, there are 9 observations available.

The out-of-sample errors are compared in two distinct horizons: right after the closure of the quarter to be predicted, when payments data are already known; and about 15 days before the release of the GDP, when not only payments data but also other coincident economic indicators have already been released. The errors are calculated considering real-time *vintages* for the GDP and payments data and pseudo-real-time *vintages* for the other economic indicators.

The results reveal payments data contribute to significantly improving Brazilian GDP nowcasting. These contributions are higher when other relevant coincident economic indicators are still unknown, just after the closure of the quarter to be predicted. In about 15 days before the GDP release, payments data still contribute to reducing the nowcasting errors, but in this horizon, the accuracy improvement is smaller, which additionally suggests that part of the relevant information contained in coincident indicators can also be found in the payment series. Supplementary tests reveal, moreover, that the accuracy gains of using payments data in GDP models can be still observed in two-step-ahead forecasts.

Prospectively, these findings suggest it will be interesting to test how payments data contribute to improving the nowcasting accuracy of other relevant Brazilian economic indicators and of regional GDPs. Another possible strand of research consists of using the factors that result from DFMs with payments data as inputs for the calculation of potential GDP (such as done with unemployment rate, capacity utilization etc.).

For Brazil, the use of credited electronic payments data in a DFM to nowcast GDP is innovative as far as known. This paper, therefore, contributes to the nowcasting literature and to central banks' efforts to anticipate the pace of economic activity.

References

Andreou, E., Ghysels, E., Kourtellos, A. (2010). Regression models with mixed sampling frequencies. Journal of Econometrics, 158, 246–261.

Aprigliano, V., Ardizzi, G., Monteforte, L. (2019). Using Payment System Data to Forecast Economic Activity. International Journal of Central Banking, vol. 15(4), pages 55-80, October.

Baffigi, A., Golinelli, R., Parigi, G., (2004). Bridge models to forecast the euro area GDP. International Journal of Forecasting 20 (3), 447–460

Banbura, M., Giannone, D., & Reichlin, L. (2010). Nowcasting. Working paper 1275, European Central Bank, Frankfurt.

Banbura, M., M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. Journal of Applied Econometrics, 29(1):133–160.

Barnett, W., M. Chauvet, D. Leiva-Leon, and L. Su. (2016). Nowcasting Nominal GDP with the Credit-Card Augmented Divisia Monetary Aggregates. Working Papers Series in Theoretical and Applied Economics No. 201605, University of Kansas, Department of Economics.

Bragoli, D., Metelli, L., Modugno, M. (2015). The importance of updating: Evidence from a Brazilian nowcasting model. OECD Journal: Journal of Business Cycle Measurement and Analysis, 2015/1.

Boivin, J., S. Ng. (2006). Are More Data Always Better for Factor Analysis? Journal of Econometrics 132 (1):169‑94.

Camacho, M., Perez-Quiros, G. (2008). Introducing the Euro-STING: short-term indicator of euro-area growth. Working Paper, Bank of Spain.

Carlsen, M., and P. E. Storgaard. (2010). Dankort Payments as a Timely Indicator of Retail Sales in Denmark. Technical Report.

Clements, P., Galvão, A. (2008). Macroeconomic forecasting with mixed frequency data: forecasting output growth in the United States. Journal of Business and Economic Statistics, 26(4), 546–554.

D'Agostino, A., Giannone, D. (2006). Comparing alternative predictors based on large-panel dynamic factor models. Working Paper Series 680, European Central Bank.

da Gama Machado, V., Nadal, R., Kawaoka, F. (2020). A Data-Rich Measure of Underlying Inflation for Brazil. Working Paper Series 516. Banco Central do Brasil.

Dahlhaus, T., Guénette, J.-D., Vasishtha, G. (2017). Nowcasting BRIC+M in real-time. International Journal of Forecasting, 33(4):915–935.

Diebold, F.X., Mariano, R.S. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Duarte, C., P. M. Rodrigues, A. Rua. (2017). “A Mixed Frequency Approach to the Forecasting of Private Consumption with ATM/POS Data.” International Journal of Forecasting 33(1): 61‑75.

Esteves, P. S. (2009). “Are ATM/POS Data Relevant When Nowcasting Private Consumption?” Technical Report.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., (2005). The generalized dynamic factor model: one-sided estimation and forecasting. Journal of the American Statistical Association 100 (471), 830–840.

Galbraith, J. W., G. Tkacz. (2007). “Electronic Transactions as High-Frequency Indicators of Economic Activity.” Technical Report.

Galbraith, J. W., G. Tkacz. (2009). “A Note on Monitoring Daily Economic Activity via Electronic Transaction Data.” CIRANO Working Paper No. 2009s-23.

Galbraith, J. W., G. Tkacz. (2018). “Nowcasting with Payments System Data. International Journal of Forecasting 34 (2): 366‑76.

Ghysels, E., Santa-Clara, P., Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regressions. Discussion paper, UNC and UCLA.

Giannone, D., Reichlin, L., Small, D. (2008). “Nowcasting: the real-time informational content of macroeconomic data.” Journal of Monetary Economics, 55, 665-676.

Gomes, G. B. (2018). “Nowcasting Brazilian GDP: a performance assessment of dynamic factor models.” FGV Scientific Journals.

Kitchen, J., Monaco, R.M. (2003). Real-time forecasting in practice: the U.S. treasury staff’s real-time GDP forecast system. Business Economics pp. 10–19.

Kuzin, V., Marcellino, M., Schumacher, C. (2011). Midas versus mixed frequency var: Nowcasting GDP in the Euro area. International Journal of Forecasting, 27, 529–542.

Mariano, R. S., Murasawa, Y. (2003): A new coincident index of business cycles based on monthly and quarterly series. Journal of Applied Econometrics, 18(4):427–443, pp. 2-6.

McCracken, M. W. (2007). Asymptotics for out-of-sample tests of Granger causality. Journal of Econometrics, 140, 719–752.

Mitchell, W. C., Burns, A. F. (1938). Statistical indicators of cyclical revivals. G. H. Moore (Ed.), Business cycle indicators: Vol. 1 (pp. 184–260). Princeton: Princeton University Press, reprinted 1961.

Nunes, C. (2005). Nowcasting quarterly GDP growth in a monthly coincident indicator model. Journal of Forecasting, 24, 575–592.

Padoa-Schioppa, T. (2004). Shaping the Payment System: A Central Bank's Role. Speech delivered at the Bank of Korea's Conference on Payment Systems, Seoul.

Stock, J., Watson M. W. (1991). A probability model of the coincident economic indicators. Leading economic indicators: New approaches and forecasting records (pp. 63-85). Cambridge: Cambridge University Press.

Valk, S., de Mattos, D., Ferreira, P. (2019). Nowcasting: An R Package for Predicting Economic Variables Using Dynamic Factor Models. The R Journal, 11(1), 230-244.

Zou, H., Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society 67(2), 301-320.

Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association 101 (476), 1418-1429.

Appendix I – Other economic indicators used as candidates for selection in Lasso and lags of release regarding the closure of quarter, in days

SGS Series

Series (SGS)	Lag	Name	Code
1453	15	Indicator of Trade Movement - IMC	Comercio
21637	45	PMS - Total Nominal Revenue	PMS_T
21638	45	PMS - Nominal Revenue of Services provided to households	PMS_SFamilia
21639	45	PMS - Nominal Revenue of Information and Comunication Services	PMS_InfCom
21640	45	PMS - Nominal Revenue of Professional, Administrative and Complemetary Services	PMS_SProf
21641	45	PMS - Nominal Revenue of Transport and Mail Services	PMS_Transp
21642	45	PMS - Nominal Revenue of Other Services	PMS_OServ
23982	45	PMS - Total Volume of Services	PMS_Vol
1455	45	Retail Sales	VVarTotal
1483	45	Retail Sales - Fuels	VVarComb
1496	45	Retail Sales - Food Products and Tobbaco	VVarAlim
1509	45	Retail Sales - Clothing and Fabric	VVarVest
1522	45	Retail Sales - Furniture	VVarMov
1548	45	Retail Sales - Cars	VVarAuto
1561	45	Retail Sales - Supermarkets	VVarHiper
20099	45	Retail Sales - Pharmaceutical Products and Cosmetics	VvarFarm
20102	45	Retail Sales - Office Supplies and Communication	VVarEquip
20104	45	Retail Sales - Other Personal and Household Items	VVarOutros
20105	45	Retail Sales - Contruction Materials	VVarConstr
20106	45	Retail Sales - Trade	VVarComAmp
4393	15	Consumer Confidence Index	IndConfCon
4394	15	Current Economic Conditions Index	IndConfConAt
4395	15	Future Expectations Index	IndConfFut
17660	10	Services Confidence Index	ICS
17661	10	Atual Conditions Index - Services	ISAS
17662	10	Expectation Index - Services	IES
1391	45	Production of Oil and derivates	PPetroleo
1393	45	Gasoline Consumption	CPetroleo
1406	75	Energy Consumption	CEnergia
21859	35	Industrial Production - General	PIMG
21861	35	Industrial Production - Mineral Extraction	PIM_Ext
21862	35	Industrial Production - Transformation Industry	PIM_Trans
21863	35	Industrial Production - Capital Goods	PIM_Capital
21864	35	Industrial Production - Intermediary Goods	PIM_BI
21865	35	Industrial Production - Consumer Goods	PIM_BConsumo
21866	35	Industrial Production - Durable Consumer Goods	PIM_ConsDur
21867	35	Industrial Production - Semi-durable and Non-durable Consumer Goods	PIM_ConsSemiNao
21868	35	Industrial Production - Inputs for Construction	PIM_Construcao
24349	45	Real Payroll - Manufacturing Industry	MassaRCNI
24351	45	Real Payroll - Manufacturing Industry	UCICNI
28558	45	Salário real na indústria de transformação (2016=100)	SalarioCNI
7357	45	Produção de aço bruto (1992=100)	PAco
24352	45	Installed Capacity Utilisation	UCI
1373	0	Total Production of Vehicles	PVeiculos
1375	0	Truck Production	PCaminhao
1378	0	Vehicle sales	VVeiculos
7386	0	Truck sales	VCaminhao
1388	10	Production of Agricultural Machinery	PMaquinasAgro
24363	45	IBC-Br	IBC
28763	30	Caged - Total	Caged
24379	50	Employed People - PNADC	Ocupadas
24380	50	Unemployed People - PNADC	Desocupadas
24369	50	Unemployment Rate - PNADC	TaxaOcup
1788	60	Restricted Monetary Base	BM
1786	60	Currency Paper	PMPP
1833	30	Expanded Monetary Base	BMAmp
7535	60	Total Federal Government Bonds	TTPF
27810	60	M2	M2
27813	60	M3	M3
22701	50	Current Transactions	TransCorrente

IBGE Series

Table (IBGE)	Lag	Name	Code
6588	10	LSPA - Annual Estimate of Planted Area	Aplantada
	10	LSPA - Annual Estimate of Harvested Area	Acolhida
	10	LSPA - Annual Estimate of Cereal Production	Cereal
	10	LSPA - Annual Estimate of Sugar Production	Acucar
	10	LSPA - Annual Estimate of Cashew Nut Production	Castanha
	10	LSPA - Annual Estimate of Tobacco Production	Fumo
	10	LSPA - Annual Estimate of Orange Production	Laranja
	10	LSPA - Annual Estimate of Cassava Production	Mandioca
	10	LSPA - Annual Estimate of Tomato Production	Tomate
	10	LSPA - Annual Estimate of Grape Production	Uva
	10	LSPA - Annual Estimate of Banana Production	Banana
	10	LSPA - Annual Estimate of Potato 1st. Crop Production	Batata1
	10	LSPA - Annual Estimate of Potato 2nd. Crop Production	Batata2
	10	LSPA - Annual Estimate of Potato 3rd. Crop Production	Batata3
	10	LSPA - Annual Estimate of Cocoa Production	Cacau

Appendix II – From quarterly to monthly series[32]

Let Y_t^M be a monthly level series of unobservable values of GDP and Y_t^Q be equal to the quarterly value of GDP, partially observable in the monthly series. Supposing yet that GDP level is observable only in the third month of the quarter, it is possible to write:

$$Y_t^Q = \begin{cases} Y_t^M + Y_{t-1}^M + Y_{t-2}^M, & t = 3, 6, 9, \ldots \\ unobservable & others. \end{cases} \tag{6}$$

GDP monthly change is given by:

$$y_t^Q = Y_t^Q - Y_{t-3}^Q. \tag{7}$$

Replacing (6) in (7), it is possible to express the quarterly change as a function of the differences of the month variable, $y_t = Y_t^M - Y_{t-1}^M$:

$$y_t^Q = Y_t^Q + Y_{t-1}^Q - Y_{t-1}^Q + Y_{t-2}^Q - Y_{t-2}^Q - Y_{t-3}^Q \tag{8}$$

$$= y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4}, \qquad t = 6, 9, \ldots$$

Suppose also that the variable to be predicted is the GDP monthly growth rate, given by:

$$x_t^Q \equiv \log(Y_t^Q) - \log(Y_{t-3}^Q). \tag{9}$$

Replacing (8) in (9) and using the approximation between arithmetic and geometric means, x_t^Q can be expressed as:

$$x_t^Q \approx \frac{1}{3}[x_t^M + 2x_{t-1}^M + 3x_{t-2}^M + 2x_{t-3}^M + x_{t-4}^M]. \tag{10}$$

In (10), the quarterly growth rate is decomposed into monthly growth rates. Next, it is enough to assume the unobservable monthly growth x_t^M admits a factor representation as in equation (1), with loadings Λ_Q, and the quarterly growth rate of GDP can be written as a function of monthly GDP:

$$x_t^Q = \overline{\Lambda_Q}\,[f_t' \ldots f_{t-4}']' + [1\ 2\ 3\ 2\ 1][\varepsilon_t^M \ldots \varepsilon_{t-4}^M] \tag{11}$$

with $\overline{\Lambda_Q} = \left[\Lambda_Q\ 2\Lambda_Q\ 3\Lambda_Q\ 2\Lambda_Q\ \Lambda_Q\right]$ (restricted matrix of loadings on the factors and their lags).

[32] Inspired in Valk, de Mattos and Ferreira (2019).

Appendix III –Variables selected by Lasso for each GDP series

	Total GDP	Agricultural Sector	Industrial Sector	Services Sector
Payment Series*	Antecip._Div_Q5K.21	Antecip._Div_Q5K.1	Antecip._Div_Q5K.11	Antecip._Div_Q5K.49
	Antecip._Div_Q5K.33	Antecip._Div_Q5K.3	Antecip._Div_Q5K.17	Antecip._Div_Q5K.52
	Antecip._Div_Q5K.49	Antecip._Div_Q.1	Antecip._Div_Q5K.20	Antecip._Div_Q5K.53
	Antecip._Div_Q.5	Antecip._Div_Q.2	Antecip._Div_Q5K.21	Antecip._Div_Q5K.62
	Antecip._Div_Q.50	Antecip._Div_Q.3	Antecip._Div_Q5K.24	Antecip._Div_Q5K.65
	Antecip._Div_Q.91	Antecip._DivVLR.1	Antecip._Div_Q5K.26	Antecip._Div_Q5K.68
	Antecip._DivVLR.50	Antecip._DivVLR.2	Antecip._Div_Q5K.27	Antecip._Div_Q5K.71
	Antecip._DivVLR.59	Antecip._DivVLR.3	Antecip._Div_Q5K.32	Antecip._Div_Q5K.78
	Antecip._DivVLR.93	Antecip._Set_Q5K.A	Antecip._Div_Q5K.33	Antecip._DivVLR.50
	Antecip._SubVLR.Cultura	Antecip._Set_Q.A	Antecip._Div_Q5K.35	Antecip._DivVLR.59
	Boleto_Div_Q5K.30	Antecip._SetVLR.A	Antecip._Div_Q5K.38	Antecip._DivVLR.85
	Boleto_Div_Q5K.59	Antecip._T_Q5K	Antecip._Div_Q5K.41	Antecip._DivVLR.90
	Boleto_Div_Q5K.79	Antecip._TVLR	Antecip._Div_Q5K.8	Antecip._DivVLR.93
	Debit_Div_Q5K.38	Boleto_Div_Q5K.3	Antecip._DivVLR.17	Antecip._Sub_Q5K.AdmPub
	Debit_Div_Q5K.50	Boleto_Div_Q.3	Antecip._DivVLR.29	Antecip._Sub_Q5K.Atimob
	Debit_Div_Q5K.52	Boleto_DivVLR.3	Antecip._DivVLR.30	Antecip._Sub_Q5K.Cultura
	Debit_Div_Q5K.55	Debit_Div_Q5K.2	Antecip._DivVLR.31	Antecip._Sub_Q5K.infCom
	Debit_Div_Q5K.73	Debit_Div_Q5K.3	Antecip._DivVLR.32	Antecip._SubVLR.Cultura
	Debit_Div_Q5K.80	Debit_DivVLR.1	Antecip._DivVLR.33	Antecip._SubVLR.ServDom
	Debit_Div_Q5K.91	Debit_DivVLR.2	Antecip._DivVLR.35	Boleto_Div_Q5K.55
	Debit_Div_Q5K.94	Debit_DivVLR.3	Antecip._DivVLR.38	Boleto_Div_Q5K.59
	Debit_Div_Q.64	Debit_SetVLR.A	Antecip._DivVLR.41	Boleto_Div_Q5K.79
	Debit_Div_Q.7	Debit_T_Q5K	Antecip._Sub_Q5K.Constr	Boleto_Div_Q5K.99
	Debit_Div_Q.72	Credit_Div_Q5K.2	Antecip._Sub_Q5K.EletGas	Boleto_DivVLR.51
	Debit_Div_Q.88	Credit_Div_Q5K.3	Antecip._Sub_Q5K.Ext	Boleto_DivVLR.60
	Debit_Div_Q.91	Credit_DivVLR.2	Antecip._SubVLR.EletGas	Boleto_DivVLR.72
	Debit_Div_Q.93	Credit_DivVLR.3	Boleto_Div_Q5K.30	Boleto_DivVLR.91
	Debit_DivVLR.12	Credit_Set_Q5K.A	Boleto_Div_Q5K.31	Boleto_Sub_Q5K.OrgInt
	Debit_DivVLR.58	TED&PIX_Div_Q.3	Boleto_Div_Q5K.9	Boleto_Sub_Q5K.ServDom
	Debit_Sub_Q5K.AlojAlim	TED&PIX_DivVLR.2	Boleto_SubVLR.Ext	Debit_Div_Q5K.50
	Debit_Sub_Q5K.Educ	TED&PIX_DivVLR.3	Debit_Div_Q5K.11	Debit_Div_Q5K.52
	Debit_SubVLR.OrgInt	TED&PIX_T_Q	Debit_Div_Q5K.13	Debit_Div_Q5K.55
	All_Div_Q5K.30		Debit_Div_Q5K.14	Debit_Div_Q5K.71
	All_Div_Q5K.55		Debit_Div_Q5K.15	Debit_Div_Q5K.80
	All_Div_Q.35		Debit_Div_Q5K.16	Debit_Div_Q5K.91
	All_Div_Q.74		Debit_Div_Q5K.18	Debit_DivVLR.50
	All_DivVLR.72		Debit_Div_Q5K.20	Debit_DivVLR.52
	AllSub_Q5K.AlojAlim		Debit_Div_Q5K.22	Debit_DivVLR.58
	Credit_Div_Q5K.91		Debit_Div_Q5K.23	Debit_DivVLR.72
	Credit_Div_Q.59		Debit_Div_Q5K.27	Debit_DivVLR.79
	TED&PIX_Div_Q5K.5		Debit_Div_Q5K.29	Debit_DivVLR.82
	TED&PIX_DivVLR.37		Debit_Div_Q5K.30	Debit_Sub_Q5K.Educ
	TED&PIX_DivVLR.7		Debit_Div_Q5K.32	Debit_SubVLR.AtAdm
			Debit_Div_Q5K.38	Debit_SubVLR.Cultura
			Debit_DivVLR.12	Debit_SubVLR.OrgInt
			Debit_DivVLR.14	Credit_Div_Q5K.90
			Debit_DivVLR.18	Credit_Div_Q5K.91
			Debit_DivVLR.20	Credit_DivVLR.51
			Debit_DivVLR.26	Credit_DivVLR.90
			Debit_DivVLR.28	Credit_DivVLR.91
			Debit_DivVLR.32	Credit_SubVLR.Transp
			Debit_DivVLR.35	TED&PIX_Div_Q5K.55
			Debit_DivVLR.36	TED&PIX_DivVLR.79
			Debit_DivVLR.9	
			Debit_Sub_Q5K.EletGas	
			Debit_SubVLR.EletGas	

(…continues)

	Total GDP	Agricultural Sector	Industrial Sector	Services Sector
Payment Series*			All_Div_Q5K.30	
			All_DivVLR.19	
			All_DivVLR.30	
			All_DivVLR.36	
			All_DivVLR.37	
			All_DivVLR.6	
			All_DivVLR.7	
			AllSubVLR.Ext	
			Credit_Div_Q5K.11	
			Credit_Div_Q5K.28	
			Credit_Div_Q5K.35	
			Credit_Div_Q5K.38	
			Credit_DivVLR.11	
			Credit_DivVLR.12	
			Credit_DivVLR.19	
			Credit_DivVLR.31	
			Credit_DivVLR.35	
			Credit_DivVLR.7	
			Credit_DivVLR.9	
			Credit_Sub_Q5K.EletGas	
			TED&PIX_Div_Q5K.29	
			TED&PIX_Div_Q5K.5	
			TED&PIX_DivVLR.11	
			TED&PIX_DivVLR.16	
			TED&PIX_DivVLR.21	
			TED&PIX_DivVLR.23	
			TED&PIX_DivVLR.30	
			TED&PIX_DivVLR.37	
			TED&PIX_DivVLR.39	
			TED&PIX_DivVLR.6	
			TED&PIX_DivVLR.7	
			TED&PIX_SubVLR.EletGas	
Other economic series (codes)**	BM	Acucar	CEnergia	Comercio
	Comercio	Banana	CPetroleo	ICS
	CPetroleo	Batata1	PAco	IES
	IndConfCon	Batata2	PIM_BConsumo	IndConfCon
	IndConfFut	Batata3	PIM_BI	IndConfFut
	PAco	Cacau	PIM_Capital	ISAS
	PIM_BConsumo	Castanha	PIM_ConsDur	PMS_InfCom
	PIM_Capital	Cereal	PIM_Construcao	PMS_OServ
	PIM_ConsDur	Laranja	PIM_Ext	PMS_SFamilia
	PIM_ConsSemiNao	Tomate	PIM_Trans	PMS_SProf
	PIM_Ext	Uva	PIMG	PMS_Transp
	PMaqAgro		PPetroleo	PMS_T
	PMS_InfCom		PVeiculos	PMS_Vol
	PMS_OServ		UCI	VVarAuto
	PMS_SFamilia			VVarComb
	PMS_SProf			VVarConstr
	PMS_Transp			VVarEquip
	PPetroleo			VVarFarm
	TaxaOcup			VVarMov
	VCaminhao			VVarVest
	VVarAuto			VVeiculos
	VVarConstr			
	VVarEquip			
	VVarVest			
	VVeiculos			

*With Div = CNAE Division, Sub = CNAE Subsection, Q5 = Number of transaction above BRL 5000.00, Q = Total transactions, VLR = Deflated Value, All = All Payments Instruments.

**See Annex I