# Dissecting Brazilian business agricultural cycle in high-dimensional and mixed-frequency context*

André Nunes Maranhão†
Nicole Rennó Castro‡

**Resumo**

A análise de ciclos de negócios tem se deparado nos anos recentes com o desafio de base de dados em alta-dimensão e com frequências mistas em termos do inicio das séries temporais disponíveis. O presente estudo trata essas questões para o caso do ciclo agropecuário brasileiro, propondo um teste entrópico de informação relativa para ambiente de hiper-dimensão e uma versão de frequência mista para o modelo dinâmico fatorial generalizado apresentado por Forni et al. [2000]. Os resultados mostram um conjunto maior de séries temporais selecionadas no período corrente em relação às demais defasagens, concentradas nas categorias relacionadas ao crédito agropecuário e à produção agropecuária. O ciclo agropecuário brasileiro estimado evidencia a alta comunalidade da categoria climáticas, pró-cíclicas e antecedentes, e de crédito, com efeitos tanto pró-cíclicos quanto contra cíclicos, defasados e antecedentes.

**Palavras Chave**: Ciclo agropecuário brasilieiro, Teste entropico de informação relativa para hiper-dimensão, Modelo dinâmico fatorial generalizado, Frequência mista em séries temporais.
**JEL Code**: C58, G15, G17.
**Indicação da Área**: 11 - Economia Agrícola e do Meio Ambiente.

**Abstract**

Business cycle analysis has in recent years been faced with the challenge of a high-dimensional database with mixed frequencies in terms available time series starting points. The present study addresses these issues in the case of the Brazilian agricultural cycle, proposing an hyper-dimension environment entropic test of relative information, and proposes a mixed-frequency version for the generalized factorial dynamic model presented by Forni et al. [2000]. The results show a larger set of time series selected in the current period in relation to the other lags, concentrated in the categories related to agricultural credit and agricultural production. The Brazilian agricultural cycle has evidenced the high commonality of the categories: climatic, pro-cyclical and leading, and credit with both pro-cyclical and counter-cyclical, lagged and leading effects.

**Key Words**: Brazilian agriculture cycle, Hyper-Dimensional entropic relative information test, Generalised dynamic factor model, Time series mixed-frequency.

*This is a preliminary version, without proper english text revision. Any errors in this context should be disregarded at this time in favor of the presented results.

†Economista, Estatístico, Doutor em Economia pelo Programa de Pós-Graduação em Economia da Universidade de Brasília (UNB). Assessor Sênior da Diretoria de Crédito do Banco do Brasil. Email: andrenmaranhao@gmail.com

‡Economista, Doutora em Economia Aplicada - Esalq/USP, Pesquisadora Equipe Macroeconomia CEPEA-Esalq/USP

# 1 Introduction

The classical business cycle considers the fluctuations of the level of economic activity, see Harding and Pagan [2002], while the deviation cycle considers the fluctuations around some trend. A third approach to business cycle representation is the so-called growth-rate cycle. The calculation of growth rates can, however, also be interpreted as a detrending device, see for a discussion Harding and Pagan [2006]. The recession in the classical cycle is characterized by an absolute decline in the level of economic activity, i.e. negative growth rates. The recession in the deviation cycle is characterized by economic growth rates below potential growth. The classic recession phases are therefore always a subsample of the recession phases of deviation cycles. Deviation cycles have gained popularity as periods of negative growth rates have been rare in industrialized countries since World War II. Thus, deviation cycles were more naturally related to fluctuations in employment and unemployment. Moreover, the concept of a deviation cycle as the gap between actual and potential output has gained political relevance through a stronger focus on Taylor-rule driven monetary policy and cyclically adjusted government balances.

Over the years, studies of economic business cycles have become the basis of theoretical instruments for empirical models; This happened slowly, with the emergence of theoretical advances and new estimation methodologies in different periods. There has been important advance in the literature on the subject since the presentation of the theoretical basis for the cycles by Kydland and Prescott in the work "*Time to build and aggregate fluctuations*" (Kydland and Prescott [1982]). Two main views were counterbalanced, one related to market failures and another to failures in the information mechanism. In the first case, the hypothesis is that the cycles are the result of market failures, and that when the economy is in recession times, the economic agents would not reach equilibrium – see Summer [1986]. The other line argues that the cycles are the result of a dynamic equilibrium; it is argued that economic agents have complete information on economic shocks (Kydland and Prescott [1990]) and incomplete information on real wages (Lucas Jr [1972]; Lucas [1973]). In this framework, the recession would occur when workers underestimate the value of real wages and the cycles would be the result of a failure in the information mechanism.

Applied studies on business cycles have emerged more intensely since the 1990s, and the new methodology proposed by Hamilton [1989] became widely used. In his paper, Hamilton [1989] considers that the American GDP series does not follow a linear process, that is, that it is subject to discrete changes in its data generating process that follow non-observed Markov Switching chain. At the same time, the econometrics of business cycles began to be studied with the factorial dynamic models that seek to capture the co-movement of a set of variables, see Stock and Watson [2011]. The econometric advances were incorporating new topics to the proposed models, and the Markov Switching was included in DFM by Chauvet [1998], Kim and Nelson [1998] and Kim and Yoo [1995]. As databases became increasingly large, econometric methods had to be adapted to this new reality. An unexplored situation was the context of high-dimensional time series, in which the number of variables may exceed the number of observations. In the econometric field of study of business cycles, this adaptation to the high dimension data in applied context occurred recently with the generalized factorial dynamic models, although the model was presented by Forni et al. [2000]. In young economies, the number of time series available for study of the economy has only recently been consolidated (at least 30 years or less). This creates a new challenge, related to the fact that the data series begin to have information available at different points in time. Econometric models have also recently adjusted to this challenge in the field of economic business cycles, with the study of Bańbura and Modugno [2014].

The Brazilian agricultural business cycle (growth) currently faces these challenges, high-dimensional database and series with mixed-frequencies starting point. The present study proposes an alternative solution to these questions, with special focus on the relation between climate and agriculture. Given

the high-dimensional characteristics of the database, the study goes through the process of selection of variables that considers not only the contemporary effects but the cyclical characteristic to be estimated. The issues addressed in this study make it a pioneer in literature.

Until the early years of the twentieth century, scientists who analyzed the growth of nations understood that the climate was among its main determinants; but from the middle of this century the climate would have disappeared from the literature of economic development, giving rise to analyzes of investments, trade policies and education (Nordhaus [1993]). According to Nordhaus (1993), in the years before and close to 1993, the climate reemerged in the literature on international environmental issues due to growing concern about climate change and global warming. Gradually, climate has returned to appear in the literature with an economic focus. Nordhaus (1993) points out that climate change vulnerability should be greater for those sectors that depend on naturally occurring rainfall and temperatures. Therefore, agriculture is likely the most vulnerable sector to climate variability and change (Nordhaus [1993]; Mendelsohn et al. [1996]; Mendelsohn et al. [1996]; Rosenzweig et al. [2014]). Many uncertainty factors affect agricultural production and climate and weather are among those factors, having major impacts on agricultural productivity (Gornall et al. [2010]; Anwar et al. [2013]).

For several reasons, Brazil is an important case study to assess the subject. The country has significant participation in the world commodity market – the gross value of Brazil's agricultural products is the fourth-largest in the world and the country is a major exporter of several food products (Food and Agriculture Organization of the United Nations Statistic Division (FAOstat), 2014). In addition, the negative effects of climate on agriculture have more severe negative consequences in countries whose economy is most heavily tied to this sector (Belloumi, 2014); and agribusiness has a major impact on Brazil's Gross Domestic Product (GDP), its trade balance, its employment level, and its level of technological innovation and adaptation. Namely, in 2017, agribusiness was responsible for about 20% of the county's GDP and employment (Center for Advanced Studies on Applied Economics - Cepea, 2016) and 45% of the total exports revenue in 2014 (Ministry of Agriculture, Livestock, and Food Supply - MAPA, 2014).

Besides, the vulnerability of agriculture to climate is accentuated in countries with a predominantly poorer population due to the limited adaptability of producers, with restrictions on access to technologies and resources to adapt to or mitigate the impacts caused by climate variables (Morton [2007]; Millner and Dietz [2015]). Brazil has an extremely heterogeneous agriculture, so that negative effects of climate on the sector contribute to increase inequality between poor and capitalized farmers. The relationship between climate and agricultural supply is also important in terms of its effect on the level of consumer prices and, therefore, on the purchasing power of the population, especially lower income. Focusing on climate change, Wheeler and Von Braun [2013] highlight that food security is strongly affected by staple food prices for the poor, and depending on the world food equation level, small supply shocks can have large impacts on prices. Data from the Consumer Expenditure Survey (POF) - IBGE (2008) show that, in 2008, food accounted for 19.8% of family consumption expenditures in Brazil, reaching 30% for the poorest families. Finally, agricultural production can be more sensitive to climate in low latitude areas, as most of Brazilian territory (Gornall et al. [2010]; Rosenzweig et al. [2014]).

Several questions will be answered at each stage of the study. In Section 2.1 we present the chronology of Brazilian agriculture events; in section 3, focused on the methodology, we present the procedures of selection of variables and estimation of the cycle in the high-dimensional and mixed frequency context. In Section 4 we describe in detail the treatments applied to the high-dimensional database. The results are presented in Section 5, in which we present, sequentially, the results of the test of hyper-dimensional entropic relative information test, of the estimation of the Brazilian agricultural cycle and of the dissection of this cycle by category level. Finally, we concluded in section 6.

# 2 Brazilian agriculture chronology

The Brazilian agricultural GDP had a consistent growth trend between 1979 and 2017, growing 281 % in the whole period. Of the 38 years analyzed, in 31 the agricultural GDP expanded Estatística [2018a]. The most intense growth occurred in the 1990s, and continued in the 2000s. In the 1980s, the sector's GDP grew relatively little from 1985 to 1990, although it grew rapidly in the first half of the decade (Table 1).

| Year | 5-years | 10-years |
|---|---|---|
| 1985 | 20,6% | |
| 1990 | 5,6% | 27,4% |
| 1995 | 22,6% | |
| 2000 | 17,4% | 44,0% |
| 2005 | 26,9% | |
| 2010 | 17,4% | 49,0% |
| 2015 | 17,8% | |

Table 1: Agriculture GDP growth.

The conditions for the expansion of modern agriculture in Brazil arose in the 1960s, with the institution of the Rural Credit System and policies of subsidized credit, rural extension and agricultural research by public institutions Buainain et al. [2013]. This modernization process was based on the conception of the Green Revolution between the 1960s and 1970s, characterized by crop genetic improvement and the intensive use of fertilizers and agrochemicals.Garcia [2014].

Beginning in 1965, the Brazilian agriculture underwent advanced technification and industrialization, with the mechanization of the sector and the internalization of the sectors producing inputs and machinery and equipment, so that their use was no longer limited to import capacity (Kageyama et al. [1987] and Staduto et al. [2004]). During this period, the foundation of institutions that played an important role in the mentioned processes took place: the Brazilian Agricultural Research Corporation (Embrapa) in 1973 and the Brazilian Entity for Technical Assistance and Rural Extension (Embrater) in 1974.

From the 1980s, the modernization of agriculture spread throughout the territory and traditional productive systems began to be replaced by new practices and organizational forms; a process that resulted in productivity increases for the country's agriculture Garcia [2014]. But at the same time, from 1987, macroeconomic and sectoral policies started to reduce the incentive to agriculture in the face of the government fiscal crisis Bacha [2004]. The contractionary policies led to the reduction of subsidies, affected the rural credit policy (with reduction of volume and increase of interest rates) and the Minimum Price Guarantee Policy (PGPM), and also implied in a reduction of public services of rural extension and of agricultural research Bacha [2004]. In general, between 1987 and 1989, the federal government's applications for agriculture fell by 46 % Barros et al. [2002].

The 1980s was also characterized by a significant reduction in real agricultural prices, following the international trend, and reversing the scenario of high prices of the 1970s Barros [2014]. The continuity of agricultural growth in the 1980s, even in the face of this scenario, is explained by the increase in productivity Barros [2014]. In addition, in the first half of the 1990s, the set of reforms related to the Washington Consensus, aimed at economic liberalization, led to trade liberalization and the agricultural sector started to face international competition and the high subsidies of developed countries Garcia [2014], Barros [2014].

The Brazil's Real Plan was also a milestone for agriculture and livestock, since with inflation control the sector was able to plan and not restrict itself to the differences between the evolution of its financial commitments (inflation-adjusted) and the evolution of agricultural prices Gasques et al.

[2004]. The negotiation of agricultural debt in 1995 also impacted on a more favorable scenario for the sector from that period Gasques et al. [2004]. However, despite the economic stability and record harvests, the appreciation of the exchange rate promoted by the Real Plan, and maintained until 1998, implied a loss of competitiveness of the national export sector with negative impacts on the agricultural area Bacha [2004], Garcia [2014].

Also in the 1990s, the mechanization of agriculture and livestock was generalized to traditional and important crops in Brazil, such as coffee, sugar cane and cotton; before that, the focus of mechanization was mainly on the production of grains, since the same technologies developed for the developed countries were implanted in Brazilian agriculture Staduto et al. [2004]. From the end of the 1990s onwards, agricultural development changed radically, and the significant results obtained by the agricultural sector resulted from the improvement of an environment of innovations, with the diffusion of knowledge and technical apparatus and the continuous search for productivity Buainain et al. [2013]. Also in the 1990s, another important milestone that positively influenced the process of technological modernization in the sector was the adoption of the Kandir Law in 1996 Garcia [2014].

Since the 2000s, the favorable situation in the international commodity market, coupled with new stimuli of agricultural policy and the maturity achieved by some agricultural chains that were restructured in the 1990s, boosted agriculture and livestock Garcia [2014]. In more recent years, there has been the widening and deepening of the innovation process, characterized by the introduction of harvesting and post-harvest technologies, genetically modified seeds, precision agriculture, increased confinement / semi-confinement practices, genetic improvement animals, among others Garcia [2014].

Summarizing, data from Gasques et al. [2014] show that the growth of the Brazilian agricultural product accelerated between 1990 and 2012, with the average annual growth rate going from 3.38% in the period from 1980 to 1989 to 4.71% between 2000 and 2012. These data also show that this result was mainly due to the expansion of the sector Total Factor Productivity (TFP), whose growth rate also grew at increasing rates during the period: 2.17 % in 1980 to 1989, 3.13 % in 1990 to 1999 and 4.06 % in 2000 to 2012.

About specific annual shocks it can be said that greater deviations from the growth trend of agricultural GDP are usually explained by climatic events or health issues Barros and Castro [2017]. From the perspective of the negative shocks of the 38 years between 1979 and 2017, agricultural GDP fell only in 1982 e 1983, 1986, 1990, 2009, 2012 and 2016. On the other hand, intense crop growth, with annual rates exceeding 8% occurred in 1980, 1985, 1987, 2002 and 2003, 2013 and 2017 (results also explained by a small harvest in previous years for the relevant growths of 1987, 2013 and 2017).

Data from Estatística [2006] show that the period 1981-1983 was characterized by crop loses for products such as cassava, maize, soybean, rice, cashew, coffee, potato and beans. In 1990, the reduction of agricultural production mainly responded to the grain dynamics. From the regional perspective, wheat production in the South decreased 44 %; for soybeans, reductions were around 30 % in the Southeast and Midwest and almost 5 % in the South (the main producer at the time); in the case of corn, the crop loss was 30 % in the Southeast region (second largest producer) Estatística [2018b]. Marquetti et al. [1991] point out that this reduction was due to climatic problems in the summer crops.

The drop in agricultural GDP in 2009 resulted mainly from the reduction in production of grain and coffee crops Estatística [2018b]. For the grains, the most intense losses were verified in the South of the Country, due to climatic problems with prolonged drought in critical period. In the case of coffee production, 2009 is a year of decline in the biennial crop cycle, which occurs in practically all producing states.

In 2012, the strong loss in Brazilian soybean production occurred due to the long drought in South America during periods of grain development, a fact related to the occurrence of La Ninã (Cepea/Soja [2013]). The Research Center also points out that the occurrence of heavy rains and hail damages the rice and wheat harvests in Rio Grande do Sul in the same year (Cepea/Arroz [2012]). In 2016, agricultural GDP had the biggest drop in 20 years. According to information from the Brazilian

Agriculture and Livestock Confederation (CNA), this fall was related to the climatic problems that affected several regions and crops in Brazil, with negative effects of the prolonged drought at the beginning of the year in Mato Grosso do Sul, Mato Grosso, Goiás and in the Matopiba region, as well as excessive rains in the Rio Grande do Sul State. Excessive heat in Minas Gerais and São Paulo also undermined the orange crops, according to information from the Fund for Citrus Protection (Fundecitrus).

In relation to the most expressive crop growth observed in the period, in 1980, important production increases for sugarcane, oranges, and grains influenced the results. A similar scenario was repeated in 1985 and 1987, but with increases in coffee production as well Estatística [2018b]. In 2002, sugarcane, orange and soybeans stood out in the boost to agricultural production, and in 2003, soybeans, corn, wheat and sugarcane presented important expansions.

In 2013, increases in area and productivity, especially for soybeans and the winter corn crops, were responsible for the increase in production in the year, according to Conab [2013]. In that year, according to the Company, although climatic conditions were not favorable, they did not affect the productivity of these crops. By 2017, the area planted with grains was the largest in history, which, coupled with favorable climatic conditions for most crops and producing regions, has led to a record level of production of tons of grains Conab [2017]. The chronology of this section will be used to evaluate the quality of the dating process and the agriculture cycles estimated by different models. In this way it will serve as a qualitative reference for the obtained empirical results.

# 3 Methodology

## 3.1 Hyper-Dimensional entropic relative information test procedure

This section we propose relative information measures for mixed-frequency data, connected to Kullback-Leibler numbers, this measure was use in a formal statistical test. By ordering the series of the data set according to these measures, we are able to obtain a subset of the data set that is most informative to model a variable of interest. The method can be used as a first step in the construction of a dynamic factor model. The objective is confine attention to a subset of the series instead of having to monitor all series in a data set. The question seems especially relevant for factor models, which exploit the idea that movements in a large number of series are driven by a limited number of common 'factors'. For a recent overview see Bai et al. [2008]. Although convergence of factor estimates requires large cross-sections and large time dimensions, see e.g. Forni and Lippi [2001] and Bai [2003], the data set need however not be very large to obtain reasonably precise factor estimates. Bai and Ng [2002] also conclude that the number of series need not be very large to get precise factor estimates.

Oversampling refers to the situation in which the data are more informative about some factors than the other ones. Including more variables in an oversampled data set could result in more precise factor estimates, which do however not improve the forecasting performance for the target variables that depend on the less dominant factors. Building upon **?**, this paper exploits concepts from information theory, in particular Kullback-Leibler numbers, to analyse information in the data. We propose a relative information measures based on gaussian distributed data with a clear link to Kullback-Leibler numbers for mixed-frequency data. We follow the same idea of Jacobs and Otter [2008], they apply similar information concepts to derive a formal test for the number of common factors and the lag order in a dynamic factor model, however in mixed-frequency context. Ordering the series of the data set according to these measures enables us to identify a subset of the data set that is most informative to modelling a variable of interest. The method can be used as a first step in the construction of a dynamic factor model.

### 3.1.1 Defining hyper-dimension.

The definition of hyper-dimension and high-dimension is not consensual in time series econometrics. The simple rule that $N > T$ was adopted by Belloni and Chernozhukov [2011] and Stock and Watson [2014]. However, the databases treated by these authors in the applications had the $\frac{N}{T}$ ratio a little above 1. Other authors like Bai et al. [2008] denominate $N > T$ as being large panel data, while other authors like Song and Bickel [2011] name it as a rich data environment. In the computational field of Big Data the definition of hyper-dimension deals with data with millions (or even billions) of record, according to Mohapatra and Majhi [2015]. Thus, for our study we will define hyper-dimension for time series analysis as being:

$$Y_{i,t}^{Hyper-dimensional}, \quad \text{if} \quad d_{HyD} = \frac{N}{T} \geq 5$$

with

$$i = 1, ..., N; \;\; t = 1, ..., T$$

Our case, we'll work with $d_{HyD} = \frac{2571}{152} > 16$ or $d_{HyD} = \frac{2571}{95} > 27$.

## 3.2 Entropy relative information

First, defining $f_1(\tilde{x}) : \tilde{x} \sim N_N(0, \Gamma), \text{with} \Gamma = C\Lambda C'$ be the density function of an N-dimensional data vector $x$[1], so $f_1(x) : x \sim N_N(0, \Lambda)$, let's also define $f_2(\tilde{x}) : \tilde{x} \sim N_N(0, I)$, then $f_2(x) : x \sim N_N(0, I)$, in all cases $x = C'\tilde{x}$. The so-called Kullback-Leibler numbers are defined as:

$$G_1 = E_{f_1}\left(\log\left(\frac{f_1(x)}{f_2(x)}\right)\right) \text{ and } G_2 = E_{f_2}\left(\log\left(\frac{f_2(x)}{f_1(x)}\right)\right) \tag{1}$$

with $G = G_1 + G_2$ is the measure of information for discriminating between the two density functions with $G = 0$ when $f_1(x) = f_2(x)$ and $G = \infty$ when we have perfect discrimination, as we can see in Golan and Maasoumi [2008], Young and Calvert [1974] and Burnham and Anderson [2002]. Note that, $tr(\Lambda) = tr(\Gamma) = N$, in this case we have $G_1 = -\log\det(\Lambda)$ and $G_2 = \log\det(\Lambda) + \frac{1}{2}(tr(\Lambda^{-1}) - N)$, so

$$\begin{aligned} 2G_2 &= 2\log\det(\Lambda) + +(tr(\Lambda^{-1}) - N) \\ 2G_1 &= -2\log\det(\Lambda) \end{aligned}$$

Then

$$2G = tr(\Lambda^{-1}) + N = tr(\Lambda^{-1}) + tr(\Lambda) = \sum_{j=1}^{N} \frac{(1 - \lambda_j^2)}{\lambda_j} \tag{2}$$

Therefore, $G$ is small (not discriminating) if the eigenvalues $\lambda_j$ are close to 1, but becomes large (discriminating) for "small" eigenvalues. When we considered Gaussian case, alternative measures of entropy and information can be used, for this propose, we define $x_t$ as N-dimensional vector of observed data at time $t = 1, ..., T$[2] normalized, and normally distributed with mean zero and variance,

[1]The index of time was suppressed without loss of generality.

[2]Nothing was said about data with mixed frequencies, however we will present the general case, and then its adjustment for this particular case.

so $x \sim N(0,\Gamma)$ with $E(x_t x_t^{'}) = \Gamma$, $tr(\Gamma) = N$.The entropy as measure of disorder for a stationary, normally distributed vector can be defined as Golan and Maasoumi [2008]:

$$2\mathbb{E}_x = cN + \log\det(\Gamma)$$

with $c = \log(2\pi) + 1 \approx 2.84$, where $2\mathbb{E}_{x,max} = cN$ when $\Gamma = I_N$ as we ca see in Golan and Maasoumi [2008] and **?** Therefore, the information or negentropy is defined as

$$2Inf_x = 2(\mathbb{E}_{x,max} - \mathbb{E}_x) = -\log\det(\Gamma) \geq 0 \tag{3}$$

If $\Gamma = I_N$ then we have zero value. Considering all the details, we can then define relative information (RI) as

$$\begin{aligned} RI_{(N,t)} &= \frac{2\mathbb{E}_{x,max} - \mathbb{E}_{x(N)}}{2\mathbb{E}_{x,max}} \\ &= \frac{2Inf_x}{2\mathbb{E}_{x,max}} \\ &= \frac{2Inf_x}{cN} \end{aligned} \tag{4}$$

Note that if $\mathbb{E}_{x(N)} = \mathbb{E}_{x,max}$ then $RI_N = 0$, as well as $\mathbb{E}_{x(N)} = 0 \Rightarrow RI_N = 1$.

### 3.3 Entropic relative information measure in factor model context

We will present the link between $IR_N$ and factor models (static or dynamic) context, therefore consider $x_{it} = (x_{1t}, ..., x_{nt})^{'}; n \in N, t \in T$ stationary N-dimensional vector process with zero mean be driven by $k$ factors, as

$$x_{it} = \chi_{it} + \xi_{it} = \sum_{j=1}^{N} b_{ij}(L)u_{jt} + \xi_{it} = B_N F_t + \epsilon_t \tag{5}$$

$$x_{it} \in \mathbb{R}^N,\ F_t \sim N_k(0, I_k),\ \epsilon_t \sim N_k(0, \Psi_{11})$$

where $\chi_{it}$ is the common component, $\xi_{it}$ is the idiosyncratic component, $b - ij(L) = B_N = B_0^n + B_1^n L + ... + B_s^n L^s$, represents the (dynamic) loadings of order s, $u_{ij}; j = 1, ..., q; t = 1, ..., T$ are common shocks mutually orthogonal white noise processes with unit variance. The variance between the first $N$ elements of $x_{it}$ is equal to $\Gamma(N) = B_N B_N^{'} + \Psi_{11}$ as we can see Bai et al. [2008].

When we add a variable $x_{N+1,t}$, we have

$$\begin{pmatrix} x_{it} \\ x_{N+1,t} \end{pmatrix} = \begin{pmatrix} B_N \\ b_{N+1} \end{pmatrix} F_t + \begin{pmatrix} \epsilon_t \\ \epsilon_{N+1,t} \end{pmatrix} \tag{6}$$

with covariance

$$\Gamma(N+1) = \begin{pmatrix} \Gamma(N) & \Gamma_{12} \\ \Gamma_{21} & \Gamma_1 \end{pmatrix}$$

where $\Gamma_{12} = B_N b_{N+1}^{'} + \Psi_{12}$ and $\Psi_{12} = E(\epsilon_t \epsilon_{N+1,t})$, as $x_{it}$ is normalized we have $b_{N+1} b_{N+1}^{'} + \sigma_{N+1}^2 =$

1, with $\sigma^2_{N+1} = E(\epsilon^2_{N+1,t})$. Using the rule of determinants for partitioned matrices we get

$$\det(\Gamma(N+1)) = \det(\Gamma(N))(1 - a_{N+1})$$

where $a_{N+1} = (b_{N+1}B^{'}_{N} + \Psi_{12})\Gamma(N)^{-1}(B_N b^{'}_{N+1} + \Psi_{12})$ and $0 \leq (1 - a_{N+1}) \leq 1$. as we have the results

$$RI_{(N,t)} = \frac{-logdet(\Gamma(N))}{cN}$$

$$RI_{(N+1,t)} = \frac{-logdet(\Gamma(N+1))}{c(N+1)} = \frac{-logdet(\Gamma(N) + (1 - a_{N+1}))}{c(N+1)}$$

In this way, we have the relation between $RI_{(N+1,t)}$ and $RI_{(N,t)}$ as

$$RI_{(N+1,t)} = RI_{(N,t)} - \frac{1}{N+1}\left(\frac{\log(1 - a_{N+1})}{c} + RI_{(N,t)}\right) \tag{7}$$

Therefore, there is only addition of relative information if $RI_{N+1,t} > RI_{(N,t)}$, or in the condition $-log(1 - a_{N+1}) > cRI_{(N,t)} \Rightarrow \mathrm{E}(\mathrm{x}_{N+1,t}x^{'}_{N,t}) = (b_{N+1}B^{'}_{N} + \Psi_{12} \neq 0)$ this is the condition that will be used to apply a formal statistical test. From Equation 7 we have $RI_{(N+1,t)} = RI_{(N,t)}$ if $a_{N+1} = 1 - \exp(-cRI_{(N,t)})$, Whenever $RI_{(N,t)}$ is close to zero, $RI_{(N+1,t)}$ increases for relative small values of $a_{N+1}$ whereas if $RI_{(N,t)}$ is close to one, $a_{N+1}$ should be close to one to allow $x_{N+1,t}$ to add relative information. We can simplify the Equation 7 considering $\Gamma(N) = C\Lambda C^{'}$ and the linear transformation $\tilde{x}_t = U^{'}\Lambda^{-\frac{1}{2}}C^{'}x_t$ and $\tilde{x}_{N+1,t} = \upsilon^{-1}x_{N+1,t}$, being $U$ orthogonal and $\upsilon^2 = 1$ obtained by the singular value decomposition with $\Lambda^{-\frac{1}{2}}C'\Gamma = U\Sigma\upsilon$ where $\Sigma = (\phi, 0, ..., 0)^{'}$, from where do we also have $\Gamma_{12} = 0 \Rightarrow \Sigma = 0$. Therefore, we have

$$\begin{pmatrix} \tilde{x}_{N,t} \\ \tilde{x}_{(N+1,t)} \end{pmatrix} = N_N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tilde{\Gamma}(N+1)\right)$$

where

$$\tilde{\Gamma}(N+1) = \begin{pmatrix} I(N) & \Sigma \\ \Sigma & 1 \end{pmatrix}$$

with $det(\tilde{\Gamma}(N+1)) = det(I_N)(1 - \phi^2_1) \Rightarrow RI_{(N+1,t)} = \frac{-\log(1-\phi^2_1)}{c(N+1)}$, where $\phi_1 \in [0, 1]$ is a coefficient of canonical correlation. Because $RI_{(N,t)} = 0$ by hypothesis is zero in Equation 7, so that there is gain of relative information, we have $\tilde{RI}_{(N+1,t)} = \frac{-log(1-\phi^2_1)}{c(N+1)}$.

## 3.4 Hyper-Dimensional entropic relative information test procedure

### 3.4.1 Formal entropic relative information test

The first step is replace $\tilde{\Gamma}(N+1)$ by a consistent estimate $\hat{\tilde{\Gamma}}(N+1)$ and applying the same procedure yields $\hat{\tilde{RI}}_{(N+1,t)} = \frac{-\log(1-\hat{\phi}^2_1)}{2}$ under $H_0 : \phi_1 = 0$, to this hypothesis we use Bartlett test statistic

$$t_{(,N)} = -\left[T - \frac{1}{2}(N+2)\log(1 - \hat{\phi}^2_1)\right] = \left[T - \frac{1}{2}(N+2)\right]2\tilde{\hat{RI}}_{N+1,t} \tag{8}$$

$t_{(,N)}$, where $\alpha$, is the significante level, follows asymptotically a $\chi^2$ distribution with $N$ degrees of freedom, see e.g. Muirhead [1982]. Testing the hypothesis $H_0 : \phi_1 = 0$ is similar to test whether

the transformed vector $(\tilde{x}'_{(N,t)}\tilde{x}_{(N+1,t)})'$ has maximum entropy. If the null hypothesis is rejected, the estimated relative information of the transformed variables equals

$$\hat{\tilde{RI}}_{(N+1,t)} = \frac{-\log(1-\hat{\phi}_1^2)}{c(N+1)}$$

### 3.4.2 Hyper-dimensional entropic relative information test procedure

The series are transformed by taking logarithms and/or differencing when necessary to assure approximate stationarity, the series should be adjusted for outliers by replacing the observations of the transformed variables using some robust method. With previous treatments, the test procedure follows the following steps:

1. We estimated the canonical correlation coefficient $\hat{\phi}_1^*$ in the context of mixed-frequency;
2. The relative information measure is calculated: $\hat{\tilde{RI}}^*_{(N+1,t)} = \frac{-\log(1-\hat{\phi}_1^{*2})}{c(N+1)}$;
3. We order the data set according to the relative information measures with respect to target variable using the following procedure:
   - the initial variable of the ordered data set is the target variable;
   - the variable that maximizes the respective relative information from the remaining data is added to the ordered data set, and so on.
4. We test if $\hat{\tilde{RI}}^*_{(N+1,t)}$ is statistically different from zero with $t^*_{(,N)} = \left[T^* - \frac{1}{2}(N+2)\right] 2\tilde{\hat{RI}}^*_{N+1,t} \sim \chi^2_N$. The null hypothesis is that an additional variable is not correlated with the variables already included in the set;
5. We can investigate the existence of relevant relative information in other lags (or leads), for which it is calculated $\hat{\tilde{RI}}^*_{(N+1,t-k)}$ and and apply the same steps.

We call attention to the fact that we estimate $\hat{\phi}_1^*$ using frequency-mixed approach but in the calculation of the test statistic we consider the different time sample sizes of each series. Although this small difference did not alter the overall results, for a small set of series that are on the threshold between having statistically nonzero relative information, the $T$ version instead of $T^*$ eventually ended up with a larger number of series. For a descriptive model, the selection of variables can use series present in different lags (considering only once present in several lags.), in order to increase the descriptive capacity of the factorial model to be estimated. Alternatively we can implement some backdating method for the mixed-frequency time series as Bańbura and Modugno [2014] presents and apply the test procedureIn this thesis the two forms were tested, as the divergence of selected series was small, we opted for first backdating the series (with the data already normalized) and later applying the relative information test..

## 3.5 Generalised dynamic factor models

Recently serious progress has been made in the theory of factor models through the Generalised Dynamic Factor Model (GDFM) of Forni, Hallin, Lippi and Reichlin, hence-forth FHLR (Forni et al. [2000]; Forni and Lippi [2001]; Forni et al. [2001, 2004, 2005]). The model differs from the classic factor model in that it allows the idiosyncratic errors to be weakly serial and cross-sectional correlated to some extent apart from being a non-parametric approach. It thereby combines the so-called "approximate static factor model" of Rothschild and Chamberlain [1982], widely applied

in financial econometrics (e.g. Arbitrage Pricing Theory, APT) and the Dynamic Factor Model of Geweke [1977], Sargent et al. [1977] for which respectively cross-sectional and serial correlation was allowed. The model is dynamic since the common shocks can hit the series at different times as opposed to the static model. The common shocks and components, which are a linear combination of them, are inherently unobservable and are estimated by means of dynamic principal components. While the familiar static principal components are based on an eigenvalue decomposition of the contemporaneous covariance matrix, dynamic principal components are based on the spectral density matrix (i.e. dynamic covariations) of the data and consequently are averages of the data weighted and shifted through time.

## 3.6 Estimating GDFM

As we presented in Section 2.4 we assume that the $N$ time series included in our panel are, after suitable transformations, a realization of a real-valued stationary $N$-dimensional vector process with zero mean $x_{it} = (x_{1t}, ..., x_{Nt})'$ Under the GDFM, satisfying the necessary conditions and assumptions, it is shown that each time series can be decomposed into two components:

$$x_{it} \quad = \quad \chi_{it} + \xi_{it} = \sum_{j=1}^{q} b_{ij}(L)u_{jt} + \xi_{it} \tag{9}$$

$\chi_{it}$ is the common component and $\xi_{it}$ the idiosyncratic component. $b_{ij}(L) = B_n(L) = B_0 + B_1 L + ... + B_s L^s$ represents the (dynamic) loadings of order $s$ which are allowed to differ in coefficient and lags across the series. The $q$ common shocks $u_{jt}; j = 1, ..., q$ are assumed to be mutually orthogonal white noise processes (at all leads and lags) with unit variance, this vector process has a non-singular spectral density matrix, equal to the first $q$ dynamic eigenvalues of the data. The idiosyncratic component is driven by variable-specific shocks, for which the GDFM allows a certain amount of correlation. The dynamic factor structure implies that the idiosyncratic component of any series is orthogonal to all common shocks at any lead or lag. The common shocks $u_{jt}$ are latent and need to be estimated. This is done through the estimation of dynamic principal components. These are obtained by the dynamic eigenvalues and eigenvectors decomposition of the spectral density matrix of $x_{it}$ which is a generalization of the orthogonalization process of the variance-covariance matrix of $x_{it}$ in case of static principal components.

## 3.7 Proposed GDFM model

Even the GDFM model considering $(N, T)$ tending to infinity, computational processing does not exceed a certain number of iterations due to physical memory limitations. In the context of time series of very high dimension they have no alternative but a step of selection of variables. In a context of mixed frequency data, the use of the GDFM model is also compromised. To address all of the issues we propose in this section, changes to the basic GDFM model. The first step is the use of the entropic test of relative information for selection of sub-set of high-dimensional time series.

### 3.7.1 Mixed-frequency generalised dynamic factor model

As previously mentioned Doz et al. [2012] show that maximum likelihood is consistent, robust and computationally feasible also in the case of large cross-sections. To maximise the likelihood over the high-dimensional parameter space they propose to use the Expectation-Maximisation (EM) algorithm. The EM algorithm was first applied for a dynamic factor model by Watson and Engle [1983] on a small cross-section. They cast the model in a state space form and derive the EM steps in the case without missing data. Shumway and Stoffer [1982] show how to implement the EM

algorithm for a state space form with missing data, however only in the case in which the matrix linking the states and the observable is known. Bańbura and Modugno [2014] proposed a general treat data sets with arbitrary pattern of missing data allowing the idiosyncratic component be correlated, the essential idea of the algorithm is to write the likelihood as if the data were complete (with draws of $N(0,1)$) and to iterate between two steps: in the expectation step we'fill in' the missing data in the likelihood, while in the maximization step we re-optimize this expectation. Under some regularity conditions, the EM algorithm converges towards a local maximum of the likelihood. A direct maximisation of the likelihood (or state-space formulation) is computationally not feasible for large $N$. However, as argued in Doz et al. [2012], the computational complexity can be circumvented by means of the Expectation-Maximisation (EM) algorithm. Bańbura and Modugno [2014] offers a solution to problems for which incomplete or latent data yield the likelihood intractable. The essential idea is to write the likelihood as if the data was complete and to 'fabricate" the missing data in the expectation step.

### 3.7.2 Algorithm for estimation of MF-GDFM

The estimation of the MF-GDFM model follows the following steps:

1. All treatments that precede the selection procedure are applied: the series available on a monthly basis were transformed to a quarterly basis by taking averages; deflation of monetary variables; de-seasonalization (when necessary); first-difference (when necessary to guarantee stationary) and finally normalization[3];

2. In order to obtain initial values for the parameters, $\theta(0)$ replace the missing observations in $x_{it}$ (observe that $x_{it}$ already normalized) by draws from $N(0,1)$ distribution;

3. Obtain the back data estimate for the series using Bańbura and Modugno [2014] EM algorithm;

4. The relative information test is applied to select, in different lags, the set of series to be used;

5. GDFM is estimated[4].

Some questions were left for future research, such as the possible bias using reconstructed data, the power of relative information testing among others.

# 4 Data and treatments

The challenges related to the database of young economies are always present in any study. The studies with high time series increase these challenges whereas these data have problems of mixed frequency, data in temporal frequency that requires a disaggregation procedure that will also demand a wide set of covariables, seasonal adjustment procedures that should be applied on a large scale but considering the particular issues of the country, the discontinuities of methodologies that generate old and new series of the same phenomenon.

[3]All the treatments applied and the description of the database used are detailed in the next section.

[4]The number of factors to estimate the model was determined by the test proposed by Alessi et al. [2010].

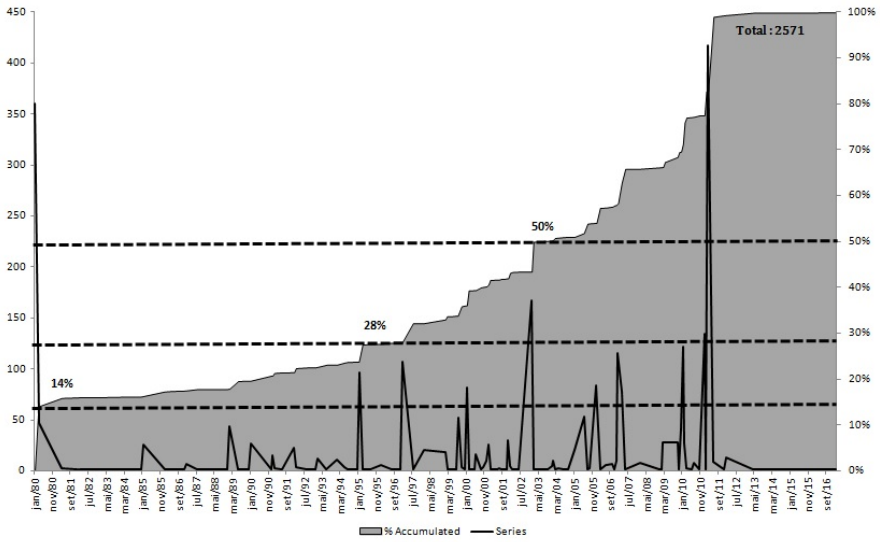


Figure 1: Temporal evolution of Brazilian series available for economic studies.

The treatment for all these issues was presented in this chapter. In order to have a database to cycles study, several steps were taken to make the models estimable. The series were made compatible with the purpose of guaranteeing the maximum possible information; the annual series were disaggregated in a high-dimensional context for quarterly frequency; the series that did not have published their seasonally adjusted versions were submitted to a sequential seasonally adjusted procedure, considering several individual issues of the series besides the Brazilian stylized facts; how the study will focus on growth cycles, we focuses on the growth cycle concept of the business cycle (unlike for instance the NBER method, which measures cycles in the level of the series, see Burns and Mitchell [1947]), defined as the quarter-on-quarter variation of the underlying variables. Being measured at a quarterly frequency, the series available on a monthly basis were transformed to a quarterly basis by taking averages, leaving 152 observations between 1980Q1 and 2017Q4, and 95 observations between 1995Q1 to 2017Q4, for different research's proposes. Furthermore, all activity variables are expressed in real terms. These are obtained by deflating nominal variables by the CPI[5] index. For all other variables (e.g. interest rates and exchange rates) both nominal and real concepts were included in the data set.

The models that will be estimated requires stationary time series. We opted to apply the same stationary procedure to all series. We first-differenced the series' levels by taking percentage changes compared to the previous quarter and by a simple difference when the level possibly exhibits negative values, for price variables (consumer prices, stock prices, ...) percentage changes with respect to the previous quarter of the index were taken. We also applied this procedure to the variables which were stationary from the outset. The reason for this is twofold i) having all variables defined in quarter-on-quarter variations enables to capture the growth cycle concept of the business cycle and ii) taking on variables in their level, even when stationary, would seriously disturb the mutual relations in the frequency domain causing phase shifts and thus invalid deduced time lags as we can see in Cohen [2001] deducing and comparing time lags from both concepts would be improper to do. Interest rate spreads, which were taken on in levels, are the only exception to this rule. These levels are however stationary and are the result of a cross-sectional difference instead of a difference in time. Given their widely illustrated covariation with the growth cycle concept of the business cycle, Estrella and Mishkin [1997], this is common practice. After, the series were normalised in order to have a zero sample mean and unit variance by subtracting their mean and dividing by their standard deviation. This standardisation is necessary to avoid overweighting of the series with large variance

[5]Each case was studied: IGP-DI used for series in general, IPCA used for consumer-related series, US CPI for the case of variables of that country and so respectively

when estimating the spectral density matrix. Afterwards, the common component is denormalised, so as to correspond to the actual series.

The final step, after the normalized data, the series with different temporal mixed-frequencies (dataset with an arbitrary pattern of missing data) were back dating using Bańbura and Modugno [2014] procedure and with the database within all these specifications, the models were estimated.

# 5 Results

## 5.1 Variable selection and models for cycle estimation - Relative information test in high-dimensional for agricultural cycle.

The first step of this section is the application of the test of relative information to identify, in different lags, the most relevant variables for estimation of the agricultural cycle. The test as detailed in chapter 1 establishes the greatest information gain in relation to the variable of interest, which in this case is the agricultural GDP. The test was applied to the 2524 variables from the current period up to the third lag (ie a full year of dynamic effects were considered). The results are shown in Figure 2. Hence, low p-values indicate that an additional variable adds information.

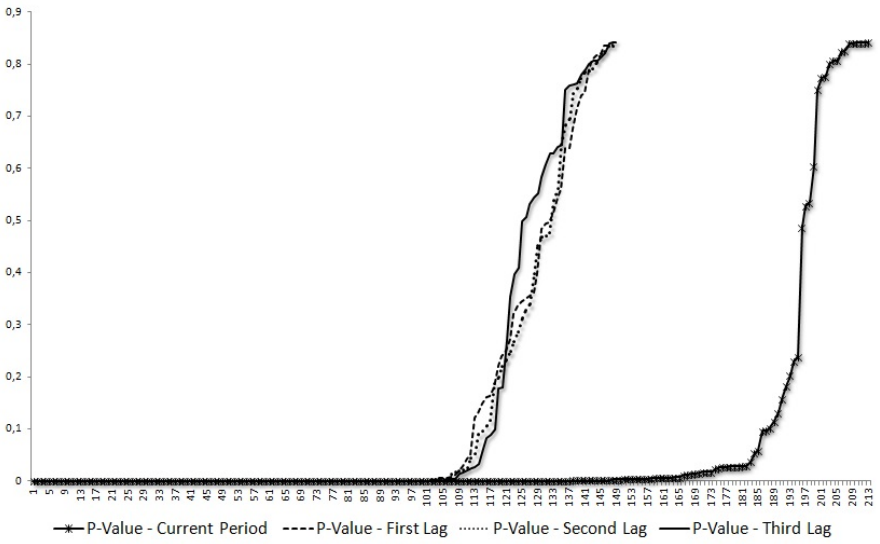


Figure 2: Relative information test - results for mixed-frequency agriculture cycle study.

The results indicate that the number of variables in the short term (current period) is considerably higher (192) than for the other lags. The number of variables selected for different lags was practically the same (112,115 and 118 respectively).This result indicates that agriculture is a more complex phenomenon in current events than in its temporal dynamics.

Given the high-dimension characteristic of the database, it is necessary to categorize the variables so that a more complete analysis of the agricultural cycle is possible. With twenty categories[6], this categorization for the agricultural cycle is the same one used by ESALQ-CEPEA[7] to study agribusiness GDP and is described in Table 2.

[6]All time series used and their respective categories are available up request.

[7]School of Agriculture "Luiz de Queiroz" (ESALQ) and Center for Advanced Studies in Applied Economics (CEPEA).

Table 2: Categories for agriculture cycle study

| Categories - Mixed-Frequency - Agriculture cycle study | | |
|---|---|---|
| Class | Series | % |
| Credit and Interest Rates | 549 | 0,21 |
| Global | 352 | 0,14 |
| Others Internal Markets | 314 | 0,12 |
| Industry | 226 | 0,09 |
| Emploment and Salary | 205 | 0,08 |
| External Sector | 163 | 0,06 |
| Governmental Finance | 124 | 0,05 |
| Agriculture Prices | 92 | 0,04 |
| Retail and Services | 88 | 0,03 |
| Agroindustrial Production | 79 | 0,03 |
| Prices and Inflation | 73 | 0,03 |
| Agriculture Credit and Interest Rates | 67 | 0,03 |
| Agriculture Production | 39 | 0,02 |
| Agriculture External Sector | 39 | 0,02 |
| Agriculture Employment and Salary | 38 | 0,01 |
| Precipitation | 32 | 0,01 |
| Maximum Temperature | 32 | 0,01 |
| Minimum Temperature | 32 | 0,01 |
| Monetary | 15 | 0,01 |
| Agricultural Inputs | 12 | 0,00 |
| Total | 2571 | 1,00 |

| Categories - Mixed-Frequency - Without National Accounts - Agriculture cycle study | | |
|---|---|---|
| Class | Series | % |
| Credit and Interest Rates | 549 | 0,22 |
| Global | 352 | 0,14 |
| Others Internal Markets | 301 | 0,12 |
| Industry | 221 | 0,09 |
| Emploment and Salary | 205 | 0,08 |
| External Sector | 145 | 0,06 |
| Governmental Finance | 124 | 0,05 |
| Agriculture Prices | 92 | 0,04 |
| Retail and Services | 80 | 0,03 |
| Agroindustrial Production | 78 | 0,03 |
| Prices and Inflation | 73 | 0,03 |
| Agriculture Credit and Interest Rates | 67 | 0,03 |
| Agriculture Production | 39 | 0,02 |
| Agriculture Employment and Salary | 38 | 0,02 |
| Agriculture External Sector | 37 | 0,01 |
| Maximum Temperature | 32 | 0,01 |
| Minimum Temperature | 32 | 0,01 |
| Precipitation | 32 | 0,01 |
| Monetary | 15 | 0,01 |
| Agricultural Inputs | 12 | 0,00 |
| Total | 2524 | 1,00 |

Notes: For the variable selection procedure, the national accounts variables are not considered, even though they are included in the estimated models.

From this categorization we can have a more detailed analysis of the relative information test for variable selection. These results are presented in Table 3.

Table 3: Variables selection - Relative information test for high-dimensional time series - Agriculture GDP.

| Class | $RI_t$ | $Selected_t$ | $RI_{t-1}$ | $Selected_{t-1}$ | $RI_{t-2}$ | $Selected_{t-2}$ | $RI_{t-3}$ | $Selected_{t-3}$ | Final Selection | 2º Component | Confirmatory | Total Sample |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Agricultural Inputs | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 2 | 0 | 0 | 12 |
| Agriculture Credit and Interest Rates | 0 | 20 | 0 | 4 | 1 | 3 | 1 | 3 | 23 | 1 | 0 | 67 |
| Agriculture Employment and Salary | 0 | 7 | 1 | 4 | 1 | 4 | 2 | 3 | 13 | 4 | 0 | 38 |
| Agriculture External Sector | 0 | 4 | 0 | 1 | 1 | 0 | 0 | 1 | 6 | 1 | 0 | 37 |
| Agriculture Prices | 2 | 5 | 0 | 5 | 1 | 1 | 0 | 2 | 11 | 3 | 0 | 92 |
| Agriculture Production | 0 | 10 | 2 | 4 | 0 | 6 | 2 | 7 | 17 | 4 | 39 | 39 |
| Agroindustrial Production | 0 | 13 | 2 | 6 | 2 | 4 | 1 | 4 | 23 | 5 | 7 | 78 |
| Credit and Interest Rates | 0 | 11 | 1 | 15 | 2 | 8 | 0 | 9 | 34 | 3 | 0 | 549 |
| Emploment and Salary | 4 | 15 | 3 | 11 | 3 | 12 | 3 | 12 | 36 | 11 | 0 | 205 |
| External Sector | 4 | 7 | 3 | 5 | 1 | 9 | 3 | 7 | 20 | 11 | 0 | 145 |
| Global | 3 | 16 | 4 | 12 | 6 | 18 | 2 | 18 | 46 | 14 | 0 | 352 |
| Governmental Finance | 1 | 5 | 1 | 3 | 0 | 2 | 2 | 2 | 10 | 4 | 0 | 124 |
| Industry | 4 | 13 | 3 | 16 | 2 | 15 | 5 | 14 | 40 | 13 | 0 | 221 |
| Maximum Temperature | 0 | 8 | 1 | 0 | 0 | 2 | 0 | 2 | 9 | 1 | 0 | 32 |
| Minimum Temperature | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 32 |
| Monetary | 0 | 1 | 2 | 0 | 0 | 2 | 1 | 1 | 2 | 3 | 0 | 15 |
| Others Internal Markets | 1 | 24 | 8 | 16 | 7 | 22 | 6 | 20 | 58 | 20 | 0 | 301 |
| Precipitation | 0 | 10 | 1 | 1 | 0 | 1 | 0 | 1 | 11 | 1 | 0 | 32 |
| Prices and Inflation | 2 | 2 | 0 | 2 | 0 | 3 | 1 | 3 | 6 | 3 | 0 | 73 |
| Retail and Services | 0 | 12 | 1 | 9 | 4 | 5 | 3 | 5 | 22 | 6 | 0 | 80 |
| Total | 21 | 192 | 33 | 115 | 31 | 118 | 32 | 115 | 397 | 108 | 46 | 2524 |

Percentage

| Class | $RI_t$ | $Selected_t$ | $RI_{t-1}$ | $Selected_{t-1}$ | $RI_{t-2}$ | $Selected_{t-2}$ | $RI_{t-3}$ | $Selected_{t-3}$ | Final Selection | 2º Component | Confirmatory | Final Selection Sample |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Agricultural Inputs | 0,00 | 0,01 | 0,00 | 0,01 | 0,00 | 0,01 | 0,00 | 0,01 | 0,01 | 0,00 | 0,00 | 0,17 |
| Agriculture Credit and Interest Rates | 0,00 | 0,10 | 0,00 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 | 0,06 | 0,01 | 0,00 | 0,34 |
| Agriculture Employment and Salary | 0,00 | 0,04 | 0,03 | 0,03 | 0,03 | 0,03 | 0,06 | 0,03 | 0,03 | 0,04 | 0,00 | 0,34 |
| Agriculture External Sector | 0,00 | 0,02 | 0,00 | 0,01 | 0,03 | 0,00 | 0,00 | 0,01 | 0,02 | 0,01 | 0,00 | 0,16 |
| Agriculture Prices | 0,10 | 0,03 | 0,00 | 0,04 | 0,03 | 0,01 | 0,00 | 0,02 | 0,03 | 0,03 | 0,00 | 0,12 |
| Agriculture Production | 0,00 | 0,05 | 0,06 | 0,03 | 0,00 | 0,05 | 0,06 | 0,06 | 0,04 | 0,04 | 0,85 | 0,44 |
| Agroindustrial Production | 0,00 | 0,07 | 0,06 | 0,05 | 0,06 | 0,03 | 0,03 | 0,03 | 0,06 | 0,05 | 0,15 | 0,29 |
| Credit and Interest Rates | 0,00 | 0,06 | 0,03 | 0,13 | 0,06 | 0,07 | 0,00 | 0,08 | 0,09 | 0,03 | 0,00 | 0,06 |
| Emploment and Salary | 0,19 | 0,08 | 0,09 | 0,10 | 0,10 | 0,10 | 0,09 | 0,10 | 0,09 | 0,10 | 0,00 | 0,18 |
| External Sector | 0,19 | 0,04 | 0,09 | 0,04 | 0,03 | 0,08 | 0,09 | 0,06 | 0,05 | 0,10 | 0,00 | 0,14 |
| Global | 0,14 | 0,08 | 0,12 | 0,10 | 0,19 | 0,15 | 0,06 | 0,16 | 0,12 | 0,13 | 0,00 | 0,13 |
| Governmental Finance | 0,05 | 0,03 | 0,03 | 0,03 | 0,00 | 0,02 | 0,06 | 0,02 | 0,03 | 0,04 | 0,00 | 0,08 |
| Industry | 0,19 | 0,07 | 0,09 | 0,14 | 0,06 | 0,13 | 0,16 | 0,12 | 0,10 | 0,12 | 0,00 | 0,18 |
| Maximum Temperature | 0,00 | 0,04 | 0,03 | 0,00 | 0,00 | 0,02 | 0,00 | 0,02 | 0,02 | 0,01 | 0,00 | 0,28 |
| Minimum Temperature | 0,00 | 0,04 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,02 | 0,00 | 0,00 | 0,25 |
| Monetary | 0,00 | 0,01 | 0,06 | 0,00 | 0,00 | 0,02 | 0,03 | 0,01 | 0,01 | 0,03 | 0,00 | 0,13 |
| Others Internal Markets | 0,05 | 0,13 | 0,24 | 0,14 | 0,23 | 0,19 | 0,19 | 0,17 | 0,15 | 0,19 | 0,00 | 0,19 |
| Precipitation | 0,00 | 0,05 | 0,03 | 0,01 | 0,00 | 0,01 | 0,00 | 0,01 | 0,03 | 0,01 | 0,00 | 0,34 |
| Prices and Inflation | 0,10 | 0,01 | 0,00 | 0,02 | 0,00 | 0,03 | 0,03 | 0,03 | 0,02 | 0,03 | 0,00 | 0,08 |
| Retail and Services | 0,00 | 0,06 | 0,03 | 0,08 | 0,13 | 0,04 | 0,09 | 0,04 | 0,06 | 0,06 | 0,00 | 0,28 |
| Total | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 0,16 |

Note: The Table presents the result of the relative information test applied to the total set of variables considered (2524). The results for each period represent $RI_{t-k} + Selected_{t-k}$ all variables which presented non-zero relative information.

The $Selected_{t-k}$ variables represent those that have a statistically non-zero value in each period. The series in Final Selection represent the series if they were statistically significant in some periods $(t-k)$, without repetitions, as well as the values presented in 2º Component, however with the series not being selected in some periods. Confirmatory represent only series directly related with agriculture GDP. The values in Final Selection Sample represent the ratio between Final Selection and Total Sample. The 47 series of the National Accounts category were not related to the process of selecting variables because they belong to the same category in which the cycle to be studied was defined.

Considering the Final Selection in the table 3, we observed the predominance of five categories: Other Internal Markets (economic activity's series of different sectors), global variables (Interna-

tional economic activity), industrial variables, employment and wages and credit and interest rates, representing together more than sixty percent of all selected series.

Dynamically the variables of the category of agricultural credit and interest rates are very representative in the current period but with few series present in other lags, the opposite happens with the series of the category of Other Internal Markets, in which the quantity of variables present in all the lags is high.

Monetary variables and agricultural inputs are the categories with the lowest number of series, while the global and Other Internal Market variables are the categories with the highest number of selected series.

In terms of sample representativity, we can observe in the Final Selected Sample that eight categories have great participation within their own category: Agricultural Production, Agricultural Credit and Interest Rates, Agricultural Employment and Salary, State Precipitations, Agroindustrial Production, Retail and Services, State Maximum Temperatures and State Minimum Temperatures.

From the total of 2524 variables tested, 505 (20%) presented non-zero relative information, of which 397 (16%) presented values statistically different from zero. Thus, each set of variables (397 of the Final Selection, 108 of the Second Component and 46 of the Confirmatory) will be used for a specific type of model to estimate the agricultural cycle.

Results presented in the Final Selected and Final Selection Sample already represent a first outline of the agricultural cycle that will be estimated in the next steps. The number of selected series gives us a relative idea of the importance of each category but its true importance will only be known with the estimation of each model.

## 5.2 Mixed-frequency generalized dynamic factor model results for agriculture cycle.

Figure 3 illustrates the growth cycle of agriculture GDP and its common component. From this figure it can be seen that the common component of agriculture GDP generally comoves with the variation in agriculture GDP, although its variation is milder and flattens out the sharp growth peaks and troughs of agriculture GDP. Later on, we will evaluate each variable with respect to this reference cycle by analysing the mutual relation between the common component of the series and the common component of agriculture GDP, representing the essential business cycle relation.

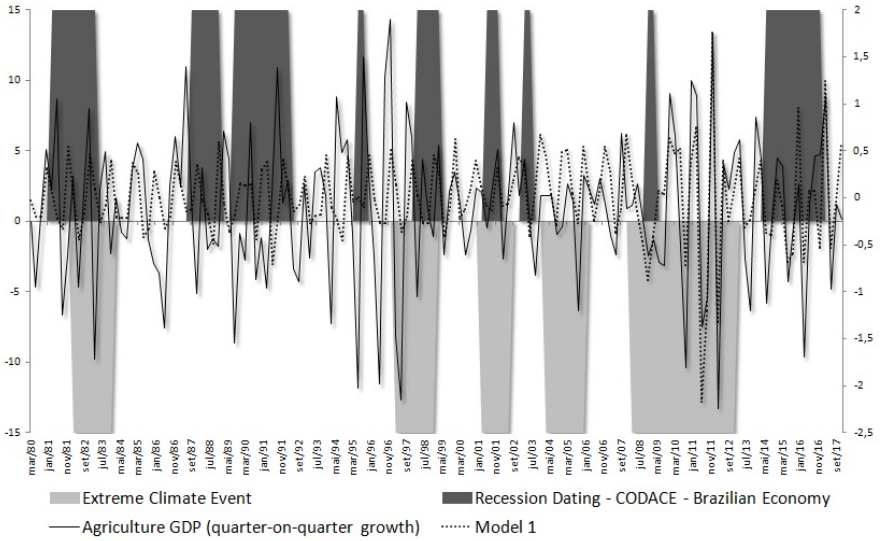


Figure 3: Quarter-on-quarter agriculture GDP growth and MF-GDFM common component.

We can observe that from 1995 the movements of the agricultural cycle become more adjusted

to the movements of agricultural GDP. We can observe that the sharpest crises of agricultural GDP are more associated with extreme climatic events than with the economic crises dated by CODACE. The co-movements of agricultural GDP and the cycle are more synchronized in the 2008 CODACE crisis and in the long period of extreme climatic events from 2008 to 2012.

### 5.2.1 Dissecting the Brazilian agriculture cycle

The results of Table 4 dissect the Brazilian agricultural cycle estimated according to the characteristics of its phases and moments by each category used. We can know from the established categories, the major degree of commonalities with the agricultural cycle, the portion of this commonality associated with the Cycle Phase Indicator (CPI): pro-cyclical or counter-cyclical, the degree of commonality associated with each moment of the Cycle Moment Indicator (CMI): lagging, coincident or leading.

In Table 4 we have the average of the commonality ratio by category which allows us to analyze more generally. The relative importance of categories is also presented in the quantities and percentages of series by categories.

### 5.2.2 Aggregate analysis by category

In a general analysis, the degree of average communality was 17% among all categories[8], with the maximum average communality of the state minimum temperatures of 82% and with lower average commonality of the series that compose the agricultural production.

The first highlight is the high average commonalities of climatic variables. There are 28 series (6% of the total), but with commonalities above 50%. The minimum temperatures and precipitations present the highest pro-cyclical and antecedent commonalities (82% and 64%), which means that less severe winters and more rainy periods are associated to expansions of the agriculture cycle, and these conditions occur before the cycle, or either periods with milder winters or wetter periods are associated in the period following an expansion of the agriculture cycle. The effects of maximum temperatures are divided between pro-cyclical and counter-cyclical movements in all cases in an antecedent way. In this way, more severe or less severe summers can indicate expansions or contractions in the agriculture cycle.

In terms of importance, after the climatic variables, the variables of credit, agriculture and general conditions are the most relevant in terms of average commonality. The agriculture credit conditions have average commonality (18%) slightly higher than the general credit conditions (15%). In both cases the degree of this commonality divides pro-cyclically and counter-cyclically, yet the highest average commonality is counter-cyclical (19% with 19 series and 18% with 16 series for agriculture credit and general credit respectively.) Occurring lagging behind agriculture cycle movements.

Other seven categories (Governamental Finance, Others Internal Markets, Retail and Services, Industry, Global, Agriculture Inputs and External Sector) have predominantly pro-cyclical and antecedent average commonality. The other categories (Employment and Salary, Prices and Inflation, Agriculture Prices, Agroindustrial Production, Agriculture External Sector, Agriculture Employment and Salary, Monetary and Agriculture Production) have dispersed average commonality in a large number of series, being this average commonality divided between pro-cyclical and counter-cyclical phases and between lagged and antecedent effects.

The conclusions we obtain from aggregated analysis by categories highlight the importance of climate variables and credit conditions for the Brazilian agriculture cycle. Among the climatic variables, all of them presented antecedent pattern to the agricultural cycle, which is natural and happens in the process of planting and harvesting or herd's pregnant and birthing periods. Better

[8]Note that the total of variables included in the model refers to the total of the process of selection of the previous section (397) plus the variables of the National Accounts (47) re-classified according to the categorization adopted, resulting in a total of 444 variables used in the model.

climatic conditions (winters and milder summers with more rainy periods) precede expansions of the Brazilian agriculture cycle . Regarding credit conditions, both agricultural and general credit have diffuse effect in the agriculture cycle, part of this credit is counter-cyclical and antecedent may indicate that producers in financial difficulty in the stage of planting or herd's pregnant stage, on the other hand, the counter-cyclical and lag commonality may indicate financial difficulty in the harvest period or herd's birth stage. The portion of credit conditions that has pro-cyclical and antecedent characteristics may indicate an increase in the investment capacity at the stage of planting or herd's pregnancy stage, in the same way that the pro-cyclical and lagged portion may indicate an increase in the capacity of investment in the harvesting phase or in the herd's birth stage.

Table 4: Average Degree of Commonality - Brazilian Agriculture Cycle

| Class | Average | Counter-Cyclical | Pro-Cyclical | Lagging | Coincident | Leading | Series | Counter-Cyclical | Pro-Cyclical | Lagging | Coincident | Leading | Series | Counter-Cyclical | Pro-Cyclical | Lagging | Coincident | Leading |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Minimum Temperature | 0,812 | 0,000 | 0,812 | 0,000 | 0,000 | 0,812 | 8 | 0 | 8 | 0 | 0 | 8 | 0,02 | 0,00 | 1,00 | 0,00 | 0,00 | 1,00 |
| Precipitation | 0,643 | 0,000 | 0,643 | 0,000 | 0,000 | 0,643 | 11 | 0 | 11 | 0 | 0 | 11 | 0,02 | 0,00 | 1,00 | 0,00 | 0,00 | 1,00 |
| Maximum Temperature | 0,506 | 0,253 | 0,707 | 0,000 | 0,000 | 0,506 | 9 | 4 | 5 | 0 | 0 | 9 | 0,02 | 0,44 | 0,56 | 0,00 | 0,00 | 1,00 |
| Agriculture Credit and Interest Rates | 0,188 | 0,195 | 0,151 | 0,195 | 0,000 | 0,151 | 23 | 19 | 4 | 19 | 0 | 4 | 0,05 | 0,83 | 0,17 | 0,83 | 0,00 | 0,17 |
| Credit and Interest Rates | 0,156 | 0,180 | 0,135 | 0,180 | 0,000 | 0,135 | 34 | 16 | 18 | 16 | 0 | 18 | 0,08 | 0,47 | 0,53 | 0,47 | 0,00 | 0,53 |
| Governmental Finance | 0,129 | 0,047 | 0,211 | 0,047 | 0,000 | 0,211 | 10 | 5 | 5 | 5 | 0 | 5 | 0,02 | 0,50 | 0,50 | 0,50 | 0,00 | 0,50 |
| Others Internal Markets | 0,122 | 0,064 | 0,165 | 0,062 | 0,121 | 0,167 | 71 | 30 | 41 | 30 | 1 | 40 | 0,16 | 0,42 | 0,58 | 0,42 | 0,01 | 0,56 |
| Retail and Services | 0,115 | 0,096 | 0,131 | 0,096 | 0,102 | 0,131 | 30 | 14 | 16 | 13 | 1 | 16 | 0,07 | 0,47 | 0,53 | 0,43 | 0,03 | 0,53 |
| Industry | 0,109 | 0,024 | 0,143 | 0,024 | 0,000 | 0,143 | 45 | 13 | 32 | 13 | 0 | 32 | 0,10 | 0,29 | 0,71 | 0,29 | 0,00 | 0,71 |
| Global | 0,102 | 0,085 | 0,119 | 0,085 | 0,000 | 0,119 | 46 | 23 | 23 | 23 | 0 | 23 | 0,10 | 0,50 | 0,50 | 0,50 | 0,00 | 0,50 |
| Agricultural Inputs | 0,095 | 0,088 | 0,103 | 0,088 | 0,000 | 0,103 | 2 | 1 | 1 | 1 | 0 | 1 | 0,00 | 0,50 | 0,50 | 0,50 | 0,00 | 0,50 |
| Emploment and Salary | 0,076 | 0,071 | 0,082 | 0,071 | 0,000 | 0,082 | 36 | 18 | 18 | 18 | 0 | 18 | 0,08 | 0,50 | 0,50 | 0,50 | 0,00 | 0,50 |
| External Sector | 0,070 | 0,042 | 0,080 | 0,042 | 0,000 | 0,080 | 38 | 10 | 28 | 10 | 0 | 28 | 0,09 | 0,26 | 0,74 | 0,26 | 0,00 | 0,74 |
| Prices and Inflation | 0,054 | 0,044 | 0,059 | 0,044 | 0,000 | 0,059 | 6 | 2 | 4 | 2 | 0 | 4 | 0,01 | 0,33 | 0,67 | 0,33 | 0,00 | 0,67 |
| Agriculture Prices | 0,052 | 0,037 | 0,060 | 0,037 | 0,000 | 0,060 | 11 | 4 | 7 | 4 | 0 | 7 | 0,02 | 0,36 | 0,64 | 0,36 | 0,00 | 0,64 |
| Agroindustrial Production | 0,045 | 0,033 | 0,057 | 0,033 | 0,071 | 0,056 | 24 | 12 | 12 | 12 | 1 | 11 | 0,05 | 0,50 | 0,50 | 0,50 | 0,04 | 0,46 |
| Agriculture External Sector | 0,041 | 0,028 | 0,063 | 0,028 | 0,000 | 0,063 | 8 | 5 | 3 | 5 | 0 | 3 | 0,02 | 0,63 | 0,38 | 0,63 | 0,00 | 0,38 |
| Agriculture Employment and Salary | 0,040 | 0,020 | 0,058 | 0,020 | 0,000 | 0,058 | 13 | 6 | 7 | 6 | 0 | 7 | 0,03 | 0,46 | 0,54 | 0,46 | 0,00 | 0,54 |
| Monetary | 0,032 | 0,060 | 0,003 | 0,060 | 0,000 | 0,003 | 2 | 1 | 1 | 1 | 0 | 1 | 0,00 | 0,50 | 0,50 | 0,50 | 0,00 | 0,50 |
| Agriculture Production | 0,013 | 0,008 | 0,018 | 0,006 | 0,013 | 0,020 | 17 | 8 | 9 | 7 | 3 | 7 | 0,04 | 0,47 | 0,53 | 0,41 | 0,18 | 0,41 |
| Mean/Total | 0,170 | 0,088 | 0,157 | 0,056 | 0,015 | 0,180 | 444 | 194 | 250 | 185 | 6 | 253 | 1,00 | 0,44 | 0,56 | 0,42 | 0,01 | 0,57 |

Note: Mixed frequency Generalized Dynamic Factor Model (MF-GDFM) - In-sample period: 1980Q1 to 2017Q4. The Table shows average degree of commonality for each class by Cycle Phase Indicator (CPI): Counter-Cyclical and/or Pro-Cyclical; Cycle Moment Indicator (CMI): Lagging, Coincident and/or Leading.

The number of series and their percentages are displayed respectively.

# 6 Conclusions

This study was based on the use of techniques and models applied to the context of high-dimensional and mixed frequencies databases, aiming at estimating the Brazilian agricultural business cycle and dissecting it to understand which categories of variables have a high commonality with the cycle. In the selection of variables for cycle estimation, eight categories stood out with a high level of selection of variables within the category. As expected, the level of variable selection was high for categories directly associated with the dynamics of agriculture itself, such as agricultural production indicators, rural credit indicators, as well as employment and wages indicators in the sector. Given the intrinsic dependence of agriculture on climate, the level of selection within the categories was also high for those related to climatic variables (maximum temperatures, minimum temperatures and rainfall in the Brazilian states).

From the cycle estimation, we observed that the most intense crises of the agricultural cycle are more associated to climatic events than to economic crises. Dissecting the cycle, we found an average commonality of 17% for the categories of variables, and the maximum level of commonality found was for the minimum state temperature category (82%). Following the minimum temperatures, the greatest commonalities were verified for the categories of precipitation (65%) and maximum temperatures (50%). This important result evidences the fundamental role of the climatic cycles for the agricultural performance in the states and, therefore, in the Country.

We found that the minimum temperatures and the precipitations presented pro-cyclical and antecedent commonalities, which means that less severe winters and rainier periods are associated to future expansions of the agricultural cycle. In addition to the important role of the climate, another important result of the study was the high commonality of the agricultural cycle with the variables related to credit, be it rural credit or general conditions. For these categories, the highest commonality occurs in countercyclical and antecedent indicators.

# References


Alessi, L., Barigozzi, M. and Capasso, M.: 2010, Improved penalization for determining the number of factors in approximate factor models, *Statistics & Probability Letters* **80**(23-24), 1806–1813.

Anwar, M. R., Li Liu, D., Macadam, I. and Kelly, G.: 2013, Adapting agriculture to climate change: a review, *Theoretical and applied climatology* **113**(1-2), 225–245.

Bacha, C. J. C.: 2004, *Economia e política agrícola no Brasil*, 1 ed. edn, Atlas, São Paulo.

Bai, J.: 2003, Inferential theory for factor models of large dimensions, *Econometrica* **71**(1), 135–171.

Bai, J. and Ng, S.: 2002, Determining the number of factors in approximate factor models, *Econometrica* **70**(1), 191–221.

Bai, J., Ng, S. et al.: 2008, Large dimensional factor analysis, *Foundations and Trends® in Econometrics* **3**(2), 89–163.

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

Barros, G. S. d. C.: 2014, Agricultura e Indústria no Desenvolvimento Econômico Brasileiro, *in* A. M. Buainain, E. Alves, J. M. da Silveira and Z. Navarro (eds), *O Mundo Rural no Brasil do Século 21*, 1ed edn, Embrapa, Brasília, pp. 79–116.

Barros, G. S. d. C., Bacchi, M. R. P. and Burnquist, H. L.: 2002, Estimação de equações de oferta de exportação de produtos agropecuários para o Brasil (1992/2000), *Technical report*, IPEA, Brasília.

Barros, G. S. d. C. and Castro, N. R.: 2017, Produto Interno Bruto Do Agronegócio E a Crise Brasileira, *Revista de Economia e Agronegócio* **15**(2), 156–162.

Belloni, A. and Chernozhukov, V.: 2011, High dimensional sparse econometric models: An introduction, pp. 121–156.

Buainain, A. M., Alves, E., da Silveira, J. M. and Navarro, Z.: 2013, Sete teses sobre o mundo rural, *Revista de Política Agrícola* **XXII**(2), 105–121.

Burnham, K. and Anderson, D.: 2002, *Model Selection and Multi-model Inference: A pratical Information-Theoretic Approach. 2nd edn.(Springer: New York.)*.

Burns, A. F. and Mitchell, W. C.: 1947, *Measuring business cycles.*

Cepea/Arroz: 2012, Agromensal Cepea/Esalq - Arroz, *Technical report*, Cepea, Piracicaba.

Cepea/Soja: 2013, Quebra da safra marca 2012 com preços recordes, *Technical report*, Cepea, Piracicaba.
**URL:** *https://www.cepea.esalq.usp.br/br/diarias-de-mercado/soja-cepea-quebra-da-safra-marca-2012-com-precos-recordes.aspx*

Chauvet, M.: 1998, An econometric characterization of business cycle dynamics with factor structure and regime switching, *International economic review* pp. 969–996.

Cohen, D. S.: 2001, Linear data transformations used in economics, *Technical Report 2001-59*.

Conab: 2013, Acompanhamento de safra brasileira: grãos, décimo segundo levantamento, setembro 2013, *Technical report*.

Conab: 2017, Acompanhamento de safra brasileira: grãos, décimo segundo levantamento, setembro 2017, *Technical report*.

Doz, C., Giannone, D. and Reichlin, L.: 2012, A quasi–maximum likelihood approach for large, approximate dynamic factor models, *Review of economics and statistics* **94**(4), 1014–1024.

Estatística, I. B. d. G. e.: 2006, Censo agropecuário - instituto brasileiro de geografia e estatística.

Estatística, I. B. d. G. e.: 2018a, Contas nacionais trimestrais - instituto brasileiro de geografia e estatística.

Estatística, I. B. d. G. e.: 2018b, Produção agrícola municipal - instituto brasileiro de geografia e estatística.

Estrella, A. and Mishkin, F. S.: 1997, The predictive power of the term structure of interest rates in europe and the united states: Implications for the european central bank, *European economic review* **41**(7), 1375–1401.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L.: 2000, The generalized dynamic-factor model: Identification and estimation, *Review of Economics and statistics* **82**(4), 540–554.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L.: 2001, Coincident and leading indicators for the euro area, *The Economic Journal* **111**(471), C62–C85.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L.: 2004, The generalized dynamic factor model consistency and rates, *Journal of Econometrics* **119**(2), 231–255.

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

Forni, M. and Lippi, M.: 2001, The generalized dynamic factor model: representation theory, *Econometric theory* **17**(6), 1113–1141.

Garcia, J. R.: 2014, Trabalho rural: tendências em face das transformações em curso, pp. 559–590.

Gasques, J., Bastos, E., Valdes, C. and Bacci, M. R. P.: 2014, Produtividade da agricultura: resultados para o Brasil e estados selecionados, *Revista de Política Agrícola* pp. 87–98.

Gasques, J. G., de Rezende, G. C., Verde, C. M. V., Salerno, M. S., da Conceição, J. C. P. R. and Carvalho, J. C. d. S.: 2004, Desempenho e crescimento do agronegocio no Brasil, *Technical report*, IPEA, Brasília.

Geweke, J.: 1977, The dynamic factor analysis of economic time series, *Latent variables in socio-economic models* .

Golan, A. and Maasoumi, E.: 2008, Information theoretic and entropy methods: An overview, *Econometric Reviews* **27**(4-6), 317–328.

Gornall, J., Betts, R., Burke, E., Clark, R., Camp, J., Willett, K. and Wiltshire, A.: 2010, Implications of climate change for agricultural productivity in the early twenty-first century, *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**(1554), 2973–2989.

Hamilton, J. D.: 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica: Journal of the Econometric Society* pp. 357–384.

Harding, D. and Pagan, A.: 2002, Dissecting the cycle: a methodological investigation, *Journal of monetary economics* **49**(2), 365–381.

Harding, D. and Pagan, A.: 2006, Synchronization of cycles, *Journal of Econometrics* **132**(1), 59–79.

Jacobs, J. P. and Otter, P. W.: 2008, Determining the number of factors and lag order in dynamic factor models: A minimum entropy approach, *Econometric Reviews* **27**(4-6), 385–397.

Kageyama, A. et al.: 1987, O novo padrão agrícola brasileiro: do complexo rural aos complexos agroindustriais, *Campinas: Unicamp* .

Kim, C.-J. and Nelson, C. R.: 1998, Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching, *Review of Economics and Statistics* **80**(2), 188–201.

Kim, M.-J. and Yoo, J.-S.: 1995, New index of coincident indicators: A multivariate markov switching factor model approach, *Journal of Monetary Economics* **36**(3), 607–630.

Kydland, F. E. and Prescott, E. C.: 1982, Time to build and aggregate fluctuations, *Econometrica: Journal of the Econometric Society* pp. 1345–1370.

Kydland, F. E. and Prescott, E. C.: 1990, The econometrics of the general equilibrium approach to business cycles, *Real Business Cycles* pp. 219–236.

Lucas Jr, R. E.: 1972, Expectations and the neutrality of money, *Journal of economic theory* **4**(2), 103–124.

Lucas, R. E.: 1973, Some international evidence on output-inflation tradeoffs, *The American Economic Review* **63**(3), 326–334.

Marquetti, A., Silveira, F. G. and da Silva, P. R. N.: 1991, Agricultura: Quebras de safra significam elevações de preço, importações e pacotes agrícolas, *Indicadores Econômicos FEE* **19**(3).

Mendelsohn, R., Nordhaus, W. and Shaw, D.: 1996, Climate impacts on aggregate farm value: accounting for adaptation, *Agricultural and Forest Meteorology* **80**(1), 55–66.

Millner, A. and Dietz, S.: 2015, Adaptation to climate change and economic growth in developing countries, *Environment and Development Economics* **20**(3), 380–406.

Mohapatra, S. and Majhi, B.: 2015, An evolutionary approach for load balancing in cloud computing, *Handbook of Research on Securing Cloud-Based Databases with Biometric Applications*, IGI Global, pp. 433–463.

Morton, J. F.: 2007, The impact of climate change on smallholder and subsistence agriculture, *Proceedings of the national academy of sciences* **104**(50), 19680–19685.

Muirhead, R. J.: 1982, *Aspects of multivariate statistical analysis.*

Nordhaus, W. D.: 1993, Reflections on the economics of climate change, *Journal of economic Perspectives* **7**(4), 11–25.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N. et al.: 2014, Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison, *Proceedings of the National Academy of Sciences* **111**(9), 3268–3273.

Rothschild, M. and Chamberlain, G.: 1982, Arbitrage, factor structure, and mean-variance analysis on large asset markets.

Sargent, T. J., Sims, C. A. et al.: 1977, Business cycle modeling without pretending to have too much a priori economic theory, *New methods in business cycle research* **1**, 145–168.

Shumway, R. H. and Stoffer, D. S.: 1982, An approach to time series smoothing and forecasting using the em algorithm, *Journal of time series analysis* **3**(4), 253–264.

Song, S. and Bickel, P. J.: 2011, Large vector auto regressions, *arXiv preprint arXiv:1106.3915* .

Staduto, J. A. R., Shikida, P. F. A. and Bacha, C. J. C.: 2004, Alteração na composição da mão-de-obra assalariada na agropecuária brasileira, *Revista de Economia Agrícola* **51**(2).

Stock, J. H. and Watson, M.: 2011, *Dynamic factor models*, Oxford University Press.

Stock, J. H. and Watson, M. W.: 2014, Estimating turning points using large data sets, *Journal of Econometrics* **178**, 368–381.

Summer, L.: 1986, Some skeptical observations on real business cycles theory, *Federal Reserve Bank of Minneapolis Quarterly Review* **10**, 23–27.

Watson, M. W. and Engle, R. F.: 1983, Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models, *Journal of Econometrics* **23**(3), 385–400.

Wheeler, T. and Von Braun, J.: 2013, Climate change impacts on global food security, *Science* **341**(6145), 508–513.

Young, T. Y. and Calvert, T. W.: 1974, *Classification, estimation and pattern recognition*, North-Holland.