SMEs, Regional Economic Growth and Cycles in Brazil

Túlio Antonio Cravo

Department of Economics
Loughborough University

Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy, Department of Economics, Loughborough University.

Abstract

This thesis presents an examination of the importance of Small and Medium Enterprises (SMEs) for economic growth and examines how sensitive employment in SMEs is to business cycle fluctuations in Brazil.

The thesis uses different empirical techniques to investigate the role of SMEs in the Brazilian regional economic growth, using a panel dataset from 1980 to 2004 for 508 Brazilian micro-regions. It first uses standard panel data estimators (OLS, LSDV, system and first differenced GMM) to analyse the (augmented) Solow growth model encompassing the importance of the relative size of the SME sector measured by the share of the SME employment in total formal employment and the level of human capital in SMEs measured by the average years of schooling of SME employees. The results show that the size of the SME sector is not significantly important for regional economic growth, but that human capital embodied in SMEs is more important in this process.

Standard panel data regressions are likely to produce biased results since they ignore the potential spatial dependence. Therefore, we present an analysis of growth regressions encompassing the SME sector considering the spatial dependence through the use of spatial econometrics. The empirical results reveal strong spatial dependence in the regional economic growth process in Brazil and confirm that human capital embodied in SMEs is more important than the size of this sector to economic growth. Nevertheless, the results indicate that the presence of SMEs generates economic growth through spatial interactions, small businesses in a given region benefit from a larger SME sector in its surrounding area. Conversely, there are no human capital spillovers in the SME sector. This analysis is complemented by a panel data spatial filtering approach that suggests that the values of the conditioning variables (including SME sector variables) across regions seem to be intrinsically linked with geographical location, supporting the view that conditioning variables carry strong spatial information with them.

In addition, we analyse the sensitivity of the employment series to business cycle fluctuations. We find that smaller establishments are more cyclically sensitive than larger ones in Brazil. Furthermore, the VAR impulse response analysis suggests the effect of small firms hiring cheaply from unemployment proportionally more than large ones during recessions. However, innovations in credit constraints hit small firms harder and help to explain the empirical regularity that small establishments are more cyclically sensitive.

Keywords: Firm size, market structure, regional economic growth, human capital, spatial effects, jobs flows, business cycles.

Acknowledgements

My foremost thanks goes to my supervisors, Dr. Adrian Gourlay and Dr. Bettina Becker, for their meticulous supervision and encouragement throughout the research for this thesis. I am also particularly grateful to my Director of Research, Dr. Paul Turner, who provided me personal support in many occasions during my research. A full studentship grant from the Department of Economics is gratefully acknowledged. I would also like to thank the staff of the Department of Economics, Student Support Centre and Research Office, who provided a supportive environment in which I could conduct my research.

I would like to thank Dr. Elias Soukiazis and Dr. Helena Marques for the encouragement to pursue my PhD. Also, I am grateful to all my PhD colleagues for making my time in Loughborough even more interesting and pleasant. I cannot forget to mention the Brazilian community in Loughborough and all my friends scattered around many lands that also supported me during this journey.

Finally, I owe my greatest thanks for my family back in Brazil, my mother Maria Cristina do Carmo Cravo, my father Jose Antonio Cravo, my siblings Telio and Thais. Also, I am gratefull for having Elisabete Simões by my side throughout this period. I truly thank you for their love and support without which this thesis would not have been possible.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Research Objectives and Contributions of the Thesis

The Small and Medium Enterprise (SME) sector is responsible for the majority of employment generation in developed and developing countries. Specifically, for the Brazilian case, around 60% of the formal employment was generated by the SME sector in 2004. This explains the large amount of financial support given to the development of the SME sector by multilateral organizations and national and regional development agencies. For instance, in 2003 the World Bank approved US$ 1.3 billion in programmes related to SMEs support (Beck et al. 2005a) and the budget of the Brazilian small business support service (SEBRAE) amounted to US$ 1.25 billion in 2009.[1] However, there is a paucity of studies exploring the importance of the SME sector for economic growth and its behaviour during the business cycle.

In order to contribute to the literature, this thesis provides the first comprehensive analysis of the formal SME sector for the Brazilian economy. In general, the contribution of the thesis lies in the application of the analysis to study the SME sector for the Brazilian case, as described later in this section. First, it investigates the importance of the SME sector for regional economic growth in Brazil during the period 1980-2004 using a panel data set with 508 Brazilian micro-regions. Second, it presents an analysis of whether formal employment series are more sensitive to business cycle fluctuations in SMEs than in large enterprises (LEs) using monthly time series data from January 2000 to July 2009 for Brazil and its states. Unfortunately, due to lack of data, the analysis does not incorporate the

[1] This information is available at www.sebrae.com.br. Conversion is based on the official exchange rate of 31/12/2009 provided by the Brazilian Central Bank. Approximately 75% of SEBRAEs' budget comes directly from compulsory contributions collected from enterprises.

informal sector and this omission might influence the results presented in the thesis.[2] Nevertheless, La Porta and Shleifer (2008) use cross-country survey to provide evidence that the informal sector is less productive than the formal one and as argued in Cravo (2010), this result suggests that the analysis is unlikely to change regarding the importance of SMEs on growth in Brazil if the informal sector is considered.

Regarding the former strand of research, the objective is to provide empirical evidence on the importance of the SME sector for regional economic growth in Brazil. This analysis draws on Beck et al. (2005a) who incorporated the SME sector in the growth regression framework for a cross-section of countries that includes Brazil; a review of the growth literature as well as how the SME sector is incorporated in this framework is provided in Chapter 2. We follow the Beck et al. (2005a) approach and construct a unique dataset for 508 Brazilian micro-regions (described in detail in Chapter 3) to analyse for the first time the importance of the SME sector to regional economic growth in Brazil. Furthermore, this thesis contributes to the literature by extending the analysis presented in Beck et al. (2005a) through the inclusion of the human capital embodied in SMEs in the model. Therefore, the thesis seeks to provide an analysis of how the size of the SME sector and the human capital embodied in this sector affect regional economic growth in Brazil. To achieve this objective, in Chapter 4, we present the results generated by standard panel data estimators; ordinary least squares (OLS), least square dummy variables (LSDV), first differenced and system GMM. The results of Chapter 4 are summarised in Cravo (2010) and Cravo et al. (2010).

The discussion of the results presented in Chapter 4 is followed by the examination of two further issues. As the omission of the spatial dependence in economic growth regressions can produce biased results as argued by Rey and Montouri (1999), we use two strategies to control for this phenomenon. Firstly, in Chapter 5 we use cross-section spatial econometrics estimators as suggested by Anselin (1988) to analyse how space affects the relationship between economic growth and SMEs. Following the estimation of cross-section spatial regressions, we also use recent developments in spatial panel data econometrics based on Elhorst

[2] As the data considers only the formal sector, informal self-employment is not considered in the analysis.

(2010) to provide for the first time empirical evidence on the process of economic growth in Brazil using spatial panel data econometrics. In addition, Chapter 5 contributes to the literature by providing evidence on how spatial interactions in the SME sector affect economic growth. The aim here is to see how spatial interactions affect the impact of the SME sector on economic growth. In other words, how does the SME sector size and human capital in neighbouring regions impact on economic growth in a given region? Secondly, instead of considering the spatial structure in the growth regressions, in the latter part of Chapter 5 we analyse for the first time the impact on growth of the explanatory variables, particularly the SME variables, using Brazilian data after removing the spatial dependence from the data using a spatial filtering as proposed by Getirs and Ord (1992) and applied to growth regressions as in Badinger et al. (2004).

Following the extensive analysis of the importance of the SME sector for the regional economic growth process in Brazil, in Chapter 6 we turn to an examination of the sensitivity of SME sector employment series to business cycle fluctuations. This discussion addresses a recent debate in the literature on whether small businesses are more sensitive to business cycles. In an influential paper, Gertler and Gilchrist (1994) suggest that SMEs are more sensitive to cyclical conditions and monetary shocks. On the other hand, recent papers by Moscarini and Postel-Vinay (2009, 2010) appear to contradict this view, based on the argument that SMEs hire proportionally more from unemployment and on empirical evidence from U.S., UK, Denmark, Canada and Brazil. We use another unique constructed dataset for Brazil to provide a comprehensive analysis on this issue for a developing country. Using this data, we provide stylised facts about the relative performance of SMEs' employment during the business cycles. Besides, using a SVAR approach as in Moscarini and Postel-Vinay (2010), we also investigate how shocks in macroeconomic variables affect the behaviour of the relative performance of SMEs employment series. A contribution in our SVAR is that we explicitly incorporate a proxy for credit constraints to shed more light on how this factor influences the behaviour of SMEs' employment during the business cycles.

Therefore, the thesis provides a comprehensive analysis of the behaviour of the SME sector in Brazil in the long-run based on the macroeconomic growth literature in Chapters 4 and 5 and in the short-run based on the macroeconomic business cycle literature in Chapter 6. This is important as there is a link between the short-run and long-run as suggested by Campbell and Mankiw (1987); they argue that fluctuations in the economic activity have a negative impact on the long-run growth performance. Therefore, understanding the business cycles in the SME sector in order to promote policies that dampen fluctuations in this sector might also improve long-run growth prospects.

Chapter 2

Literature Review

2.1 Introduction

In recent years the interest in the study of the importance of small and medium enterprises (SMEs) has attracted increasing attention from researchers and policy-makers alike. Therefore, the aim of this chapter is to provide a review of the literature in the context of the importance of SMEs to the process of economic growth and about the behaviour of this sector during the business cycle fluctuations. To achieve this, the remaining sections of the chapter are organized as follows. Section 2.2 provides a literature review of the economic growth literature from the Solow's (1956) model to a more flexible framework that allows us to incorporate other determinants of economic growth into the model as argued in Sala-i-Martin (1996). Recently, economists started to consider SMEs as an important growth determinant and Sub-section 2.2.2 provides a review of the literature that incorporates explicitly the Small and Medium Enterprise (SME) sector into the growth model and presents some possible channels through which SMEs can impact on economic growth performance. Additionally, Sub-section 2.2.3 discusses the phenomenon of spatial dependence in the process of regional economic growth, and reviews the spatial econometrics literature used to account for the spatial dependence in order to provide more reliable estimations in this context. Finally, Section 2.2 provides a review of the empirical economic growth literature in Brazil that uses both standard and spatial econometrics in Sub-section 2.2.4. Section 2.3 reviews the scarce literature on the behaviour of SMEs during business cycle fluctuations and the last section concludes.

2.2 An Overview of the Economic Growth Literature

This section presents a brief historical overview of the development of the economic growth literature. Also, it provides a review of the literature that provides support to the inclusion of the SME sector in growth models and reviews the empirical growth literature on the Brazilian economy.

Broadly speaking, two main approaches can be distinguished, the neo-classical growth models and the endogenous growth models. The former models are based on the work of Solow (1956), and the latter models are based on the tradition of Lucas (1988) and Romer (1986, 1990).

Solow (1956) constructed his model based on two main equations, the production function and the capital accumulation equation. The first equation is the neo-classical production function and the second determines the path of capital accumulation. Together these equations determine the GDP per capita steady-state. Since the model assumes decreasing marginal returns to capital, this approach predicts that poorer economies tend to grow faster than richer ones in earlier stages (due to the lower capital stock they hold), in other words, this model predicts absolute GDP per capita convergence.[3] The greater the distance of an economy from its own steady-state the faster the growth becomes. Convergence is found when the inverse relationship between the growth of GDP per capita and its initial level is confirmed and this result is more likely to occur for a set of economies with similar economic and institutional characteristics.

However, the empirical evidence did not confirm the prediction of absolute convergence, especially for a group of countries with dissimilar levels of development. Hence, two alternative approaches have emerged to reconcile the theory with the empirical evidence. The first alternative is based on augmented versions of Solow Model with decreasing marginal returns to capital; the other is what is known as endogenous growth models, which are based on specifications that present constant or increasing returns to scale.

[3] The model implicitly suggests that absolute convergence occurs only if structural factors of different economies are the same.

In the endogenous growth models, human capital (or knowledge) is often the starting point to increasing returns to scale and considered one of the most fundamental economic growth determinants. Romer (1986, 1990), for example, formalizes the relationship between economic growth and the stock of knowledge and technical progress. According to Romer, new ideas have special characteristics, they are non-rival commodities.[4] This characteristic can generate positive externalities and increasing returns to scale properties. Ideas and knowledge are non-rival goods because they can be used by one person which in no way limits its use by another.

Another contribution to the endogenous growth literature is found in Lucas (1988). He emphasizes that human capital accumulation can be considered as an alternative source of sustained growth (alternative to technological change). In his model, growth is primarily driven by human capital accumulation, and the differences observed in growth rates across countries can be explained by differences in the rates of human capital accumulation over time. According to Lucas, *education* and *learning by doing* are the main sources of human capital accumulation (or skill acquisition), with skilled labour being allocated to research activities to enhance growth performance.

More recently, Vandenbussche et al. (2005) develop a theoretical endogenous growth model, in which technological improvement is the result of a combination of innovation and imitation. They show that the composition of human capital is important in explaining the growth pattern of countries in different stages of development. The growth-enhancing properties of human capital depend on both its composition and the distance to the technological frontier. Low-skill labour associated with lower education contributes to catching-up towards the frontier in the less developed countries through the process of imitation, and high-skill labour associated with higher education is growth-enhancing in the more advanced countries through the process of innovation. In particular they showed that higher levels of human capital expressed by tertiary education have higher growth effects in

[4] More precisely, Romer (1990) argues that the ideas and knowledge are non-rival goods but human capital itself is rival.

the OECD economies, since this level of human capital is responsible for technological improvements due to higher innovation.

On the other hand, Mankiw et al. (1992) suggest an augmented Solow model that includes human capital accumulation in the production function, recognizing explicitly the role of this capital in the economic growth process. In this model, output is produced from physical capital, human capital and labour. Based on a cross-section of countries, they advocate that the empirical results are consistent with the proposed production function and that their interpretation of the evidence of convergence contrasts sharply with the endogenous growth argument. They argue that after controlling for the growth determinants suggested by the Solow model and human capital, there is substantial convergence in GDP per capita. This is the concept of conditional convergence.

A generalization of the conditional convergence concept is known as “Barro Regressions”, after Barro (1991) seminal work. Once these informal regressions include the investment ratio and initial income, they can be seen as an extension of Mankiw et al. (1992) and interpreted in the same terms. This more *ad hoc* framework was first used by Barro (1991) and can be interpreted as a mixture of neo-classical and endogenous growth models because it considers the hypothesis of convergence together with variables with increasing returns to capital characteristics.

The empirical results based on the work of Mankiw et al. (1992) and Barro (1991) reconciled the theory with the empirical evidence. After their work the tradition of growth studies based on the Solow Model was reborn and Durlauf et al. (2005) listed an extensive number of empirical studies using different sets of control variables in growth regressions. In addition, Sala-i-Martin (2002) states that empirical research demonstrates that the conditional convergence hypothesis is one of the strongest and most robust empirical regularities in macroeconomics data. The neo-classical growth framework is described in detail in the next section.

2.2.1 The Neo-Classical Growth Model

The theoretical basis for our empirical analysis is the Solow (1956) neo-classical model based on the production function with labour-augmenting technical progress, assuming decreasing returns to capital:[5]

$$Y(t) = K(t)^{\alpha}\left[A(t)\ L(t)\right]^{1-\alpha}, \qquad \text{with} \quad 0 < \alpha + \beta < 1 \tag{2.1}$$

where Y is output, K and L are factor inputs, physical capital and labour, respectively. The term A is the level of technology and α is the physical capital elasticity with respect to output.

The model assumes that L and A grow exogenously at constant rates n and g, given by $L(t) = L(0)\,e^{nt}$ and $A(t) = A(0)\,e^{gt}$, respectively. Therefore, the number of effective units of labour, $A(t)L(t)$, grows at rate $n+g$.

We assume that a constant fraction of output, s, is invested and K depreciates at a constant exogenous rate, δ, therefore, the change in the stock of capital is given by the following expression:[6]

$$\dot{K} = \frac{dK(t)}{dt} = sY - \delta K \tag{2.2}$$

If we rewrite the model in terms of effective units of labour, the production function becomes:

$$\overline{y}(t) = \overline{k}(t)^{\alpha} \tag{2.3}$$

And the dynamic specification of the capital stock per effective unit labour that drives the capital stock over time takes the following form:

[5] The description of the model follows closely Mankiw et al. (1992) and Islam (1995).

[6] The expression for the level of physical capital is given by: $S = sY,\ 0 < s < 1$.

$$\dot{\bar{k}}(t) = s\bar{y}(t) - (n + g + \delta)\,\bar{k}(t) \qquad (2.4)$$

with $\bar{y}(t) = \frac{Y(t)}{AL(t)}$ and $\bar{k}(t) = \frac{K(t)}{AL(t)}$ being the units of output and capital in terms effective units of labour.

In the steady-state, the rate of growth of the capital stock is set to zero ($\dot{\bar{k}}$ = 0); and $\bar{k}^*$ satisfies this condition as follows:

$$\bar{k}^* = \left(\frac{s}{n + g + \delta}\right)^{\frac{1}{1-\alpha}} \qquad (2.5)$$

Substituting the expressions found for $\bar{k}^*$ into the production function (2.3), we derive the steady-state value of output:

$$\bar{y}^* = \left(\frac{s}{n + g + \delta}\right)^{\frac{\alpha}{1-\alpha}} \qquad (2.6)$$

Taking natural logs from both sides we get an expression for the effective GDP per capita as follows:

$$\ln \bar{y}^* = \left(\frac{\alpha}{1-\alpha}\right)\ln(s) - \left(\frac{\alpha}{1-\alpha}\right)\ln(n + g + \delta) \qquad (2.7)$$

This equation shows that effective GDP per capita is negatively related to population growth and positively related to the accumulation of physical capital.

Given that $\bar{y}^*$ is the steady-state level of effective income per capita given by equation (2.7) and $\bar{y}(t)$ is its value at time *t*, approximating around the steady-state, the speed of convergence close to the steady state is given by:

$$\frac{d\ln \bar{y}(t)}{dt} = \lambda\left[\ln\left(\bar{y}^{*}\right) - \ln\left(\bar{y}(t)\right)\right] \tag{2.8}$$

where λ = $(n+g+\delta)(1-\alpha)$ is the rate of convergence dependent on the population growth, technology growth and capital depreciation rates and the output elasticity with respect to physical and human capital. The solution for this differential equation implies that:

$$\ln \bar{y}(t) = \left(1-e^{-\lambda T}\right)\ln \bar{y}^{*} + e^{-\lambda T}\ln \bar{y}(0) \tag{2.9}$$

where $\bar{y}(0)$ is GDP per effective units of labour at the initial point in time.

Subtracting $\ln \bar{y}(0)$ from both sides of equation (2.9) we obtain a specification that represents a partial adjustment process:

$$\ln \bar{y}(t) - \ln \bar{y}(0) = \left(1-e^{-\lambda T}\right)\left[\ln \bar{y}^{*} - \ln \bar{y}(0)\right] \tag{2.10}$$

In this model the growth of GDP per effective units of labour between the initial and final period is determined by the distance between the initial GDP level and its steady-state value. Substituting for $\bar{y}^{*}$ from equation (2.7) we obtain the following expression:

$$\ln \bar{y}(t) - \ln \bar{y}(0) = \left(1-e^{-\lambda T}\right)\left[+\left(\frac{\alpha}{1-\alpha}\right)\ln(s) - \left(\frac{\alpha}{1-\alpha}\right)\ln(n+g+\delta) - \ln \bar{y}(0)\right] \tag{2.11}$$

where the growth of GDP is explained by the determinants of the steady-state and by its initial level.

However, estimating the convergence equation (2.11) does not take into account that economies may differ in their production functions which is reflected in the unobservable $A(0)$. A GDP per capita formulation (rather than the GDP in terms of effective units of labour) would allow us to specify an equation that considers the economy technological specific effect explicitly.

From the initial definition of income per effective worker; $\bar{y}(t)=\frac{Y(t)}{A(t)L(t)}=\frac{Y(t)}{L(t)A(0)e^{gt}}$, after taking natural logs we obtain:

$$\ln \bar{y}(t)=\ln\left[\frac{Y(t)}{L(t)}\right]-\ln A(t)\Leftrightarrow \ln \bar{y}(t)=\ln y(t)-\ln A(0)-gt \qquad (2.12)$$

Substituting (2.12) into equation (2.11) we obtain the usual convergence equation in terms of GDP per capita:

$$\begin{aligned}\ln y(t)-\ln y(0)=-\left(1-e^{-\lambda T}\right)\ln y(0)+\left(1-e^{-\lambda T}\right)\frac{\alpha}{1-\alpha}\ln(s)\\ -\left(1-e^{-\lambda T}\right)\frac{\alpha}{1-\alpha}\ln(n+g+\delta)+\left(1-e^{-\lambda T}\right)\ln A(0)+gt\end{aligned} \qquad (2.13)$$

where $\left(1-e^{-\lambda T}\right)\ln A(0)$ is the economy time-invariant individual effect. Equation (2.13) is the framework that has been used for empirical analysis of the standard economic growth model. It states that GDP per capita growth is inversely related to the initial level of GDP per capita, a phenomenon known as convergence. Additionally, growth is determined by the determinants of the steady-state; physical capital, population growth, the initial level of technology and the technological progress.

However, Mankiw et al. (1992) found that the original Solow model does not correctly predict the magnitude of the effect of physical capital on growth. To overcome this anomaly, they augmented the model to encompass human capital into the analysis. The introduction of human capital boosted the performance of the Solow model and became the standard specification for growth analysis.

2.2.1.1 Adding Human Capital Accumulation to the Solow Model

Mankiw et al. (1992) claim that human capital is an important factor omitted in the original Solow model. To avoid ignoring this important factor they propose an augmented Solow model that encompasses human capital. The augmented production function is given by:

$$Y(t) = K(t)^{\alpha} H(t)^{\beta} \left[A(t)\ L(t)\right]^{1-\alpha-\beta}, \quad \text{with} \quad 0 < \alpha + \beta < 1 \tag{2.14}$$

where Y is output, K physical capital, H human capital, L labour, A is the level of technology and α and β are the physical and human capital elasticities with respect to output.

The model is derived under the same assumptions as the original Solow model and assumes that L and A grow exogenously at constant rates n and g, given by $L(t) = L(0)\,e^{nt}$ and $A(t) = A(0)\,e^{gt}$, respectively. Therefore, the number of effective units of labour, that is, $A(t)L(t)$, grows at rate $n+g$.

Analogous to Section 2.2.1 and, under the same standard neo-classical assumptions, the augmented production function, in terms of effective units of labour, is given by

$$\bar{y} = \bar{k}^{\alpha} \bar{h}^{\beta} \tag{2.15}$$

The dynamic specification of the model that now incorporates human and physical capital takes the following form:

$$\dot{\bar{k}}(t) = s_k\ \bar{y}(t) - (n + g + \delta)\,\bar{k}(t) \tag{2.16}$$

$$\dot{\bar{h}}(t) = s_h\ \bar{y}(t) - (n + g + \delta)\,\bar{h}(t) \tag{2.17}$$

Where s_k is the fraction of income invested in physical capital and s_h the fraction invested in human capital.

The steady-state for the human and physical capital stock ($\dot{\bar{k}}$ =0 and $\dot{\bar{h}}$ = 0) are represented by:

$$\bar{k}^* = \left(\frac{s_k^{1-\beta} s_h^{\beta}}{n+g+\delta} \right)^{\frac{1}{1-\alpha-\beta}} \tag{2.18}$$

$$\bar{h}^* = \left(\frac{s_k^{\alpha} s_h^{1-\alpha}}{n+g+\delta} \right)^{\frac{1}{1-\alpha-\beta}} \tag{2.19}$$

Substituting the expressions found for $\bar{k}^*$ and $\bar{h}^*$ into the production function (2.15), we derive, analogously, the steady-state value of output:

$$\bar{y}^* = \left(\frac{s_k^{1-\beta} s_h^{\beta}}{n+g+\delta} \right)^{\frac{\alpha}{1-\alpha-\beta}} * \left(\frac{s_k^{\alpha} s_h^{1-\alpha}}{n+g+\delta} \right)^{\frac{\beta}{1-\alpha-\beta}} \tag{2.20}$$

Taking natural logs from both sides we get the following expression:

$$\ln \bar{y} = \left(\frac{\alpha}{1-\alpha-\beta} \right) \ln(s_k) - \left(\frac{\alpha+\beta}{1-\alpha-\beta} \right) \ln(n+g+\delta) + \left(\frac{\beta}{1-\alpha-\beta} \right) \ln(s_h) \tag{2.21}$$

Substituting (2.21) into the expression that represents the adjustment process towards the steady-state (2.10) we get:

$$\ln \bar{y}(t) - \ln \bar{y}(0) = \left(1 - e^{-\lambda T}\right) \left[\begin{array}{l} \left(\frac{\beta}{1-\alpha-\beta} \right) \ln(s_h) + \left(\frac{\alpha}{1-\alpha-\beta} \right) \ln(s_k) \\ - \left(\frac{\alpha+\beta}{1-\alpha-\beta} \right) \ln(n+g+\delta) - \ln \bar{y}(0) \end{array} \right] \tag{2.22}$$

Finally, re-arranging the expression (2.22) we obtain the convergence equation in terms of GDP per capita:

$$\begin{aligned} \ln y(t) - \ln y(0) = &-\left(1-e^{-\lambda T}\right) \ln y(0) + \left(1-e^{-\lambda T}\right) \frac{\beta}{1-\alpha-\beta} \ln(s_h) + \left(1-e^{-\lambda T}\right) \frac{\alpha}{1-\alpha-\beta} \ln(s_k) \\ &- \left(1-e^{-\lambda T}\right) \frac{\alpha+\beta}{1-\alpha-\beta} \ln(n+g+\delta) + \left(1-e^{-\lambda T}\right) \ln A(0) + gt \end{aligned} \tag{2.23}$$

Similar to equation (2.13), GDP per capita growth is inversely related to the initial level of GDP per capita and growth is determined by the determinants of the steady-state that now incorporates human capital.

2.2.1.2 Barro Regressions

Although Mankiw et al. (1992) provide a theoretical framework for growth regressions in line with the empirical evidence, Temple (1999) points out that the most common approach is the use of a more *ad hoc* regression that encompasses other factors that influence growth. These variables are chosen based on previous results in the literature rather than on an explicit theoretical model. Regressions of this type are known as "Barro Regressions", after Barro (1991)'s seminal work. Once these informal regressions include the investment ratios and initial income, they can be seen as an extension of Mankiw et al. (1992).

The more *ad hoc* framework is arguably more flexible and can implicitly be seen as a mixture of neo-classical and endogenous growth models, once it encompasses the hypothesis of convergence and the use of policy variables that can present increasing returns to capital characteristics.[7] Barro and Sala-i-Martin (1992) and Sala-i-Martin (1996), for example, estimate a simplified representation of the growth framework, based on equation (2.23) and Barro's tradition given by the following equation:

$$gr_{it} = a_i - b \ln y_{i,t-1} + \psi X_{it} + v_{it} \qquad (2.24)$$

where *gr* denotes the GDP per capita growth, $\ln y_{t-1}$ is the initial GDP per capita, *i* denotes each individual economy, *t* represents each period of time considered, the constant term represents the time invariant economy specific effect, and *X* represents a vector of variables that hold constant the steady-state of economies. This vector encompasses the growth determinants suggested by the original Solow model, as in

[7] This framework is more flexible in the sense that it does not need a formal theoretical derivation to include control variables into the model.

equation (2.23). However, in the Barro (1991) tradition, the vector X is flexible enough to encompass other growth determinants that come from outside the Solow model (e.g. rule of law, trade, inflation, SMEs). If the coefficient of the initial GDP per capita is negative (b>0) and $\psi \neq 0$ the GDP per capita exhibits conditional convergence.

Empirical research based on "Barro Regressions" flourished, for instance, Durlauf et al. (2005) list more than a hundred growth determinants included in this type of growth regressions. In addition, according to Sala-i-Martin (2002) an important contribution made by the growth literature that follows the Mankiw et al. (1992) and Barro (1991) tradition is that it has exerted influence on other economic literatures such as development, economic geography, macroeconomics, econometrics and industrial organisation. Recently, this influence was also extended to the importance of SMEs and entrepreneurship for economic growth (e.g. Beck et al. 2005a; Audretsch and Keibach 2004; Mueller 2007) and the next section discusses this in more detail.

2.2.2 SMEs as Determinants of Economic Growth

In recent years the study of SMEs and economic growth has attracted increasing attention. However, although this sector employs the majority of the labour force in developed and developing countries alike, little has been done to study the importance of SMEs for economic growth. In this section we review the literature that incorporates SMEs and entrepreneurship in the neo-classical growth framework, arguing that these factors are important growth determinants that have been omitted in the growth framework.

Researchers have been arguing that small firms and entrepreneurship are important growth determinants that have been omitted in the neo-classical growth framework (e.g. Audretsch and Keilbach 2004; Audretsch 2007). Furthermore, Solow (2007) recognizes entrepreneurship as an important force that drives a wedge between knowledge and total factor productivity. It can bridge the gap between specific pieces of technological knowledge and innovations through the creation of new firms. Therefore, explaining how SMEs affect growth could add to the explanatory power of growth theory.

Though the interest in SMEs has been increasing lately, there is a lack of studies relating SMEs and economic growth, particularly in developing countries. Only in a few papers we can observe the inclusion of the SME sector into economic growth framework, which provides a limited literature review about this subject but clearly indicates that further studies on this topic are needed to fulfil the knowledge gap concerning SME and economic growth.

A seminal study using growth regressions to analyse SMEs and growth is provided by Beck et al. (2005a). They use the generic representation of the growth regression presented earlier (equation 2.24) to study the impact of the size of the SME sector on growth in a cross section of countries of different levels of economic development. They include a new variable in the model that represents the share of SME employment in manufacturing (using 250 employees as the cut-off for the definition of a SME) along with other variables that control for the steady-state of the countries.[8] Beck et al. (2005a) use a sample of 45 countries to estimate a cross-section regression for the period 1990-2000 using OLS and 2SLS.[9] The results from OLS regressions provide evidence that there is a positive relation between the importance of the relative size of the SME sector (given by the relative size of the SME sector defined as the share of the manufacturing labour force in firms with 250 or fewer employees in total manufacturing labour force) and GDP per capita growth. However, they also point out that this relationship is not robust when controlling for simultaneity bias (when using the 2SLS and instrumental variables) and additional research is necessary to understand better the impact of the SME sector on economic growth.

The European Commission report "The 2003 Observatory of European SMEs" illustrates the role of SMEs in the economy stressing the importance of the

[8] This new variable was constructed by Ayyagari et al (2003, 2007). Beck et al. (2005a) use 250 employees as cut-off because a large number of countries use this cut-off to classify SMEs.

[9] OLS stands for ordinary least squares and 2SLS (two-steps least squares estimator) is used to control for endogeneity. To control for simultaneity bias, it is also appropriate to use instrumental variables (IV) to extract the exogenous component of SME250. Beck et al (2005a) focus on exogenous national characteristics that theory and past empirical findings suggest influence the business environment. Their core instrument set include an indicator of ethnic diversity and dummy variables for transition, African and Latin American countries.

emergence of the entrepreneur economy to economic growth in Europe. It provides results that support the view that SMEs have a positive impact on European economic performance. The report also suggests that three mechanisms are important to generate growth. First, SMEs are seen as a vehicle for knowledge spillovers. Second, SMEs exert positive influence on economic output by increasing the amount of competition. Finally, entrepreneurship provides a great amount of diversity among enterprises and this degree of diversity influences the economic growth potential.[10] Similarly, Carree and Thurik (1998) suggest that a greater employment share of SMEs led to better economic performance in 13 European countries in the early 1990s. The results indicate that an industry with a greater presence of small firms, relative to the same industries in other countries, performed better in terms of output growth.

Audretsch et al. (2002) estimate the impact of deviation from optimal industrial structure on growth using data for 18 European countries. The overall idea is that, if small firm presence is above its optimal level, adding new firms would reduce economic growth, if otherwise, the entry of new small firms would enhance economic performance.[11] Hence, the deviation from optimal industry structure generates a growth penalty. The results show that, in general, a larger shift toward small firms is associated with higher growth acceleration.

Audretsch and Keilbach (2004) suggest that a key factor has been omitted from the neo-classical production function proposed by Solow (1956): the entrepreneurial capital[12]. They propose an augmented production function that encompasses entrepreneurial capital as a determinant of output and provide at least three reasons to believe that this factor presents a positive effect on growth. First, through the spillover mechanism, entrepreneurial capital can generate knowledge via either the appropriation of knowledge from other firms or through a knowledgeable entrepreneur (from the university or another firm) that decides to begin a new firm.

---

[10] This report was produced by David B. Audretsch, Roy Thurik, Ton Kwaak and Niels Bosma.

[11] Small firm presence is measured in terms of value of shipments. Audretsch et al. (2002) use the difference between the growth rate of small and large firms to test this hypothesis. Their empirical results suggest that a higher growth of small firms is positively related with economic growth. Therefore, they argue that the presence of small firms is below its optimal level and the economy's industrial structure deviates from its optimal.

[12] They used the star-up rate as an indicator of entrepreneurship capital.

Second, entrepreneurship capital can exert a positive influence on growth through competition by an increased number of enterprises competing for market share. Finally, the third reason is by providing diversity across firms. Their estimations are based on a cross-section of 327 West German *Kreise* (regions) for the year 1992. The augmented production function including the entrepreneurial capital (measured by the start-up enterprise rates) suggests that this capital is important to economic performance across Germany Regions.

Mueller (2007) follows Audretsch and Keilbach (2004) closely but considers an augmented production function that encompasses the entrepreneurial capital along with R&D employment in adjacent regions to capture knowledge spillovers. To estimate the augmented function, Mueller (2007) employs a panel data of 72 planning regions in West Germany using data from 1990 to 2003. The estimations are in line with the results obtained by Audretsch and Keilbach (2004), and provide evidence that entrepreneurial capital is important to growth. Additionally, Mueller (2007) finds that the additional variable that captures the knowledge stock in adjacent regions affects economic growth as well, indicating the presence of positive regional spillovers stemming from R&D activities.

In another study using a cross-section of countries, Van Stel et al. (2005) provide evidence using data for 36 countries that suggests a positive impact of the total entrepreneurship rate on economic growth for developed countries and the opposite result for developing ones.[13] They argue that a possible explanation for the negative effect in poorer countries is that the entrepreneurs have lower human capital levels compared to entrepreneurs in developed countries. According to them, "it is likely that the negative effect reflects the presence of many marginal entrepreneurs (shopkeepers) in small crafts who may be more productive as wage-earner in a bigger firm." However, they did not have a proxy for the entrepreneurs' human capital to extend their analysis.

Acs et al. 2008 also offer a discussion on why entrepreneurship affects economic growth differently in developed and developing countries. The different

[13] The measure of entrepreneurship used by Van Stel et al (2005) is the total entrepreneurial activity, defined as that percentage of adult population (18–64 years old) that is either actively involved in starting a new venture or is the owner/manager of a business that is less than 42 months old (which can be formal or informal).

impact of entrepreneurship on growth could be attributed to different entrepreneurship responses to the institutional set up and level of development. For instance, they argue that the relationship between entrepreneurship and economic development is most likely negative in developing countries and most likely to be positive in developed countries. This is because the opportunity to necessity entrepreneurship ratio is greater in developed countries.[14] In aggregate terms, the SME sector in developing countries is dominated by labour intensive and low-tech firms that are more likely to be related with necessity entrepreneurship. Acs and Amorós (2008), for instance, show that a high proportion of the entrepreneurial activity in Latin America is driven by necessity entrepreneurship. Because entrepreneurship seems to respond differently according to the level of development, the effect of SMEs on growth is likely to be different due to different characteristics of the SME sector. For example, Table 1 in Beck et al. (2005a) shows that SMEs' employment share in Brazil is 58.8% while in Finland it is 58.15%. Therefore, a proxy for the presence of SMEs alone might not be able to show us the true impact of SMEs on growth because the nature of the SME sector is likely to be different.

Before continuing the discussion about the importance of institutions for the performance of SMEs and entrepreneurship, a note on the direct link between these two concepts is useful. Importantly, while start-up rates are used in Audretsch and Keilbach (2004), Mueller (2007) and Van Stel et al. (2005) represent dynamic measures of entrepreneurship, the stock of SMEs' employment used in Beck et al. (2005a) and in Carree and Thurik (1998) is a static measure. They represent different dimensions of entrepreneurship but are clearly related concepts. As observed in Wennekers and Thurik (1999), small firms are vehicles by which individuals channel their entrepreneurial ambitions. Higher small businesses start-up rates will lead to a higher share of workers in the SME sector, while a higher mortality rate will adjust the size of this sector in opposite direction. In this sense, the employment share of the small business sector also adjusts for failures in the entrepreneurial activity (Cravo et al. 2010). Also, Reynolds et al. (1994) and Audretsch and Fritsch (2002) found that a

[14] Necessity entrepreneurship is the act of starting a business as the last resort. In contrast, opportunity entrepreneurship expresses the other motivation for starting a business to exploit a perceived business opportunity (Acs et al. 2008).

greater share of employment in small businesses is positively correlated with a higher level of start-up rates.

The channels through which the institutional arrangements can impact on entrepreneurship and growth are discussed in Baumol (1990, 2008). Baumol (1990)'s basic hypothesis suggests that while the total supply of entrepreneurs differs across economies, the productive contribution of the society's entrepreneurial activities varies much more due to their allocation between productive and unproductive activities. Hence, policy makers should worry about the allocation of entrepreneurship by providing a good institutional arrangement that promotes productive entrepreneurship at the expense of rent seeking[15]. Similarly, Baumol (2008) discusses the role of innovative entrepreneurs in the economic growth process and how public policy can stimulate the entrepreneur's contribution to productivity. He also stresses the importance of small firms to innovation, arguing that major breakthroughs that are vital for growth tend to come from small new enterprises. He argues that institutions (along with other factors) are important to entice entrepreneurs away from their previous unproductive activities and leading them to productive undertakings. He suggests that the handicaps to the creation of a new firm, such as taxes, should be removed and institutional arrangements should discourage unproductive entrepreneurship. The activity of innovative entrepreneurship should also consider the importance of technology transfer through the absorptive capacity. He concludes that the success of one economy depends on its success in promoting productive entrepreneurship, innovation and importation of technology. Additionally, he states that small firms are important for all of these activities and play a critical role in the growth of the economy and that it is a serious mistake to overlook measures that encourage productive entrepreneurship.

Institutions may also be responsible for inciting human capital formation for productive entrepreneurs. Dias and McDermott (2006) propose a model where structural changes towards a modern economy depend on the role of entrepreneurs, human capital and institutions. In their model, entrepreneurs come from a pool of

[15] According to Baumol (1990), unproductive entrepreneurship takes many forms. Litigation and takeovers, tax evasion and avoidance efforts seem to constitute the prime threat to productive entrepreneurship.

individuals that belong to the managerial class, which is specialized in two activities; rent-seeking and entrepreneurship. The important point is that more (productive) entrepreneurs lead to more human capital formation. Workers will look for education that suits a more productive job (that requires a higher level of human capital) offered by the entrepreneurs. Therefore, barriers that prevent productive entrepreneurship to develop should be removed and better institutional policies are required to create more productive entrepreneurs and improve economic performance. If unproductive entrepreneurship dominates, educational improvements will be neutralized and will have little long-run effect.

The importance of institutions for entrepreneurship is assessed by Nystrom (2008). Using a panel of 23 OECD countries over the period 1972-2002, Nystrom uses self employment as a proxy for entrepreneurship and measures of economic freedom (e.g. size of the government, legal structure, access for money, freedom to trade and regulation of credit) to account for the institutional quality. The empirical results support the idea that institutional quality is important for the entrepreneurial activity. If institutions fail, it is likely that the SMEs' performance will fail as well. In addition, Beck et al. (2008) argue that institutional failure in the form of financial constraint prevents SMEs to fully develop and can be a burden to their ability of generating economic growth. In the same line, Michelacci and Silva (2007) provide evidence that financial markets are a constraint to local entrepreneurship.

Nevertheless, due to differences in institutions, human capital and rent-seeking levels, the presence of SMEs in a developing economy probably does not have the same meaning as in a knowledge based economy. Thus, this reveals a caveat in Beck et al. (2005a) because the pure presence of SMEs alone might be difficult to interpret. Differences in the level of institutional development provide different incentives to productive entrepreneurship and we do not know whether the effect of SMEs on growth comes from the structure itself or from another factor related to SMEs, such as, the level of human capital embodied in SMEs.

Human capital is an important factor to the process of economic growth and development. In Nelson and Phelps (1966), the rate of increase in technology level is an increasing function of educational attainment. Education speeds up the process of technological diffusion, and a higher level of education tends to incite the adoption of innovations earlier. Similarly, Barro (2001) argues that human capital facilitates the absorption of superior technologies and generates growth. Besides, human capital can also encourage growth by stimulating the process of innovation. Similarly, Griffith et al. (2004) argue that R&D has two faces and stimulates growth directly through innovation and indirectly through facilitating the imitation process. They found evidence that educational attainment has an important role in stimulating both, the absorption of superior technologies (absorptive capacity) and innovation. As a result, a higher level of human capital in SMEs might be important for the absorptive capacity of this sector and is an important factor to be considered when studying the relationship between SMEs and growth. Conversely, lower levels of human capital do not stimulate innovation and the absorptive capacity, and are more likely to be associated with SMEs motivated by necessity entrepreneurship.

Therefore, a SME proxy that encompasses human capital can shed additional light on the relationship between SMEs and growth once it takes into account the ability of the SME sector to appropriate knowledge from more productive firms. If the SME sector improves its productivity, through innovation or imitation, a positive effect on growth is expected from the SMEs' human capital level.

2.2.3 Econometric Issues and Spatial Dependence

In the empirical growth literature using panel data the most commonly used estimation methods are the LSDV, the first differenced GMM and the system GMM. The former estimator controls for the individual level of technology in each economy as for example in Islam (1995). The latter two are designed to also control for endogeneity as for example in Caselli et al. (1996) and Hoeffler (2002).

However, economic growth models have been criticised by scholars that argue that we should not treat economies independently, especially in cases where the economies are in a contiguous space. For instance, Temple (1999) discusses the evidence for spatial correlation in economic growth estimations and argues that the standard errors in most growth regressions should be treated with a certain degree of mistrust due to the presence of spatial dependence. Nevertheless, he also argues that the adjustment for spatial correlation raises formidable statistical problems, and the identification of genuine spillovers is problematic and needs further investigation. Thus, the analysis of economic growth should consider the existence of spatial dependence to avoid misspecification of the models.

A seminal work on spatial econometrics can be found in Anselin (1988). He discusses the importance of spatial dependence in the estimation process and proposes alternatives approaches to consider the spatial effects in an econometric perspective. According to Anselin (1988), spatial effects have been ignored by the traditional econometricians and were left to the separate field of spatial econometrics.

He argues that spatial dependence can emerge in the form of spatial autocorrelation and spatial heterogeneity. Spatial autocorrelation is expressed by the lack of spatial independence which is often observed among observations in cross-sectional data sets. He argues that this dependence can be considered to lie at the core of the disciplines of regional science and geography, as expressed in Tobler (1970)'s first law of geography, in which everything is related to everything else, but near things are more related than distant things. In other words, there is positive spatial autocorrelation when values of a random variable tend to be more similar the closer the two observations are in space. In its turn, spatial heterogeneity means that economic behaviour is not stable across geographic space, generating spatial regimes with clusters of high or low values of a variable. The omission of spatial dependence in a growth regression can be reflected in the error term and produce biased results. Two popular strategies to account for the presence of spatial dependence are proposed in Anselin (1988). The first considers the structure of the spatial dependence in the error term (spatial error model) and the second is a spatial autoregressive model (spatial lagged model) where the spatial dependence is modelled in the form of a spatially lagged dependent variable. However, he shows that the estimation of models with spatial dependence using OLS produces unreliable estimations. In the case of a spatial error model, due to non-spherical residuals, the inference might be misleading and in the case of a spatial lagged model due to the simultaneity introduced through the use of the spatially lagged dependent variable the spatial parameters estimations are biased. Therefore, he suggests and derives the maximum-likelihood estimators that are able to provide consistent estimations in the presence of spatial dependence.

Rey and Montouri (1999) consider the process of income convergence from a spatial econometric perspective using cross sectional data for the U.S states for the period of 1929-1994. They estimate a growth model without considering the spatial dependence and detected the presence of spatial dependence in the error term, suggesting that the convergence model is misspecified when spatial effects are not considered. To consider the presence of spatial dependence, they estimate the spatial error model and the spatial lagged model using the maximum-likelihood estimators proposed by Anselin (1988). Their results show that models that incorporate spatial effects perform better and the spatial dependence seems to be better explained by the spatial error model, suggesting that random shocks to one state are propagated throughout the U.S states.

A detailed survey provided by Abreu et al. (2005) reviews the literature and shows that spatial dependence is an important aspect of the economic growth process. The empirical evidence from the literature confirms that spatial dependence is a regular phenomenon, therefore, the use of spatial econometrics to estimate regional growth models should be the standard procedure. For instance, Henley (2005) considers the spatial effects in the convergence equation in Great Britain and concludes that the regional GDP data are exposed to substantially spatial autocorrelation.[16] The use of cross-sectional spatial econometrics models show that Great Britain appears to be an economy of growing spatial inequality, where the dynamics of the underlying growth process are highly spatial in pattern with only some counties acting as regional growth poles that impact on their surrounding areas' growth.

López-Bazo et al. (2004) use cross-sectional data for 108 European regions to test for the interdependencies between regions using spatial econometrics. Their empirical results suggest that growth and the initial GDP per capita level of neighbouring regions affect the GDP per capita growth in a given region. This

[16] Henley (2005) uses two regional datasets that provide similar qualitative results; the first is the GDP for 62 British counties and estimated Gross Value Vdded (GVA) for 128 Nomenclature of Units for Territorial Statistics of the European Union (NUTS III). Data are from the UK Office for National Statistics.

suggests the presence of spillovers in the regional growth process in Europe which can be detected through the use of spatial econometrics.

Ertur and Koch (2007) use cross-sectional data with 91 countries to test a growth model augmented with technological spillovers. Their results using spatial econometrics show that their spatially augmented Solow model provides a better understanding of the important role played by spillover effects in international growth and convergence processes.

Arbia et al. (2010) use spatial econometrics to access the importance of institutions to economic growth in Europe. They estimate a growth model using a cross-section of 271 European regions (NUTS II) and found the presence of spillovers stemming from neighbouring regions. Also, their results suggest that regions that are neighbours to regions with more developed institutions tend to converge more rapidly.

Dall'erba and Le Gallo (2008) evaluate the impact of structural funds on the growth convergence process among 145 European regions over 1989–1999. Their results suggest the existence of spatial spillovers in the economic growth process but no effect of structural funds on growth.

Ramajo et al. (2008) also use spatial econometrics to assess the impact of EU cohesion policies on growth convergence controlling for spatial heterogeneity and interregional spillovers. Their results indicate the presence of significant geographic spillovers in the EU regional growth process. Also, they suggest that countries subject to cohesion policies are converging at a faster rate. Thus, the results provide support for policies designed to promote regional growth in the less-developed regions.

However, the empirical results mentioned above are based on cross-sectional spatial regressions. Panel data models have advantages over the cross-section ones as they are generally more informative, control for individual and time fixed effects, and contain more variation and less collinearity among the variables in the model. The use of panel data also results in greater availability of degrees of freedom and therefore increases the efficiency in the estimation (Elhorst 2010). Also, Islam (1995) in the context of growth models uses panel estimations to allow for time specific and individual (country or region) specific levels of technology, the same argument is valid in the spatial context. The need to account for spatial heterogeneity stems from the fact that spatial units have individual characteristics that are space-specific and time invariant (Elhorst 2005), and cross-section estimates cannot control for spatial and time fixed effects and their results might be biased (Elhorst and Fréret 2009).

Nevertheless, because the estimation of spatial panel models is more complex, the application of this setting in regional convergence studies is at a very early stage (Arbia et al. 2008). However, new developments in the estimation of spatial panel models made, for instance, by Elhorst (2003, 2005 and 2010) provide procedures to estimate this type of model. In these papers, he shows that maximum likelihood estimations of spatial fixed effects model can be carried out with standard techniques developed by Anselin (1988) after demeaning the variables in the model.

For instance, Mohl and Hagen (2010) apply a spatial panel econometric estimator as proposed by Elhorst (2010) to analyse the impact of the EU structural funds on 126 EU regions over the period 2000-2006. The results suggest that a given region benefits from economic growth in neighbouring regions, as evidenced by the coefficient of the spatially weighted dependent variable.

Arbia et al. (2008) performs a growth exercise with regional European data to check whether the evidence on regional economic convergence depends on the estimation strategy. They estimate cross-section and panel data models using both standard econometrics and spatial econometrics. Their results indicate that all models consistently find convergence. However, the spatial panel regressions can account for the spatial fixed effect and control for spillovers.

Badinger et al. (2004) considered a different strategy to control for the spatial dependence in a panel data setting. Instead of modelling the spatial relationships they filtered them from the growth model. They provide evidence that the European regional data (NUTS II) for the period 1985-1999 exhibit a high degree of spatial correlation. Then, to estimate a consistent growth regression and take this spatial correlation into account, they use the spatial filter developed by Getis and Ord (1992). After isolating the spatial effects they claim that the estimated results are more plausible to a convergence process than the previous panel data convergence analyses for EU regions. Maza and Villaverde (2009) and Battisti and Di Vaio

(2008) also use spatial filtering to analyse economic growth in the Spanish regions and in the European regions, respectively. The disadvantage with this methodology is that it removes the space from the analysis and cannot say anything about the degree of spatial spillovers in the model.

Hence, the literature provides evidence about the existence of interactions and regional spillovers in growth regressions and an important way of considering these effects is through the use of spatial econometrics.

2.2.4 The Economic Growth Literature on Brazil

This section presents a review of the empirical evidence on the economic growth literature in Brazil.

Ferreira and Diniz (1995) present evidence of absolute convergence in the income per capita growth in the Brazilian states over the period 1970-1985. In a more detailed study, Ferreira (2000) provides evidence about the convergence process among the Brazilian states over the period 1970-1995 using panel data estimations. The results suggest the existence of conditional convergence in the Brazilian states and that the implied convergence rate is approximately 3% per annum for the period of analysis.

Azzoni (2001) analysed the growth process in the Brazilian states over the period 1939 – 1995. He estimated the standard growth convergence equation and found convergence in the GDP per capita for the whole period of analysis. However, the results indicate that the rate of convergence varied a lot in different sub-periods of analysis and he argues that the overall result that Brazilian states converge is due to the strong convergence process that took place between 1975 and 1990.

More recently, Nakabashi and Salvato (2007) provide evidence of conditional convergence for the Brazilian states using panel data estimations controlling for specific individual state effects.[17] They also stress the importance of human capital

[17] In their paper, convergence is conditional on structural factors, such as, population growth, human and physical capital. This is similar to the Equation 2.23 presented earlier in this section.

and the results indicate that both quality and quantity of human capital are important to explain the disparity in income level across States.[18]

In addition, Laurini et al. (2005) used a non-parametric analysis to confirm the existence of conditional convergence and suggest the existence of club convergence for the Brazilian municipalities.[19] Alternatively, using a "regression tree" analysis, Coelho and Figueiredo (2007) also find a similar pattern.[20] Their results suggest the existence of club convergence and confirm the regional pattern that the northeast region belongs to the poorest club while the south and southeast states belong to the richest one. These studies suggest the existence of spatial heterogeneity in the process of economic growth in Brazil.

Importantly, none of the above mentioned studies considered explicitly the existence of spatial dependence. Following the evidence on the need to consider the spatial dependence, the recent literature on regional economic growth in Brazil started to use cross-sectional spatial econometrics to account for the importance of spatial spillovers.

Silveira-Neto and Azzoni (2006) estimate growth regressions using spatial econometrics to analyse the economic growth process in Brazil. They use a cross-sectional data for the Brazilian states for the period between 1985 and 2001. Their results suggest the existence of spatial heterogeneity in the country expressed by a low-income cluster in the northern part of the country and a high-income cluster in the southern part. In addition they found the presence of spatial spillovers operating in Brazil and suggest that spillovers might operate through human capital.

Resende (2009) provides an analysis of the economic growth process in four different spatial scales in Brazil (states, micro-region, spatial cluster and

[18] The human capital proxy that used is years of schooling multiplied by an index of education quality, made up of variables related to infra-structure, teaching and student performance quality. The variables are the percentage of teachers holding an undergraduate degree, student performance (pass rate) and number of students per classroom.

[19] In the club convergence hypothesis (polarisation, clustering), per capita incomes of countries that are identical in their structural characteristics converge to one another in the long-run provided that their initial conditions are similar as well (Galor, 1996). For instance, Quah (1996, 1997a, 1997b) argues that polarization in the dynamics of the distribution of GDP per capita across regions indicates the existence of club convergence.

[20] The regression tree procedure sorts the sample into ascending order and divides the sorted sample in various sub-samples in order to allow the intercept and slope of the growth regression to vary across the sub-samples. The procedure then tests the hypothesis that the parameters of the growth regression do not vary across the sub-samples. The rejection of this hypothesis is consistent with the existence of multiple basins of attraction as implied by the club convergence hypothesis (see Johnson and Takeyama, 2001).

municipalities). His results suggest that spatial dependences are operating in all spatial scales but are particularly stronger in the most disaggregated geographic level of municipalities.

Therefore, the literature on economic growth in Brazil provides evidence about the existence of convergence among Brazilian regions and indicates the presence of spatial spillovers in this process. Yet, to the best of our knowledge, there is no study about the importance of the SME sector for the Brazilian economic growth and no study using a more sophisticated panel data analysis to study in more detail the role of spatial dependence in the country.

2.3 Business Cycles and SMEs' Performance

The output of economies does not grow smoothly, and movements about the trend in output, or business cycles, in any country can be characterized by a stochastic disturbed difference equation of very low order (Lucas, 1977). Since the early days of the study of business cycles in the U.S. to more recent times, it has been an empirical regularity that the pattern of total employment is pro-cyclical and follows the fluctuation of real GDP very closely (e.g. Burns and Mitchell 1946; Moore 1961; Kydland and Prescott 1990; Hodrick and Prescott 1997; Veracierto 2008). The literature on cycles in Brazil also suggests that employment is pro-cyclical with GDP (e.g. Ellery et al. 2002; Issler et al. 2009). However, little is known about the behaviour of the level of employment in SMEs during business cycles and further studies could provide valuable information about the best policy to be followed in order to dampen employment and economic fluctuations. The paucity of evidence on the behaviour of SMEs during business cycles is surprising given that the small enterprise sector employs the majority of the labour force, both in developed or in developing countries[21].

[21] See Ayyagary et al. (2007) and Beck et al. (2005a).

Brock and Evans (1989), for instance, argue that research on the role of small firms in the labour market during business cycles can provide critical information for the sensible formulation of government policy. They conjecture that different behaviour of small businesses during cycles is likely to be related with the stronger financial liquidity constraints that they face. In the same line, the European Commission report "The 2003 Observatory of European SMEs" argues that in the existing economic literature, little has been published about the behaviour of SMEs as a response to business cycles in terms of employment and output. For example, the process of entry and exit is highly relevant for the development of the SME sector, but so far not linked to existing business cycle indicators.

In a rare and very influential paper on this subject, Gertler and Gilchrist (1994) provide evidence indicating that small firms are more sensitive to cyclical conditions and account for a significantly disproportionate amount of the resultant decline in manufacturing that follows tightening of monetary policy. They study three sets of variables: sales, inventories, and short-term debt. For each variable they construct time series of growth rates for small and large firms. The growth rates of sales and inventories in small firms exhibit a similar pattern and decline sharply relative to large firms after episodes of tight money and during recessions. For instance, their evidence suggests that small firms' sales drop substantially more than large firms' sales. On the other hand, inventories and short term debt increase relatively more in large firms after a tight money episode; these firms appear to borrow heavily to smooth the impact of declining sales on the business.

They also analyse the effect of shocks to tight money events on sales, inventories and debt using vector autoregressions (VARs). The episodes of tight money have a substantial larger effect on small firms. Small firms' sales drop more than four percent per year faster than large firms' sales for a period of ten quarters after a tight money shock. Even sharper differences emerge in the inventory behaviour. Small firms' inventories decline at a faster pace than sales decline, while in large firms inventories drift up for a considerable period. In terms of short term debts, the short-term debt in small firms drops rapidly along with inventories. Overall, large firms appear to borrow to smooth the impact of a downturn, but small firms do not. Additionaly they provide evidence on the asymmetric behaviour of small firms during the cycles. Interestingly, in booms the response of small and large firms looks quite similar. They conclude that small firms contract substantially relative to large firms after tight money and that they account for a significantly disproportionate amount of the ensuing decline in manufacturing. They suggest that the difference in terms of behaviour across firm size class may be due to financial factors at work, supporting the view that after tight money and at the onset of recessions, credit flows to small firms contract relatively more than credit flows to large firms.

Recently, Moscarini and Postel-Vinay (2009) present new empirical evidence for a set of countries (U.S., Canada, Denmark, UK and Brazil) and extensive evidence at regional and sectoral level for the U.S. that suggests that large firms are more sensitive than small ones to business cycle conditions. Using a measure of relative employment performance given by the difference in employment growth rates between large and small employers, their evidence suggests that larger employers shrink faster, or expand more slowly, during a typical recession, and create more of their new jobs late in the following expansion. They argue that after a recession, when unemployed workers are abound, firms hire mostly and cheaply from unemployment. As the reservoir of unemployment dries out, more productive, larger firms find it profitable to start raising wages to raid workers from less productive competitors. Workers quit mostly from small, less productive, low-paying firms moving to large, high-paying firms. The growth in the employment of large firms is fuelled by the stock of employment at small firms, which takes some time to replenish after a recession. Hence, employment in small firms grows faster and peaks earlier than in large firms. This differential source of hiring, unemployment for smaller firms and poaching for larger firms, is the core of their theoretical explanation for the patterns that are documented in their paper and discussed in detail in Moscarini and Postel-Vinay (2008).

Moscarini and Postel-Vinay (2010) complement the argument put forward by Moscarini and Postel-Vinay (2009). They estimate a SVAR with the federal funds rate, inflation rate, unemployment and a relative measure of the performance of small and large businesses given by the differential in average returns between benchmark portfolios of small cap stocks and portfolios of large cap stocks. Their results show that larger firms are more sensitive to shocks in unemployment. Also, they show that shocks to interest rates impact negatively on the performance of smaller businesses. They argue that this is because the interest rate rises during booms reduces small firms' growth and this effect is lessened during recessions (due to lower interest rates).

Pinto (2011) provides evidence on how different firm size and age influence the cyclical properties of aggregate job relocation. Using data for the Portuguese economy, his empirical results show that small and young firms have lower coefficients of variation of job destruction and job creation. As a result, larger firms are expected to influence proportionally more the cyclical variation of aggregate job flows. He argues that his evidence is in line with Moscarini and Postel-Vinay (2009).

Nevertheless, the evidence on the importance of small firms to the business cycle is limited and this is also expressed by the scant evidence on this issue for developing countries. Hence, more attention to the documentation of the sensitivity of small firms to cycles is also needed in developing countries. This is particularly important because the existing literature has documented that business cycle fluctuations and labour market conditions differ significantly in developing and developed countries. Rand and Tarp (2002), Neumeyer and Perri (2005) and Aguiar and Gopinath (2007), for instance, point out some important differences in the pattern of business cycles in developing economies. These countries tend to be prone to sudden crises, often making it difficult to discern economic regularities. Their economies are substantially more volatile when compared with developed countries because the volatilities of output, real interest rates, and net exports are higher. Developing economies are often characterized by frequent regime shocks due to dramatic reversals in fiscal, monetary, and trade policies observed in these economies. Hence, the evidence drawn from developing country data might also be important to construct theoretical models that take into account features of the small firm sector that are particularly important for the business cycle in developing countries. In addition, this evidence might also be important to devise social public policies designed to deal with higher levels of unemployment during a downturn.

2.4 Conclusion

In this chapter we review the literature about the role of SMEs in the economic growth process and in the business cycles. Regarding the relationship between SMEs and growth, firstly, this chapter covers the economic growth literature in a way that allows us to understand how we can consider the SME sector in this framework. We started from the textbook Solow model based on the ideas presented in Solow (1956) and followed developments in the literature (e.g., Mankiw et al. 1992; Barro 1991) that allow us to include the SME sector in a growth model framework as in Beck et al. (2005a). Secondly, we discuss the use of spatial econometrics to control for the presence of spatial dependence and provide better estimations. Thirdly, we review the existing literature that incorporates SMEs and entrepreneurship in the growth model framework. Noticeably, we found scarce evidence in this literature analysing the importance of SMEs in developing countries.

In the second part of this chapter, we review the limited literature on the importance of SMEs for business cycle fluctuations. This literature provides conflicting results. The well-know paper by Gertler and Gilchrist (1994) argues that small firms are more sensitive to cycles mainly due to financial constraints. On the other hand, new evidence provided by Moscarini and Postel-Vinay (2008, 2009, 2010) and Pinto (2011) argue otherwise. Again, there is a paucity of studies for developing countries and empirical evidence for these countries can help to provide a better understanding about the relationship between small firms and business cycles.

Chapter 3

Data Description

3.1 Introduction

The aim of this chapter is to provide a detailed description of RAIS (*Relação Anual de informações Sociais*) and CAGED (*Cadastro Geral de Empregados e Desempregados*) databases.[22] The former database is used in Chapters 4, 5 and 6, and the latter in Chapter 7. These two databases are particularly important because they provide the information about Small and Medium Enterprises (SMEs) necessary to perform this research. In addition, this chapter also provides more detail on the supplementary data used in this research, particularly, GDP per capita. It is important to note that a specific discussion about the data constructed and used in each chapter will be provided in the respective chapter.

To provide more details on the data used in this research, this chapter is set up as follows. Section 3.2 describes RAIS and CAGED in detail. Section 3.3 provides a discussion of complementary data. Finally, Section 3.4 draws a conclusion from the provided information.

3.2 What are RAIS and CAGED?

RAIS is the Brazilian annual census data of the Ministry of Labour and Employment that covers the formal sector of the economy. It is a comprehensive administrative annual record of formal establishments and workers across sectors and regions of Brazil. CAGED is a monthly database of the Ministry of Labour and Employment that records all monthly job flows in the formal sector. By law, they cover all

[22] RAIS stands for the Brazilian Annual Report of Social Information and CAGED for the General Registry of Employed and Unemployed. I take the opportunity to thank the Brazilian Ministry of Labour for providing access to both databases. Also, I would like to thank Rosangela Farias and Simone Taouk, both from the Ministry of Labour, that provided valuable assistance in the construction of the employment series based on CAGED used in Chapter 7.

formally employed workers, therefore, no information about the informal sector employment is provided. The implication for the lack of data for the informal sector is that the results of the thesis are only directly related to the formal sector of the Brazilian economy.

RAIS was established by the Law nº 76.900 of 23/12/1975 to provide labour market information for the government and research purposes. Originally, RAIS was designed to monitor the entry of foreign labour and to control the registry related to the *Fundo de Garantia por Tempo de Serviço* (FGTS), which is the government severance employment fund.[23] It was also used to provide information for the tax collection process and for the concession of benefits by the Ministry of Social Security. Similarly, CAGED was created in the form of the Law nº of 23/12/1965, when it was made compulsory for the formal sector to report information on job flows.

Currently, RAIS and CAGED are the only governmental instruments that regulate the concession of the *Abono Salarial*, the minimum-wage supplement program. If an establishment fails to report the information required by RAIS or CAGED, it faces automatic fines that are proportional to the workforce size and the length of the delay. However, because the payment of the annual wage supplement is exclusively based on them, employers and workers have strong incentives to fulfil RAIS and CAGED records. The Ministry of Labour and Employment estimates that, currently, around 97% of all formal workers in Brazil are covered by RAIS and CAGED.

These databases evolved slowly and RAIS became a well established set of data in 1985. Its comparable historical series is available from that year onwards. In the case of the complex task of collecting the monthly data of CAGED, comparable data are available only from 1996 onwards. In addition, if one wants to use data disaggregated by size class, the data starts only from 2000 onwards. During the nineties, they benefited from important advances regarding the data quality. A government campaign and improvements in the data collection process expanded

[23] FGTS can be translated into English as Guarantee Fund for Time of Service. Each month, 8 percent of the employee's salary must be deposited by employers in this fund.

their coverage and quality. Since 1997, for instance, the data has been collected via the internet, which makes the data collection quicker and more reliable.

An important difference between RAIS and CAGED is that the former considers all formal employment links ("celetistas" and "estatutários") and the latter considers only "celetistas". The workers that have their contracts regulated according to CLT (Consolidação das Leis do Trabalho) are called "celetistas".[24] Briefly, this formal employment link regulates the labour market in the private sector and comprises more than 80% of the formal employment links in Brazil. On the other hand, the workers that have their contracts regulated by the Federal Law 8112 are called "estatutários", or statutory servers. They are public servants of municipalities, states or federal government. In practice, that means that CAGED does not consider the job flows related to public servants.

The RAIS and CAGED systems are broadly used by Brazilian researchers but its use in the international literature in economics is limited. The data is available only for authorized researchers and the difficulties to obtain the data from abroad may partially explain its limited use. Examples of papers that used RAIS are: Muendler (2007), Menezes-Filho et al. (2008), Esteves (2007), and Aguayo-Tellez et al. (2010). The fact that RAIS and CAGED have just started to appear in international publications and because they are available only in Brazilian Portuguese justify our efforts to provide a more extensive description of this database.

3.2.1 Information about Establishments and Workers

RAIS provides information about the number of employees and establishments. It also provides a broad range of information about the characteristics of the linkage between employers and employees, personnel details and occupational classification of the workforce. The information available can be analysed according to many criteria from municipalities to the country level, while the sectoral distribution can be

[24] CLT can be translated as the Consolidation of Labor Laws, which are the laws that govern the labour market relations in Brazil.

observed using the IBGE (*Instituto Brasileiro de Geografia e Estatística*) and CNAE (*Classificação Nacional de Atividades Econômicas*) classifications of activities.[25] Importantly, all of the information contained in RAIS (and CAGED) can be retrieved according to the size of the establishments with size measured in number of workers. This is the most important characteristic of RAIS and CAGED that allows us to construct the data to analyse the importance of this sector in the Brazilian economy. For Chapters 4, 5 and 6 we constructed data for two dimensions of the SME sector. Firstly, we collected data to construct a measure of the SME sector size given by the share of the SME employment in total employment. Secondly, we collected data about the average years of schooling of workers in SMEs to construct a measure of the level of human capital in the SME sector.

Alternatively, the monthly CAGED database provides important information about job flows but is not as detailed in terms of demographic information about the firms and workers as RAIS. Importantly, CAGED can also be retrieved using the same sectoral and geographical classifications. The information of job flows retrieved from CAGED is used to construct the employment stocks for small and large businesses across Brazilian states and sectors to analyse the sensitivity of SMEs to business cycle fluctuations in Chapter 7. To construct these employment series, we use the two-step procedure of the Brazilian Ministry of Labour to build the monthly employment series from the information of job flows retrieved from CAGED.[26] Firstly, at the beginning of each year, the updated stock of workers across sectors and regions is provided by the Brazilian Ministry of Labour. The data of the stock of worker by firm size bins is not directly provided and a special request should be made to the Ministry of Labour. Subsequently, we use CAGED database to retrieve the monthly net employment variation according to the firm size bins to construct the employment series to analyse the relative performance of employment in SMEs during business cycles.

[25] IBGE is the acronym for the Brazilian Institute for Geography and Statistics, and CNAE can be translated as the Brazilian National Classification of Economic Activities.
[26] For instance, this procedure is used to produce the main monthly employment statistic about employment stock and fluctuation in Brazil.

To provide a better understanding of the production process of RAIS and CAGED, the collection process, the law requirements and the means of RAIS dissemination are explained in the following section.

3.2.2 RAIS/CAGED Data Collection Process and Dissemination

RAIS data collection is nowadays based on GD RAIS software and the yearly declaration must be submitted only by the internet. Each year, the establishments of the formal sector are obliged to submit the RAIS declaration to the Ministry of Labour and Employment. The data collection takes place until March of the subsequent year of observation and all formal enterprises must submit a RAIS declaration. The process of data collection for CAGED is similar. The formal enterprises that registered any job creation or job destruction in a given month must report this information to the Ministry of Labour using the CAGED WEB until the 7^{th} day of the subsequent month.

After the completion of the data collection, the data is processed and released in CD format accompanied by the software SGT micro under the *Programa de Disseminação de Estatísticas do Trabalho* (PDET), which means the dissemination program of labour statistics. To obtain the software SGT micro with CAGED and RAIS, a written formal request should be made to the Brazilian Ministry of Labour. This software allows the user to retrieve information cross combining the information available in the data by the establishment size bins.

3.2.3 The Sectoral Classifications of RAIS and CAGED

The sectoral classification of economic activities used in this thesis available from RAIS was elaborated by IBGE, the official national statistics office. It was used alone in Brazil until 1994, when, the *Comissão Nacional de Classificação* (Concla), which is a governmental body designed to improve the Brazilian classification system, proposed many changes in the classification of activities to make Brazilian data internationally comparable.

A new classification of activities called CNAE was created based on the United Nation's International Standard Industrial Classification of All Sectors (ISIC) Rev.3. The ISIC is a product of international agreements among national authorities represented at the United Nations (UN) statistical commission. Therefore, it is a reference for many other national classifications and for the General Industrial Classification of Economic activities within the European Communities (NACE). Hence, CNAE has a straightforward comparability with ISIC and NACE, for instance.

This first CNAE classification is known as CNAE/95. In 2002, CNAE/95 was updated and labelled CNAE 1.0. These classifications have the same structure and small differences only appear at the third and fourth digit level[27]. Later, in 2006, a new international classification structure, ISIC 4.0, was released. Again, to follow the new international standards, a new revision took place and CNAE 2.0 emerged. An extensive conversion table that links CNAE 2.0 to CNAE 1.0 can be found on Concla's web page[28]. We are not going to concentrate on the changes brought by this new classification because all CNAE versions are linked to the IBGE Classification through the CNAE95/1.0.

Importantly, the IBGE sectoral classification continues to be an important classification used to study aggregate sectoral issues and is widely used in Brazil. To preserve the time series information and at the same time allow for a direct comparison with international data when necessary, RAIS reports both IBGE and CNAE classifications.

Table 3.1 summarizes the information about the number of sectors that each classification covers and the availability of their time series length in each case. The detailed sectoral structure and translation of all sectors of IBGE and CNAE 1.0 classifications are provided throughout the Appendices 3.5.1, 3.5.2 and 3.5.3.

---

[27] These differences can be observed in the Appendix 3.5.3. The original table of conversion in Portuguese is available at the following web address:
http://www.ibge.gov.br/concla/cnae/correspondencias/CNAExCNAE1.0.pdf.

[28] http://www.ibge.gov.br/concla/pub/revisao2007/PropCNAE20/CNAE20_Correspondencia20x10.xls.

Table 3.1 – Sectoral Classifications and Number of Sectors in each Digit Level.

Digit Level	Number of Sectors		
	IBGE	CNAE-95 / CNAE 1.0	CNAE 2.0[a]
1	5	17	21
2	8	59	87
3	25	223	285
4		564/581	673
Time Length Available	1985-Present	1995-Present	2006-Present

Source: Brazilian Ministry of Labour and Employment.

For the IBGE sectoral classification used in this research, RAIS provides information that spans from 1985 to the present date. Using any of the CNAEs' classification, the length of the available time series would be substantially reduced.

3.2.4 Virtues and Problems of RAIS/CAGED

RAIS and CAGED are administrative datasets, and for that reason they are not so detailed in some aspects. For example, it does not provide information about Research and Development (R&D) expenditure, innovative activity, or details about the commercialisation of a new product or service. In spite of this, it constitutes a very rich dataset to investigate aspects related to the formal market in Brazil.

The strength of this data is its coverage; RAIS encompasses around 97% of the formal market in Brazil and comprises information of around 39 million workers in 2009. This information can be cross combined and retrieved by municipalities, sectors, occupation, size of the enterprises, level of schooling of workers and a considerable array of criteria. Another advantage is that the data is released very quickly compared to other macroeconomics series. In the case of CAGED, that does not include public servants and establishments as discussed earlier, it comprises information of 32.5 million workers in the beginning of 2009.

However, the Ministry of Labour and Employment recommends caution when using RAIS and CAGED information, bearing in mind the advantages and disadvantages of these databases. According to the Ministry of Labour, problems related to omission of declarations and misleading information are more frequent in small municipalities, agriculture, public administration and building related activities. In addition, RAIS and CAGED are filled in on a self-classification basis, which means that the establishment responsible for the information has to complete the RAIS and CAGED forms by itself according to the reference guidelines. As a result, establishments might classify themselves into a wrong sectoral group, for instance. However, we believe that this problem is offset by the census nature of RAIS and CAGED, in the sense that random misfiling would be compensated by the size of the database. As RAIS and CAGED are administrative records about the formal sector, they do not record any data regarding the informal sector.

Therefore, even though some problems can emerge from RAIS/CAGED, its coverage of the Brazilian formal sector and emerging use of this datasets in international journals encourage us to use this administrative registry in this thesis.

3.3 Supplementary Data

A range of supplementary data was used in this research. In Chapters 4, 5 and 6, a substantial effort was put in the construction of a regional panel data for the Brazilian micro-regions. Brazil has 558 micro-regions and this task was made easier because IBGE and IPEADATA (*Instituto de Pesquisas Economica Aplicada Database*) provide sophisticated data mining tools that allow us to retrieve the information for the Brazilian micro-regions in a straightforward fashion.[29] Particularly, IPEADATA is a data repository that compiles nearly all economic data produced in the main institutions of the country (e.g. IBGE, IPEA, Central Bank). Similarly, for Chapter 7,

[29] IPEADATA stands for Institute of Applied Economic Research Database.

the time series necessary to perform the analysis on the sensitivity of SMEs to cycles were collected mainly from IPEADATA.

Overall, the collection of supplementary data was a straightforward process but we always made an extra effort to understand in detail the data collected. In this process, the construction of the GDP per capita series deserves a more detailed discussion.

The GDP per capita series is a central piece of important information necessary to our analysis of the relationship between regional economic growth and SMEs performed in Chapters 4, 5 and 6. Importantly, the period of analysis is restricted by the availability of the GDP series.

During the process of compiling and checking the GDP per capita data we were able to reconstruct the GDP per capita series at factor cost and report a mistake in the regional aggregation of the data to IPEADATA. To construct the series of the real GDP at factor price (GDP) we collected the information available for Brazilian municipalities and aggregate it into micro-regions. The available data points in the range from 1980 to 1996 (1980, 1985 and 1996) were constructed by IPEA based on the IBGE measurement and were collected from IPEADATA. Alternatively, between 1999 and 2004, the IBGE provides annual information for the GDP of Brazilian municipalities that was aggregated into micro-regions.[30] Originally, IBGE reports the GDP at market prices but also provides results for value added and taxes, which allows us to recalculate the GDP at factor prices. We were able to check our micro-regional aggregation results with the results available on IPEADATA. We found different results from 2001 onwards and reported this fact to IPEA. They acknowledged that they have failed to consider the correct key of correspondence linking new municipalities created after 2001 to their respective micro-regions, confirming that our results are correct.

Despite the fact that the GDP at municipality level, was calculated from different institutes (IPEA and IBGE), they are based on similar methodologies and upon IBGE measurement. It was calculated allocating the GDP of each state across

[30] From 2004 onwards IBGE started to report this information based on a new methodology to calculate the GDP for the country and the data based on this methodology is available only from this period onwards. Therefore, our analysis does not include data after 2004 because the GDP data is calculated using a different methodology.

its municipalities according to the observable production factors in each municipality. The main difference is that they use different proxies to allocate the state's GDP across municipalities because the same information for some production factors is not available every year. Although the data processing described above may create some measurement errors, this is the best series that could be constructed to analyse growth at micro-regional level. Problems with regional data are not restricted to Brazil. For instance, many regional studies for the European Union are based on the compilation of information from Cambridge Econometrics that put together data comprised from different statistics offices that are not fully compatible, therefore, being subject to measurement errors (e.g. Fingleton 2000; Bosker 2007; Dallerba and LeGallo 2008).

3.4 Conclusion

This chapter has provided an overview about several aspects of the RAIS and CAGED databases. It has explained their nature and their contents, describes their data construction process and provides a thorough explanation of their sectoral classifications. Being administrative data, it is clear that they are not so detailed in some aspects but are very useful due to their coverage.

Besides, due to the fact that the databases allow us to retrieve information by the size of the establishment, we believe that they constitute a great source of information to study the SME sector, which is the aim of this research. Data regarding SMEs are difficult to be obtained and the use of RAIS and CAGED can be seen as an important source of information to study this type of establishment in Brazil. Furthermore, their use in academic international journals provides an additional incentive to use RAIS and CAGED. Therefore, although we can find shortcomings, we see them as a reliable and rich source of information to study the state of the SME sector in Brazil.

The supplementary data used in this research to support the analysis of the importance of the SME sector for economic growth in Brazil and the study about the behaviour of SMEs' employment during business cycles was collected directly from

IBGE, IPEADATA, and the Central Bank. In the process of the data checking, some problems in the geographic aggregation of GDP data were found. We identified the problem and reported it to the data provider who rectified the information available according to our calculation.

3.5 Appendix

3.5.1 IBGE Sectoral Classification

This appendix presents the structure of the IBGE classification available in RAIS and CAGED databases[31]. Apart from the residual sector called "others", we can observe that the more aggregated hierarchical level ("major sector") comprises 5 sectors while the most disaggregated hierarchical level ("sub sector") encompasses 25 sectors.

Table 3.2. Structure of the IBGE Classification by Major sector, Sector and Sub Sector

Major Sector	***Sector***	***Sub sector***
1. Industry	***1. Mining and Quarrying***	1 Mining and quarrying
	2. Manufacturing	2 Manufacture of non-metallic mineral products 3 Manufacture of metallic products 4 Manufacture of machinery, equipment and instruments 5 Manufacture of electrical and telecommunications equipment 6 Manufacture of transport equipment 7 Manufacture of wood products and furniture 8 Manufacture of paper and paperboard, and publishing 9 Manufacture of rubber, tobacco, leather, and products 10 Manufacture of chemical and pharmaceutical products 11 Manufacture of apparel and textiles 12 Manufacture of footwear 13 Manufacture of food, beverages, and ethyl alcohol
	3. Industrial Public Services - Electricity, gas and water supply	14 Electricity, gas and water supply (industrial Public Services)
2. Construction	***4. Construction***	15 Construction
3. Commerce	***5. Commerce***	16 Retail trade 17 Wholesale trade
4. Services (Great Sector)	***6. Services (Sector) Sub sectors:***	18 Financial intermediation and insurance 19 Real estate and business services 20 Transport, storage and telecommunications 21 Hotels and restaurants, repair and maintenance services 22 Medical, dental and veterinary services 23 Education
	7. Public Administration (Sector)	24 Public administration and social services
5. Primary (Great Sector)	***8. Primary (Sector)***	25 Agriculture, farming, hunting, forestry and fishing
6. Othesr (Great Sector)	***9. Others (Sector)***	26 Others.

Source: Ministry of Labour and Employment.

[31] The original classification in Portuguese can be found at the following web address: http://www.mte.gov.br/EstudiososPesquisadores/Pdet/tabelas/default.asp.

3.5.2 CNAE Sectoral Classification

The table below provides the structure of the 1st and 2nd digit level of CNAE-95/CNAE 1.0[32]. CNAE classifies the economical activities in Brazil until the fourth digit level.

Table 3.3 Structure of Section and Division of CNAE-95/CNAE 1.0

Section	Division	Description
A	01 .. 02	Agriculture, farming and forestry
B	05 .. 05	Fishing
C	10 .. 14	Mining and quarrying industry
D	15 .. 37	Manufacturing
E	40 .. 41	Electricity, gas and water supply
F	45 .. 45	Construction
G	50 .. 52	Trade; repair of motor vehicles and personal and household goods
H	55 .. 55	Hotels and restaurants
I	60 .. 64	Transport, storage and communications
J	65 .. 67	Financial intermediation, insurance, pension funding and related services
K	70 .. 74	Real estate, renting and business activities
L	75 .. 75	Public administration, defence and compulsory social security
M	80 .. 80	Education
N	85 .. 85	Health and social work
O	90 .. 93	Other community, social and personal service activities
P	95 .. 95	Private households services with employed persons
Q	99 .. 99	Extra-territorial organizations and bodies

Source: Ministry of Labour and Employment, Concla.

3.5.3 Correspondence between IBGE and CNAE

This appendix presents the correspondence between the IBGE classification and CNAE-95/CNAE 1.0, at the fourth digit level. The full list of CNAE classes presents 614 classes because this table encompasses both, CNAE-95 and CNAE 1.0. There are 564 codes for CNAE-95 and 581 classes for CNAE 1.0. The latter (CNAE 1.0) is an update of the former (CNAE-95) in which 50 new classes were created and 33 classes dropped. The classes that were dropped out or created are indicated in the table and the codes that are not marked remain the same under both classifications.

[32] These two classifications are presented together because the main structure did not change. Differences can be found only at the third and fourth digit levels. In Appendix 3.5.3 we can observe the differences between CNAE-95 and CNAE 1.0.

Table 3.4 – Correspondence between CNAE and IBGE

Class	CNAE-95/CNAE1.0	Sub sector (IBGE)	Sector (IBGE)	Major sector (IBGE)
10006	Mining of coal	1- Mining and quarrying	1- Mining and quarrying	1- Industry
11100	Extraction of crude petroleum and natural gas			
11207	Service activities incidental to oil and gas extraction excluding surveying			
13102	Mining of iron ores			
13218	Mining of non-ferrous metal ores (aluminium)			
13226	Mining of non-ferrous metal ores (tin)			
13234	Mining of non-ferrous metal ores (manganese)			
13242	Mining of non-ferrous metal ores (precious metal)			
13250	Mining of radioactive ores			
13293	Mining of others non-ferrous metal ores			
14109	Quarrying of stone, sand and clay			
14214	Mining of chemical and fertilizer minerals			
14222	Extraction of salt			
14290	Other mining and quarrying of non-metallic ores			
23302	Processing of nuclear fuel	2- Manufacture of non-metallic mineral products	2- Manufacture	
26115	Manufacture of plain glass and glass for security reasons (sheets, plates, laminated).			
26123	Manufacture of glass containers (lids, bottles, etc).			
26190	Manufacture of glass products			
26204	Manufacture of cement			
26301	Manufacture of articles of concrete, fibre-cement, asbestos, cement and plaster.			
26417	Manufacture of structural non-refractory clay and ceramic products			
26425	Manufacture of refractory ceramic products			
26492	Manufacture of non-structural non-refractory ceramic ware			
26913	Cutting, shaping and finishing of stone (non-quarrying activities).			
26921	Manufacture of lime and plaster			
26999	Manufacture of other non-metallic mineral products.			
27111	Manufacture of laminated steel (DROPPED OUT)	3- Manufacture of metallic products		
27120	Manufacture of non-laminated steel (DROPPED OUT)			
27138	Manufacture of pig iron (CREATED)			
27146	Manufacture of Ferro-alloys (CREATED)			
27219	Manufacture of pig iron (DROPPED OUT)			
27227	Manufacture of basic iron, steel, and Ferro-alloys (DROPPED OUT)			
27235	Production of semi-finished products of iron (CREATED)			
27243	Manufacture of drawn steel (CREATED)			
27251	Manufacture of long drawn (CREATED)			
27260	Manufacture of drawn wire steel products (CREATED)			
27294	Manufacture of drawn wire steel products (DROPPED OUT)			
27316	Manufacture of steel tubes with open seam			
27391	Manufacture of other tubes of iron and steel			
27413	Metallurgical Production of alloys with aluminium base			
27421	Metallurgical Production of the precious metal			
27499	Metallurgical Production of alloys with other non-ferrous metal base			
27510	Casting of iron and steel			
27529	Casting of non-ferrous metals and their alloys.			
28118	manufacture of metal structures for bridges, buildings, towers, masts, columns, etc.			
28126	Manufacture of metal doors, windows and their frames			
28134	Manufacture of parts of heavy steam generators			
28215	Manufacture of tanks, reservoirs and containers of metal for central heating			
28223	Manufacture of steam generators, except for central heating hot water boilers and for vehicles			
28312	Manufacture of cast wrought iron			
28320	Manufacture of cast wrought non-ferrous metal and their alloys			
28339	Manufacture of stamped metal products			
28347	Production of metal objects directly from metal powders			
28398	Treatment and coating of metals, welding and general mechanical engineering			
28410	Manufacture of cutlery			
28428	Manufacture of products for saw – except metal frames			
28436	Manufacture of hand tools			
28819	Manufacture of tanks, reservoirs and containers of metal (CREATED)			
28827	Manufacture and repair of steam generators, except central heating hot water boilers (CREATED)			
28916	Manufacture of containers of metal			
28924	Manufacture of wired artefacts			
28932	Manufacture of funnel and artefacts of metal for domestic use			
28991	Manufacture of other products of metal			
37109	Recycling of metal waste and scrap			
29114	Manufacture of engines and turbines of internal combustion, except aircraft, vehicle and cycle engines	4 Manufacture of machinery, equipment and instruments		
29122	Manufacture of hydraulic power engines and motors which consist of powerful pumps.			
29130	Manufacture of taps, cocks, valves and similar appliances.			
29149	Manufacture of compressors			
29157	Manufacture of bearings, gears, gearing and driving elements for industry purpose.			
29211	Manufacture of non-electric ovens, furnaces, furnace burners and equipments			
29220	Manufacture of electric furnaces or ovens			
29238	Manufacture of lifting and handling equipment			
29246	Manufacture of refrigerating or freezing equipment for industrial purposes			
29254	Manufacture of unit air-conditioners			
29297	Manufacture of general purpose machinery			
29319	Manufacture of agricultural and forestry machinery			
29327	Manufacture of tractors used in agriculture			

Table 3.4 – Correspondence between CNAE and IBGE - *continued*

29408	Manufacture of machine-tools			
29513	Manufacture of machinery for extraction and prospecting of crude petroleum			
29521	Manufacture of machinery for mining, quarrying and construction			
29530	Manufacture of track-laying tractors and tractors used in construction or mining.			
29548	Manufacture of machinery for surfacing (graders, levellers, scrapers, excavators and road rollers).			
29610	Manufacture of machinery for metallurgy			
29629	Manufacture of machinery for food, beverage and tobacco processing			
29637	Manufacture of machinery for textile industry			
29645	Manufacture of machinery for apparel, footwear and leather production			
29653	Manufacture of machinery for the pulp, paper and paperboard industries			
29696	Manufacture of other specific purpose machinery			
29718	Manufacture of fire guns and ammunition			
29726	Manufacture of heavy weapons			
29815	Manufacture of stoves refrigerators and freezers, dishwashers, laundry equipment			
29890	Manufacture of other domestic appliances			
29912	Maintenance and repair of engines, pumps, compressors transmission equipments (CREATED)			
29920	Maintenance and repair of machinery and equipments for general purposes (CREATED)			
29939	Maintenance and repair of machinery and equipments for agriculture purposes (CREATED)			
29947	Maintenance and repair of machine-tools (CREATED)			
29955	Maintenance and repair of machinery for mining, quarrying and construction (CREATED)			
29963	Maintenance and repair of other specific purpose machinery (CREATED)			
30112	Manufacture of non-electronic office machinery			
30120	Manufacture of electronic office, accounting and computing machinery			
30210	Manufacture of computers			
30228	Manufacture of peripheral equipment for electronic machinery			
33502	Manufacture of watches and clocks			
31119	Manufacture of generators	5 Manufacture of electrical and telecommunications equipment		
31127	Manufacture of transformers and similar equipments			
31135	Manufacture of electric motors			
31216	Manufacture of electricity distribution and control apparatus			
31224	Manufacture of electric material for installations			
31305	Manufacture of insulated wire and cable			
31410	Manufacture of accumulators, primary cells and primary batteries (except for vehicles)			
31429	Manufacture of accumulators, primary cells and primary batteries for vehicles			
31518	Manufacture of electric lamps			
31526	Manufacture of lighting equipment (except for vehicles)			
31607	Manufacture of electrical equipment for vehicles			
31810	Maintenance and repair of electric motors, generators and transformers (CREATED)			
31828	Maintenance and repair of accumulators, primary cells and primary batteries (CREATED)			
31917	Manufacture of electrodes and similar products of coal and graphite for electric and magnet purposes			
31925	Manufacture of equipments for signalling and alarm			
31992	Manufacture of other electrical equipment			
32107	Manufacture of basic electronic components			
32212	Manufacture of television and radio transmitters and apparatus for line telephony and line telegraphy			
32220	Manufacture of telephones and similar goods			
32301	Manufacture of television and radio receivers, sound or video recording or reproducing apparatus, and associated goods			
32905	Maintenance and repair of telephone communication system equipment and related equipments (CREATED)			
34100	Manufacture of motor vehicles (cars, pickups, vans)	6 Manufacture of transport equipment		
34207	Manufacture of trucks and buses			
34312	Manufacture of bodies (coachwork), trailers and semi-trailers for trucks			
34320	Manufacture of bus bodies			
34398	Manufacture of bodies (coachwork), trailers and semi-trailers for other vehicles			
34410	Manufacture of parts and accessories for the motor system			
34428	Manufacture of parts and accessories for transmission and gears			
34436	Manufacture of parts and accessories for brakes			
34444	Manufacture of parts and accessories for steering system and suspension			
34495	Manufacture of other metal accessories for motor vehicles not classified in other class			
34509	Recondition and repair of motors (car, vans, buses, etc)			
35114	Building and repairing of ships and boats			
35122	Building and repairing of leisure ships and boats			
35211	Manufacture of railway and tramway locomotives and rolling stock			
35220	Manufacture of parts and accessories of railway and tramway locomotives and rolling stock			
35238	Maintenance and repair of railway vehicles			
35319	Manufacture of aircraft			
35327	Maintenance of aircraft			
35912	Manufacture of motorcycles			
35920	Manufacture of bicycles and non-motorised vehicles			
35998	Manufacture of other transport equipment			
20109	Sawmilling of wood	7 Manufacture of wood products and furniture		
20214	Manufacture of veneer sheets, plywood and laminboard			
20222	Manufacture of wooden frames and wooden house pieces			

Table 3.4 – Correspondence between CNAE and IBGE - *continued*

20230	Manufacture of wooden containers			
20290	Manufacture of other products of wood; manufacture of articles of cork, straw and plaiting materials			
36110	Manufacture of furniture (predominant made by wood)			
36129	Manufacture of furniture (predominant made by metal)			
36145	Manufacture of mattress			
36978	Manufacture of brushes, brooms and paintbrushes			
21105	Manufacture of pulp for paper and paperboard	8 Manufacture of paper and paperboard, and publishing		
21210	Manufacture of paper			
21229	Manufacture of paper and paperboard			
21318	Manufacture of paper containers			
21326	Manufacture of corrugated paperboard and of containers of paperboard			
21415	Manufacture of other articles of paper and paperboard for office use			
21423	Manufacture of tapes and forms			
21490	Manufacture of other items of pulp, paper or paperboard			
22110	Publishing of newspapers (DROPPED OUT)			
22128	Publishing of magazines(DROPPED OUT)			
22136	Publishing of books (DROPPED OUT)			
22144	Publishing of recorded media (cassettes, discs and other recorded materials)			
22152	Publishing of books, magazines, newspapers and other publications (CREATED)			
22160	Publishing of books (CREATED)			
22179	Publishing of newspaper (CREATED)			
22187	Publishing of magazines (CREATED)			
22195	Other publishing			
22217	Printing of newspapers, magazines and books			
22225	Printing of educational contents and material for industrial and commercial uses			
22292	Other printings			
16004	Manufacture of tobacco products	9 Manufacture of rubber, tobacco, leather, and products		
19100	Tanning and dressing of leather			
19216	Manufacture of luggage, handbags and other similar items			
19291	Manufacture of other leather items			
22314	Reproduction of recorded media (cassettes and discs)			
22322	Reproduction of recorded media (VHS)			
22330	Reproduction of recorded movies (DROPPED OUT)			
22349	Reproduction of software in floppy discs and tapes			
25119	Manufacture of rubber tyres and tubes			
25127	Rethreading and rebuilding of rubber tyres			
25194	Manufacture of other rubber products			
33103	Manufacture of medical and surgical equipment and orthopaedic appliances			
33200	Manufacture of instruments and appliances for measuring, checking, testing, navigating and other purposes, except industrial process control equipment			
33308	Manufacture of industrial process control equipment			
33405	Manufacture of optical instruments and photographic equipment			
33502	Manufacture of watches and clocks			
33910	Maintenance of medical and surgical equipment and orthopaedic appliances (CREATED)			
33928	Maintenance of medical instruments and appliances for measuring, checking, testing, navigating and other purposes, except industrial process control equipment (CREATED)			
33936	Maintenance of industrial process control equipment (CREATED)			
33944	Maintenance of optical instruments and photographic equipment (CREATED)			
36919	Manufacture of jewellery and related articles			
36927	Manufacture of musical instruments			
36935	Manufacture of sports goods (including sport fishing)			
36943	Manufacture of games and toys			
36951	Manufacture of pens, pencils and other office items			
37206	Recycling of non-metal waste and scrap			
92118	Motion picture and video production			
23108	Manufacture of coke oven products	10 Manufacture of chemical and pharmaceutical products		
23205	Manufacture of refined petroleum products (DROPPED OUT)			
23213	Manufacture of refined petroleum products (CREATED)			
23299	Manufacture of petroleum products using other methods (other than oil refining) (CREATED)			
24112	Manufacture of chlorine and alkalis			
24120	Manufacture of fertilizers inputs			
24139	Manufacture of fertilizers, nitrogen compounds and other similar products			
24147	Manufacture of industrial gases			
24198	Manufacture of other inorganic products			
24210	Manufacture of basic petrochemical products			
24228	Manufacture of inputs for resin and fibre			
24295	Manufacture of other organic products			
24317	Manufacture of thermoplastic resin			
24325	Manufacture of thermofixed resin			
24333	Manufacture of elastomer			
24414	Manufacture of artificial fibre, wire, cable and filaments			
24422	Manufacture of synthetic fibre, wire, cable and filaments			
24511	Manufacture of pharma-chemicals			
24520	Manufacture of medicines for human purposes			
24538	Manufacture of medicines for animal purposes			
24546	Manufacture of medicinal chemicals			
24619	Manufacture of fungicide			
24627	Manufacture of herbicide			
24635	Manufacture of insecticide			

Table 3.4 – Correspondence between CNAE and IBGE - *continued*

24694	Manufacture of other agro-chemical products			
24716	Manufacture of soap and detergents			
24724	Manufacture of cleaning and polishing preparations			
24732	Manufacture of perfumery and cosmetics			
24813	Manufacture of paints, varnishes and similar coatings			
24821	Manufacture of printing inks			
24830	Manufacture of waterproofing, solvents and similar products			
24910	Manufacture of sealants and adhesive			
24929	Manufacture of explosives			
24937	Manufacture of catalyst			
24945	Manufacture of additives for industrial purpose			
24953	Manufacture of chemical products for photographic purposes			
24961	Manufacture of discs and tapes			
24996	Manufacture of other chemical products			
25216	Manufacture of plastics products (laminated and tubular)			
25224	Manufacture of plastics products (container)			
25291	Manufacture of other plastics products			
36994	Other manufacturing			
17116	Preparation of cotton	11 Manufacture of apparel and textiles		
17191	Preparation of other natural fibres			
17213	Spinning of cotton			
17221	Spinning of other natural fibres			
17230	Spinning of artificial or synthetic fibres			
17248	Manufacture of line to weave and embroider			
17310	Weaving of cotton			
17329	Weaving of artificial and synthetic textiles fibres			
17337	Weaving of other natural textiles fibres			
17418	Manufacture of textile articles for domestic use			
17493	Manufacture of other textile articles			
17507	Finishing of textiles			
17612	Manufacture of made-up textile articles, except apparel			
17620	Manufacture of carpets and rugs			
17639	Manufacture of cordage, rope, twine and netting			
17647	Manufacture of special textile articles			
17698	Manufacture of other textiles			
17710	Manufacture of knitted articles			
17728	Manufacture of socks			
17795	Manufacture of other knitted and crocheted fabrics and articles			
18112	Manufacture of underwear			
18120	Manufacture of apparel articles			
18139	Manufacture of professional apparel			
18210	Manufacture of clothing accessories			
18228	Manufacture of clothing accessories for industrial and personnel safety			
36960	Manufacture of other articles for sewing			
19313	Manufacture of leather footwear	12 Manufacture of footwear		
19321	Manufacture of shoes (any material)			
19330	Manufacture of plastic footwear			
19399	Manufacture of footwear made by other materials			
15113	Slaughter of cattle, production and processing and of meat and meat products	13 Manufacture of food, beverages, and ethyl alcohol		
15121	Slaughter of birds and other small animals, production and processing and of meat and meat products			
15130	Processing of meat, lard and other products			
15148	Processing and preserving of fish and fish products			
15210	Processing and preserving of fruit			
15229	Processing and preserving of vegetables			
15237	Processing of fruit and vegetables juices			
15318	Manufacture of vegetable oils			
15326	Manufacture of refined vegetable oils			
15334	Manufacture of margarine and other animal oils			
15415	Manufacture of milk			
15423	Manufacture of dairy products			
15431	Manufacture of ice cream			
15512	Manufacture of rice and related products			
15520	Manufacture of wheat and related products			
15539	Manufacture of manioc flour and related products			
15547	Manufacture of corn flour and related products			
15555	Manufacture of starches and starch products and manufacture of corn oil			
15563	Manufacture of prepared animal feeds			
15598	Manufacture of other grain mill products			
15610	Manufacture of sugar			
15628	Refine and milling of sugar			
15717	Milling of coffee bean			
15725	Manufacture of soluble coffee			
15814	Manufacture of bakery products			
15822	Manufacture of bakery products (biscuits and wafers)			
15830	Manufacture of cocoa, chocolate and sugar confectionery			
15849	Manufacture of macaroni, noodles, couscous and similar farinaceous products			
15857	Preparation of herbs and spices, curries and seasonings			
15865	Preparation of diet products, food for children and other products			
15890	Manufacture of other food products			
15911	Distilling, rectifying and blending of spirits			
15920	Manufacture of wines			
15938	Manufacture of malt liquors and malt			
15946	Production of mineral waters			
15954	Manufacture of soft drinks			
23400	Ethyl alcohol production			
55247	Ready meal providers			
40100	Production and distribution of electricity (DROPPED OUT)	14 Electricity, gas and water supply (industrial Public Services)	3. Industrial Public Services - Electricity, gas and water supply	
40118	Production of electricity (CREATED)			
40126	Transmission of electricity (CREATED)			

Table 3.4 – Correspondence between CNAE and IBGE - *continued*

40134	Wholesale of electricity (CREATED)			
40142	Distribution of electricity (CREATED)			
40207	Manufacture of gas and distribution of gaseous fuels through mains			
40304	Steam and hot water supply			
41009	Collection, purification and distribution of water			
90000	Sewage and refuse disposal, sanitation and similar activities			
45110	Demolition and site preparation	15 Construction	4. Construction	2- Construction
45128	Drilling and building foundations for construction			
45136	Site preparation			
45217	Building (residences, industry, and for commercial purposes)			
45225	Road construction			
45233	Construction of great structures			
45241	Urban development and landscaping (DROPPED OUT)			
45250	Building assembly			
45292	Other types of construction			
45314	Construction for electricity generation (ex: dam)			
45322	Construction for electricity distribution (DROPPED OUT)			
45330	Construction for telephone and communication systems			
45349	Construction for environment protection (DROPPED OUT)			
45411	Building installation of electricity			
45420	Building installation of air conditioning and similar activities			
45438	Building installation of hydraulic and drain systems and similar activities			
45497	Other types of building installation			
45500	Building completion (CREATED)			
45519	Masonry and plastering (DROPPED OUT)			
45527	Sealing and general painting services (DROPPED OUT)			
45594	Other construction related services (DROPPED OUT)			
45608	Renting of construction or demolition equipment with operator			
50105	Wholesale and retail trade of motor vehicles	16 Retail trade	5. Commerce	3- Commerce
50202	Maintenance and repair of motor vehicles			
50300	Wholesale and retail trade of motor vehicle parts and accessories			
50415	Wholesale and retail trade of motorcycles and related parts and accessories			
50423	Maintenance and repair of motorcycles			
50504	Wholesale and retail trade of automotive fuel			
52116	Retail sale in non-specialized hypermarket (more than 5000 squared metres) with food, beverages or tobacco predominating.			
52124	Retail sale in non-specialized stores (between 300 and 5000 squared metres) with food, beverages or tobacco predominating.			
52132	Retail sale in non-specialized stores (less than 300 squared metres) with food, beverages or tobacco predominating.			
52140	Retail sale in non-specialized stores with food, beverages or tobacco predominating – convenience stores/corner stores			
52159	Other retail sale in non-specialized stores			
52213	Retail sale of foodstuffs in bakery			
52221	Retail sale of candy, chocolate and related goods in specialized stores			
52230	Retail sale of meat – butcher shop			
52248	Retail sale of beverages			
52299	Retail trade of other foodstuffs in specialized stores			
52310	Retail sale of textiles in general			
52329	Retail sale of clothing, footwear and complements			
52337	Retail sale of leather goods, luggage and similar goods.			
52418	Retail sale of pharmaceutical and medical goods, cosmetic and toilet articles			
52426	Retail sale of household appliances (including musical instruments)			
52434	Retail sale of household furniture and similar and lightening equipment for residence			
52442	Retail sale of hardware, paint and glass			
52450	Retail sale of hardware and equipment for offices			
52469	Retail sale of books, newspapers, magazines and related activities			
52477	Retail sale of liquefied gas from oil.			
52493	Other retail sale non-specified before			
52507	Retail sale of second-hand goods in stores			
52612	Retail sale via mail order houses (DROPPED OUT)			
52620	Retail sale via mobile stalls in public streets. (CREATED)			
52698	Other non-store retail sale			
52710	Repair of personal and household goods (household appliances)			
52728	Repair of footwear			
52795	Repair of other personal and household goods			
51110	Wholesale on a fee or contract basis (sales representative) of raw materials, live animals and semi-manufactured products.	17 Wholesale trade		
51128	Wholesale on a fee or contract basis (sales representative) of fuels, metal and chemical products.			
51136	Wholesale on a fee or contract basis (sales representative) of wood, hardware and similar.			
51144	Wholesale on a fee or contract basis (sales representative) of industrial equipments, ships and aircrafts.			
51152	Wholesale on a fee or contract basis (sales representative) of furniture and similar for domestic use			
51160	Wholesale on a fee or contract basis (sales representative) of textiles, clothing and footwear			
51179	Wholesale on a fee or contract basis (sales representative) of food, beverages and tobacco			
51187	Wholesale on a fee or contract basis (sales representative) of other products non-specified before.			
51195	Wholesale on a fee or contract basis (sales representative) of other products in general (non-specialized)			
51217	Wholesale of agricultural raw materials			
51225	Wholesale of live animals			
51314	Wholesale of milk and derivatives			
51322	Wholesale of cereal, flour and starch			
51330	Wholesale of vegetables and fruits			

Table 3.4 – Correspondence between CNAE and IBGE - *continued*				
51349	Wholesale of meat and meat products			
51357	Wholesale of fishing			
51365	Wholesale of beverages			
51373	Wholesale of tobacco			
51390	Wholesale of other meals/food non-specified before			
51411	Wholesale of textiles in general			
51420	Wholesale of clothing, footwear and complements			
51438	Wholesale of footwear			
51446	Wholesale of household appliances			
51454	Wholesale of pharmaceutical and medical goods.			
51462	Wholesale of cosmetic and perfumery			
51470	Wholesale of office products, stationery books, newspapers, magazines and other publications.			
51497	Wholesale of other household goods			
51519	Wholesale of fuels and related products			
51527	Wholesale of products of mineral origin			
51535	Wholesale of construction materials, wood and hardware			
51543	Wholesale of chemical products			
51551	Wholesale of waste and scrap			
51594	Wholesale of other intermediate products (non-agricultural)			
51616	Wholesale of machinery, equipment and supplies for agricultural purpose			
51624	Wholesale of machinery, equipment and supplies for commercial purpose (DROPPED OUT)			
51632	Wholesale of machinery, equipment and supplies for office use (DROPPED OUT)			
51640	Wholesale of machinery and equipment for commercial purpose (CREATED)			
51659	Wholesale of machinery and equipment for office purpose (CREATED)			
51691	Wholesale of machinery, equipment and supplies for professional purpose non-specified before			
51918	Wholesale of general goods (non-specialized)			
51926	Wholesale specialized in other goods (not described before)			
65102	Central Bank	18 Financial intermediation and insurance	6. Services	4. Services
65218	Commercial banks			
65226	Universal banks (with commercial characteristics)			
65234	Special saving banks, general funds ("Caixas economicas")			
65242	Credit unions			
65315	Universal banks (without commercial characteristics)			
65323	Investment banks			
65331	Development banks			
65340	Mortgage banks			
65358	Savings and loan associations			
65404	Financial leasing			
65510	Development agencies			
65595	Other financial lender institutions			
65919	Investment funds			
65927	Societies			
65935	Management of non-financial assets (CREATED)			
65994	Other financial intermediation			
66117	Life insurance			
66125	Non-life insurance			
66133	Reinsurance			
66214	Pension funding (closed)			
66222	Pension funding (open)			
66303	Health insurance plan			
67113	Administration of financial markets (secondary market)			
67121	Activities of financial intermediation in transactions of securities			
67199	Activities auxiliary to financial intermediation			
67202	Activities auxiliary to insurance and pension funding			
2135	Forestry, logging and related service activities	19 Real estate and business services		
22314	Reproduction of recorded media (discs and cassettes)			
22322	Reproduction of recorded media (VHS)			
22330	Reproduction of movies (DROPPED OUT)			
22349	Reproduction of software in floppy discs and tapes			
51110	Wholesale on a fee or contract basis (sales representative) of raw materials, live animals and semi-manufactured products.			
51128	Wholesale on a fee or contract basis (sales representative) of fuels, metal and chemical products.			
51136	Wholesale on a fee or contract basis (sales representative) of wood, hardware and similar.			
51144	Wholesale on a fee or contract basis (sales representative) of industrial equipments, ships and aircrafts.			
51152	Wholesale on a fee or contract basis (sales representative) of furniture and similar for domestic use			
51160	Wholesale on a fee or contract basis (sales representative) of textiles, clothing and footwear			
51179	Wholesale on a fee or contract basis (sales representative) of food, beverages and tobacco			
51187	Wholesale on a fee or contract basis (sales representative) of other products non-specified before.			
51195	Wholesale on a fee or contract basis (sales representative) of other products in general (non-specialized)			
63118	Cargo handling			
63126	Storage and warehousing			
63215	Other supporting transport activities (terrestrial transport)			
63231	Other supporting transport activities (air transport)			
63304	Activities of travel agencies and tour operators			
63401	Activities of cargo transport			
67113	Administration of financial markets (secondary market)			
67121	Activities of financial intermediation in transactions of securities			
67199	Activities auxiliary to financial intermediation			
67202	Activities auxiliary to insurance and pension funding			

Table 3.4 – Correspondence between CNAE and IBGE - *continued*				
70106	Real estate activities (including buy and sale activities)			
70203	Renting of properties			
70319	Real estate activities on a fee or contract basis (including property evaluation)			
70327	Other arrangement for properties administration			
70408	Condominium of building			
71102	Renting of automobiles			
71218	Renting of other land transport equipment			
71226	Renting of water transport equipment			
71234	Renting of air transport equipment			
71315	Renting of agricultural machinery and equipment			
71323	Renting of construction and civil engineering machinery and equipment			
71331	Renting of office machinery and equipment			
71390	Renting of other machinery and equipment			
71404	Renting of personal and household goods			
72109	Hardware consultancy			
72206	Software development (DROPPED OUT)			
72214	Development and edition of software ready for use (CREATED)			
72290	Software development and consultancy (by order) (CREATED)			
72303	Data processing			
72400	Data base activities and on-line distribution of electronic information			
72508	Maintenance and repair of office, accounting and computing machinery			
72907	Other computer related activities			
73105	Research and development on natural sciences and engineering			
73202	Research and development on social sciences and humanities			
74110	Legal activities			
74128	Accounting, book-keeping and auditing activities			
74136	Market research and public opinion polling			
74144	Management of holdings			
74152	Firm headquarters and administrative units			
74160	Business and management consultancy activities			
74209	Architectural and engineering activities and related technical consultancy			
74306	Technical testing and quality analysis			
74403	Advertising			
74500	Labour recruitment and provision of personnel			
74608	Investigation and security activities			
74705	Building-cleaning activities			
74918	Photographic activities			
74926	Packaging activities			
74993	Other business activities (service sector)			
92401	News agency activities			
99007	Extra-territorial organizations and bodies			
60100	Inter cities transport via railways	20 Transport, storage and telecommunications		
60216	Urban transport via railways			
60224	Urban transport via railways (metro)			
60232	Regular urban land transport of passenger (Bus and similar)			
60240	Regular inter cities land transport of passenger (Bus and similar)			
60259	Non-regular land transport of passenger (by contract/order)			
60267	Freight transport by road			
60275	Freight transport by road (dangerous products)			
60283	Freight transport by road (by removal van and similar)			
60291	Other regular means of transport used for touristy purposes (tram, cable car, etc).			
60305	Transport via pipelines			
61115	Water transport (cabotage)			
61123	Water transport (long-distance)			
61212	Passengers inland water transport			
61220	Cargo inland water transport			
61239	Urban water transport			
62103	Scheduled air transport			
62200	Non-scheduled air transport			
62308	Space transport			
63118	Cargo handling			
63126	Storage and warehousing			
63215	Other supporting transport activities (terrestrial transport)			
63223	Other supporting transport activities (water transport)			
63231	Other supporting transport activities (air transport)			
63304	Activities of travel agencies and tour operators			
63401	Activities of cargo transport			
64114	National post activities			
64122	Courier activities other than national post activities			
64203	Telecommunications			
31895	Manufacture of other electrical equipment (CREATED)	21 Hotels and restaurants, repair and maintenance services		
36137	Manufacture of furniture of other materials (other than wood and metal)			
50202	Maintenance and repair of motor vehicles			
50423	Maintenance and repair of motorcycles			
52710	Repair of personal and household goods (household appliances)			
52728	Repair of footwear			
52795	Repair of other personal and household goods			
55115	Hotels with restaurants (DROPPED OUT)			
55123	Hotels without restaurants (DROPPED OUT)			
55131	Hotels (CREATED)			
55190	Other provision of short-stay accommodation			
55212	Restaurants			
55220	Snack bars and similar			
55239	Canteens (private)			
55247	Ready meal providers			
55298	Other food activities			
66214	Pension funding (closed)			
66222	Pension funding (open)			

Table 3.4 – Correspondence between CNAE and IBGE - *continued*				
70408	Condominium of building			
71404	Renting of personal and household goods			
74705	Building-cleaning activities			
74918	Photographic activities			
75302	Social security			
85316	Social work with accommodation			
85324	Social work without accommodation			
90000	Sewage and refuse disposal, sanitation and similar activities			
91111	Activities of business and employers organizations			
91120	Activities of professional organizations			
91200	Activities of trade unions			
91910	Activities of religious organizations			
91928	Activities of political organizations			
91995	Activities of other membership organizations			
92118	Motion picture and video production			
92126	Motion picture and video distribution			
92134	Motion picture projection			
92215	Radio activities			
92223	Television activities			
92312	Dramatic arts, music and other arts activities			
92320	Management of theatres and similar			
92398	Other entertainment activities			
92401	News agency activities			
92517	Library and archives activities			
92525	Museums activities and preservation of historical sites and buildings			
92533	Botanical and zoological gardens and nature reserves activities			
92614	Sporting activities			
92622	Other recreational activities			
93017	Washing, and (dry-) cleaning of textile and dyeing			
93025	Hairdressing and other beauty treatment			
93033	Funeral and related activities			
93041	Other activities of body maintenance			
93092	Other personal service activities			
95001	Private households with employed persons			
85111	Hospital activities	22 Medical, dental and veterinary services		
85120	Hospital activities (in emergencies)			
85138	Medical and dental practice activities			
85146	Complementary activities of diagnostic and treatment			
85154	Activities of other health professionals (physiotherapist, physic education professionals and similar)			
85162	Other human health activities			
85200	Veterinary activities			
80110	Pre-school education (DROPPED OUT)	23 Education		
80128	Primary education (DROPPED OUT)			
80136	Pre-school education - nursery education (CREATED)			
80144	Pre-school education (CREATED)			
80152	Primary education (CREATED)			
80209	Secondary education (CREATED)			
80217	General secondary education (DROPPED OUT)			
80225	Technical and vocational secondary education (DROPPED OUT)			
80306	Higher education (DROPPED OUT)			
80314	Higher education – undergraduates (CREATED)			
80322	Higher education - undergraduates and postgraduates (CREATED)			
80330	Higher education - postgraduates and diplomas (CREATED)			
80918	Adult and other education - driving school and flying school (DROPPED OUT)			
80926	High school senior education (DROPPED OUT)			
80934	Continuous professional learning (DROPPED OUT)			
80942	Distance learning (DROPPED OUT)			
80950	Special education (DROPPED OUT)			
80969	Professional education of technical level (CREATED)			
80977	Professional education of technological level (CREATED)			
80993	Other education (CREATED)			
75116	General public service activities	24 Public administration and social services	7. Public Administration	
75124	Regulation of the activities of agencies that provide health care, education, cultural services and other social services			
75132	Regulation of business activities			
75140	Ancillary service activities for the government as a whole			
75213	Foreign affairs			
75221	Defense activities			
75230	Justice			
75248	Public order and safety activities			
75256	Civil defense (disasters assistance)			
75302	Social security			
1112	Growing of cereals	25 Agriculture, farming, hunting, forestry and fishing	8. Primary	5. Primary
1120	Growing of cotton			
1139	Growing of cane sugar			
1147	Growing of tobacco			
1155	Growing of soy			
1198	Growing of other temporary (seasonal) crops			
1210	Growing of vegetables, horticultural specialties			
1228	Growing of fruit, ornamental plants and nursery products			
1317	Growing of citrus fruit			
1325	Growing of coffee bean			
1333	Growing of cacao			
1341	Growing of grapes			
1392	Growing of other permanent crops			
1414	Farming of cattle			
1422	Farming of other big animals			
1430	Farming of sheep			
1449	Farming of pig			

Table 3.4 – Correspondence between CNAE and IBGE - *continued*				
1457	Farming of bird			
1465	Farming of other animals			
1503	Growing of crops combined with farming of animals (mixed farming)			
1619	Agricultural husbandry service activities			
1627	Animal husbandry service activities, except veterinary activities			
1708	Hunting and game propagation and related service activities			
2119	Forestry (silviculture)			
2127	Forestry			
5118	Fishing and related activities			
5126	Aquaculture activities and related activities			
	IGNORED	26 Others.	9. Others	6. Others

Source: Ministry of Labour and Employment

3.5.4 List of Micro-regions

Table 3.5 provides the list of Brazilian micro-regions with the region they belong to. In addition, Figure 3.1 provides a map of Brazil with the indication of the regions.

Table 3.5 – List of Brazilian Micro-regions

	Micro-region	Region		Micro-region	Region		Micro-region	Region
1	Porto Velho	North	187	Brejo Pernambucano	North-East	373	Araraquara	South-East
2	Guajará-Mirim	North	188	Mata Setentrional Pernambucana	North-East	374	São Carlos	South-East
3	Ariquemes	North	189	Vitória de Santo Antão	North-East	375	Rio Claro	South-East
4	Ji-Paraná	North	190	Mata Meridional Pernambucana	North-East	376	Limeira	South-East
5	Alvorada D'Oeste	North	191	Itamaracá	North-East	377	Piracicaba	South-East
6	Cacoal	North	192	Recife	North-East	378	Pirassununga	South-East
7	Vilhena	North	193	Suape	North-East	379	São João da Boa Vista	South-East
8	Colorado do Oeste	North	194	Fernando de Noronha	North-East	380	Moji Mirim	South-East
9	Cruzeiro do Sul	North	195	Serrana do Sertão Alagoano	North-East	381	Campinas	South-East
10	Tarauacá	North	196	Alagoana do Sertão do São Francisco	North-East	382	Amparo	South-East
11	Sena Madureira	North	197	Santana do Ipanema	North-East	383	Dracena	South-East
12	Rio Branco	North	198	Batalha	North-East	384	Adamantina	South-East
13	Brasiléia	North	199	Palmeira dos Índios	North-East	385	Presidente Prudente	South-East
14	Rio Negro	North	200	Arapiraca	North-East	386	Tupã	South-East
15	Japurá	North	201	Traipu	North-East	387	Marília	South-East
16	Alto Solimões	North	202	Serrana dos Quilombos	North-East	388	Assis	South-East
17	Juruá	North	203	Mata Alagoana	North-East	389	Ourinhos	South-East
18	Tefé	North	204	Litoral Norte Alagoano	North-East	390	Itapeva	South-East
19	Coari	North	205	Maceió	North-East	391	Itapetininga	South-East
20	Manaus	North	206	São Miguel dos Campos	North-East	392	Tatuí	South-East
21	Rio Preto da Eva	North	207	Penedo	North-East	393	Capão Bonito	South-East
22	Itacoatiara	North	208	Sergipana do Sertão do São Francisco	North-East	394	Piedade	South-East
23	Parintins	North	209	Carira	North-East	395	Sorocaba	South-East
24	Boca do Acre	North	210	Nossa Senhora das Dores	North-East	396	Jundiaí	South-East
25	Purus	North	211	Agreste de Itabaiana	North-East	397	Bragança Paulista	South-East
26	Madeira	North	212	Tobias Barreto	North-East	398	Campos do Jordão	South-East
27	Boa Vista	North	213	Agreste de Lagarto	North-East	399	São José dos Campos	South-East
28	Nordeste de Roraima	North	214	Propriá	North-East	400	Guaratinguetá	South-East
29	Caracaraí	North	215	Cotinguiba	North-East	401	Bananal	South-East
30	Sudeste de Roraima	North	216	Japaratuba	North-East	402	Paraibuna/Paraitinga	South-East
31	Óbidos	North	217	Baixo Cotinguiba	North-East	403	Caraguatatuba	South-East
32	Santarém	North	218	Aracaju	North-East	404	Registro	South-East
33	Almeirim	North	219	Boquim	North-East	405	Itanhaém	South-East
34	Portel	North	220	Estância	North-East	406	Osasco	South-East
35	Furos de Breves	North	221	Barreiras	North-East	407	Franco da Rocha	South-East
36	Arari	North	222	Cotegipe	North-East	408	Guarulhos	South-East
37	Belém	North	223	Santa Maria da Vitória	North-East	409	Itapecerica da Serra	South-East
38	Castanhal	North	224	Juazeiro	North-East	410	São Paulo	South-East
39	Salgado	North	225	Paulo Afonso	North-East	411	Mogi das Cruzes	South-East
40	Bragantina	North	226	Barra	North-East	412	Santos	South-East
41	Cametá	North	227	Bom Jesus da Lapa	North-East	413	Paranavaí	South
42	Tomé-Açu	North	228	Senhor do Bonfim	North-East	414	Umuarama	South
43	Guamá	North	229	Irecê	North-East	415	Cianorte	South
44	Itaituba	North	230	Jacobina	North-East	416	Goioerê	South
45	Altamira	North	231	Itaberaba	North-East	417	Campo Mourão	South
46	Tucuruí	North	232	Feira de Santana	North-East	418	Astorga	South
47	Paragominas	North	233	Jeremoabo	North-East	419	Porecatu	South
48	São Félix do Xingu	North	234	Euclides da Cunha	North-East	420	Floraí	South
49	Parauapebas	North	235	Ribeira do Pombal	North-East	421	Maringá	South

Table 3.5 – List of Brazilian Micro-regions - *continued*

50	Marabá	North	236	Serrinha	North-East	422	Apucarana	South
51	Redenção	North	237	Alagoinhas	North-East	423	Londrina	South
52	Conceição do Araguaia	North	238	Entre Rios	North-East	424	Faxinal	South
53	Oiapoque	North	239	Catu	North-East	425	Ivaiporã	South
54	Amapá	North	240	Santo Antônio de Jesus	North-East	426	Assaí	South
55	Macapá	North	241	Salvador	North-East	427	Cornélio Procópio	South
56	Mazagão	North	242	Boquira	North-East	428	Jacarezinho	South
57	Bico do Papagaio	North	243	Seabra	North-East	429	Ibaiti	South
58	Araguaína	North	244	Jequié	North-East	430	Wenceslau Braz	South
59	Miracema do Tocantins	North	245	Livramento do Brumado	North-East	431	Telêmaco Borba	South
60	Rio Formoso	North	246	Guanambi	North-East	432	Jaguariaíva	South
61	Gurupi	North	247	Brumado	North-East	433	Ponta Grossa	South
62	Porto Nacional	North	248	Vitória da Conquista	North-East	434	Toledo	South
63	Jalapão	North	249	Itapetinga	North-East	435	Cascavel	South
64	Dianópolis	North	250	Valença	North-East	436	Foz do Iguaçu	South
65	Litoral Ocidental Maranhense	North-East	251	Ilhéus-Itabuna	North-East	437	Capanema	South
66	Aglomeração Urbana de São Luís	North-East	252	Porto Seguro	North-East	438	Francisco Beltrão	South
67	Rosário	North-East	253	Unaí	South-East	439	Pato Branco	South
68	Lençois Maranhenses	North-East	254	Paracatu	South-East	440	Pitanga	South
69	Baixada Maranhense	North-East	255	Januária	South-East	441	Guarapuava	South
70	Itapecuru Mirim	North-East	256	Janaúba	South-East	442	Palmas	South
71	Gurupi	North-East	257	Salinas	South-East	443	Prudentópolis	South
72	Pindaré	North-East	258	Pirapora	South-East	444	Irati	South
73	Imperatriz	North-East	259	Montes Claros	South-East	445	União da Vitória	South
74	Médio Mearim	North-East	260	Grão Mogol	South-East	446	São Mateus do Sul	South
75	Alto Mearim e Grajaú	North-East	261	Bocaiúva	South-East	447	Cerro Azul	South
76	Presidente Dutra	North-East	262	Diamantina	South-East	448	Lapa	South
77	Baixo Parnaíba Maranhense	North-East	263	Capelinha	South-East	449	Curitiba	South
78	Chapadinha	North-East	264	Araçuaí	South-East	450	Paranaguá	South
79	Codó	North-East	265	Pedra Azul	South-East	451	Rio Negro	South
80	Coelho Neto	North-East	266	Almenara	South-East	452	São Miguel do Oeste	South
81	Caxias	North-East	267	Teófilo Otoni	South-East	453	Chapecó	South
82	Chapadas do Alto Itapecuru	North-East	268	Nanuque	South-East	454	Xanxerê	South
83	Porto Franco	North-East	269	Ituiutaba	South-East	455	Joaçaba	South
84	Gerais de Balsas	North-East	270	Uberlândia	South-East	456	Concórdia	South
85	Chapadas das Mangabeiras	North-East	271	Patrocínio	South-East	457	Canoinhas	South
86	Baixo Parnaíba Piauiense	North-East	272	Patos de Minas	South-East	458	São Bento do Sul	South
87	Litoral Piauiense	North-East	273	Frutal	South-East	459	Joinville	South
88	Teresina	North-East	274	Uberaba	South-East	460	Curitibanos	South
89	Campo Maior	North-East	275	Araxá	South-East	461	Campos de Lages	South
90	Médio Parnaíba Piauiense	North-East	276	Três Marias	South-East	462	Rio do Sul	South
91	Valença do Piauí	North-East	277	Curvelo	South-East	463	Blumenau	South
92	Alto Parnaíba Piauiense	North-East	278	Bom Despacho	South-East	464	Itajaí	South
93	Bertolínia	North-East	279	Sete Lagoas	South-East	465	Ituporanga	South
94	Floriano	North-East	280	Conceição do Mato Dentro	South-East	466	Tijucas	South
95	Alto Médio Gurguéia	North-East	281	Pará de Minas	South-East	467	Florianópolis	South
96	São Raimundo Nonato	North-East	282	Belo Horizonte	South-East	468	Tabuleiro	South
97	Chapadas do Extremo Sul Piauiense	North-East	283	Itabira	South-East	469	Tubarão	South
98	Picos	North-East	284	Itaguara	South-East	470	Criciúma	South
99	Pio IX	North-East	285	Ouro Preto	South-East	471	Araranguá	South
100	Alto Médio Canindé	North-East	286	Conselheiro Lafaiete	South-East	472	Santa Rosa	South
101	Litoral de Camocim e Acaraú	North-East	287	Guanhães	South-East	473	Três Passos	South
102	Ibiapaba	North-East	288	Peçanha	South-East	474	Frederico Westphalen	South
103	Coreaú	North-East	289	Governador Valadares	South-East	475	Erechim	South
104	Meruoca	North-East	290	Mantena	South-East	476	Sananduva	South
105	Sobral	North-East	291	Ipatinga	South-East	477	Cerro Largo	South
106	Ipu	North-East	292	Caratinga	South-East	478	Santo Ângelo	South
107	Santa Quitéria	North-East	293	Aimorés	South-East	479	Ijuí	South
108	Itapipoca	North-East	294	Piuí	South-East	480	Carazinho	South
109	Baixo Curu	North-East	295	Divinópolis	South-East	481	Passo Fundo	South
110	Uruburetama	North-East	296	Formiga	South-East	482	Cruz Alta	South
111	Médio Curu	North-East	297	Campo Belo	South-East	483	Não-Me-Toque	South
112	Canindé	North-East	298	Oliveira	South-East	484	Soledade	South
113	Baturité	North-East	299	Passos	South-East	485	Guaporé	South
114	Chorozinho	North-East	300	São Sebastião do Paraíso	South-East	486	Vacaria	South
115	Cascavel	North-East	301	Alfenas	South-East	487	Caxias do Sul	South
116	Fortaleza	North-East	302	Varginha	South-East	488	Santiago	South
117	Pacajus	North-East	303	Poços de Caldas	South-East	489	Santa Maria	South
118	Sertão de Cratéus	North-East	304	Pouso Alegre	South-East	490	Restinga Seca	South
119	Sertão de Quixeramobim	North-East	305	Santa Rita do Sapucaí	South-East	491	Santa Cruz do Sul	South
120	Sertão de Inhamuns	North-East	306	São Lourenço	South-East	492	Lajeado-Estrela	South
121	Sertão de Senador Pompeu	North-East	307	Andrelândia	South-East	493	Cachoeira do Sul	South
122	Litoral de Aracati	North-East	308	Itajubá	South-East	494	Montenegro	South
123	Baixo Jaguaribe	North-East	309	Lavras	South-East	495	Gramado-Canela	South
124	Médio Jaguaribe	North-East	310	São João Del Rei	South-East	496	São Jerônimo	South
125	Serra do Pereiro	North-East	311	Barbacena	South-East	497	Porto Alegre	South

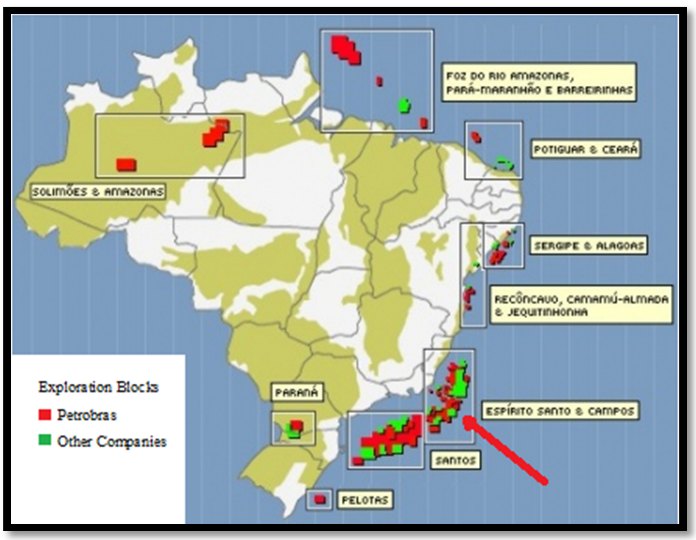
Table 3.5 – List of Brazilian Micro-regions - *continued*

126	Iguatu	North-East	312	Ponte Nova	South-East	498	Osório	South
127	Várzea Alegre	North-East	313	Manhuaçu	South-East	499	Camaquã	South
128	Lavras da Mangabeira	North-East	314	Viçosa	South-East	500	Campanha Ocidental	South
129	Chapada do Araripe	North-East	315	Muriaé	South-East	501	Campanha Central	South
130	Caririaçu	North-East	316	Ubá	South-East	502	Campanha Meridional	South
131	Barro	North-East	317	Juiz de Fora	South-East	503	Serras de Sudeste	South
132	Cariri	North-East	318	Cataguases	South-East	504	Pelotas	South
133	Brejo Santo	North-East	319	Barra de São Francisco	South-East	505	Jaguarão	South
134	Mossoró	North-East	320	Nova Venécia	South-East	506	Litoral Lagunar	South
135	Chapada do Apodi	North-East	321	Colatina	South-East	507	Baixo Pantanal	Centre-West
136	Médio Oeste	North-East	322	Montanha	South-East	508	Aquidauana	Centre-West
137	Vale do Açu	North-East	323	São Mateus	South-East	509	Alto Taquari	Centre-West
138	Serra de São Miguel	North-East	324	Linhares	South-East	510	Campo Grande	Centre-West
139	Pau dos Ferros	North-East	325	Afonso Cláudio	South-East	511	Cassilândia	Centre-West
140	Umarizal	North-East	326	Santa Teresa	South-East	512	Paranaíba	Centre-West
141	Macau	North-East	327	Vitória	South-East	513	Três Lagoas	Centre-West
142	Angicos	North-East	328	Guarapari	South-East	514	Nova Andradina	Centre-West
143	Serra de Santana	North-East	329	Alegre	South-East	515	Bodoquena	Centre-West
144	Seridó Ocidental	North-East	330	Cachoeiro de Itapemirim	South-East	516	Dourados	Centre-West
145	Seridó Oriental	North-East	331	Itapemirim	South-East	517	Iguatemi	Centre-West
146	Baixa Verde	North-East	332	Itaperuna	South-East	518	Aripuanã	Centre-West
147	Borborema Potiguar	North-East	333	Santo Antônio de Pádua	South-East	519	Alta Floresta	Centre-West
148	Agreste Potiguar	North-East	334	Campos dos Goytacazes	South-East	520	Colíder	Centre-West
149	Litoral Nordeste	North-East	335	Macaé	South-East	521	Parecis	Centre-West
150	Macaíba	North-East	336	Três Rios	South-East	522	Arinos	Centre-West
151	Natal	North-East	337	Cantagalo-Cordeiro	South-East	523	Alto Teles Pires	Centre-West
152	Litoral Sul	North-East	338	Nova Friburgo	South-East	524	Sinop	Centre-West
153	Catolé do Rocha	North-East	339	Santa Maria Madalena	South-East	525	Paranatinga	Centre-West
154	Cajazeiras	North-East	340	Bacia de São João	South-East	526	Norte Araguaia	Centre-West
155	Sousa	North-East	341	Lagos	South-East	527	Canarana	Centre-West
156	Patos	North-East	342	Vale do Paraíba Fluminense	South-East	528	Médio Araguaia	Centre-West
157	Piancó	North-East	343	Barra do Piraí	South-East	529	Alto Guaporé	Centre-West
158	Itaporanga	North-East	344	Baía da Ilha Grande	South-East	530	Tangará da Serra	Centre-West
159	Serra do Teixeira	North-East	345	Vassouras	South-East	531	Jauru	Centre-West
160	Seridó Ocidental Paraibano	North-East	346	Serrana	South-East	532	Alto Paraguai	Centre-West
161	Seridó Oriental Paraibano	North-East	347	Macacu-Caceribu	South-East	533	Rosário Oeste	Centre-West
162	Cariri Ocidental	North-East	348	Itaguaí	South-East	534	Cuiabá	Centre-West
163	Cariri Oriental	North-East	349	Rio de Janeiro	South-East	535	Alto Pantanal	Centre-West
164	Curimataú Ocidental	North-East	350	Jales	South-East	536	Primavera do Leste	Centre-West
165	Curimataú Oriental	North-East	351	Fernandópolis	South-East	537	Tesouro	Centre-West
166	Esperança	North-East	352	Votuporanga	South-East	538	Rondonópolis	Centre-West
167	Brejo Paraibano	North-East	353	São José do Rio Preto	South-East	539	Alto Araguaia	Centre-West
168	Guarabira	North-East	354	Catanduva	South-East	540	São Miguel do Araguaia	Centre-West
169	Campina Grande	North-East	355	Auriflama	South-East	541	Rio Vermelho	Centre-West
170	Itabaiana	North-East	356	Nhandeara	South-East	542	Aragarças	Centre-West
171	Umbuzeiro	North-East	357	Novo Horizonte	South-East	543	Porangatu	Centre-West
172	Litoral Norte	North-East	358	Barretos	South-East	544	Chapada dos Veadeiros	Centre-West
173	Sapé	North-East	359	São Joaquim da Barra	South-East	545	Ceres	Centre-West
174	João Pessoa	North-East	360	Ituverava	South-East	546	Anápolis	Centre-West
175	Litoral Sul	North-East	361	Franca	South-East	547	Iporá	Centre-West
176	Araripina	North-East	362	Jaboticabal	South-East	548	Anicuns	Centre-West
177	Salgueiro	North-East	363	Ribeirão Preto	South-East	549	Goiânia	Centre-West
178	Pajeú	North-East	364	Batatais	South-East	550	Vão do Paranã	Centre-West
179	Sertão do Moxotó	North-East	365	Andradina	South-East	551	Entorno de Brasília	Centre-West
180	Petrolina	North-East	366	Araçatuba	South-East	552	Sudoeste de Goiás	Centre-West
181	Itaparica	North-East	367	Birigui	South-East	553	Vale do Rio dos Bois	Centre-West
182	Vale do Ipanema	North-East	368	Lins	South-East	554	Meia Ponte	Centre-West
183	Vale do Ipojuca	North-East	369	Bauru	South-East	555	Pires do Rio	Centre-West
184	Alto Capibaribe	North-East	370	Jaú	South-East	556	Catalão	Centre-West
185	Médio Capibaribe	North-East	371	Avaré	South-East	557	Quirinópolis	Centre-West
186	Garanhuns	North-East	372	Botucatu	South-East	558	Brasília	Centre-West

3.5.5 Maps of Brazil

Figure 3.1: Regions and States in Brazil

Figure 3.2: Oil Exploration Blocks in Brazil



Source: Petrobras

Chapter 4

SMEs and Economic Growth in Brazil

4.1 Introduction

In Chapter 2 we derived the Mankiw et al. (1992) convergence growth model and augmented it to obtain a more general and flexible model that allows the inclusion of additional policy and structural variables. Regressions of this type of model are known as “Barro Regresions”, due to Barro (1991) seminal paper. After Mankiw et al. (1992) and Barro (1991) influential works, the growth literature based on Solow (1956) tradition flourished and considered a plethora of policy and structural variables in the growth framework. Levine and Renelt (1992) and Durlauf et al. (2005), for example, list an extensive number of variables that were used in growth regressions and linked many other fields of economics to the literature of economic growth. In this context, Sala-i-Martin (2002) argues that one important contribution made by the growth literature that follows the Mankiw et al. (1992) and Barro (1991) tradition is that it has exerted influence on other economic literatures such as development, economic geography, macroeconomics, econometrics and industrial organisation. Recently, this influence was also extended to study the importance of Small and Medium Enterprises (SMEs) and entrepreneurship for economic growth as, for example, in Beck et al. (2005a), Audretsch and Keiback (2004) and Mueller (2007). In addition, Solow (2007) recognizes entrepreneurship as one important force that drives a wedge between knowledge and total factor productivity, the study of entrepreneurship and small firms could add to the explanatory power of growth theory.

However, the available empirical evidence is either based on studies focusing on cross-country regressions or studies for developed regions. Little has been done to study SMEs and entrepreneurship in this tradition in developing countries.

The aim of this chapter is to examine the relationship between the SME sector and economic growth in Brazil using a panel data setting covering 503 micro-regions for the period 1980-2004, providing additional evidence for the role of the SME sector in the context of a developing country. The chapter also examines how these aspects of the SME sector influence economic growth in different regions with dissimilar level of development. The chapter focuses on the analysis of two important aspects of the SME sector. Firstly, it deals with the importance of the relative size of the SME sector measured by the share of the SME sector in the total formal employment. Secondly, it provides an extended analysis of the relationship between SME and growth focusing on the stock of SMEs' human capital measured by a new constructed variable for the average years of schooling in SMEs.

4.2 The Model and Data

4.2.1 The Model

The specification used in this chapter to study SMEs and economic growth using Brazilian micro-regional data is the general growth equation (2.24) presented in Chapter 2, which will be first estimated using a proxy for the size of the SME sector as in Beck et al. (2005a) and afterwards extended to encompass a proxy for human capital of SMEs. The specification representing explicitly the SME sector becomes:

$$gr_{it} = a_i - b \ln y_{i,t-1} + \beta \ln SMER_{it} + \alpha \ln SMEH_{it} + \psi \ln X_{it} + v_{it} \qquad (4.1)$$

where i denotes each individual economy, t represents each period of time considered, gr denotes the annual GDP per capita growth in each cross-section, $\ln y_{t-1}$ is the initial GDP per capita, a_i represents the time invariant economy specific effect, and v is the error term. Furthermore, *SMER* is the relative size and *SMEH* is the human capital of SME sector, respectively. The vector X encompasses other growth determinants suggested in the economic growth literature as discussed in Chapter 2.[33] Explaining the channels through which SMEs can influence growth, this framework also allows us to put some light on the importance of this sector for developing economies.

4.2.2 Data

To evaluate the role of SMEs in Brazil a new dataset combining regional data for GDP, population and education with the aggregate results of individual establishment data was constructed to study the relationship between growth and SMEs. The data set used to construct the data used in this chapter is described in detail in Chapter 3. We use a balanced panel dataset from 1980 to 2004 with 508 cross-sectional observations organised in intervals as close as possible to 5-years intervals (according to data availability) to avoid business cycle influence as argued in Islam (1995), Caselli et al. (1996), Temple (1999), Bond et al. (2001) and Badinger et al. (2005). The data was collected from three sources: IBGE (Brazilian Institute of Geography and Statistics), IPEADATA (Institute of Applied Economic Research data system) and RAIS (Brazilian Annual Report of Social Information). The information from IBGE and IPEA were used to collect the micro-regional level data for variables other than SMEs' measures, and RAIS provided the information necessary to construct our SMEs' measures. What follows is a description of the variables and their sources.

1. Real GDP at factor price data for each micro-region (GDP) was collected from IBGE and IPEADATA. As a result, based on these two sources, our data points for this series are: 1980, 1985, 1996, 2000 and 2004. This series constitutes our main constraint in terms of data availability for the construction of the panel

[33] In this case the vector X encompasses the overall level of human capital ln(School) and the term $\ln(n+\delta+g)$, under the usual assumption that $\delta+g$ equals 0.05.

dataset.[34] From these data points, the average annual growth rates for each time span were calculated for the GDP per capita.

2. The data on population, used to calculate the GDP per capita and population growth (*n*), were collected from IPEA.

3. The average years of schooling of the population over 25 years old (*School*) were taken from IPEADATA. We used 5-year lagged School information from 1980 to 2000 to construct our panel. The data for this variable is available only in a ten year time span interval from 1980 to 2000, and data points in between these years were constructed by interpolation. We used this variable lagged five years because 2000 was the most recent information available.

Additionally, two measures to capture the SMEs' dimension in Brazil were used. The information necessary to construct these variables was retrieved from RAIS (using the software SGT micro provided by the Ministry of Labour). However, before describing the variables that represent the SMEs dimension, we briefly discuss the criteria used to define SMEs. The classification of SMEs varies across and within countries. Ayyagari et al. (2007), for instance, provide a thorough discussion about the difficulties of collecting data and finding a common measure for SMEs. They show that the most commonly used criterion to classify SMEs is based on employment information. The cut-off used to define SMEs generally varies between 100 and 500 employees, with a large number of sources using 250 employees as a cut-off. For instance, the European Union and Beck et al. (2005) adopt 250 employees as a cut-off to classify SMEs. Therefore, the cut-off of 250 employees seems to be a reasonable choice based on existing classifications and is also in line with the literature on SMEs. In addition, we used the cut-off of 500 employees to provide alternative results using a different cut-off.[35] The variables related to the SME sector were constructed as follows.

[34] See Chapter 3 for further details.

[35] The cut-off of 500 employees is particularly important because despite the fact that there is no legal classification for SMEs based on the number of employees in Brazil, SEBRAE uses this cut-off to define SMEs. This criterion is also adopted by the U.S Small Business Administration.

4. The relative size of the small and medium enterprise sector (*SMER*) is the share of the SME sector in the total formal labour force in manufacturing. We also constructed this measure extending the definition of an SME sector, incorporating commerce and services (*SMER2*). As discussed above, we constructed all the SME variables considering 250 and 500 employees as cut off points to define SMEs.

5. Finally, we constructed the variable for the average years of schooling in SMEs (*SMEH*) manufacturing from the information about educational attainment provided by RAIS. We also constructed this variable incorporating commerce and services sectors (*SMEH2*). We follow Muendler (2007) and attribute a number of years of schooling to each level of educational attainment provided by RAIS to generate a continuous series, Table 4.1 reports this correspondence. This variable that captures the human capital dimension in SMEs is constructed as follows:

$$SMEH = \frac{\sum_{i=0}^{N} (E^{i} \times S^{i})}{E}$$

where E^i is the number of SME employees in each range of education, S^i the number of years of schooling for that particular range and E is the total number of SME employees.

Table 4.1. Educational Range and Years of Schooling

RAIS Educational Range Level	**Imputed years of Schooling**
Illiterate	0
Primary School Dropout	1
Primary School Graduate	4
Middle School Dropout	5
Middle School Graduate	8
High School Dropout	9
High School Graduate	11
College Dropout	12
College Graduate	15

Note: The Brazilian educational system is similar to the US system in terms of its structure. Education is divided into three levels, with several grades in each division. Fundamental (or primary) education is equivalent to the US Elementary School. Middle education is equivalent to the US Middle School. High school is also equivalent to the US high school. The main difference is that high school in Brazil has only 3 years as opposed to 4 in the US.

The final sample comprises 508 out of 558 Brazilian micro-regions for each of the four cross sections, generating a panel with 2032 observations. The sample loses 53 regions that did not present data for all variables of interest at all data points.[36] A drawback in our data is the paucity of information about physical capital. Unfortunately, this type of information is not available for Brazil, even at state level. Researchers usually use industrial electricity consumption as a proxy for physical capital when studying growth at state level. However, this solution is not possible at the more disaggregated level, such as micro-regions. Energy suppliers by law are not obliged to provide this information by municipality to the Brazilian Ministry of Energy and Mines or to any other regulator.[37]

Our data about SMEs are not without caveats as already discussed in Chapter 3. As stated by Beck et al. (2005a), SME measures are static in the sense that they do not account for the entry of new firms, graduation of successful SMEs into large and the exit of failing ones. In terms of measurement, accordingly to the Ministry of Labour, omissions of information are more frequent in small municipalities, agriculture, public administration and building related activities.[38] In addition, the RAIS is completed on a self classification basis, which means that the establishment responsible for the information has to complete the RAIS form by itself accordingly to the RAIS reference guide. As a result, establishments might classify themselves into a wrong sectoral group, for instance. However, it is reasonable to assume that

---

[36] The sample loses 50 micro-regions that did not present data at all data points. These micro-regions are a set of small municipalities usually located in poor and isolated areas, most of them in the Amazon Basin region. In 2004, this set of micro-regions encompassed only 2% of the Brazilian population and it is reasonable to assume that the omission of these regions due to lack of data does not generate a serious bias; this omission would only affect the results in the unlikely situation where the regions not considered have a substantially different dynamics when compared to neighbouring regions of the sample. Also, these regions that could be part of an unbalanced panel would reduce the number of cross-sections available to calculate the GMM diagnostic tests. We need to use of five cross sections in the GMM-DIFF estimates, since two cross-sections are lost to control for the lagged GDP per capita and to take the first difference. Using unbalanced panel, we would not have the minimum of three cross sections necessary to calculate the diagnostic tests for the GMM estimates (see Arellano and Bond 1991).

[37] We thank João Antonio Moreira Patusco from the Brazilian Ministry of Energy and Mines for this information.

[38] This is an additional reason for the use of micro-regions. This more aggregated territorial unit encompasses many municipalities and reduces bias from inaccurate data in small municipalities. We do not rely on information from these problematic sectors (agriculture, public administration and construction) to construct our SMEs' variables.

most of these problems are offset by the census nature of RAIS, in the sense that stochastic misfiling would be compensated by the size of the data base.

4.3 Panel Data Estimators

A panel data approach is used to estimate equation (4.1). First, the model is estimated without *SMEH* as in Beck et al. (2005a) and then extended to encompass aspects of SMEs' human capital. As discussed in the previous section, the period of analysis spans from 1980 to 2004 and the data is organized in intervals as close as possible to 5-year intervals, generating four cross-sectional points with 508 observations each. The data for the right hand side variables correspond to the last year of the respective time spans, unless specified in the previous section.

Three familiar methods of estimation with panel data are used. First, we estimate an OLS model under the assumption that all regions have the same level of technology. In this case the constant term is the same for all economies and we make the assumption that the individual technological term is decomposed into a constant term and an economy-specific shock term that is independent of the explanatory variables.[39] This assumption is made in Mankiw et al. (1992) to justify the use of their OLS regression. However, Arellano (2003) suggests that a major motivation for using panel datasets has been the ability to control for possibly correlated, time invariant heterogeneity, without observing it. Clearly the OLS regression fails to acknowledge this effect. In the context of growth models, Islam (1995) argues that it is difficult to believe that the error term is not correlated with explanatory variables when we are not considering the economic specific technological level explicitly. To allow for differences in the level of technology, we introduce micro-regional fixed effects expressed in the specific dummy variables estimated by Least Squares Dummy Variables (LSDV) as in Islam (1995). Nevertheless, although LSDV explicitly incorporates the economy micro-regional specific effect, it fails to consider the endogeneity problem. The estimated coefficients from the growth equation are

[39] The technological term in the equation (2.13) in Chapter 2 becomes $\ln A(0)=a+v$.

biased and inconsistent due to the fact that $\mathrm{E}(v_{i,t}|x_i, y_{t-1}) \neq 0$, meaning that independent variables are correlated with the error term. To take into account this endogeneity, the first differenced GMM Arellano and Bond (1991) estimator (GMM-DIFF), such as first applied to the convergence regression by Caselli et al. (1996) and the system GMM Blundell and Bond (1998) estimator (GMM-SYS) are the alternatives to tackle this problem. Simulations in Arellano and Bond (1991) and in Blundell and Bond (1998) showed that these estimators perform well in panels with "small *T* (time periods) and large *N* (number of cross-sectional observations)" and are less prone to suffer from the bias demonstrated by Nickell (1981) when we estimate a dynamic model using fixed effects.

4.3.1 First Differenced GMM Estimation

The growth equation (4.1) is first differenced to eliminate the regional specific effects and the dynamic equation to be estimated by GMM-DIFF becomes:

$$\begin{aligned} gr_{it} = \gamma_t - b\Delta \ln y_{i,t-1} + \beta\Delta \ln SMER_{it} + \\ + \alpha\Delta \ln SMEH_{it} + \psi\Delta \ln X_{it} + \Delta v_{it} \end{aligned} \quad (4.2)$$

where the left hand side is the log difference growth of GDP per capita, *X* is a vector that control for the conditional convergence, *SMER* and *SMEH* represent the SME sector share and human capital stock of SMEs, respectively, and γ_t reflects the period specific intercepts that capture productivity changes that are common to all regions.[40]

Assuming the absence of serial correlation, i.e. $\mathrm{E}(v_{it}) = \mathrm{E}(v_{it}v_{is}) = 0$, values of a dependent variable *y* lagged two periods or more are valid instruments in the equation in first differences. This is because the "lagged" dependent variable is

[40] In this case the vector *X* encompasses the overall level of human and population growth. Our constant term γ_t is proxied by time dummies for 1996, 2000 and 2004.

predetermined with respect to v_{it}, meaning it is independent of current disturbances but may be influenced by a past one.[41]

As in Bond et al. (2001), we do not expect our control variables to be strictly exogenous (uncorrelated to past, present and future values of the error term) and consider all of them as potentially endogenous in the sense that there are feedbacks of present and past shocks to GDP growth onto the present value of control variables, a situation expressed by:

$$\mathrm{E}(x_{it}v_{is}) \neq 0 \qquad \text{for } s \leq t$$

In this case, only values of control variables lagged two periods or more are valid instruments for the differenced equation.

Assuming that transient errors are serially uncorrelated and that the lagged dependent variable is predetermined, all valid instruments will be used with the following moment restriction:

$$\mathrm{E}(y_{i,t-2}\Delta v_{it}) = 0 \qquad \text{for } i = 1,...,N \text{ and } t = 3,...,T$$

Additionally, assuming that our control variables are endogenous; the complementary set of moment conditions emerges:

$$\mathrm{E}(x_{i,t-2}\Delta v_{it}) = 0 \qquad \text{for } i = 1,...,N \text{ and } t = 3,...,T$$

Using the moment restrictions for $y_{i,t-1}$ and x_{it} based on all possible valid instruments, we can summarize these moment restrictions in vector form as: $\mathrm{E}(Z_i^{'}\ \Delta v_i) = 0$, where Δv_i is the vector of the differenced error terms ($\Delta v_{i3},...,\Delta v_{iT}$) and Z_i is a matrix that comprises the instruments to be used by the moment restriction. Following Hansen (1982), the optimal estimator for generalized method of moments

[41] The dynamic nature of the growth convergence equation can be better observed if we rewrite it following Islam (1995). In this case, Equation 4.1 can also be written as:

$$\ln y_{it} = (1-b)\ln y_{i,t-1} + \beta \ln SMER_{it} + \alpha \ln SMEH_{it} + \psi \ln X_{it} + v_{it}$$

minimizes the sum of the squares $\bar{v}' ZAZ' \bar{v}$, where $\bar{v}$ is the (T-2) vector (Δv_{i3}, $\Delta v_{i4,\ldots,}\Delta v_{iT}$) and A is any square symmetric matrix that needs to be chosen. For the case of equation (4.2), the solution for the optimal estimator is then given by:

$$\delta = \begin{bmatrix} \hat{b} \\ \hat{\beta} \\ \hat{\alpha} \\ \hat{\psi} \end{bmatrix} (W'ZAZ'W)^{-1} W'ZAZ'\Delta y$$

where W is the vector for all independent variables.[42]

Alternative choices of A generate the one step or two step estimators. The one step estimator is obtained by setting $A = (N^{-1} \sum_{i=1}^{N} Z_i' H Z_i)^{-1}$, where H is a (T-2) square matrix that has twos in the main diagonal, minus ones in the first subdiagonal and zeros otherwise (Arellano and Bond, 1991). The matrix H is based on the assumption that the errors are independent and identically distributed with constant variance. Hence, the standard errors estimates would not be robust to heteroskedasticity or serial correlation in the errors. To calculate the two step estimator, the residual from the preliminary one step estimation are used to calculate the optimal choice for A based on a consistent estimate of the covariance matrix of the error term when we consider $H = \Delta\hat{v}_i \Delta\hat{v}_i'$, where $\hat{v}_i$ are the residuals from an internal one step consistent estimation. Therefore, the expression for A becomes: $A = (N^{-1} \sum_{i=1}^{N} Z_i' \Delta\hat{v}_i \Delta\hat{v}_i' Z_i)^{-1}$ and generates the so called two step estimation.

Nevertheless, Arellano and Bond (1991) warn that the estimate of the asymptotic standard errors of the two-step estimation can be severely downward biased in small samples, suggesting that caution would be advisable in making inferences based on the two-step estimation alone. Motivated by this drawback, Windmeyer (2005) propose a correction for the covariance matrix of the error term for finite sample that performed well in simulations, leading to a more accurate

[42] The vector W comprises ΔY_{t-1}, ΔX, $\Delta SMER$ *and* $\Delta SMEH$ in this case.

inference for the two step estimator. All two-step estimations reported in this chapter use Windmeyer (2005)'s correction.

4.3.2 System GMM Estimation

According to Blundell and Bond (1998), instruments used in the standard first differenced GMM become less informative when the coefficient of dependent lagged variable increases and when the variance of the fixed effect η_i increases (i.e. small time series information). Therefore, they suggest estimating a system combining two sets of equations. In the first set of equations, as described in the previous section, lagged levels of y_{it} and x_{it} were used as instruments for the equation in first difference. The complementary set of equations in the proposed system is based on the equation in levels using instruments in difference as suggested by Arellano and Bover (1995). Blundell and Bond (1998) augment Arellano-Bond estimator by introducing the additional assumptions that the first difference of instruments is uncorrelated with fixed effects. These are represented by the following restrictions:

$$\mathrm{E}(\eta_i \Delta y_{i2}) = 0 \text{ for } i = 1,\ldots,N$$

and

$$\mathrm{E}(\eta_i \Delta x_{it}) = 0 \text{ for } i = 1,\ldots, N \text{ and } t = 2,\ldots,T$$

Again, as in the previous section, validity depends on the assumption that the error term is not serially correlated. Furthermore, as the lagged *y* is predetermined and *x* is endogenous, $\Delta y_{i,t-1}$ and $\Delta x_{i,t-1}$ and further lags of differenced variables can serve as instruments for the equations in levels. These additional assumptions yields *T-2* further linear moment conditions as follows:

$$\mathrm{E}(v_{it} \Delta y_{i,t-1}) = 0 \text{ for } i = 1,..,N \text{ and } t = 3,..,T$$

and

$$\mathrm{E}(v_{it} \Delta x_{i,t-1}) = 0 \text{ for } i = 1,\ldots,N \text{ and } t = 3,\ldots,T$$

Using the moment restrictions for $\Delta y_{i,t-1}$ and Δx_{it} based on all possible valid instruments, we can summarize these moment restrictions in vector form as; $\mathrm{E}\left(\Delta Z_i^{+} v_i\right) = 0$, where Z_i^{+} is a matrix that comprises the instruments to be used by the moment restrictions.

Simulations reported in Blundell and Bond (1998) show that the use of additional moment conditions for the equation in levels generates a more efficient estimator when the number of time series observations is small. In the context of growth models, Bond et al. (2001) suggest that the system GMM provides more reliable results in empirical growth estimations. Hence, to provide all ranges of results and avoid misleading conclusions based on GMM in differences alone, in this chapter we provide the results for the first differenced and system GMM estimators, providing the alternative results for both, one and two step estimations using Windmeyer (2005)'s correction.[43]

4.4 Results

This section reports the results for the two specifications we are going to use to draw some inferences about the role of SMEs in the process of economic growth in Brazil. Firstly, we follow Beck et al. (2005a) and incorporate the share of employment accounted for SMEs as in equation (4.1). Later, we extend the model and consider the SMEs' stock of human capital as an additional measure for the SME sector.

4.4.1 SMEs' Sector Share and Growth

Table 4.2 reports the results for the same specification using two different cut-offs definitions of SMEs. As mentioned earlier, as there are many different definitions of SMEs, we opted for reporting two different employment cut off points to classify an

[43] The Windmeyer (2005) correction is available in the package xtabond2 for Stata, for further details see Roodman (2009b).

establishment as a SME. The first one considers the cut off of 250 employees as used by Beck et al. (2005a), and the second uses the Brazilian (and the U.S) classification of SMEs with a cut off of 500 employees. Overall they present very similar results and do not contradict each other in the sense that they never have opposite and significant signs.[44] For each set of regressions we first report the results for the standard pooling regression. Subsequently, we incorporate the individual specific micro-regional effect through LSDV. Finally, to consider together the individual effect and treat for endogeneity, we present the results for GMM-DIFF and GMM SYS together with their respective diagnostic tests at the bottom of each table.

The OLS results in the first column, of each set of size criterion, indicate the presence of convergence among Brazilian micro-regions as suggested by the existing empirical literature discussed in Chapter 2 about growth in Brazil. The coefficients are strongly significant at one percent level and have the expected negative sign. The coefficients of human capital are positive and strongly significant as well, suggesting a positive relation between the stock of human capital and the growth process. As suggested by the theory, the variable that encompasses the population growth is negatively related to the growth rates and enters significantly in all OLS regressions. As mentioned before, unfortunately, we do not have a proxy for physical capital for the Brazilian micro-regions. This is certainly a drawback but we also believe that a substantial part of the initial stock of physical capital is captured by the initial GDP per capita level.

Concerning the SME sector, the results suggest that a larger SME sector is a sign of slower growth. For the two size criteria, *SMER* (using the cut off of 250 and 500 employees) enters with a negative sign and significantly. However, OLS regressions lack of reliability does not allow us to make any assertive statement since these regressions do not control for either the individual specific effect or endogeneity. Simulations reported by Arellano and Bond (1991) show that in a dynamic context the OLS regression tends to push the lagged dependent variable (in this case our convergence coefficient) upwards.[45]

[44] Overall we focus on the criterion of 250 employees but always comment if there is any important difference in the results from the different size criteria.
[45] See Arellano and Bond (1991), Table 1 on page 284.

To incorporate the micro-regional specific effect we use the LSDV to control for micro-regional specific time invariant heterogeneity. This allows us to control for a range of micro-regional specific characteristics (e.g. level of technology, rule of law, geography, institutions) without observing them. These results are reported in the second column of each set of regressions. We observe a stronger convergence pattern, evidenced by the convergence coefficient and a positive effect of education on growth. These coefficients are significant at one percent level. Table 4.2 also suggests a negative effect on growth stemming from the population growth when we use the LSDV estimator; however, results are not statistically significant. Again, we observe a negative effect of the size of SME sector in terms of employment on growth, and this effect is always statistically significant in all LSDV regressions.

Nevertheless, the LSDV estimations are subjected to endogeneity bias, especially when we have a dynamic panel model. Simulations provided by Arellano and Bond (1991) show that when we estimate the lagged dependent variable coefficient (convergence coefficient) using LSDV, the results tend to be biased downwards, as demonstrated by Nickell (1981). To treat endogeneity, we apply the GMM-DIFF and the GMM-SYS, reporting the results for the first and second step estimates respectively. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t-1* and earlier. The columns 3 and 4 report the first and second step estimates using the GMM in first differences. Once again the coefficients on the initial GDP per capita are strongly significant reinforcing the argument of conditional convergence in Brazil. The positive effect of education on growth is also confirmed with a high degree of confidence and the coefficients on the population growth remain without statistic significance for the GMM-DIFF estimator.

For *SMER*, using 250 employees as a cut off, coefficients are not significant statistically at conventional levels, suggesting that SMEs do not promote economic growth. Furthermore, the negative and significant effect of *SMER* (provided by the

LSDV estimation) on growth is not robust when we try to control for endogeneity.[46] But would that result be a glimpse of lack of robustness in terms of the negative effect of size of the SME sector on economic growth? If this is true, our analysis goes close to Beck et al. (2005a) in the sense that no robustness was found, suggesting a neutral effect of SMEs on growth and no causal effect from SMEs on growth. Nonetheless, as OLS and LSDV are biased in opposite directions, we would expect a reliable estimator to provide a convergence coefficient that lies between the results of those methods. However, we can observe in the Table 4.2 that the GMM-DIFF provides convergence coefficients that are ever smaller that of the LSDV regression, suggesting a downward bias. Simulations in Blundell and Bond (1998) illustrate the poor finite bias experienced by GMM-DIFF and proposed the use of GMM-SYS with additional moment conditions that proved to be more robust than GMM-DIFF in some cases.

The system GMM estimations for the initial GDP per capita now are negative and significant at one percent level and lie between the OLS and LSDV estimates. This is a sign that the results provided by the system GMM estimator can be less biased in terms of the coefficient size. The coefficient of the human capital variable is again positive and highly significant, while the population growth is in line with the theory and is negative and significant in all GMM-SYS regressions at one percent level of confidence. The negative effect of SME on growth turns positive in all set of regressions and is significant for the *SMER* (when we use 250 employees as a cut off) at five percent level in the one step estimations.

Therefore, Table 4.2 provides mixed evidence about the importance of the manufacturing SME sector for economic growth. The GMM-DIFF results implicitly suggest that manufacturing SMEs are less productive than larger ones and do not contribute to gains in productivity in Brazil. On the other hand, GMM-SYS results go opposite and tend to support the view that SMEs are important for economic growth. The difference can be associated with the validity of the additional instruments made available by that GMM-SYS that depends on the assumption that changes in instrumenting variables (as we use the difference as instruments for the

[46] The results for *SMER* 500 employees cut off are negative and significant in all regressions.

equation in levels) are uncorrelated with the fixed effects. In particular, they require that individuals (in our case micro-regions) are not too far from their steady-states, in the sense that the initial distances from steady-states are not systematically related to the fixed effects (Roodman, 2009a). If the specific individual distance from the steady-state alone is to be correlated with future growth (through catching up effect) the additional instruments will be made invalid since one individual specific effect will be related to future growth in the conditioning variables, a situation that is not difficult to imagine in the growth theory context. But how can we infer which estimator provide more reliable results given the assumptions about the moment conditions?

When we use instrumental variables' methods, it is important to test whether the instruments we are using are valid. To evaluate the assumptions made by the GMM estimators when constructing instruments, diagnostic tests for the validity of instruments developed by Arellano and Bond (1991) are presented at the bottom of all tables. The values for AR1 and AR2 are the *p*-values for the m_1 and m_2 autocorrelation test under the null of no autocorrelation.[47] We are particularly interested in the results provided by AR2, which tests for the lack of second-order serial correlation in the first differenced residuals. The first-order autocorrelation is expected and uninformative, since Δv_{it} is related to $\Delta v_{i,t-1}$ because they share the common term $v_{i,t-1}$. Thus to check for first-order autocorrelation in levels we check for second-order autocorrelation in differences (Roodman, 2009b).

We also report the Sargan and Hansen overidentification test for the joint validity of the instruments under the null that the instruments are valid.[48] The difference between these tests is that Hansen is a more consistent statistic when errors are believed to be heteroskedastic while Sargan is more appropriate when

[47] The m_l statistic for the *l*-lag order correlation proposed by Arellano and Bond (1991) is given by the following expression: $m_l = \frac{\hat{v}'_{-l}\hat{v}}{\sqrt{\hat{v}}}$, where $\hat{v}$ represents the estimated residuals of GMM estimations. The m_l order statistic is standard normal distributed and test the null that differenced errors are not *l*-order serially autocorrelated.

[48] The Hansen statistic is given by: $J = \hat{v}'Z(\sum_{i=1}^{N} Z_i'\hat{v}_i\hat{v}_i'Z_i)^{-1}Z'\hat{v}$, where $\hat{v}$ represents the two step residuals in this case. This statistic becomes the Sargan statistic when we believe errors are homoskedastic and use the first step residuals (see Arellano and Bond (1991) page 282). Sargan and Hansen statistics are distributed as chi-squared with degrees of freedom equal to the number of over-identifying restrictions. The null hypothesis is $\mathrm{E}[Z'v] = 0$. Under the null that instruments are valid.

errors are believed to be homoskedastic. Thus, if robust inference is sought one may calculate the test for overidentification via Hansen test (Baum et al. 2003).

Additionally, the Hansen test can also be used to test the validity of subsets of instruments, via the "Difference-in-Hansen" test. If we perform the estimations with and without a subset of suspect instruments, the difference in the two results is itself asymptotically chi-squared distributed as the original tests, under the same null of joint validity of instruments. This test provides the criterion we need to choose between GMM-DIFF and GMM-SYS. If we accept the null hypothesis, it means that the additional subset of instruments of the latter (GMM-SYS) is exogenous. As discussed before, these additional moment conditions require the non trivial assumption that the initial distance from steady-state across individuals are uncorrelated with fixed effects.

Using the full set of instruments available, the results for the autocorrelation test (AR2) in Tables 4.2 indicate that instruments cannot be considered invalid due to autocorrelation for the two step GMM-DIFF estimates. On the other hand, the results for one step GMM-DIFF and GMM-SYS suggest that the instruments are being made invalid because of autocorrelation. This pattern suggests that the problem of autocorrelation might stem mainly from the additional instruments provided by GMM-SYS estimation, once we get more valid instruments from the GMM-DIFF estimates. The Sargan and Hansen tests for the validity of the instruments always reject the null that instruments are valid. However, the Sargan test is not reliable when we observe heteroskedasticity and we rely more on Hansen test for inference. The Difference-in-Hansen test is also provided for the GMM-SYS to look at the additional subset of instruments created by the additional moment conditions for the equation in levels.[49] For all GMM-SYS estimates, Difference-in-Hansen test always rejects the null and suggests that instruments for the specification in levels are not independent of the error term and are also a source of additional trouble in terms of endogeneity. This indicates that the use of the GMM-SYS is adding more

[49] The "Difference-in-Hansen" test is calculated based on the residual of the GMM two-step estimations (see footnote 41 and Section 4.3.1). Therefore, the value of this test reported for both 1-step and 2step estimations are the same.

endogeneity to the system and that the use of the GMM-DIFF could be more appropriate.

Therefore, the preferred results based on GMM-DIFF estimator support the idea that SMEs are not important for growth. These results are in line with our expectations and with the discussion provided in Chapter 2 that argues that the entrepreneurial activity in Latin American countries are dominated by necessity entrepreneurship that is likely to be negatively associated with economic growth. Additionally, the results provided by the Diference-in-Hansen test do not support the additional assumption of Blundell and Bond (1998) that the specific effect and the error term are uncorrelated. In other words, cross-micro-regional differences in unexplained growth (including the fixed effects) apparently correlate with distances from the steady-states.[50]

Thus far, our analysis is based only on the manufacturing sector and this was also a shortcoming identified by Beck et al. (2005a) that suggested that it would be useful to have information on SME employment beyond manufacturing. Conversely to Beck et al. (2005a), we are able to go beyond the manufacturing sector and collect data for the commerce and services sectors to create another measure for SMEs and provide evidence for an extended SME sector. The new variable of interest is now the SME employment share in manufacturing, commerce and service (*SMER2*). Table 4.3 mimics the structure of Table 4.2 and reports the same estimations for our extended SME sector.

Overall, we can observe that considering an extended SME sector provides the same qualitative analysis when we look at the OLS and LSDV estimates in terms of the sign and significance of the estimated coefficients. When we analyse the set of GMM estimates the initial analysis (based on Table 4.2) about the effect of the initial GDP per capita, human capital and population growth still holds. The coefficients on the initial GDP per capita are strongly significant reinforcing the argument of conditional convergence in Brazil. The positive effect of education on growth is also confirmed and the coefficients on the population growth remain without statistic significance for the GMM-DIFF estimator. These results that the population growth

[50] This result is according to our expectations since it is not difficult to

is not significant could be related to the fact that income per capita is the main determinant of migration in Brazil as argued in Figueiredo and Garcia (2003). Similar insignificant results for population growth in growth regressions are found in Nakabashi and Salvato (2007) that suggested that this significant population growth endogeneity makes its coefficient insignificant. Our system GMM estimations for the initial GDP per capita are negative and significant at one percent level and the coefficient on the human capital variable is again positive and highly significant. Now, the population growth is consistent with economic growth theory and is negative and significant.

When we consider the effect of the SME sector on growth, the analysis favour the view that SMEs does not promote growth and we do not observe mixed evidence for this extended SME sector (manufacturing, commerce and services) with the GMM-DIFF and GMM-SYS providing similar results. The coefficient of SME sector is always negative and significant at ten percent level for 7 GMM estimates out of 8 results, GMM-DIFF and GMM-SYS now support the view that SMEs are less productive. This is also an indication that SMEs in manufacturing are more important and productive than SMEs in the commerce or service sectors. It is not surprising that the extended SME sector that includes commerce and services sectors consistently suggest that small businesses are negatively related to economic growth. This is because most of the commerce and services sectors are made of small high street shops that are likely not to be more productive than the manufacturing sectors.

The diagnostic tests for the GMM estimates presented in Table 4.3 indicate that instruments cannot be considered invalid due to autocorrelation for the GMM-DIFF estimates and suggests that the problem of autocorrelation stems from the additional instruments provided by GMM-SYS estimation. The Sargan and Hansen tests for joint validity of the instruments always reject the null that instruments are valid for all of the GMM estimates. Additionally, for all GMM-SYS estimates, Hansen-in-Difference test always rejects the null and suggests that instruments for the specification in levels are not independent of the error term and are also a source of additional trouble in terms of endogeneity, suggesting that the GMM-DIFF could be more appropriate.

Therefore, for Tables 4.2 and 4.3, Sargan and Hansen tests are not supportive to the hypothesis that the instruments we used are exogenous. Furthermore, the Difference-in-Hansen test also indicates that the additional set of instruments provided by the GMM-SYS estimator is also a source of endogeneity, suggesting the use of the GMM-DIFF estimator to analyse a causal relationship. However, results provided by the GMM-DIFF in this section are driven by endogeneity and do not support the view that SMEs are important to promote growth in Brazil.[51] Additionally, they indicate that the presence of SMEs of service and commerce sectors make the negative impact of SMEs on growth clearer. The results provided by the GMM-DIFF are in line with Van Stel et al. (2005) that found a negative effect of entrepreneurship on growth for developing economies and provides indirect indication that the institutions in Brazil do not incentive productive entrepreneurship. The fact that more SMEs do not foster economic growth could be related to the fact that they are not intensive in human capital and therefore are not able to contribute to the productivity growth, an explanation hinted by Van Stel et al. (2005).

Therefore, considering just the structure of the economy in terms of the share of SME employment share might not be as informative and ignores other important characteristics that shape the SME sector, such as, the level of human capital used by this sector. Drawing on Van Stel et al. (2005)' suggestion and to extend Beck et al. (2005a), we suggest the use of a proxy for SMEs' human capital, *SMEH*. This proxy represents the average years of schooling of SMEs' employees. We expect that higher levels of human capital (applied into the SME) are associated with higher levels of productivity in SMEs. When human capital level available in SMEs is higher, the productivity of the SMEs tends to be higher because the presence of higher skilled workers facilitates the adoption of a new technology used elsewhere (absorptive capacity) through a process of technological diffusion described by Nelson and Phelps (1966). Additionally, Griffith et al. (2004), argue that human capital can also constitute an important aspect of the innovation process (innovative capacity) that leads to productivity improvements and growth. Hence, the relative size of the SME

[51] Overall, the estimates presented in Tables 4.2 and 4.3 are driven by endogeneity as suggested by diagnostic tests. Therefore, it is difficult to make any robust inference about any causal relationship of SMEs on growth.

sector alone might not be informative about its interaction with growth and the level of SMEs' human capital would add a great amount of information to the analysis.

This broader approach to study the interaction between SMEs and growth could help to clarify the differences between the roles of SMEs across regions and can be an important instrument to propose an adequate public policy for them.

Table 4.2: Industry Employment Share and Growth

	SME250						SME500					
	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step
lnGDP_{t-1}	-0.0444***	-0.142***	-0.168***	-0.188***	-0.0630***	-0.0714***	-0.0436***	-0.142***	-0.168***	-0.191***	-0.0656***	-0.0702***
	(-13.13)	(-27.40)	(-11.23)	(-7.02)	(-5.18)	(-5.98)	(-12.96)	(-27.40)	(-11.39)	(-7.38)	(-5.76)	(-7.75)
ln*School*	0.0510***	0.0628***	0.0965***	0.0398	0.0793***	0.0984***	0.0505***	0.0642***	0.0929***	0.0307	0.0783***	0.0959***
	(8.60)	(4.74)	(4.19)	(1.33)	(3.85)	(4.94)	(8.51)	(4.87)	(4.69)	(1.47)	(4.26)	(6.82)
ln*(n+g+d)*	-0.0184***	-0.0086	0.0053	0.0212	-0.0260***	-0.0212***	-0.0182***	-0.0089	0.0072	0.0098	-0.0197***	-0.0141***
	(-3.20)	(-0.93)	(0.14)	(0.37)	(-5.55)	(-4.10)	(-3.16)	(-0.96)	(0.19)	(0.19)	(-5.89)	(-3.84)
ln*SMER*	-0.0081***	-0.0089**	-0.0074	0.0050	0.0599**	0.0441*	-0.0077***	-0.0110***	-0.0364*	-0.0425*	0.0307	0.0110
	(-3.76)	(-2.41)	(-0.28)	(0.15)	(2.08)	(1.66)	(-3.11)	(-2.87)	(-1.66)	(-1.71)	(1.33)	(0.42)
*Dummy*1995	-0.0672***	-0.0564***	-0.0656***	-0.0402***	-0.0806***	-0.0849***	-0.0671***	-0.0568***	-0.0624***	-0.0344***	-0.0776***	-0.0822***
	(-15.82)	(-9.04)	(-6.71)	(-3.01)	(-9.91)	(-10.48)	(-15.79)	(-9.16)	(-7.91)	(-3.79)	(-11.92)	(-13.27)
Dummy 2000	0.0086*	-0.0072	-0.0286*	0.0027	-0.0199	-0.0318**	0.0087*	-0.0078	-0.0230*	0.0126	-0.0146	-0.0246**
	(1.65)	(-0.87)	(-1.79)	(0.13)	(-1.34)	(-2.06)	(1.68)	(-0.95)	(-1.77)	(0.94)	(-1.14)	(-2.21)
Dummy 2005	-0.0403***	-0.0293***	-0.0478***	0.0003	-0.0624***	-0.0750***	-0.0402***	-0.0304***	-0.0440***	0.0071	-0.0583***	-0.0699***
	(-7.58)	(-2.92)	(-2.79)	(0.01)	(-4.38)	(-5.57)	(-7.55)	(-3.05)	(-3.13)	(0.45)	(-4.80)	(-7.25)
Observations	2032	2032	1524	1524	2032	2032	2032	2032	1524	1524	2032	2032
Adjusted R^2-	0.294	0.348					0.293	0.349				
Instruments			18	18	27	27			18	18	27	27
AR1			0.0000	0.0947	0.0000	0.0000			0.0000	0.0790	0.0000	0.0000
AR2			0.0001	0.0587	0.0106	0.0044			0.0014	0.1850	0.0040	0.0023
Hansen			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.000	0.000					0.000	0.000

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

Table 4.3: Industry, Commerce and Services Employment Share and Growth

	SME250						SME500					
	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step
lnGDP_{t-1}	-0.0455***	-0.143***	-0.171***	-0.195***	-0.0682***	-0.0716***	-0.0445***	-0.143***	-0.169***	-0.194***	-0.0681***	-0.0724***
	(-13.31)	(-27.69)	(-11.59)	(-7.45)	(-5.94)	(-8.85)	(-13.14)	(-27.67)	(-11.14)	(-8.04)	(-5.83)	(-8.39)
ln*School*	0.0528***	0.0589***	0.0700***	0.0214	0.0860***	0.102***	0.0521***	0.0613***	0.0763***	0.0119	0.0908***	0.105***
	(8.84)	(4.46)	(3.00)	(0.78)	(4.59)	(8.19)	(8.73)	(4.67)	(3.54)	(0.52)	(4.86)	(8.00)
ln*(n+g+d)*	-0.0205***	-0.0106	0.0069	0.0292	-0.0723***	-0.0555**	-0.0196***	-0.0110	0.0044	0.0165	-0.0923***	-0.0635***
	(-3.55)	(-1.15)	(0.18)	(0.55)	(-3.44)	(-2.54)	(-3.39)	(-1.20)	(0.11)	(0.31)	(-3.82)	(-2.81)
ln*SMER*	-0.0182***	-0.0292***	-0.0853**	-0.0558	-0.0369***	-0.0282**	-0.0182***	-0.0346***	-0.105***	-0.115*	-0.0494***	-0.0330**
	(-4.28)	(-4.41)	(-2.47)	(-1.04)	(-2.70)	(-1.98)	(-3.79)	(-4.85)	(-2.70)	(-1.95)	(-3.21)	(-2.28)
*Dummy*1995	-0.0674***	-0.0539***	-0.0510***	-0.0286**	-0.0771***	-0.0835***	-0.0675***	-0.0550***	-0.0539***	-0.0244**	-0.0792***	-0.0851***
	(-15.90)	(-8.63)	(-5.24)	(-2.24)	(-11.85)	(-15.09)	(-15.89)	(-8.88)	(-6.20)	(-2.26)	(-11.95)	(-14.96)
Dummy 2000	0.0079	-0.0037	-0.0074	0.0161	-0.0108	-0.0240**	0.0080	-0.0053	-0.0112	0.0233	-0.0132	-0.0257**
	(1.52)	(-0.45)	(-0.47)	(0.84)	(-0.83)	(-2.37)	(1.54)	(-0.64)	(-0.78)	(1.52)	(-0.99)	(-2.44)
Dummy 2005	-0.0408***	-0.0250**	-0.0227	0.0179	-0.0616***	-0.0741***	-0.0407***	-0.0271***	-0.0284*	0.0249	-0.0662***	-0.0771***
	(-7.67)	(-2.49)	(-1.32)	(0.86)	(-4.89)	(-8.30)	(-7.65)	(-2.72)	(-1.84)	(1.44)	(-5.20)	(-8.17)
Observations	2032	2032	1524	1524	2032	2032	2032	2032	1524	1524	2032	2032
Adjusted R^2	0.296	0.354					0.294	0.356				
Instruments			18	18	27	27			18	18	27	27
AR1			0.0000	0.0864	0.0000	0.0000			0.0000	0.0286	0.0000	0.0000
AR2			0.1650	0.5000	0.0162	0.0066			0.0425	0.7200	0.0187	0.0063
Hansen			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.000	0.000					0.000	0.000

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

4.4.2 SME and Growth: Extending the Model with SMEs Human Capital

The results for our extended specification with *SMEH* for manufacturing SMEs are shown in Table 4.4. When we introduce *SMEH*, the OLS results in the first column of each set of size criterion provides the same interpretation as before, indicating the presence of convergence among Brazilian micro-regions, a positive effect of the overall human capital on growth, and a negative effect of population growth on economic performance. The SME sector size represented by *SMER* remains negative and significant. Additionally, our new variable, *SMEH* does not influence growth and is not significant. When we include the specific individual effect through the use of the LSDV regressions in the second column of each set of regressions, we also observe the same pattern when comparing results with Table 4.2. The results indicate convergence, positive and significant coefficients for *School*, and a negative and not significant coefficient for $\ln(n+s+g)$. The coefficient for the SMEs' employment share enters negatively and is significant at five percent level (using the cut off of 250 employees).[52] Including the specific regional effect, the coefficient of *SMEH* is positive for all size criteria and is significant at one percent level for the cut off of 250 employees. Are these results provided by the LSDV estimations an indication that the qualitative aspect of SMEs is important to foster economic growth? Does this imply that the policy makers should be worried not only in promoting entrepreneurship but in promoting qualitative entrepreneurship? Nevertheless, as discussed earlier in this chapter, the LSDV estimations are not reliable to make a robust inference once these methods ignore the problem of endogeneity.

The GMM-DIFF estimations in columns 3 and 4 indicate that the inclusion of *SMEH* provide the same qualitative results found in Table 4.2 for the other control variables and support the view of the neutral or negative impact of *SMER* on growth. The coefficients for *SMER* are always insignificant statistically when using 250 employees as a cut off and negatively significant using 500 employees as a cut off, reinforcing the idea that the presence of SMEs themselves do not promote growth. On the other hand *SMEH* enters positively and is significant for the one step GMM-

[52] The same pattern was observed for this variable in Table 4.2.

DIFF estimates and is a sign that this aspect of SMEs could be important for growth. The human capital aspect of SMEs could be important to their absorptive capacity, making easier to adopt new technologies from leading companies in the country or elsewhere. It contributes to the process of technological diffusion, increases productivity, and therefore promotes economic growth. In columns five and six we find that the system GMM coefficient lies between the OLS and LSDV estimates, a sign that the size of the convergence coefficients could be less biased. The catching up effect captured by the convergence coefficient, the positive impact of human capital and the negative impact of population growth is confirmed again. The impact of the *SME* sector size in terms of employment is in line with the GMM-DIFF results and all coefficients are not statistically significant and do not support the view that the *SME* sector size promotes growth. Results for *SMEH* using this estimator do not indicate in any regression a possible positive effect of this variable on growth. The Sargan and Hansen tests again always reject the null that instruments are valid and for all GMM-SYS estimates. Moreover, Difference-in-Hansen test always rejects the null and suggests that the additional set of instruments for the specification in levels are not independent of the error term and are an additional source of endogeneity, suggesting the use of the GMM-DIFF. Considering the GMM-DIFF estimates as more appropriate to analyse causal relationships, the results are additional evidence that a larger SME sector is not a sign of faster growth in the Brazilian micro-regions and that human capital embodied in SMEs seems to influence growth positively when we use 250 employees as a cut off to define SMEs.

Table 4.5 provides the results for our extended model for the expanded SME sector (considering SMEs in the manufacturing, commerce and service sectors). The results are similar to the previous results provided in Table 4.4 with respect to the specification tests. We could not rule out endogeneity and the specification in levels is source of additional endogeneity in the system. For our preferred GMM-DIFF estimates the inclusion of an SME extended sector confirms the negative effect of SME sector size. Conversely, SMEs' human capital does not suggest a positive impact on growth in none of the regressions when we include commerce and services. These results confirm the findings from the previous tables regarding the presence of SMEs, that is, a negative effect stemming from the size of SME sector.

Additionally, an incipient effect of the *SMEH2* (SMEs' human capital for manufacturing, commerce and services) on economic growth is not observed. These results can also indicate that human capital is more important for the SME manufacturing sector than for the SME sector including commerce and services.

The lack of positive effect of *SMER* on growth could be explained by the presence of institutional failures that prevent SMEs to develop (Beck et al. 2005a). The financial constraint, for example, is one of the institutional failures that prevent SMEs to reach their optimal size and generate economic growth (Beck et al. 2008). Institutional failures can give incentives to rent seeking activities and incites unproductive entrepreneurship (Baumol 1990, 2008). This is another channel that can also help to explain the lack of positive effect of *SMER* on growth. Therefore, the presence of SMEs *per se* is not linked to economic growth.

Additionally, *SMEH* show an incipient positive impact on growth suggesting that SMEs' human capital participate in the process of productivity growth (absorptive and innovative capacity) in the SME manufacturing sector. Furthermore, *SMEH* could have presented a negative effect on growth for the whole of the SME sector (manufacturing, commerce and services) because it does not work without institutional improvement. Dias and McDermott (2006) argue that human capital formation alone does not guarantee a positive impact on growth through entrepreneurship because this additional human capital can be used in unproductive rent-seeking entrepreneurial activity and will have little long-run impact on growth. Conversely, if there are better institutions, more productive entrepreneurs will emerge and the accumulation of human capital by workers will be driven by the demand side willing to fulfil better paid (productive) position made available by the entrepreneurs. Therefore, institutional failures in the developing world, alongside with cheap and unskilled labour, are likely to promote unproductive entrepreneurship and neutralize the effect of human capital in SMEs in the commerce and services sectors.[53]

[53] It is not difficult to assume that manufacturing SME sector is less prone to rent-seeking (unproductive) entrepreneurship due to the very nature of this activity. Therefore, institutional failures of this type are more likely to be observed in the commerce and services sectors.

Therefore, this section presented the first set of results to analyse the relationship between growth and SMEs across Brazilian micro-regions. Noticeably, our results suffer from endogeneity and we cannot rely on them to make inference or give responsible advice to policy makers. The endogeneity problem is recurrent in the growth literature and sometimes is a difficult problem to solve. In our particular case, this endogeneity can be caused by omitted variables such as the stock or level of physical capital, oil producing regions, institutional quality and from spatial dependence, once development and growth follow a regional pattern in Brazil. In the following sections we try to address these issues to complement the initial analysis presented in this section.

Table 4.4: Industry Employment Share, *SMEH* and Growth

	SME250						SME500					
	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step
lnGDP_{t-1}	-0.0444***	-0.142***	-0.170***	-0.198***	-0.0661***	-0.0666***	-0.0436***	-0.142***	-0.169***	-0.195***	-0.0661***	-0.0720***
	(-13.12)	(-27.47)	(-10.97)	(-7.26)	(-5.72)	(-7.08)	(-12.97)	(-27.43)	(-11.63)	(-7.70)	(-5.91)	(-8.51)
ln*School*	0.0485***	0.0535***	0.0356	0.0197	0.0962***	0.103***	0.0473***	0.0566***	0.0696**	0.0442	0.0985***	0.137***
	(7.55)	(3.91)	(1.04)	(0.55)	(3.23)	(3.20)	(7.39)	(4.16)	(2.22)	(1.36)	(3.15)	(3.76)
ln$(n+g+d)$	-0.0175***	-0.0081	0.0051	0.0355	-0.0347*	-0.0268	-0.0169***	-0.0087	0.0052	0.0126	-0.0344*	-0.0444
	(-2.99)	(-0.88)	(0.14)	(0.64)	(-1.95)	(-1.15)	(-2.90)	(-0.94)	(0.14)	(0.26)	(-1.78)	(-1.51)
ln*SMER*	-0.0081***	-0.0090**	0.0012	-0.0134	0.0255	0.0341	-0.0077***	-0.0110***	-0.0404*	-0.0496*	-0.0006	-0.0227
	(-3.76)	(-2.45)	(0.04)	(-0.36)	(1.14)	(1.20)	(-3.12)	(-2.88)	(-1.84)	(-1.89)	(-0.03)	(-1.09)
ln*SMEH*	0.0056	0.0186***	0.115**	0.0703	-0.0364	-0.0219	0.0073	0.0157**	0.0405	-0.0104	-0.0464	-0.0875
	(1.00)	(2.59)	(2.41)	(1.19)	(-0.77)	(-0.35)	(1.31)	(2.25)	(0.90)	(-0.18)	(-0.90)	(-1.16)
*Dummy*1995	-0.0675***	-0.0569***	-0.0683***	-0.0422***	-0.0763***	-0.0821***	-0.0674***	-0.0573***	-0.0618***	-0.0351***	-0.0736***	-0.0783***
	(-15.86)	(-9.14)	(-6.59)	(-2.92)	(-11.18)	(-11.03)	(-15.85)	(-9.24)	(-7.81)	(-3.86)	(-11.74)	(-11.49)
Dummy 2000	0.0077	-0.0092	-0.0412**	-0.0126	-0.0101	-0.0252	0.0076	-0.0097	-0.0255*	0.0088	-0.0035	-0.0109
	(1.47)	(-1.11)	(-2.36)	(-0.48)	(-0.72)	(-1.63)	(1.45)	(-1.18)	(-1.84)	(0.53)	(-0.26)	(-0.70)
Dummy 2005	-0.0416***	-0.0331***	-0.0693***	-0.0198	-0.0516***	-0.0660***	-0.0420***	-0.0339***	-0.0501***	0.0045	-0.0458***	-0.0505**
	(-7.59)	(-3.27)	(-3.45)	(-0.67)	(-3.29)	(-3.62)	(-7.64)	(-3.36)	(-3.07)	(0.22)	(-2.95)	(-2.55)
Observations	2032	2032	1524	1524	2032	2032	2032	2032	1524	1524	2032	2032
Adjusted R^2	0.294	0.351					0.293	0.351				
Instruments			21	21	32	32			21	21	32	32
AR1			0.0000	0.0454	0.0000	0.0000			0.0000	0.0895	0.0000	0.0000
AR2			0.0010	0.0630	0.0093	0.0059			0.0009	0.2250	0.0186	0.0509
Hansen			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.000	0.000					0.000	0.000

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

Table 4.5: Industry, Commerce and Services Employment Share, *SMEH* and Growth

	SME250						SME500					
	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step
lnGDP_{t-1}	-0.0473***	-0.144***	-0.173***	-0.192***	-0.0629***	-0.0690***	-0.0462***	-0.143***	-0.174***	-0.202***	-0.0645***	-0.0696***
	(-13.65)	(-27.69)	(-11.84)	(-7.69)	(-5.63)	(-8.01)	(-13.41)	(-27.67)	(-11.84)	(-8.14)	(-5.98)	(-8.17)
ln*School*	0.0582***	0.0596***	0.0705***	0.0133	0.0838***	0.0973***	0.0572***	0.0625***	0.0837***	0.0222	0.0865***	0.101***
	(9.34)	(4.48)	(2.83)	(0.51)	(4.61)	(7.86)	(9.12)	(4.72)	(3.64)	(0.93)	(4.73)	(8.00)
ln*(n+g+d)*	-0.0240***	-0.0103	0.0161	-0.0022	-0.0715***	-0.0420**	-0.0225***	-0.0107	0.0218	0.0306	-0.0773***	-0.0510**
	(-4.07)	(-1.12)	(0.39)	(-0.04)	(-2.83)	(-1.96)	(-3.84)	(-1.17)	(0.53)	(0.48)	(-2.94)	(-2.25)
ln*SMER*	-0.0178***	-0.0290***	-0.0582*	-0.0164	-0.0129	-0.0200	-0.0180***	-0.0343***	-0.0803**	-0.0761	-0.0308	-0.0275
	(-4.18)	(-4.39)	(-1.71)	(-0.33)	(-0.66)	(-1.01)	(-3.74)	(-4.80)	(-2.16)	(-1.17)	(-1.53)	(-1.47)
ln*SMEH*	-0.0293***	0.0071	-0.0491	-0.107	-0.0560	0.0012	-0.0241***	0.0119	0.0013	-0.0123	-0.0222	0.0058
	(-2.97)	(0.44)	(-0.72)	(-1.33)	(-0.97)	(0.03)	(-2.60)	(0.76)	(0.02)	(-0.14)	(-0.44)	(0.14)
*Dummy*1995	-0.0675***	-0.0546***	-0.0488***	-0.0222*	-0.0747***	-0.0817***	-0.0676***	-0.0563***	-0.0568***	-0.0273**	-0.0770***	-0.0837***
	(-15.95)	(-8.47)	(-4.10)	(-1.84)	(-11.43)	(-13.28)	(-15.94)	(-8.77)	(-5.09)	(-2.39)	(-11.80)	(-14.10)
Dummy 2000	0.0090*	-0.0052	-0.0028	0.0325	-0.0030	-0.0231*	0.0089*	-0.0079	-0.0176	0.0159	-0.0083	-0.0260**
	(1.74)	(-0.58)	(-0.13)	(1.52)	(-0.22)	(-1.87)	(1.71)	(-0.89)	(-0.89)	(0.69)	(-0.62)	(-2.19)
Dummy 2005	-0.0363***	-0.0274**	-0.0102	0.0478*	-0.0472***	-0.0707***	-0.0371***	-0.0314***	-0.0333	0.0220	-0.0574***	-0.0756***
	(-6.59)	(-2.38)	(-0.35)	(1.68)	(-2.77)	(-4.46)	(-6.74)	(-2.74)	(-1.21)	(0.67)	(-3.55)	(-5.26)
Observations	2032	2032	1524	1524	2032	2032	2032	2032	1524	1524	2032	2032
Adjusted R^2	0.298	0.354					0.296	0.356				
Instruments			21	21	32	32			21	21	32	32
AR1			0.0000	0.1180	0.0000	0.0000			0.0000	0.0985	0.0000	0.0000
AR2			0.0307	0.4960	0.0137	0.0053			0.0070	0.5290	0.0149	0.0053
Hansen			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.0000	0.0000					0.0000	0.0000

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

4.5 SMEs and Growth in Two Different Regimes

The type of influence SMEs exert on economic growth that we observe when all Brazilian micro-regions are pooled together might present different results if we control for the level of development of those micro-regions. Van Stel et al. (2005) and Wennekers et al. (2005), for example, provide evidence of a different interactive dynamics between entrepreneurship and level of economic development. They argue that entrepreneurship has a negative effect for relatively poor economies and a positive effect for relatively rich economies. For that reason, it is reasonable to look for different dynamics in groups of regions with dissimilar levels of development, especially when we observe a high degree of dissimilarities in the sample. When we split the sample, we restrict the analysis to sets of more similar economies and implicitly control for the determinants of their steady-state such as institutional quality and other unobserved regional characteristics that might be important for growth.

Brazil is a country with considerable cross regional asymmetries, as reported by Ferreira (2000), Laurine et al. (2005), among others.[54] Therefore, pooling all micro-regions together could make it more difficult to draw useful inferences for public policy regarding SMEs in Brazil. Moreover, pooling regions in the same sample ignores the dynamics of the distribution of GDP per capita across regions. Quah (1996, 1997a, 1997b) observes that the traditional analysis based on the standard convergence equation says nothing about the dynamics of the distribution of GDP per capita and suggests the analysis of this distribution to identify different dynamics across economies. Laurini et al. (2005) follow Quah's analysis and provide evidence of two different regimes in Brazil stemming from the existence of two GDP per capita convergence clubs, a lower GDP per capita club formed mainly by municipalities of Northeast and North regions, and a richer club formed mainly by the municipalities of South, South east and Centre-west regions.

[54] Table 4.20 in the Appendix 4.8.3 illustrates these asymmetries across states that are even more pronounced at micro-regional level. A map with the states and macroregions is provided in Chapter 3.

Moreover, Acemoglu et al. (2001) and Rodrik et al. (2004), for instance, provide empirical evidence that GDP per capita and the quality of institutions are positively related. For the Brazilian case, Menezes-Filho et al. (2006) draw on Acemoglu et al. (2001) and provide empirical evidence that institutions are important determinants of GDP per capita of Brazilian states. Furthermore, Naritomi et al**.** (2009) investigate the long-run determinants of local institutions in Brazil and document that income per capita is positively related with institutional development. In this sense, we expect that the two sub-samples derived from the GDP per capita distribution, implicitly control for different levels of institutional quality. Summing up, the purpose of such division is twofold; to detect different convergence processes in the two groups and find out how SMEs variables contribute to the economic growth process in regions with different levels of development.

Figure 4.1. Kernel Densities for the Relative GDP per capita

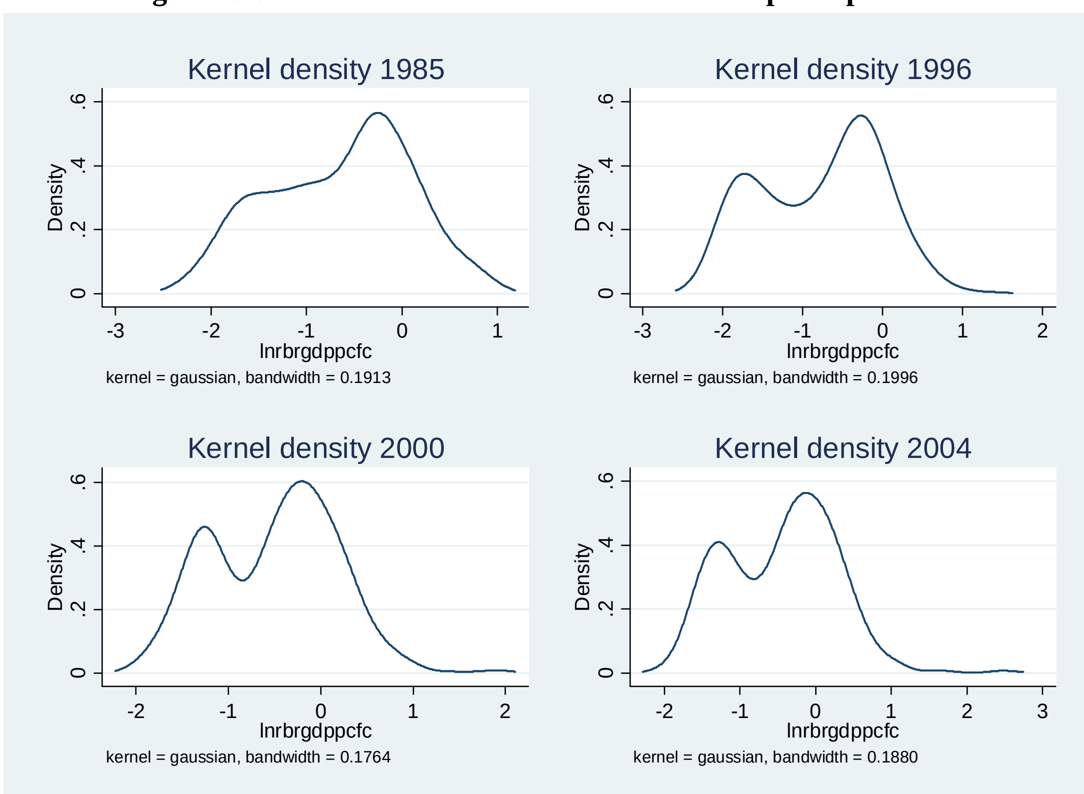


Figure 4.1 plots the Gaussian kernel densities of the natural logarithm of relative GDP per capita for each cross-section available in our panel.[55] Relative GDP is constructed by dividing the value of micro-regional per capita GDP by the mean per capita GDP for all micro-regions in the same year, and finally taking the natural logarithm of this value. In this normalization process, a zero value on the horizontal axis indicates per capita GDP equal to the national mean. This figure confirms the results of Laurini et al. (2005) for Brazilian micro-regions. It shows the poorest club encompassing mainly Northeast and North micro-regions and the richest club grouping mainly South, Southeast and Centre-west regions. From 1985 to 2004 we can observe that the polarization became more pronounced, suggesting the formation and consolidation of a bimodal distribution. A separate analysis for each group could contribute to deepen our analysis about SMEs and growth. For 2004, the lower and upper peaks of the distribution correspond to -1.2813 and -0.1218 of the logarithm of relative GDP (respective per capita GDP of 0.2776 and 0.8853 times the Brazilian mean for that year). While the upper peak is fairly close to the national GDP per capita, the lower one is just around a quarter of this value, illustrating the great Brazilian regional inequality. We can also observe that the lowest point in the valley that separates the clubs correspond to -0.8155 (respective per capita GDP 0.44 times the national mean for that year). We use this point as a criterion to divide our initial sample and to control for different growth dynamics. Using this criterion we end up with 149 micro-regions in the lower peak and 359 micro-regions in the upper one.

4.5.1 SMEs in Poor Regions

After splitting the initial sample, we perform an identical analysis and estimate equation 4.1 separately to each group of regions to see whether there is any different pattern regarding the role of SMEs in each group. Table 4.6 reports the results for the poorest income group.

---

[55] The Kernel smoothed estimations in Figure 4.1 are obtained using a Gaussian Kernel following the procedure of getting an automatic bandwidth choice suggested by Silverman (1986). The automatic bandwidth is calculated performing the command "density(x, bw.nrd())"using the software *R*. The Figures themselves are generated by STATA using the optimal bandwidth provided by *R*.

One distinctive difference for the group of poorest income regions is that the specification tests are more supportive to the validity of instruments. The second-order autocorrelation (AR2) results do not reject the null of no autocorrelation for all GMM estimates. Another difference is that now the Difference-in-Hansen test always supports the inclusion of the subset of additional instruments originated by the GMM-SYS estimator, suggesting that this estimator performs a bit better than the GMM-DIFF for this set of regions. Furthermore, the Hansen test performs better but still does not accept the null of no endogeneity. However, it is clear that this set of regressions is less subject to endogeneity than the results for the country as whole. Probably, this is because the club of poor micro-regions in more homogenous than the country as a whole. If this is the case, the grouping of these regions can better control for club convergence and for omitted variables that might be a source of endogeneity (e.g. physical capital).[56] Since we split our initial sample into two, it is more plausible now to assume that the cross-micro-regional specific differences are not correlated with the initial distances from the steady-state, and therefore, making the additional moment conditions imposed by the GMM-SYS more realistic. When we restrict the analysis to sets of economies where differences are smaller, we implicitly control for their steady-state (Sala-i-Martin, 1996), and make the additional assumptions made by the GMM-SYS more realistic.

The results for the 149 poorest micro-regions presented in Table 4.6 support the results found in Table 4.2. For all regressions, the system GMM estimations suggest convergence among the poorest regions in Brazil, a positive effect of education on growth and a negative effect coming from the population growth, all significant at least at five percent level.

The results for the size of SME sector remain negative. More SMEs is always a sign of slower growth rates for poor regions. We also present the same table for the extended SME sector, considering commerce and service along with manufacturing SMEs (Table 4.7). The inclusion of commerce and services also support the view that SMEs do not foster economic growth in the poorest regions of Brazil.

[56] For instance, if the omitted physical capital is relatively more homogeneous in the poor club, the omission of this variable is less likely to be a source of endogeneity.

When we augment our model to accommodate the SMEs' human capital the analysis remain unchanged, with a negative and significant effect of the size of SME sector on growth when using the system GMM (Table 4.8). The role of *SMEH* is unclear and the GMM-SYS does not provide a conclusive result after controlling for endogeneity, providing a positive but not statistically significant result for *SMEH*. Similar analyses apply to the extended SME sector, with the inclusion of commerce and services confirming that the SME sector is not important for growth (Table 4.9).

Hence, the evidence drawn from the poor micro-regions supports the results we find in our initial set of regressions encompassing 508 micro-regions for the role of the convergence coefficient, overall human capital, population growth and the presence of SMEs. The difference now is that we do not find any incipient positive role of *SMEH* on growth. Convergence is found for all set of regressions, human capital influences growth positively and population growth reduces economic performance. Additionally, SME sector size has no positive effect on economic growth and SMEs' human capital seems to be underemployed, with skilled and non-skilled workers performing the same type of work or engaged in unproductive rent-seeking entrepreneurial activities encouraged by institutional failures. For this set of regions, *SMEH* is not using fully its absorptive and innovative capacity and are not contributing to the process of technological diffusion and productivity growth.

Table 4.6: Industry Sectors Employment Share and Growth

	SME250						SME500					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	LSDV	DiffGMM	DiffGMM	SysGMM	SysGMM	OLS	LSDV	DiffGMM	DiffGMM	SysGMM	SysGMM
			1-step	2-step	1-step	2-step			1-step	2-step	1-step	2-step
lnGDP_{t-1}	-0.0843***	-0.149***	-0.0979***	-0.0954***	-0.0851***	-0.0769***	-0.0832***	-0.149***	-0.0983***	-0.0974***	-0.0849***	-0.0752***
	(-14.98)	(-19.01)	(-4.94)	(-6.53)	(-3.42)	(-7.51)	(-14.95)	(-18.92)	(-4.76)	(-5.58)	(-3.52)	(-6.52)
ln*School*	0.0353***	0.0772***	0.0698	0.0988*	0.0347**	0.0353***	0.0351***	0.0785***	0.0765	0.111**	0.0364***	0.0341***
	(4.34)	(2.99)	(0.83)	(1.88)	(2.46)	(2.96)	(4.30)	(3.05)	(0.94)	(2.46)	(2.76)	(2.85)
ln$(n+g+d)$	-0.0133	-0.00783	0.00704	-0.0237	-0.0146***	-0.0139***	-0.0127	-0.00765	0.00642	-0.0355	-0.0150***	-0.0141***
	(-1.40)	(-0.62)	(0.18)	(-0.54)	(-6.87)	(-6.85)	(-1.34)	(-0.60)	(0.16)	(-0.71)	(-7.42)	(-6.93)
ln*SMER*	-0.00574**	-0.00920*	-0.0168	-0.0127	-0.0187**	-0.0247**	-0.00556**	-0.00805*	-0.0237	-0.0178	-0.0190**	-0.0261**
	(-2.49)	(-1.93)	(-0.62)	(-0.42)	(-1.99)	(-2.16)	(-2.18)	(-1.73)	(-1.51)	(-0.80)	(-2.30)	(-2.11)
*Dummy*1995	-0.0656***	-0.0777***	-0.0811*	-0.0992***	-0.0662***	-0.0703***	-0.0654***	-0.0780***	-0.0836*	-0.104***	-0.0663***	-0.0684***
	(-10.14)	(-5.56)	(-1.73)	(-3.74)	(-7.04)	(-9.31)	(-10.09)	(-5.59)	(-1.83)	(-4.42)	(-7.19)	(-9.35)
Dummy 2000	0.0243***	-0.0186	-0.00280	-0.0231	0.0247*	0.0279**	0.0248***	-0.0194	-0.00667	-0.0315	0.0239*	0.0303**
	(2.96)	(-0.94)	(-0.04)	(-0.57)	(1.84)	(2.23)	(3.02)	(-0.98)	(-0.11)	(-0.87)	(1.90)	(2.51)
Dummy 2005	-0.0432***	-0.0707***	-0.0726	-0.101**	-0.0436***	-0.0453***	-0.0428***	-0.0716***	-0.0775	-0.111***	-0.0442***	-0.0426***
	(-4.75)	(-2.82)	(-0.86)	(-2.06)	(-3.05)	(-3.64)	(-4.70)	(-2.86)	(-0.94)	(-2.60)	(-3.14)	(-3.50)
Observations	596	596	447	447	596	596	596	596	447	447	596	596
Adjusted R^2	0.597	0.622					0.596	0.621				
Instruments			18	18	27	27			18	18	27	27
AR1			0.0001	0.0000	0.0003	0.0000			0.0000	0.0000	0.0002	0.0000
AR2			0.7340	0.5920	0.8340	0.8970			0.2930	0.2870	0.4110	0.3650
Hansen			0.0025	0.0025	0.0175	0.0175			0.0011	0.0011	0.0180	0.0180
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.277	0.277					0.487	0.487

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

Table 4.7: Industry, Commerce and Services Employment Share and Growth

	SME250						SME500					
	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step
lnGDP_{t-1}	-0.0863***	-0.151***	-0.102***	-0.0957***	-0.0771***	-0.0716***	-0.0847***	-0.149***	-0.102***	-0.0971***	-0.0757***	-0.0706***
	(-14.98)	(-19.22)	(-5.50)	(-6.48)	(-3.05)	(-5.32)	(-15.02)	(-19.06)	(-5.21)	(-6.24)	(-3.02)	(-5.14)
ln*School*	0.0380***	0.0771***	0.0584	0.0930**	0.0279**	0.0311***	0.0371***	0.0810***	0.0870	0.116***	0.0287**	0.0312**
	(4.61)	(3.00)	(0.69)	(2.02)	(2.04)	(2.65)	(4.51)	(3.15)	(1.07)	(2.83)	(2.16)	(2.52)
ln*(n+g+d)*	-0.0131	-0.00742	0.00461	-0.0480	-0.0438**	-0.0708**	-0.0124	-0.00814	0.00508	-0.0491	-0.0375*	-0.0588*
	(-1.38)	(-0.59)	(0.12)	(-1.02)	(-2.16)	(-2.47)	(-1.30)	(-0.64)	(0.12)	(-0.87)	(-1.82)	(-1.69)
ln*SMER*	-0.0135***	-0.0203***	-0.0400	-0.0489	-0.0177	-0.0357**	-0.0134***	-0.0189**	-0.0499	-0.0594	-0.0136	-0.0278
	(-2.89)	(-2.72)	(-1.20)	(-1.19)	(-1.35)	(-1.98)	(-2.65)	(-2.36)	(-1.35)	(-0.97)	(-1.05)	(-1.29)
*Dummy*1995	-0.0662***	-0.0766***	-0.0732	-0.0953***	-0.0629***	-0.0664***	-0.0661***	-0.0790***	-0.0881*	-0.108***	-0.0635***	-0.0668***
	(-10.23)	(-5.50)	(-1.56)	(-3.87)	(-6.41)	(-8.70)	(-10.20)	(-5.67)	(-1.93)	(-4.97)	(-6.91)	(-8.46)
Dummy 2000	0.0223***	-0.0183	0.00597	-0.0170	0.0318**	0.0356***	0.0230***	-0.0214	-0.0157	-0.0369	0.0310**	0.0343***
	(2.70)	(-0.93)	(0.10)	(-0.47)	(2.57)	(3.00)	(2.79)	(-1.08)	(-0.26)	(-1.12)	(2.54)	(2.78)
Dummy 2005	-0.0448***	-0.0694***	-0.0592	-0.0954**	-0.0381***	-0.0403***	-0.0441***	-0.0733***	-0.0858	-0.117***	-0.0388***	-0.0396***
	(-4.91)	(-2.78)	(-0.70)	(-2.14)	(-2.59)	(-3.21)	(-4.83)	(-2.93)	(-1.04)	(-2.93)	(-2.83)	(-3.07)
Observations	596	596	447	447	596	596	596	596	447	447	596	596
Adjusted R^2	0.598	0.625					0.597	0.623				
Instruments			18	18	27	27			18	18	27	27
AR1			0.0001	0.0000	0.0002	0.0000			0.0000	0.0000	0.0002	0.0000
AR2			0.7770	0.7050	0.7350	0.8120			0.5820	0.4670	0.6690	0.6900
Hansen			0.0021	0.0021	0.0080	0.0080			0.0018	0.0018	0.0048	0.0048
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.281	0.281					0.260	0.260

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

Table 4.8: Industry Employment Share, SMEH and Growth

	SME250						SME500					
	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step
$\ln GDP_{t-1}$	-0.0845***	-0.150***	-0.102***	-0.103***	-0.0879***	-0.0773***	-0.0834***	-0.149***	-0.101***	-0.105***	-0.0865***	-0.0755***
	(-15.00)	(-19.06)	(-6.20)	(-6.72)	(-3.62)	(-6.67)	(-14.98)	(-18.94)	(-5.41)	(-6.16)	(-3.53)	(-5.68)
ln*School*	0.0373***	0.0765***	0.0663	0.101	0.0283	0.0297*	0.0373***	0.0783***	0.0733	0.110**	0.0256	0.0276
	(4.42)	(2.97)	(0.81)	(1.60)	(1.32)	(1.77)	(4.43)	(3.04)	(0.94)	(2.23)	(1.28)	(1.61)
$\ln(n+g+d)$	-0.0145	-0.00735	0.0237	-0.000876	-0.00559	-0.00643	-0.0141	-0.00736	0.0144	-0.00316	-0.00339	-0.00745
	(-1.51)	(-0.58)	(0.63)	(-0.02)	(-0.38)	(-0.54)	(-1.47)	(-0.58)	(0.39)	(-0.07)	(-0.27)	(-0.75)
ln*SMER*	-0.00565**	-0.00926*	-0.0127	-0.00678	-0.0219**	-0.0290***	-0.00534**	-0.00834*	-0.0237	-0.0198	-0.0215**	-0.0290**
	(-2.45)	(-1.94)	(-0.46)	(-0.21)	(-2.21)	(-2.76)	(-2.09)	(-1.79)	(-1.52)	(-0.79)	(-2.25)	(-2.21)
ln*SMEH*	-0.00457	0.00793	0.0366	0.000430	0.0225	0.0169	-0.00538	0.00581	0.0269	-0.0127	0.0298	0.0157
	(-0.89)	(1.22)	(0.66)	(0.01)	(0.64)	(0.60)	(-1.05)	(0.90)	(0.64)	(-0.39)	(0.98)	(0.65)
*Dummy*1995	-0.0653***	-0.0795***	-0.0888	-0.0972***	-0.0690***	-0.0710***	-0.0650***	-0.0796***	-0.0892*	-0.0964***	-0.0690***	-0.0678***
	(-10.06)	(-5.66)	(-1.57)	(-2.85)	(-6.26)	(-7.90)	(-10.01)	(-5.65)	(-1.70)	(-3.53)	(-6.62)	(-7.85)
*Dummy*2000	0.0250***	-0.0220	-0.0190	-0.0243	0.0183	0.0259*	0.0257***	-0.0221	-0.0180	-0.0242	0.0171	0.0300**
	(3.03)	(-1.10)	(-0.23)	(-0.46)	(1.22)	(1.92)	(3.12)	(-1.10)	(-0.25)	(-0.59)	(1.23)	(2.20)
*Dummy*2005	-0.0420***	-0.0754***	-0.0928	-0.0992	-0.0521**	-0.0497***	-0.0412***	-0.0754***	-0.0924	-0.0973*	-0.0539***	-0.0453***
	(-4.55)	(-2.98)	(-0.84)	(-1.52)	(-2.56)	(-3.04)	(-4.46)	(-2.96)	(-0.92)	(-1.90)	(-2.84)	(-2.91)
Observations	596	596	447	447	596	596	596	596	447	447	596	596
Adjusted R^2	0.597	0.622					0.596	0.621				
Instruments			21	21	32	32			21	21	32	32
AR1			0.0001	0.0000	0.0003	0.0000			0.0000	0.0000	0.0003	0.0000
AR2			0.3010	0.5070	0.5420	0.6770			0.1620	0.3870	0.2200	0.2450
Hansen			0.0047	0.0047	0.0356	0.0356			0.0011	0.0011	0.0390	0.0390
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.397	0.397					0.752	0.752

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

Table 4.9: Industry, Commerce and Services Employment Share, *SMEH* and Growth

	SME250						SME500					
	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step
$\ln GDP_{t-1}$	-0.0898***	-0.151***	-0.101***	-0.102***	-0.0786***	-0.0744***	-0.0885***	-0.150***	-0.100***	-0.0999***	-0.0793***	-0.0740***
	(-14.82)	(-19.18)	(-5.55)	(-5.95)	(-3.22)	(-5.71)	(-14.65)	(-18.95)	(-5.76)	(-5.64)	(-3.56)	(-5.59)
ln*School*	0.0424***	0.0774***	0.0666	0.109**	0.0301**	0.0337***	0.0414***	0.0808***	0.0840	0.121***	0.0303**	0.0357***
	(4.95)	(3.01)	(0.88)	(2.25)	(2.09)	(2.75)	(4.83)	(3.14)	(1.30)	(2.83)	(2.07)	(2.64)
$\ln(n+g+d)$	-0.0165*	-0.00816	0.0164	-0.0390	-0.0502*	-0.0861***	-0.0155	-0.00883	0.0143	-0.0318	-0.0450*	-0.0853***
	(-1.71)	(-0.64)	(0.41)	(-0.81)	(-1.95)	(-2.75)	(-1.61)	(-0.69)	(0.35)	(-0.63)	(-1.75)	(-2.75)
ln*SMER*	-0.0132***	-0.0208***	-0.0292	-0.0184	-0.00694	-0.0238	-0.0132***	-0.0193**	-0.0293	-0.00974	-0.00586	-0.0215
	(-2.84)	(-2.76)	(-1.01)	(-0.42)	(-0.55)	(-1.46)	(-2.63)	(-2.40)	(-0.97)	(-0.21)	(-0.43)	(-1.14)
ln*SMEH*	-0.0196*	-0.00874	0.0169	-0.0521	-0.0341	-0.0494**	-0.0171*	-0.00919	0.00560	-0.0495	-0.0292	-0.0533**
	(-1.83)	(-0.49)	(0.27)	(-1.14)	(-1.21)	(-2.11)	(-1.73)	(-0.54)	(0.08)	(-1.05)	(-0.81)	(-2.22)
*Dummy*1995	-0.0676***	-0.0765***	-0.0779*	-0.102***	-0.0635***	-0.0669***	-0.0672***	-0.0786***	-0.0871**	-0.109***	-0.0632***	-0.0683***
	(-10.40)	(-5.49)	(-1.86)	(-3.99)	(-6.10)	(-8.67)	(-10.34)	(-5.63)	(-2.48)	(-4.85)	(-6.21)	(-8.58)
Dummy 2000	0.0198**	-0.0179	-0.00187	-0.0287	0.0318***	0.0345***	0.0207**	-0.0206	-0.0143	-0.0385	0.0314**	0.0330**
	(2.38)	(-0.91)	(-0.04)	(-0.75)	(2.58)	(2.74)	(2.48)	(-1.04)	(-0.32)	(-1.11)	(2.57)	(2.44)
Dummy 005	-0.0444***	-0.0678***	-0.0707	-0.0987**	-0.0335**	-0.0343***	-0.0436***	-0.0709***	-0.0850	-0.111***	-0.0337***	-0.0347***
	(-4.87)	(-2.69)	(-1.01)	(-2.06)	(-2.51)	(-2.87)	(-4.79)	(-2.80)	(-1.52)	(-2.59)	(-2.64)	(-2.63)
Observations	596	596	447	447	596	596	596	596	447	447	596	596
Adjusted R^2	0.600	0.624					0.599	0.623				
Instruments			21	21	32	32			21	21	32	32
AR1			0.0000	0.0000	0.0002	0.0000			0.0000	0.0000	0.0001	0.0000
AR2			0.6080	0.8360	0.8660	0.9410			0.5210	0.8000	0.8690	0.8820
Hansen			0.0011	0.0011	0.0117	0.0117			0.0004	0.0004	0.0082	0.0082
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.506	0.506					0.536	0.536

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

4.5.2 SMEs in the Richest Micro-regions

The same procedure is applied to the richest income group and results are shown from Table 4.10 to 4.13 in order to provide a comparative analysis about the importance of SMEs for economic growth in different groups of Brazilian micro-regions. In Table 4.10, Hansen test indicates the existence of endogeneity and the Difference-in-Hansen test suggest the use of the GMM-DIFF. Comparing to the poorest club, the strongest similarity is the confirmation of highly significant convergence process, and the positive effect of the overall human capital on growth. Additionally, the GMM-DIFF estimates provide insignificant coefficients for *SMER* and suggest a neutral impact of the SME sector's size on growth for the richest set of micro-regions.

For the extended SME sector (including commerce and services) in Table 4.11, the Difference-in-Hansen test is on the threshold of the acceptance of the additional set of instruments provided by the GMM-SYS estimator when we consider a conventional level of significance of ten percent for the estimation using 250 employees as a cut off. Based on this estimator, the results are in line with the view that SME sector's size does not promote growth. However, the alternative results for the GMM-DIFF are also in line with this argument that *SMER* does not promote growth.

The augmented model with SME's human capital differs from the analyses we made for the poorest regions. The Hansen-in-Difference test suggests the use of the GMM-DIFF estimator and the results support the view that the SME sector size does not have a positive effect on growth (as in the case of the poorest club) but the analysis from the impact of *SMEH* is different. The one step GMM-DIFF (for the cut off of 250 employees) reported in Table 4.12 shows that *SMEH* enters positively and significantly, suggesting that this aspect of SMEs might play an important role in the economic growth process.

For the extended SME sector in Table 4.13, Difference-in-Hansen test suggests the use of the additional GMM-SYS moments conditions. We find that the SME sector size is not statistically significant and that SMEs' human capital is negative and significant. Hence, the results suggest that the SME human capital applied in manufacturing is more important for growth when comparing to SME sector as a whole; a result also found for the country as a whole in Table 4.5. Therefore, the regressions for this group of micro-regions suggest that the SME sector size does not promote economic growth. However, at first glance, *SMEH* now seems to play an incipient positive role in the growth process for the manufacturing sector, SMEs with more human capital are more productive and contribute for economic growth. This positive effect also found in Table 4.4 seems to be driven by the richest regions.

But how does *SMEH* come to influence growth in the richest regions? And why just the richest set of micro-regions present an incipient positive relationship between growth and *SMEH*?

This incipient positive role of *SMEH* on growth for the richest regions can be related with the quality of institutions in this group of micro-regions. The states located in Southeast, South and Centre-west of Brazil tend to have better institutions that provide better business environment and positive incentives to be productive entrepreneur at the expense of rent-seeking behaviour. Menezes-Filho et al. (2006), for example, draw on Acemoglu et al. (2001) and show the importance of institutional quality for economic performance in the Brazilian states. They suggest that institutional quality differ and are an important factor to explain economic performance across Brazilian regions. Better institutions provide positive incentives to productive entrepreneurship as suggested by Baumol (1990, 2008), and create the condition that incites more human capital formation that will be better used by productive entrepreneurs (Dias and McDermott, 2006). This effective use of human capital means that you can actually use it to increase your productivity making use of its potential absorptive capacity (Nelson and Phelps, 1966) and its innovative capacity (Griffith et al. 2004).

Additionally, institutions (and the level of human capital) are also important determinants of FDI decisions across countries (Globerman and Shapiro, 2002) and can also determine the locations of FDI within a country with disparities in the quality of economic institutions. Du, Lu and Tao (2008)'s results suggest that multinationals prefer to invest in Chinese regions that provide better institutional settings. FDI and multinationals bring technology from abroad that can create spillovers for the productive entrepreneurs that make full use of human capital in SMEs. In the case of Brazil, Poole (2007) suggests the existence of spillovers from multinationals through work mobility. Therefore, better institutions can create better conditions for human capital to thrive in the whole of the economy and in SMEs. Good institutions provide at least two channels that allow *SMEH* to thrive. First, they create the conditions necessary to make better use of this capital. Second, it can also attract more FDI to the region with better institutional settings that can create spillovers for the rest of the economy.

As in the regressions for the poorest regions, the tests of specification are more supportive to the validity of instruments when compared to the results for the country as a whole (as discussed earlier in Section 4.5.1, this is related to the fact that splitting the sample better controls for unobserved factors that might be a source of endogeneity). For most regressions, we accept the null of no autocorrelation of second order, and sometimes we also accept the null that the subsets of instruments coming from the equation in levels are valid. We have used the full set of available instruments for the estimates of GMM-DIFF and GMM-SYS, and although the test of specifications (specially the Hansen test) have failed to support the validity of instruments we believe that the estimates presented in this section ease the endogeneity problem (as we observe some improvements in our diagnostic tests) when we split the sample. In our specific panel, the data restriction and few cross-sections available do not provide us the degree of freedom to test many other sets of instruments.

Brazilian regions are very diverse and present huge degree of development inequality. The fact that part of the economy presents aspects of developed economies contributes to the differences in the role of SMEs in the economic growth process across regions. The results that indicate a negative (or a non positive) role of SMEs in the poorest regions and an incipient positive effect of SMEs' human capital in the richest regions are in line with Van Stel et al. (2005) that found evidence of a negative impact of entrepreneurship in poor countries and a positive effect for rich countries. Van Stel et al. (2005) speculate that the negative effect in poorer economies is due to the lower level of human capital of entrepreneurs in those areas.

In this case, it is likely that the negative effect of the SME presence reflects the presence of many “marginal” entrepreneurs that would be more productive if they worked for a larger and more productive firm. Nevertheless, they did not have a proxy for the entrepreneurs’ human capital to deepen their analysis.[57] This is a distinctive characteristic of our work and the results suggested that the human capital aspect of SMEs should not be ignored if one wants a full picture about the importance of the SME sector in one economy.

Different regions can have different institutional performances, and therefore provide different incentives for productive entrepreneurship (e.g. Baumol 1990, 2008; Dias and McDermott 2006) that makes *SMER* influence economic performance differently. For instance, more productive entrepreneurship incites more human capital formation through more workers willing to demand education to get a better wage. This leads to a more effective use of human capital and makes *SMEH* useful to capture productivity changes in the most developed regions of Brazil. Therefore, the presence of education (mainly driven by the supply side) is not sufficient to create prosperity. From a policy perspective it may be best to develop institutions to encourage productive entrepreneurship at the expense of rent seekers to fully utilise human capital in SMEs. The adequate institutional settings can create the conditions to use fully the absorptive and innovative capacity of the *SMEH* in the process of technological diffusion and productivity growth.

[57] The discussion about human capital and entrepreneurship has parallels with the self-employment literature. For instance, Parker (2004) provides an extensive review of the role of education in self-employment. The probability of self-employment returns to self-employment and the heterogeneity in entrepreneurship might be determined by human capital and ability.

Table 4.10: Industry Employment Share and Growth

	SME250						SME500					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	LSDV	DiffGMM	DiffGMM	SysGMM	SysGMM	OLS	LSDV	DiffGMM	DiffGMM	SysGMM	SysGMM
			1-step	2-step	1-step	2-step			1-step	2-step	1-step	2-step
lnGDP_{t-1}	-0.0525***	-0.143***	-0.155***	-0.164***	-0.0752***	-0.0736***	-0.0515***	-0.143***	-0.162***	-0.175***	-0.0756***	-0.0730***
	(-11.50)	(-21.81)	(-8.99)	(-5.35)	(-4.94)	(-6.12)	(-11.31)	(-21.83)	(-8.55)	(-5.39)	(-5.22)	(-6.77)
ln*School*	0.0221***	0.142***	0.248***	0.153***	0.0345	0.0613***	0.0209**	0.142***	0.232***	0.166***	0.0400*	0.0621***
	(2.61)	(6.60)	(6.02)	(2.71)	(1.29)	(2.68)	(2.47)	(6.62)	(6.01)	(3.00)	(1.66)	(3.37)
ln$(n+g+d)$	-0.0129*	-0.0066	-0.0359	-0.0044	-0.0464***	-0.0289***	-0.0127*	-0.0067	-0.0166	0.0238	-0.0385***	-0.0260***
	(-1.94)	(-0.56)	(-0.83)	(-0.05)	(-4.49)	(-2.71)	(-1.91)	(-0.56)	(-0.37)	(0.32)	(-5.66)	(-3.71)
ln*SMER*	-0.0121***	-0.0081*	0.0063	-0.0283	0.0312	0.0055	-0.0116***	-0.0108**	-0.0388	-0.0376	0.0059	-0.0058
	(-3.86)	(-1.66)	(0.20)	(-0.54)	(0.87)	(0.19)	(-3.15)	(-2.04)	(-1.03)	(-1.33)	(0.18)	(-0.24)
*Dummy*1995	-0.0501***	-0.0778***	-0.115***	-0.0728***	-0.0569***	-0.0591***	-0.0500***	-0.0778***	-0.105***	-0.0748***	-0.0558***	-0.0584***
	(-9.51)	(-9.11)	(-7.65)	(-3.05)	(-7.04)	(-7.75)	(-9.49)	(-9.14)	(-7.39)	(-4.41)	(-7.47)	(-8.65)
Dummy 2000	0.0180***	-0.0460***	-0.101***	-0.0502	0.0016	-0.0106	0.0182***	-0.0460***	-0.0872***	-0.0574**	0.0037	-0.0086
	(2.85)	(-4.02)	(-4.45)	(-1.54)	(0.10)	(-0.75)	(2.88)	(-4.03)	(-4.26)	(-2.15)	(0.25)	(-0.70)
Dummy 2005	-0.0095	-0.0657***	-0.132***	-0.0653*	-0.0204	-0.0343***	-0.0096	-0.0661***	-0.118***	-0.0722**	-0.0205	-0.0338***
	(-1.45)	(-4.79)	(-5.30)	(-1.69)	(-1.36)	(-2.67)	(-1.45)	(-4.84)	(-5.19)	(-2.39)	(-1.47)	(-3.06)
Observations	1436	1436	1077	1077	1436	1436	1436	1436	1077	1077	1436	1436
Adjusted R^2	0.268	0.268					0.266	0.269				
Instruments			18	18	27	27			18	18	27	27
AR1			0.0000	0.0083	0.0000	0.0000			0.0000	0.0283	0.0000	0.0000
AR2			0.0065	0.4460	0.0677	0.0722			0.0966	0.3270	0.0928	0.0892
Hansen			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.028	0.028					0.008	0.008

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

Table 4.11: Industry, Commerce and Services Employment Share and Growth

	SME250						SME500					
	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step
lnGDP_{t-1}	-0.0532***	-0.143***	-0.158***	-0.163***	-0.0789***	-0.0756***	-0.0521***	-0.144***	-0.159***	-0.167***	-0.0743***	-0.0769***
	(-11.64)	(-21.97)	(-9.46)	(-5.56)	(-5.33)	(-6.22)	(-11.42)	(-22.05)	(-8.92)	(-5.43)	(-4.81)	(-6.17)
ln*School*	0.0234***	0.133***	0.224***	0.163***	0.0598**	0.0750***	0.0226***	0.133***	0.204***	0.145**	0.0594**	0.0780***
	(2.77)	(6.14)	(4.89)	(2.67)	(2.34)	(3.63)	(2.66)	(6.20)	(4.82)	(2.41)	(2.34)	(3.65)
ln$(n+g+d)$	-0.0168**	-0.0095	-0.0375	0.0039	-0.104***	-0.0783***	-0.0153**	-0.0094	-0.0321	-0.0072	-0.119***	-0.0789***
	(-2.50)	(-0.81)	(-0.93)	(0.05)	(-4.32)	(-2.98)	(-2.29)	(-0.80)	(-0.77)	(-0.09)	(-4.63)	(-2.70)
ln*SMER*	-0.0262***	-0.0316***	-0.0356	-0.0044	-0.0461***	-0.0356**	-0.0252***	-0.0408***	-0.0789	-0.0467	-0.0558***	-0.0354**
	(-4.36)	(-3.30)	(-0.81)	(-0.08)	(-2.86)	(-2.13)	(-3.58)	(-3.90)	(-1.64)	(-0.79)	(-3.43)	(-1.97)
Dummy1995	-0.0502***	-0.0737***	-0.104***	-0.0781***	-0.0598***	-0.0600***	-0.0504***	-0.0737***	-0.0950***	-0.0705***	-0.0604***	-0.0617***
	(-9.55)	(-8.54)	(-6.30)	(-3.55)	(-8.08)	(-9.41)	(-9.56)	(-8.61)	(-6.33)	(-3.28)	(-8.04)	(-9.18)
Dummy 2000	0.0179***	-0.0398***	-0.0848***	-0.0592*	0.0007	-0.0098	0.0177***	-0.0399***	-0.0724***	-0.0465	0.0013	-0.0101
	(2.85)	(-3.43)	(-3.31)	(-1.85)	(0.04)	(-0.81)	(2.82)	(-3.47)	(-3.21)	(-1.52)	(0.08)	(-0.80)
Dummy 2005	-0.0094	-0.0585***	-0.114***	-0.0731**	-0.0314**	-0.0397***	-0.0099	-0.0592***	-0.0998***	-0.0593*	-0.0337**	-0.0422***
	(-1.44)	(-4.21)	(-4.01)	(-1.97)	(-2.13)	(-3.45)	(-1.50)	(-4.31)	(-3.96)	(-1.65)	(-2.29)	(-3.53)
Observations	1436	1436	1077	1077	1436	1436	1436	1436	1077	1077	1436	1436
Adjusted R^2	0.270	0.274					0.267	0.277				
Instruments			18	18	27	27			18	18	27	27
AR1			0.0001	0.0109	0.0000	0.0000			0.0000	0.0172	0.0000	0.0000
AR2			0.0618	0.1390	0.3840	0.2090			0.1420	0.3310	0.4140	0.2010
Hansen			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.108	0.108					0.052	0.052

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

Table 4.12: Industry Employment Share, *SMEH* and Growth

	SME250						SME500					
	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step
lnGDP_{t-1}	-0.0530***	-0.142***	-0.155***	-0.156***	-0.0757***	-0.0687***	-0.0524***	-0.141***	-0.163***	-0.173***	-0.0742***	-0.0705***
	(-11.63)	(-21.66)	(-9.20)	(-6.19)	(-5.07)	(-5.19)	(-11.54)	(-21.64)	(-9.21)	(-5.34)	(-5.28)	(-5.78)
ln*School*	0.0095	0.112***	0.176***	0.128*	0.0595	0.0789	0.0068	0.118***	0.200***	0.148	0.0638*	0.0900*
	(0.99)	(4.80)	(2.99)	(1.87)	(1.63)	(1.49)	(0.72)	(5.16)	(3.36)	(1.52)	(1.69)	(1.95)
ln$(n+g+d)$	-0.0086	-0.0069	-0.0366	-0.0012	-0.0585***	-0.0508	-0.0078	-0.0076	-0.0169	0.0299	-0.0536**	-0.0532
	(-1.26)	(-0.58)	(-0.93)	(-0.02)	(-2.65)	(-1.36)	(-1.15)	(-0.64)	(-0.40)	(0.41)	(-2.21)	(-1.26)
ln*SMER*	-0.0113***	-0.0084*	0.0238	0.0035	0.0089	0.0066	-0.0103***	-0.0094*	-0.0371	-0.0412	-0.0208	-0.0164
	(-3.60)	(-1.73)	(0.82)	(0.10)	(0.26)	(0.16)	(-2.81)	(-1.79)	(-1.28)	(-1.05)	(-0.63)	(-0.67)
ln*SMEH*	0.0278***	0.0462***	0.129*	0.106	-0.0468	-0.0553	0.0325***	0.0423***	0.0577	0.0387	-0.0507	-0.0722
	(2.81)	(3.17)	(1.94)	(1.28)	(-0.88)	(-0.55)	(3.36)	(3.07)	(0.81)	(0.30)	(-0.80)	(-0.72)
*Dummy*1995	-0.0513***	-0.0772***	-0.118***	-0.0881***	-0.0540***	-0.0555***	-0.0513***	-0.0779***	-0.104***	-0.0751***	-0.0526***	-0.0540***
	(-9.73)	(-9.07)	(-8.44)	(-4.27)	(-6.81)	(-6.06)	(-9.73)	(-9.19)	(-7.68)	(-4.20)	(-6.76)	(-6.44)
Dummy 2000	0.0141**	-0.0479***	-0.115***	-0.0820**	0.0099	-0.0011	0.0133**	-0.0492***	-0.0921***	-0.0622**	0.0140	0.0034
	(2.19)	(-4.19)	(-5.34)	(-2.49)	(0.60)	(-0.06)	(2.07)	(-4.31)	(-4.42)	(-2.41)	(0.80)	(0.17)
Dummy 2005	-0.0157**	-0.0714***	-0.155***	-0.107***	-0.0106	-0.0193	-0.0170**	-0.0732***	-0.127***	-0.0803**	-0.0094	-0.0158
	(-2.27)	(-5.18)	(-6.41)	(-2.79)	(-0.58)	(-0.77)	(-2.46)	(-5.31)	(-5.19)	(-2.52)	(-0.49)	(-0.62)
Observations	1436	1436	1077	1077	1436	1436	1436	1436	1077	1077	1436	1436
Adjusted R^2	0.272	0.275					0.271	0.275				
Instruments			21	21	32	32			21	21	32	32
AR1			0.0000	0.0009	0.0000	0.0000			0.0001	0.0218	0.0000	0.0000
AR2			0.0102	0.0312	0.1670	0.1520			0.0345	0.2620	0.3130	0.2340
Hansen			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.025	0.025					0.010	0.010

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

Table 4.13: Industry, Commerce and Services Employment Share, *SMEH* and Growth

	SME250						SME500					
	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step
$\ln GDP_{t-1}$	-0.0549***	-0.143***	-0.163***	-0.165***	-0.0707***	-0.0742***	-0.0534***	-0.144***	-0.165***	-0.179***	-0.0696***	-0.0764***
	(-11.88)	(-21.95)	(-9.40)	(-5.74)	(-4.79)	(-6.03)	(-11.62)	(-22.04)	(-9.18)	(-5.89)	(-4.69)	(-6.47)
ln*School*	0.0329***	0.133***	0.230***	0.169**	0.0964***	0.0939***	0.0313***	0.132***	0.189***	0.119	0.0868***	0.0960***
	(3.52)	(6.12)	(4.71)	(2.55)	(3.06)	(3.84)	(3.32)	(6.14)	(3.90)	(1.62)	(2.65)	(4.13)
$\ln(n+g+d)$	-0.0211***	-0.0096	-0.0259	-0.0098	-0.155***	-0.103***	-0.0191***	-0.0094	-0.0030	0.0704	-0.143***	-0.0978***
	(-3.04)	(-0.81)	(-0.58)	(-0.12)	(-4.42)	(-3.58)	(-2.77)	(-0.80)	(-0.06)	(0.71)	(-4.46)	(-3.69)
ln*SMER*	-0.0269***	-0.0314***	-0.0406	-0.0043	0.0152	0.0117	-0.0260***	-0.0407***	-0.109**	-0.0903	-0.0160	0.00464
	(-4.47)	(-3.28)	(-0.89)	(-0.08)	(0.49)	(0.49)	(-3.69)	(-3.89)	(-2.06)	(-1.30)	(-0.58)	(0.18)
ln*SMEH*	-0.0373**	-0.0063	-0.0805	-0.128	-0.243**	-0.155**	-0.0323**	0.0127	0.0907	0.186	-0.150*	-0.133*
	(-2.39)	(-0.24)	(-0.77)	(-0.98)	(-2.40)	(-2.11)	(-2.11)	(0.51)	(0.89)	(1.25)	(-1.77)	(-1.78)
*Dummy*1995	-0.0503***	-0.0734***	-0.0982***	-0.0698***	-0.0580***	-0.0589***	-0.0505***	-0.0745***	-0.0945***	-0.0712***	-0.0601***	-0.0594***
	(-9.58)	(-8.42)	(-5.66)	(-2.98)	(-7.32)	(-8.14)	(-9.60)	(-8.57)	(-5.47)	(-3.06)	(-7.84)	(-8.80)
Dummy 2000	0.0200***	-0.0389***	-0.0731**	-0.0401	0.0216	-0.0003	0.0197***	-0.0419***	-0.0805***	-0.0700*	0.0141	-0.0003
	(3.16)	(-3.20)	(-2.55)	(-1.07)	(1.29)	(-0.02)	(3.10)	(-3.45)	(-2.83)	(-1.95)	(0.94)	(-0.02)
Dummy 2005	-0.00373	-0.0569***	-0.0906**	-0.0392	0.0115	-0.0118	-0.00487	-0.0628***	-0.115***	-0.0950**	-0.00640	-0.0162
	(-0.54)	(-3.70)	(-2.46)	(-0.82)	(0.51)	(-0.70)	(-0.70)	(-4.08)	(-3.16)	(-2.13)	(-0.34)	(-0.89)
Observations	1436	1436	1077	1077	1436	1436	1436	1436	1077	1077	1436	1436
Adjusted R^2	0.273	0.273					0.269	0.276				
Instruments			21	21	32	32			21	21	32	32
AR1			0.0001	0.0142	0.0000	0.0000			0.0001	0.0105	0.0000	0.0000
AR2			0.1770	0.4500	0.6790	0.3040			0.1170	0.1720	0.6070	0.3290
Hansen			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Sargan			0.0000	0.0000	0.0000	0.0000			0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff					0.210	0.210					0.053	0.053

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The AR statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. The heading SME250 indicates that the SME sector considers establishments with less than 250 employees and SME500 establishments with less than 500 employees.

4.6 Regressions with Valid Instruments

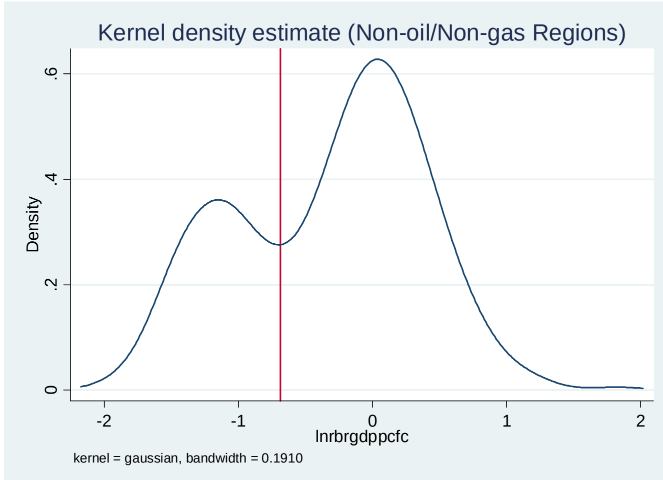
In this section, we want to address two factors that could have contributed to the high degree of endogeneity we observed in the previous sections, even after internally instrumenting our independent variables. Firstly, when we observe the top of the micro-regional GDP per capita ranking, we find many micro-regions involved in the oil economy on the top of the table, indicating that oil economies could be one important omitted variable that induces growth.[58] Secondly, the economies in the valley and in the tails of our kernel density estimation, used as a criterion to split the sample could also damage our ability to isolate the behaviour of the core part of the peaks. A common factor could be responsible for either the failure of the regions in the left end of each peak or the success of those in the right end of each peak.

We try to control for regions that benefit from the presence of oil or natural gas using the data of royalties' revenue that Brazilian municipalities receive when they are related with the extraction of oil or gas. In 2004, 124 out of 508 micro-regions received royalties and are excluded from the sample. Figure 4.2 shows the kernel density estimated for this alternative sample without oil or gas producers and illustrates the same bimodal distribution but with a smaller tail (the longer tail of Figure 4.1 probably is influenced by the micro-regions that produce oil or gas in the far right end of the distribution). To control for the outliers of each club (poor and rich regions) and control for the transitional economies in the neighbourhood of the valley, we create two alternative samples removing 20% of observations in the extremities of each group.

This strategy provides three new samples, the first with 384 Brazilian micro-regions that are not directly related with oil or gas production, the second with 109 micro-regions from the lower (poorest) peak, and the third with 197 micro-regions from the upper (richest) peak.

[58] Please refer to Appendix 4.8.2 for further details.

Figure 4.2: Kernel Densities and Club convergence – Non-Oil Sample



4.6.1 Additional Results

From Tables 4.14 to 4.16 we report the alternative results when we remove the micro-regions involved in the oil or gas economy. Results are for the full specification that consider together *SMER* and *SMEH* for the manufacturing sector and are comparable to Tables 4.4, 4.8 and 4.12.[59] We provide the results for the two criteria used in previous sections to classify SMEs and focus our analyses on the cut off of 250 employees. The results for the two criteria never contradict each other in the sense that they never have opposite and significant sign for the same type of estimator.

Table 4.14 shows the result for our sample without oil/gas regions. As in the Table 4.4, diagnostic tests indicate that the instruments in levels are additional sources of endogeneity, therefore, we prefer the GMM-DIFF estimates. We observe that results for our preferred estimator are very similar and support the views of the

[59] We also tried the specification only with *SMER* but they do not provide improvements in terms of instruments endogeneity. Before removing 20% of micro-regions from each peak of the distribution we also performed the same estimations without trimming the samples and trimming 10% of the micro-regions. Results are similar but we could not find any valid set of instruments. Regressions for the whole of the SME sector (including commerce and services) do not provide good set of instruments either.

negative or neutral impact of *SMER* on growth. The *SMER*'s coefficients are negative or insignificant statistically. On the other hand, *SMEH* enters positively and is significant for the two step GMM-DIFF, showing again a incipient sign that SMEs' human capital is important for growth. However, our strategy of removing oil/gas economies did not provide much better results in terms of instruments validity and we are unable to do assertive inferences using this set of results.

Table 4.15 reports the results for our alternative sample of poor regions and results provide better instruments when comparing to Table 3.8. Diagnostic tests suggest the inclusion of GMM-SYS instruments and Hansen test also suggests that this estimator is less subject to endogeneity and performs better than the GMM-DIFF. The GMM-SYS performs better (in terms of instruments endogeneity) than in Table 3.8 but we cannot rule out endogeneity with confidence. Results are similar to those founds in Table 4.8 and suggest that neither *SMER* nor *SMEH* are important to foster economic performance.

Finally, Table 4.16 presents the results for the alternative sample of Rich micro-regions, and the Hansen test only rules out endogeneity for the GMM-DIFF regressions, suggesting that these results might be more reliable to analyse a causal relationship. Results for the GMM-DIFF estimator indicate that the size of the SME sector is not a sign of faster growth and *SMER* is not significant for both one and two step estimates. However, the *SMEH* is positive and significant for growth. Therefore, for rich regions, the presence of SMEs is not important *per se* but their quality seems to be. These alternative results provide the best set of instruments according to the Hansen test and are less prone to the effects coming from endogeneity when we compare with results in Table 4.12. Overall, the estimates provided in this section present better results for the diagnostic tests for endogeneity and make inferences slightly more reliable. However, they do not rule out the presence of endogeneity with high degree of confidence.

Therefore, removing the oil or gas related economies and trimming the tail of the distribution reduced slightly the degree of endogeneity in the GMM estimates. However, results are still in line with the set of regressions presented in Section 4.5. The SME sector is not important for the poorest regions while the SMEs' human capital is important for the richest regions. Despite having used better instruments in this section, we still have to interpret those results with caution because the endogeneity is not ruled out of our GMM estimates with high degree of confidence.

The remaining endogeneity could be related with omitted variables (e.g. physical capital, institutional settings) but the spatial dynamics could be playing a significant part in explaining it. It is not hard to imagine that endogeneity created by spatial autocorrelation is more likely to occur at micro-regional level than at state level. For instance, Silveira-Neto and Azzoni (2006) argue that the hypothesis of spatial dependence is rejected for Brazilian states after conditioning growth regressions on variables that reflect the regional patterns. Similarly, Resende (2009) analyses the economic growth process in four different spatial scales (states, micro-region, spatial cluster and municipalities) and suggests that spatial dependences are operating mainly in the most disaggregated geographic level of micro-regions and municipalities. Hence, we mimic our regressions using the state level data in the Appendix 4.8.1 and the results provide indirect support for this argument. Using the state level data we get good results in terms of ruling out endogeneity.

Therefore, this section provides additional empirical results that are in line with the conclusions drawn from Section 4.3 but are less prone to endogeneity. However, this problem is persistent and might be related with the spatial interactions across Brazilian regions.

Table 4.14: Industry Employment Share, SMEH and Growth (Non-Oil/Gas Regions) – Brazil

	SME250				SME500			
	(1) DiffGMM 1-step	(2) DiffGMM 2-step	(3) SysGMM 1-step	(4) SysGMM 2-step	(1) DiffGMM 1-step	(2) DiffGMM 2-step	(3) SysGMM 1-step	(4) SysGMM 2-step
lnGDP_{t-1}	-0.195***	-0.226***	-0.0795***	-0.0822***	-0.188***	-0.213***	-0.0758***	-0.0785***
	(-11.04)	(-6.79)	(-6.17)	(-6.44)	(-10.66)	(-6.32)	(-6.08)	(-6.71)
ln*School*	0.114***	0.0284	0.145***	0.158***	0.121***	0.0433	0.147***	0.156***
	(3.49)	(1.01)	(4.45)	(4.64)	(3.58)	(1.51)	(4.61)	(4.78)
ln$(n+g+d)$	0.0330	0.0430	-0.0577***	-0.0635**	0.0261	0.0217	-0.0607***	-0.0640**
	(0.80)	(0.68)	(-3.06)	(-2.36)	(0.61)	(0.32)	(-3.23)	(-2.53)
ln*SMER*	-0.0328	-0.0184	-0.0652**	-0.0683***	-0.0348*	-0.0329	-0.0713**	-0.0825***
	(-1.41)	(-0.70)	(-1.99)	(-2.63)	(-1.89)	(-1.39)	(-2.13)	(-3.19)
ln*SMEH*	0.0111	0.0672*	-0.131***	-0.147**	0.00300	0.0203	-0.134***	-0.143**
	(0.21)	(1.71)	(-2.64)	(-2.26)	(0.06)	(0.52)	(-2.82)	(-2.41)
*Dummy*1995	-0.0647***	-0.0374***	-0.0620***	-0.0619***	-0.0669***	-0.0356***	-0.0641***	-0.0641***
	(-6.35)	(-3.37)	(-8.03)	(-6.58)	(-7.48)	(-3.43)	(-8.58)	(-7.54)
Dummy 2000	-0.0415**	-0.0175	0.00850	0.00743	-0.0418***	-0.00700	0.00865	0.00638
	(-2.27)	(-0.85)	(0.63)	(0.44)	(-2.61)	(-0.36)	(0.67)	(0.41)
Dummy 2005	-0.0514**	-0.0120	-0.0199	-0.0178	-0.0539***	-0.00186	-0.0213	-0.0216
	(-2.32)	(-0.54)	(-1.33)	(-0.87)	(-2.73)	(-0.09)	(-1.49)	(-1.17)
Observations	1152	1152	1536	1536	1152	1152	1536	1536
Instruments	21	21	32	32	21	21	32	32
AR1	0.0014	0.3440	0.0000	0.0000	0.0008	0.3940	0.0000	0.0000
AR2	0.0006	0.2460	0.7320	0.8280	0.0002	0.2970	0.2590	0.3080
Hansen	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sargan	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff			0.000	0.000			0.000	0.000

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.

Table 4.15: Industry Employment Share, SMEH and Growth (Non-Oil/Gas Regions), Poor

	SME250				SME500			
	(1) DiffGMM 1-step	(2) DiffGMM 2-step	(3) SysGMM 1-step	(4) SysGMM 2-step	(1) DiffGMM 1-step	(2) DiffGMM 2-step	(3) SysGMM 1-step	(4) SysGMM 2-step
lnGDP_{t-1}	-0.134***	-0.132***	-0.119***	-0.104***	-0.138***	-0.133***	-0.115***	-0.102***
	(-7.34)	(-5.54)	(-5.40)	(-5.35)	(-7.42)	(-5.49)	(-5.03)	(-4.92)
ln*School*	0.250***	0.199*	0.0580***	0.0420*	0.209***	0.189**	0.0562***	0.0469**
	(3.02)	(1.88)	(3.10)	(1.84)	(3.28)	(2.38)	(3.11)	(2.12)
ln*(n+g+d)*	0.0117	0.0330	-0.0148	-0.0116	-0.0237	-0.00975	-0.0129	-0.0109
	(0.20)	(0.36)	(-1.58)	(-1.19)	(-0.45)	(-0.10)	(-1.48)	(-1.25)
ln*SMER*	0.0691*	0.0388*	-0.0262	-0.0306*	-0.00473	-0.000604	-0.0213*	-0.0206*
	(1.82)	(1.70)	(-1.31)	(-1.71)	(-0.42)	(-0.03)	(-1.85)	(-1.67)
ln*SMEH*	0.0494	0.0441	-0.00186	0.00755	0.0312	0.0286	0.00430	0.0111
	(1.21)	(1.35)	(-0.08)	(0.30)	(1.02)	(1.01)	(0.20)	(0.51)
*Dummy*1995	-0.169***	-0.141***	-0.0603***	-0.0631***	-0.143***	-0.135***	-0.0613***	-0.0661***
	(-3.65)	(-2.70)	(-6.40)	(-5.17)	(-4.09)	(-3.48)	(-6.72)	(-5.56)
*Dummy*2000	-0.151**	-0.105	0.0186	0.0246	-0.109**	-0.0921	0.0172	0.0207
	(-2.24)	(-1.35)	(0.96)	(1.00)	(-2.13)	(-1.49)	(0.91)	(0.87)
*Dummy*2005	-0.246***	-0.194**	-0.0375**	-0.0365	-0.195***	-0.176**	-0.0401**	-0.0437*
	(-2.88)	(-2.05)	(-1.96)	(-1.48)	(-3.00)	(-2.39)	(-2.23)	(-1.87)
Observations	327	327	436	436	327	327	436	436
Instruments	21	21	32	32	21	21	32	32
AR1	0.0002	0.0001	0.0004	0.0001	0.0000	0.0002	0.0004	0.0001
AR2	0.1640	0.0975	0.5660	0.5210	0.0443	0.0779	0.2770	0.2350
Hansen	0.0054	0.0054	0.0444	0.0444	0.0006	0.0006	0.0299	0.0299
Sargan	0.0006	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Hansen-in-Diff			0.635	0.635			0.708	0.708

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.

Table 4.16: Industry Employment Share, SMEH and Growth (Non-Oil/Gas Regions), Rich

	SME250				SME500			
	(1) DiffGMM 1-step	(2) DiffGMM 2-step	(3) SysGMM 1-step	(4) SysGMM 2-step	(1) DiffGMM 1-step	(2) DiffGMM 2-step	(3) SysGMM 1-step	(4) SysGMM 2-step
lnGDP_{t-1}	-0.105***	-0.0859***	-0.110***	-0.0931***	-0.108***	-0.106***	-0.106***	-0.0847***
	(-4.84)	(-3.47)	(-8.00)	(-6.26)	(-4.70)	(-3.92)	(-7.80)	(-5.31)
ln*School*	0.0810	0.124	0.151***	0.105**	0.0547	0.154	0.146***	0.0801*
	(0.68)	(1.36)	(3.15)	(2.27)	(0.33)	(1.29)	(3.26)	(1.89)
ln*(n+g+d)*	-0.0779**	-0.0590	-0.0864***	-0.0781**	-0.0747	-0.0265	-0.0859***	-0.0607*
	(-2.11)	(-0.98)	(-3.83)	(-2.50)	(-1.59)	(-0.36)	(-3.70)	(-1.69)
ln*SMER*	0.0216	0.0279	-0.0558**	-0.0320	-0.0109	-0.0124	-0.0617**	-0.0287*
	(0.58)	(0.54)	(-2.24)	(-1.29)	(-0.32)	(-0.32)	(-2.49)	(-1.68)
ln*SMEH*	0.458***	0.362***	-0.150**	-0.112	0.535*	0.313*	-0.143**	-0.0665
	(2.80)	(2.73)	(-2.22)	(-1.47)	(1.85)	(1.76)	(-2.20)	(-0.82)
*Dummy*1995	-0.143***	-0.139***	-0.0510***	-0.0423***	-0.139***	-0.129***	-0.0528***	-0.0446***
	(-4.18)	(-3.62)	(-6.50)	(-5.46)	(-4.03)	(-3.58)	(-6.66)	(-5.70)
*Dummy*2000	-0.183***	-0.167***	0.00150	0.00836	-0.189***	-0.157***	0.0012	0.0034
	(-3.47)	(-2.84)	(0.11)	(0.57)	(-3.05)	(-2.66)	(0.09)	(0.21)
*Dummy*2005	-0.256***	-0.232***	0.0012	0.0073	-0.276***	-0.219***	-0.0015	-0.0012
	(-3.73)	(-3.06)	(0.08)	(0.39)	(-3.16)	(-2.86)	(-0.11)	(-0.06)
Observations	591	591	788	788	591	591	788	788
Instruments	21	21	32	32	21	21	32	32
AR1	0.0007	0.0002	0.0001	0.0014	0.0214	0.0007	0.0000	0.0011
AR2	0.6710	0.9650	0.9330	0.6780	0.5280	0.3050	0.8720	0.4000
Hansen	0.0508	0.0508	0.0059	0.0059	0.0711	0.0711	0.0069	0.0069
Sargan	0.0003	0.0003	0.0000	0.0000	0.0160	0.0160	0.0000	0.0000
Hansen-in-Diff			0.721	0.721			0.616	0.616

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.

4.7 Conclusion

This chapter has provided the first systematic investigation about the importance of SMEs for economic growth in Brazil. The standard neo-classical growth framework was extended as in Beck et al. (2005a) to incorporate the SME sector into it. Proxies for the employment share of SMEs and for the SMEs' human capital were used to evaluate the importance of this sector in three different samples using standard methods of panel data analysis and instrumental variable methods to account for endogeneity. The empirical results show that SME sector size has a negative or neutral effect on growth and that SMEs' human capital has no clear effect in this process for the group of the poorest micro-regions. However, the results for Brazilian micro-regions and for the group of the richest regions suggest that SMEs' human capital seems to play a role in the growth process. These findings are similar with Van Stel et al. (2005) and Wennekers et al. (2005) that provide empirical evidence indicating a U-shaped relationship between entrepreneurship and level of development. As there are indications of different institutional performances across regions in Brazil, this is in line with Baumol (1990, 2008) and Dias and McDermott (2006) that argue that different stages of institutional development will create different levels of productive entrepreneurship; and hence, will provide different incentives towards SMEs. However, these results are prone to endogeneity (that might be associated with the omission of the spatial dimension) and must be interpreted with caution. We reduced the degree of endogeneity excluding the oil/gas economies and trimming the samples. The results found are similar, increasing our confidence when interpreting them. Additionally, the inclusion of commerce and services weakens the potential causal effect of SME on growth, suggesting that SMEs in manufacturing are more important for growth.

From the public policy point of view, the results suggest that the negative association between growth and SME sector size in Brazil points out that some factors are not enticing the right type of productive small firms, in this case, growing the SME sector is not likely to increase productivity. Also, different performances of the level of human capital in small businesses indicate that different SME-oriented public policies are required in different regions with dissimilar levels of development. Hence, directly subsidizing SMEs in regions with significant institutional constraints and lower levels of human capital in SMEs does not seem to be an adequate SME policy to accelerate growth, especially in poorer regions. Rather, the focus should be on institutional improvement and educational policies that can create the necessary conditions for the development of a thriving SME sector that can contribute for the process of economic development.

4.8 Appendix

4.8.1 Additional Results for Brazilian States

The following analyses provide alternative results of our model for Brazilian states using a panel data setting covering 25 states with annual data for the period 1985-2004.[60] However, to control for business cycles we did not use the annual data directly, but the 3-year average over the sample period.[61] All series were also averaged over a 3-year period. In this set of results, it is interesting to see whether the endogeneity problem reduces when we use a more aggregated level of regional division. If this fact is observed, it implicitly supports our argument that we were unable to get valid instruments for Brazilian micro-regions partially due to spatial autocorrelation.

The data used is retrieved from two sources; IPEADATA and RAIS. Real GDP per capita for each state (GDPpc) is collected from IPEADATA. The data for population used to calculate the population growth (n), and the average years of schooling of the population over 25 years old (*School*), are also taken from IPEADATA. A limitation in our data is the absence of data on physical capital for Brazilian states. As a proxy, the average of industrial consumption of electricity (K) for each 3-year period retrieved from IPEADATA is used. This measure has been extensively used as a proxy to capital stock in Brazil (e.g. Lau et al. 1993; Ferreira 2000; Nakabashi and Salvato 2007). Lau et al. (1993), for example, argue that this measure has the advantage over the capital stock since it already embodies a rate of

[60] A longer time series is not available due to a change in the national accounts' methodology from 2004 onwards. The sample loses two states because the most recent state, Tocantins, was created in 1988. It constitutes the northern territory of the former state of Goiás, which retains the southern part of the territory and kept its original name, Goiás. For these reason it was not possible to construct a reliable series for these states and we excluded these two states from the sample.

[61] Using the annual data allows more degrees of freedom for the relative short panel time length of our study. However, to control for business cycles we used a 3-year average to reduce its influence. This approach still allows the use of five cross sections in the GMM-DIFF estimates, since two cross-sections are lost to control for the lagged GDPpc and to take the first difference. Increasing the time length to the usual 5-year average would reduce the degrees of freedom and we would not have the minimum of three cross sections necessary to calculate the diagnostic tests for the GMM estimates (see Arellano and Bond 1991).

utilization adjustment. As in the case of the analysis for the Brazilian micro-regions, RAIS provided the information necessary to construct our SMEs' measures.

The difference here is that we are able to collect annual data for all variables we used to study micro-regions plus the industrial energy usage to proxy for physical capital. Therefore, our panel has five cross-sections with 25 observations each, providing a panel of 150 observations. It also allows us to check the robustness of the results after the inclusion of a proxy for physical capital in the model.

Tables 4.17 and 4.18 present the results for the Brazilian states (similar to Tables 4.2 and 4.4) added with physical capital. They report the estimates for the LSDV and estimations for GMM-DIFF and GMM-SYS using a cut off of 250 employees to classify SMEs. In Table 4.17, from columns 1 to 5, the estimations include only *SMER* to account for the SME sector, as in Beck et al. (2005a), and we do not control for the industrial electricity consumption. Column 1 reports the results for the LSDV and indicates a strong convergence pattern, a positive effect of education on growth, and a negative but not significant effect stemming from population growth. These results are consistent with the results presented for micro-regions. For the SME sector, we observe a negative and significant effect of *SMER* on growth, indicating that SMEs are not associated with economic growth. However, the LSDV model does not address endogeneity, and to take this into account, the GMM in first difference (GMM-DIFF) and the system GMM (GMM-SYS) are alternative estimators to tackle this potential problem.

Results for the one-step and two-step GMM-DIFF and GMM-SYS estimates are reported from columns 2 to 5 respectively. The results for both estimators confirm the existence of conditional convergence and the positive impact of education on growth with a high degree of significance. Again, the coefficients on the population growth indicate no statistical significance. The coefficients for the size of the SME sector in manufacturing is always negative and is significant for both one-step estimations, consistent with the initial evidence found by the LSDV estimator, and therefore supporting the view that SMEs are negatively related to economic growth in Brazil. In addition, diagnostic tests for the validity of instruments presented at the bottom of the table do not indicate any problem with the validity of the internal instruments. The values for m_2 are the *p*-values for the autocorrelation test under the null of no autocorrelation and values for Hansen test are the *p*-values for the joint validity of the instruments under the null that the instruments are valid. Unlike the regressions for micro-regions, the Hansen test suggests that instruments are exogenous and allows us to make inference about the causal impact of the conditioning variables on growth.

The results with the inclusion of the industrial electricity usage are reported from columns 6 to 10. The presence of SMEs in manufacturing is robust to the inclusion of this variable, increasing our confidence about the sign of this effect.

Table 4.17 SMEs Employment and Growth - Manufacturing

	(1) LSDV	(2) GMM-DIFF 1-step	(3) GMM-DIFF 2-step	(4) GMM-SYS 1-step	(5) GMM-SYS 2-step	(6) LSDV	(7) GMM-DIFF 1-step	(8) GMM-DIFF 2-step	(9) GMM-SYS 1-step	(10) GMM-SYS 2-step
lnGDP_{t-1}	-0.528***	-0.653***	-0.726***	-0.192***	-0.185***	-0.586***	-0.758***	-0.797***	-0.173**	-0.170***
	(-7.61)	(-3.97)	(-2.65)	(-3.89)	(-3.96)	(-8.23)	(-5.32)	(-3.88)	(-2.46)	(-2.59)
ln*School*	0.421***	0.552***	0.538***	0.435***	0.401***	0.448***	0.558***	0.534***	0.429***	0.412***
	(4.00)	(3.73)	(3.46)	(4.25)	(2.87)	(4.34)	(3.41)	(2.73)	(3.98)	(2.84)
ln$(n+g+d)$	-0.125	0.368	0.365	-0.0984	-0.0770	-0.000571	0.507	0.360	-0.176	-0.0811
	(-0.80)	(0.99)	(0.66)	(-1.21)	(-0.76)	(-0.00)	(1.18)	(1.16)	(-1.10)	(-0.42)
ln*SMER*	-0.146**	-0.500**	-0.347	-0.154**	-0.128	-0.122**	-0.415**	-0.272	-0.178**	-0.150
	(-2.37)	(-2.03)	(-1.11)	(-2.22)	(-1.25)	(-2.01)	(-2.13)	(-1.16)	(-2.15)	(-1.24)
ln(*K*)						0.0666***	0.110*	0.0879*	-0.00957	0.00252
						(2.63)	(1.74)	(1.81)	(-0.55)	(0.11)
Observations	150	125	125	150	150	150	125	125	150	150
Adjusted R^2	0.483					0.508				
Instruments		30	30	35	35		30	30	35	35
m_1		0.194	0.497	0.00916	0.00890		0.201	0.545	0.00759	0.00940
m_2		0.259	0.192	0.507	0.557		0.330	0.203	0.504	0.561
Hansen		0.637	0.637	0.961	0.961		0.700	0.700	0.961	0.961

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The m_2 statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. Hansen statistics tests for endogeneity under the null that instruments are valid. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. All results include time dummies.

The results for the extended specification with *SMEH* are shown in Table 4.18. Columns 1 to 5 present results without physical capital. The inclusion of *SMEH* does not change the initial analysis drawn from the results of Table 4.17. For example, *SMER* is again negatively related to growth. Results with the addition of the variable *SMEH* indicate that higher levels of human capital do not reduce economic growth and in fact has a positive and significant coefficient in the one-step

GMM-DIFF estimation (column 2), supporting the view that this characteristic of the SME sector is more important than the size of this sector for economic growth in Brazil. The inclusion of the physical capital proxy again provides very similar results, in the sense that the the size of the SME sector is negatively related with growth and the human capital embodied in SMEs does not harm economic performance.

Table 4.18. SMEs' Human Capital and Growth - Manufacturing

	(1) LSDV	(2) GMM-DIFF 1-step	(3) GMM-DIFF 2-step	(4) GMM-SYS 1-step	(5) GMM-SYS 2-step	(6) LSDV	(7) GMM-DIFF 1-step	(8) GMM-DIFF 2-step	(9) GMM-SYS 1-step	(10) GMM-SYS 2-step
lnGDP_{t-1}	-0.532***	-0.708***	-0.744***	-0.215***	-0.219*	-0.584***	-0.759***	-0.672**	-0.196***	-0.201**
	(-7.68)	(-4.40)	(-3.06)	(-3.40)	(-1.73)	(-8.15)	(-5.15)	(-2.21)	(-2.65)	(-2.32)
ln*School*	0.393***	0.408***	0.414**	0.392***	0.399***	0.436***	0.456***	0.435***	0.387***	0.366**
	(3.67)	(3.29)	(2.26)	(4.18)	(2.63)	(4.08)	(3.10)	(2.79)	(4.18)	(2.01)
ln$(n+g+d)$	-0.0762	0.526	0.342	0.0198	0.109	0.00939	0.570	0.434	-0.0570	-0.0127
	(-0.48)	(1.47)	(0.54)	(0.16)	(0.54)	(0.06)	(1.43)	(1.26)	(-0.40)	(-0.06)
ln*SMER*	-0.136**	-0.434*	-0.278	-0.138*	-0.141	-0.120*	-0.393**	-0.292	-0.162**	-0.146
	(-2.20)	(-1.91)	(-0.75)	(-1.91)	(-1.33)	(-1.96)	(-2.16)	(-1.05)	(-2.12)	(-1.39)
ln*SMEH*	0.283	0.636*	0.697	0.206	0.352	0.104	0.367	0.130	0.207	0.263
	(1.30)	(1.66)	(1.33)	(1.20)	(1.09)	(0.46)	(0.98)	(0.26)	(1.26)	(0.46)
ln(*K*)						0.0624**	0.0713	0.0385	-0.00968	-0.00931
						(2.31)	(1.12)	(0.45)	(-0.58)	(-0.21)
Observations	150	125	125	150	150	150	125	125	150	150
Adjusted R^2	0.486					0.504				
Instruments		35	35	41	41		35	35	41	41
m_1		0.427	0.649	0.00946	0.0184		0.434	0.510	0.00852	0.0158
m_2		0.198	0.175	0.476	0.481		0.239	0.219	0.479	0.486
Hansen		0.810	0.810	0.985	0.985		0.880	0.880	0.993	0.993

Notes: *t* statistics in parentheses.* Significant at 10%; **significant at 5%; ***significant at 1%.The m_2 statistic is the autocorrelation test under the null that differenced errors are not serially autocorrelated. Hansen statistics tests for endogeneity under the null that instruments are valid. In all GMM instrumented estimations, all growth determinants are treated as potentially endogenous; therefore, for GMM-DIFF the instrument set contains observations on growth determinants dated *t*-2 and earlier, and for GMM-SYS the added set of instruments are the differenced growth determinants dated *t*-1 and earlier. All results include time dummies.

Therefore, the objective of this section is support the argument that regional interdependence is one important omitted factor in our micro-regional regressions. Results for the more aggregate regional level, states, seem to suffer less from endogeneity and the instruments created for our estimates are more likely to be exogenous. Also, the empirical results based on a panel of Brazilian states also indicate that the size of the SME sector has a negative effect on regional growth and that the SMEs' human capital seems to be more important for the process of economic growth.

4.8.2 Ranking of Micro-regions

Table 4.19 provides the rankings ordered by GDP per capita and royalties received by each micro-region. The micro-region of Macaé has the largest GDP per capita and is the second in terms of receipts from oil activities. Noticeably, we can see that its GDP per capita is about 13 times the average micro-regional GDP per capita and the huge gap separating this region from the twentieth region ranked in terms of GDP per capita. This illustrates how the oil economies can influence the GDP per capita distribution and influence the tail of Kernel Densities in Figure 4.1.

Table 4.19: Rankings of GDP per capita and Royalties

Rank for GDPpc	Micro-regions	GDPpc	Rank for Royalties	Micro-regions	Oil Royalties
1	Macaé	80.01	1	Campos dos Goytacazes	735500832.26
2	Bacia de São João	69.21	2	Macaé	450988019.25
3	São Jerônimo	34.53	3	Bacia de São João	335648682.87
4	Campos dos Goytacazes	30.30	4	Lagos	221820988.11
5	Parecis	19.23	5	Rio de Janeiro	202935301.05
6	Alto Araguaia	18.24	6	Caraguatatuba	77864509.83
7	Macau	18.16	7	Salvador	48996738.17
8	Suape	17.42	8	Mossoró	47596906.72
9	Vale do Paraíba (RJ)	17.09	9	Macau	45755604.74
10	Alto Teles Pires	16.50	10	Joinville	28488542.39
11	Lagos	16.03	11	Catu	28214759.60
12	Guaporé	14.36	12	Baía da Ilha Grande	26815396.66
13	Concórdia	13.94	13	Itaperuna	25979436.70
14	São José dos Campos	13.85	14	Linhares	25786149.46
15	Andradina	13.59	15	São Mateus	25648025.68
16	Não-Me-Toque	12.92	16	Japaratuba	23927455.23
17	Caxias do Sul	12.72	17	Aracaju	22437954.57
18	Quirinópolis	12.46	18	Osório	22227338.68
19	Brasília	12.43	19	Santo A. de Pádua	21924842.52
20	Paranaguá	12.27	20	Itaguaí	20381943.10
	Brazil (Average)	6.15			

Notes: GDPpc is the Real GDP per capita at constant prices in R$(thousands), 2000 as base year.
Revenues are the current value in R$(Reais) of received royalties from oil and natural gas for 2005.
The data for royalties can be obtained at http://www.royaltiesdopetroleo.ucam-campos.br/

4.8.3 An Overview of Brazilian States

Table 4.20: Overview of Brazilian Regions

States	GDPpc R$ (1000)	Electricity (per MWh)	Population (per million)	School (years)	*SMEH*250	*SMER*250
Acre	4.28	24.10	0.67	5.35	6.31	91.80
Alagoas	2.95	1841.97	3.00	4.23	7.47	13.01
Amazonas	6.48	1315.42	3.23	6.89	7.93	99.99
Amapá	4.63	23.23	0.59	7.38	9.69	35.15
Bahia	4.15	9370.87	13.78	5.06	8.38	62.82
Ceará	3.18	1862.19	8.10	5.05	7.87	49.24
Distrito Federal	21.75	357.69	2.33	8.96	8.80	68.80
Espírito Santo	8.73	3802.84	3.40	6.83	8.10	73.48
Goiás	5.67	2746.93	5.61	6.31	8.14	64.96
Maranhão	2.61	6633.42	6.10	4.50	7.18	80.59
Minas Gerais	6.31	23418.87	19.24	6.20	7.31	65.26
Mato Grosso do Sul	6.02	750.30	2.26	6.44	7.74	55.17
Mato Grosso	8.42	1134.82	2.75	6.23	7.75	64.23
Pará	3.54	9773.53	6.97	5.76	5.87	67.31
Paraíba	2.96	1019.87	3.60	4.95	7.33	53.07
Pernambuco	3.74	2227.20	8.41	5.57	8.37	61.83
Piauí	2.33	181.97	3.00	4.46	7.91	77.25
Paraná	7.78	8671.58	10.26	6.78	7.44	61.81
Rio de Janeiro	10.12	6532.66	15.20	7.73	8.29	59.73
Rio Grande do Norte	3.75	1019.96	3.00	5.53	7.45	51.93
Rondônia	5.30	220.93	1.53	5.46	7.87	87.03
Roraima	5.12	13.57	0.39	6.30	6.91	100.00
Rio Grande do Sul	8.39	8846.39	10.78	6.92	8.17	60.86
Santa Catarina	9.16	6969.10	5.87	7.14	8.05	60.22
Sergipe	4.30	1132.74	1.97	5.48	8.75	58.86
São Paulo	11.33	45733.05	40.44	7.55	7.33	54.29
Tocantins	4.38	111.02	1.31	5.63	7.79	82.23

Sources: IPEA and RAIS.

Chapter 5

SMEs and Economic Growth in Brazil: a Spatial Approach

5.1 Introduction

In Chapter 4 we use the augmented growth model to analyse the importance of the SME sector for economic growth at the micro-regional geographic level in Brazil. Overall, the size of the SME sector is not important for growth and the impact coming from human capital of SMEs seems to be the most important aspect of this sector affecting positively the economic performance. However, in Chapter 4, we ignore the potential spatial dependence, and the regressions are still prone to endogeneity after the estimation of models using instrumental variables, a result that can also be related with the omission of the spatial dimension.

The seminal work of Anselin (1988) demonstrate the need for modelling the spatial dependence using spatial econometrics due to the existence of spatial effects. As discussed in Anselin (1988), spatial effects in the form of spatial autocorrelation and spatial heterogeneity are the two main aspects that merit particular attention. Spatial autocorrelation can be defined as the coincidence of value similarity with location similarity, and spatial heterogeneity means that economic behaviour is not stable across geographic space, generating spatial regimes with clusters of high or low values of a variable. The omission of this spatial phenomenon in economic growth regressions can produce biased results as argued for instance in Rey and Montouri (1999). After these seminal works, spatial econometrics has been growing and empirical evidence suggests that it is an essential tool when we analyse economic growth for a set of regions. The empirical evidence from the literature shows that spatial dependence is a regular phenomenon, therefore, the use of spatial econometrics for growth regressions should be the standard procedure. For instance, the survey by Abreu et al. (2005) reviews empirical and theoretical works and shows that spatial dependence is an important aspect of the economic growth process.

Therefore, the aim of this chapter is to re-examine the importance of the SME sector for economic growth considering the importance of the space for regional economic growth in Brazil taking into account the possible spatial dependence. We will therefore examine whether the use of spatial econometrics has any effect on our previous results. The conclusions drawn from this chapter will be based on results that account for regional spillovers and feedbacks among regions and will increase our confidence when interpreting the relationship between SMEs and regional economic growth in Brazil. Also, a robustness test about the importance of space for regional economic growth is provided in the Appendix 5.9.8. An analysis of panel growth regressions after removing space from the data is provided and the results provided in the Appendix 5.9.8 provide additional evidence of the importance of the spatial structure for the process of economic growth in Brazil.

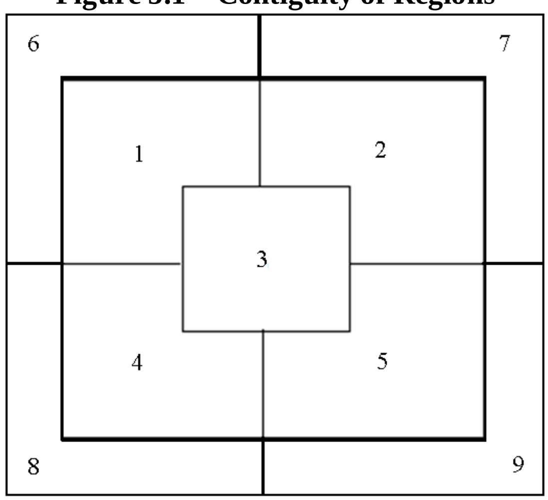
This chapter is structured as follows. In Section 5.2 we discuss how to model the spatial structure. In Section 5.3 we focus on the initial analysis of the spatial dependence. This is followed by Section 5.4 that explains how to consider the spatial dependence using spatial econometrics. In Section 5.5 we apply spatial econometrics in a cross-section set up and in Section 5.6 we present the model specification for a panel data setting. Section 5.7 discusses the results and Section 5.8 concludes.

5.2 Spatial Structure

To assess the existence of spatial effects, the first task is to quantify the spatial structure for Brazilian micro-regions in a form of spatial weight matrix. As noted by LeGallo and Ertur (2003) and Abreu et al. (2005), the choice of the spatial matrix is one of the most controversial issues in the spatial econometrics literature and should be made with caution as the choice of spatial weight can have a substantive impact on results. This is because the spatial matrix imposes a strong restriction on the data, assuming a rigid spatial pattern for the spatial dependence. It is difficult to find the spatial matrix that corrects for the true spatial dependence and the choice of a spatial weight is usually guided by the reasonable assumption that the spatial dependence declines with distance. The construction of the spatial structure should reflect the fundamental theorem of regional science, distance matters. Therefore, the strength of spatial dependence should decline with the geographical distance between observations.[62]

There are numerous ways of modelling the spatial structure and the simplest one is based on a binary matrix that captures the way regions are spatially connected. The matrix designed to model the spatial relationships is called spatial weight. The most frequently used type of binary weights is the Queen contiguity weights. In these spatial weights, regions that share a common border or vertices with other regions of interest are assigned the value of 1. On the other hand, regions that do not share borders or vertices are assigned the value of 0. Figure 5.1 below illustrates a contiguity of regions from which an example of the first-order Queen's contiguity spatial weight W is created.

Figure 5.1 – Contiguity of Regions



From this figure, consider first only 5 regions (from 1 to 5) to construct a 5x5 spatial weight matrix using the first-order contiguity relations. We associate rows of the matrix with the index i, and columns with the index j representing neighbouring regions to region i. For row one we set a value of one in columns 2, 3 and 4,

[62] This is the expression of the first law of geography (e.g. Tobler, 1970), where everything is related to everything else but near things are more related than distant ones.

reflecting the fact that regions two, three and four are first-order contiguous to region one. We repeat this procedure to each region, resulting in the matrix C shown below.

$$C = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

This is not the final spatial matrix W yet. The elements w_{ii} on the diagonal of the weight matrix are set to zero while the elements w_{ij} indicate the way region i is spatially connected to region j. For the purpose of forming a spatial lag of values from neighbours, we can normalize the matrix C to have a row-standardized matrix W, where the weight matrix is standardized such that the elements of a row sum up to one. The final row-standardized spatial weight matrix that has rows adding up to 1 is shown below.

$$W = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}$$

If we want to go further and assume that regions that are further away also have influence in a given region, we construct a Queen's contiguity spatial weight based on a second-order spatial autocorrelation. Now consider the outer ring of the regions represented in Figure 5.1 by regions 6, 7, 8 and 9. The same process assigning ones to neighbours of neighbours is considered to construct the spatial structure with two spatial lags to model space. In this case, for instance, the region 3 would be neighbour to all regions. Further order spatial matrices are constructed following the same procedure but taking further spatial lags to model the spatial structure. We constructed four spatial weights based on the Queen contiguity structure from the first to the fourth spatial lag order for the cumulative form of lags.

However, spatial weights constructed using this criterion can result in considerable variability in terms of number of neighbours for each region and create methodological problems (e.g. LeGallo and Ertur 2003). Therefore, the concept of *k-nearest* neighbours calculated from the distance between regions centroids is used to construct the row-standardised spatial weights as in LeGallo and Ertur (2003):

$$W = \begin{cases} w_{ij}^*(k) = 0 \text{ if } i = j \\ w_{ij}^*(k) = 1 \text{ if } d_{ij} \leq d_i(k) \text{ and } w_{ij}(k) = w_{ij}^*(k) \Big/ \sum_j w_{ij}^*(k) \\ w_{ij}^*(k) = 0 \text{ if } d_{ij} > d_i(k) \end{cases} \tag{5.1}$$

where $d_i(k)$ is a critical cut-off distance defined for each region *i*. It is the k^{th} order smallest distance between regions *i* and *j* in a manner that each region *i* has exactly *k* neighbours. In other words, we assign ones to the *k-nearest* neighbours of each region. Matrices based on 20, 40 and 60 nearest neighbours are constructed to check the robustness of our results. The spatial weight matrices for Queen contiguity and *k-nearest* neighbours are created using the software GeoDa 0.9.5-i.[63] All versions of spatial weight matrices are row-standardised. We have got the shapefile used to construct these weights for the Brazilian micro-regions from the Regional and Urban Economics Lab (NEREUS) from the University of São Paulo (USP).[64]

5.3 Methods of Exploratory Spatial Data Analysis

Following the construction of many versions of spatial weights, we now turn our attention to the Exploratory Spatial Data Analysis (ESDA). This is a set of measures that allows the detection of patterns of spatial dependence. To explore the spatial pattern of the data, two measures will be employed to check for global and local spatial autocorrelation. The global autocorrelation will be assessed by the Moran's *I* statistic and the Local Indicator for Spatial Autocorrelation (LISA) based on the

[63] GeoDa is a free software program for spatial data analysis developed by Dr. Luc Anselin. It can be downloaded for free at: http://geodacenter.asu.edu/software.

[64] The shapefile with all 558 micro-regions is edited to fit our data with 508 micro-regions according to data availability.

decomposition of the global Moran's *I* for each region will be used to check the local autocorrelation.

The Moran's *I* statistic gives a formal indication of the degree of spatial autocorrelation of a given variable and is given by the following expression:

$$I_G = \frac{\sum_i \sum_j w_{ij} z_{ij} z_{ij}}{\sum_i z_i^2} = \frac{Z'WZ}{Z'Z} \qquad (5.2)$$

where *Z* is the vector of a given variable in deviation from its mean and *W* is the spatial weight matrix with the elements w_{ii} on the diagonal set to zero whereas the elements w_{ji} indicate the way region *i* is connected to region *j*. This index gives a formal indication of the association between the original vector of variables *Z* and its spatially lagged transformation *WZ*. Table 5.1 reports the results of Moran's *I* statistic for all variables defined and used in the estimations in Chapter 4, using seven different spatial weight matrices described in Section 5.2. Regardless of the spatial structure imposed, all variables present a positive association between the original variable and its spatial lagged version. The spatial autocorrelation is stronger in the GDP per capital and human capital measures but is positive for all variables. These results are in line with the idea that the physical capital (intrinsically related to GDP per capita) and human capital spillover. It is also related to the neoclassical model prediction that physical and human capital move to areas where they can get higher rates of return, generating convergence (Chapter 2). What the spatial autocorrelation suggests is that, in general, physical and human capitals move following the spatial structure of the country.

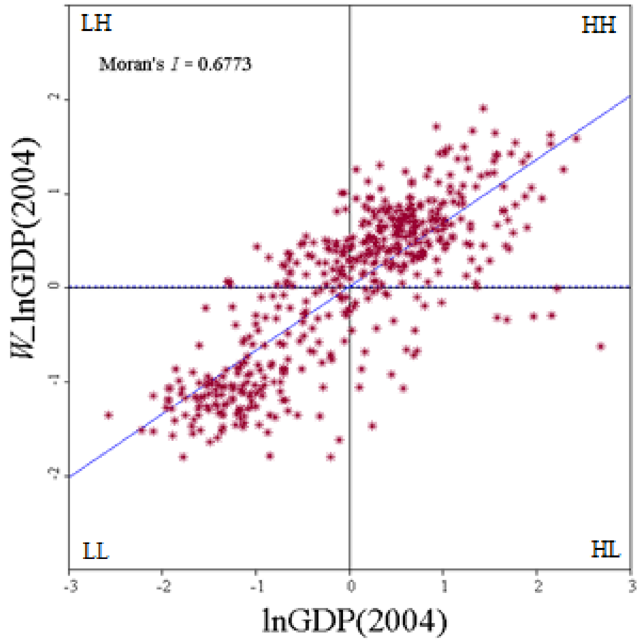
Table 5.1: Global Spatial Autocorrelation (Moran's *I*)

Weight Structure	lnGDP_{1980}	lnGDP_{2004}	ln(n+g+d)	ln*School*	ln*SMEH*	ln*SMER*
Queen 1st	0.6773	0.6723	0.4862	0.6270	0.3398	0.2901
Queen 2nd	0.6494	0.6143	0.3929	0.6066	0.3153	0.1925
Queen 3nd	0.6182	0.5642	0.3113	0.5802	0.2883	0.1594
Queen 4th	0.5930	0.5315	0.2324	0.5571	0.2582	0.1101
K-nearest (*k*=20)	0.6458	0.6143	0.3813	0.6110	0.3464	0.1529
K-nearest (*k*=40)	0.6231	0.5732	0.3021	0.5837	0.2991	0.1170
K-nearest (*k*=60)	0.6111	0.5553	0.2345	0.5700	0.2798	0.0801

Notes: Results are for the GDP per capita of 1980 and 2004. The results for the remaining variables are for their average between 1985 and 2004. The calculations were carried on using the software ***R***.

The global spatial autocorrelation may also be visualized graphically in the Moran scatterplot. Moran's *I* is formally equivalent to the slope coefficient of the linear regression of WZ on Z. A preliminary assessment of local spatial instability can also be studied by means of the Moran scatterplot (Anselin, 1996; Anselin and Bao, 1997). The four different quadrants of the scatterplot correspond to the four types of local spatial association between a region and its neighbors: (HH) represents a region with a high value surrounded by regions with high values in quadrant I; (LH) a region with low value surrounded by regions with high values in quadrant II; (LL) a region with a low value surrounded by regions with low values in quadrant III; and (HL) a region with high value surrounded by regions with low values in quadrant IV. Quadrants HH and LL refer to positive spatial autocorrelation indicating spatial clustering of similar values. Conversely, quadrants LH and HL refer to negative spatial autocorrelation, indicating spatial clustering of dissimilar values. The Moran scatterplot may thus be used to visualize atypical localizations and provide a visual impression on the overall stability of the global spatial pattern of dependence. Figure 5.2 illustrates the Moran scatterplot for the GDP per capita in 1980 using the first-order Queen contiguity weight matrix.

Figure 5.2 – Moran Scatterplot



We can observe the positive spatial association between the values of the log of GDP per capita in 1980. Relatively high (low) income regions tend to be located nearby other high (low) income regions more often than would be expected by random chances. Therefore, each region should not be viewed as an independent observation, as has been implicitly assumed in studies of regional growth convergence (Rey and Montouri, 1999). This pattern can be observed for all variables in the model and for all different spatial weight matrices.[65]

However, Moran's *I* ignores the existence of spatial instability in the sense that it can not assess the significance of the regional structure of spatial autocorrelation. To investigate formally whether there are local spatial clusters of high or low values that can contribute more (or less) to the global spatial autocorrelation, a measure of local autocorrelation is required. The local indicators of spatial association (LISA) allows for the decomposition of the global Moran's *I* into the contribution of each observation (Anselin, 1995). In this index, only neighbouring values of the regions are included. The index takes the following form:

[65] Note that the slope of Figure 5.2 matches the Moran's *I* reported in Table 5.1.

$$I_i = \frac{z_i \sum_j w_{ij} z_j}{\sum_i z_i^2} \qquad (5.3)$$

where variables are defined as in equation (5.2). LISA can be used to assess the significance of local spatial clustering around an individual location and for the identification of pockets of spatial nonstationarity (Anselin, 1995). Local spatial clusters can be identified as those locations or set of locations where LISA is significant.[66] Since there is a link between the local indicators and the global statistic, LISA outliers will be associated with regions which exert significant influence on Moran's *I*. Finally, combining the information of a Moran scatterplot with the significance of LISA yields the so called "Moran significance map", showing the regions with significant LISA and indicating by a colour code the quadrants they belong to in the Moran scatterplot (see Anselin 1996 and Anselin and Bao 1997). A positive value for I_i indicates spatial clustering of similar values, and negative values indicate cluster of dissimilar values. The cluster is classified as such when the value at a given location is more similar to its neighbours than would be the case under spatial randomness.

[66] Inference of LISA (local Moran) is based on the permutation approach (Anselin, 1995). It works applying an approach in terms of conditional permutation where the value z_i of the region *i* is fixed and other values are permuted on all other locations of the sample.

Figure 5.3. Local Moran's Significance Map for ln*GDP*(1980)

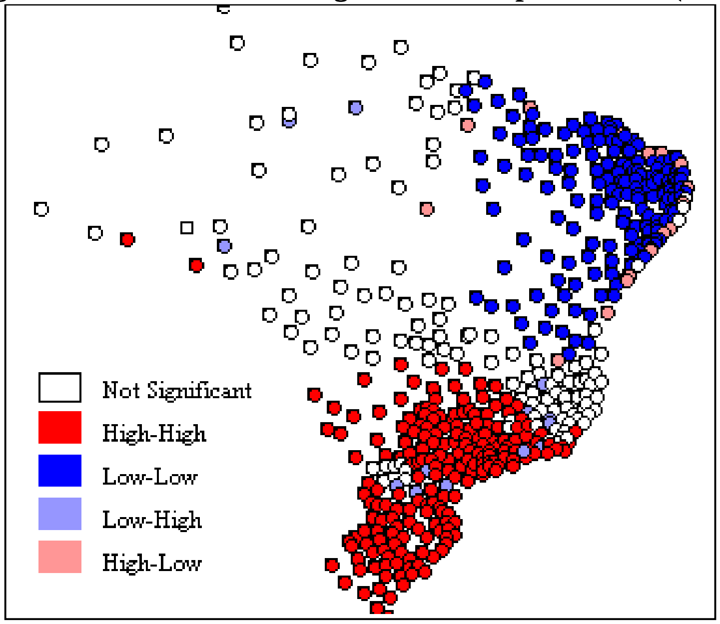


Figure 5.3 presents the significance map for the GDP per capita in 1980, and one important thing to notice is the presence of two strong regional clusters. The first is the presence of the cluster of high-high values in the southern part of Brazil.[67] The second cluster of low-low values dominates the spatial pattern in the northern part of the territory. The overall pattern of the local measures of spatial autocorrelation is positive and is driven by the majority of micro-regions falling in quadrants I and III in Figure 5.2. There seems to be an intrinsic polarization in the layout of the spatial regimes resulting in a clear north-south polarization[68]. These results for micro-regions support findings of Silveira-Neto and Azzoni (2006) that use a local indicator of spatial association and suggest the existence of two geographical income clusters in Brazil, one cluster encompassing South and Southeast states, and another cluster of low-low values grouping states in the Northeast part of the territory.

[67] To compare the LISA map with the location of the oil production in Brazil, Figure 3.2 in Chapter 3 provides a map of oil exploration in Brazil. The largest portion of the oil is off-shore and the red arrow in the south-east indicates the region that concentrates the largest oil reserves of the country.

[68] Alternative LISA maps for the GDP per capita in 1980 and 2004 using different spatial weights are reported in the Appendix 5.9.1 and confirm the regularity of the pattern observed in Figure 5.3. LISA maps are constructed using GeoDa 0.9.5-I.

These exploratory results suggest that care must be taken when we estimate the standard growth regression presented in Chapter 4 due to the presence of spatial autocorrelation and spatial heterogeneity. In this case, standard regressions that do not consider spatial effects are likely to be biased and spatial econometrics is required.

5.4 Model Specification with Spatial Dependence

The previous section provides evidence of the existence of spatial autocorrelation and spatial heterogeneity that can lead to misleading estimations when spatial econometrics is not considered. In this section the attention turns to a spatial econometric analysis of growth regressions in order to provide insights to this question. Conventional growth regressions assume that regional observations are independent. However, there is a growing consensus that regional income growth models exhibit spatial dependence (LeSage and Fisher, 2008). Therefore, spatial econometrics is necessary to provide consistent estimations. What follows is an exposition of the specifications and interpretations of spatial econometrics models that will be considered to re-examine the relationship between regional economic growth and SMEs in case of spatial dependence. This exposition is based mainly on Anselin (1988) and LeSage and Pace (2009).

A large number of model specifications for spatial processes have been suggested in the literature but the focus in this chapter is on the most common models in spatial econometrics according to LeSage and Pace (2009): the spatial error model (SEM), the spatial autoregressive model (SAR) and the spatial Durbin model (SDM).

5.4.1 The Spatial Error Model

We can use the Spatial Error Model (SEM) to reflect spatial dependence working through the disturbance process. The SEM specification based on the baseline Equation (4.1) presented in Chapter 4 to analyse the importance of SMEs for economic growth takes the following form:[69]

$$gr_t = \alpha - \ln(y_0)\beta + X\psi_t + \varepsilon_t \qquad (5.4)$$
$$\varepsilon_t = \lambda W \varepsilon_t + v_t$$

where *gr* denotes the annual GDP per capita growth in a cross-section setting, $\ln(y_0)$ is the initial GDP per capita, and the vector *X* encompasses growth determinants as discussed in Chapter 2 and 4. These growth determinants are the overall level of human capital ln(*School*), population growth adjusted for depreciation and technological progress $\ln(n+\delta+g)$ (under the usual assumption that $\delta+g$ equals 0.05), the relative size of the SME sector (*SMER*) and the human capital of the SME sector (*SMEH*). The term ε_t is the error term where λ is a scalar spatial error coefficient, and $v \sim N(0, \sigma^2 I_n)$. Concerning the error process, this specification means that a random shock introduced into a specific region will not only affect the growth rate in that region but, through the spatial transformation in the error term, will impact the growth rates of other regions (Rey and Montouri 1999, Ertur et al. 2006).

The spatial econometric literature has shown that OLS estimation is inappropriate for models incorporating spatial effects. For the SEM specification, Anselin (1988) shows that the effect of spatial residual autocorrelation on the properties of OLS estimation is in line with the effect of time series residual autocorrelation. Parameters will be unbiased, but inefficient due to the non-spherical structure of the disturbance variance matrix. Therefore, inferences about the convergence process should be based on the SEM specification estimated via the maximum likelihood estimator. Anselin (1988) derives the maximum likelihood function for the SEM specification that is reported in the Appendix 5.9.2.1.

---

[69] Equation (4.1) is given by: $gr_{it} = a_i - b\ln y_{i,t-1} + \beta \ln SMER_{it} + \alpha \ln SMEH_{it} + \psi \ln X_{it} + v_{it}$

5.4.2 The Spatial Autoregressive Model

In time series models, lagged values of the dependent variable are often included to account for missing explanatory variables. A similar motivation can be used for spatial lags of the dependent variable (e.g. Abreu et al. 2005). In this specification, the spatial autoregressive structure (where a spatial lag of the dependent variable is included in the right-hand side of the specification) can be combined with a conventional regression model to produce a spatial extension of the linear regression model called Spatial Autoregressive Model (SAR) that takes the following form:

$$gr_t = \alpha + \rho W gr_t - \ln(y_0)\beta + X\psi_t + \nu_t \qquad (5.5)$$

where ρ is the scalar spatial autoregressive parameter and all other terms are as previously defined in equation 5.4.

Regarding the interpretation of this specification, Abreu et al. (2005) provide a guide on how to interpret SAR results. They demonstrated that the total marginal effect of an increase in one dependent variable in the case of a SAR model can be decomposed into a direct marginal effect that is not region specific and an indirect and induced marginal effect that is region specific. The latter effect is generated by the spatial interactions introduced by the spatial weight W that links all regions in the system. In the case of a SAR model, Abreu et al. (2005) show that only the direct marginal effect on the dependent variable (growth rate) of a marginal change in the initial GDP per capita is represented by β, and not the total marginal effect as in the case of a non-spatial model where β would represent the total marginal effect directly. Arbia et al. (2008), LeSage and Fisher (2008) and LeSage and Pace (2009) also discuss the interpretation of SAR type models. Additionally, the presence of the lagged dependent variable incorporates a sort of spatial conditional convergence as argued in Arbia et al. (2008); convergence would be conditioned and influenced by the spatial structure. LeSage and Pace (2009) argue that this specification is a hallmark of spatial econometrics, and this set up can be used to provide extended versions of SAR models, such as the Spatial Durbin Model (SDM) described in the next sub-section.

Anselin (1988) and LeSage and Pace (2009) demonstrate that the OLS estimator will be biased and inconsistent for the parameters of a SAR model due to the simultaneity in the nature of the spatial autocorrelation process caused by the introduction of the spatial lag. The spatial lag is an average of neighbouring values and leads to a connectivity relation where the covariance of the error term between two regions is not zero. As an alternative, we estimate equation 5.5 using the maximum likelihood function for the SAR specification proposed by Anselin (1988) that is reported in the Appendix 5.9.2.1.

5.4.3 The Spatial Durbin Model

The Spatial Durbin Model (SDM) provides a generalization of the conventional growth regression incorporating the spatial lag of the dependent variable and all other conditioning variables. The model provides the basis for a growth regression that is sufficiently general to allow for three types of spatial interdependencies in the growth process: first, spatial effects working through the dependent variable, second, spatial effects working through the initial income variable, and third, spatial effects working through a set of conditioning variables (LeSage and Fisher, 2007). This specification takes the following form:

$$gr_t = \alpha + \rho W gr_t - \ln(y_0)\beta_1 + W\ln(y_0)\beta_2 + X_t\psi_1 + WX_t\psi_2 + \nu_t \quad (5.6)$$

where all variables are defined as in Section 5.4.2, with the addition of vector ***WX*** used to account for the spatial lagged values of conditioning variables in vector ***X***. This set of variables represents the explanatory variables constructed as averages from neighbouring observations. As a result, this specification includes spatial lags of the dependent and independent variables and is suitable to capture externalities and spillovers arising from different sources (e.g. Ertur and Koch 2007). As before, ρ is the scalar spatial autoregressive parameter, and we expect this coefficient to be positive indicating that regional growth rates are positively related to growth rates of neighbouring regions. Following the interpretation of the SDM model by Elhorst and Fréret (2009), the coefficient of the spatially lagged dependent variable in the SDM model indicates the magnitude by which economic growth rates change in reaction to the growth rates in the neighbouring regions. The coefficients on the original independent variables represent the direct effect of those variables on growth and the spatially lagged independent variables capture the spillover effects of those variables on growth (the spatial effect working through the conditioning variables). The coefficients capture the interaction effects of a given independent variable with the values for this variable in the neighbouring regions.

The SDM is the unrestricted form of a model with cross-sectional dependence in the errors and appears as the nesting model in a more general approach of model selection. The SDM is an attractive specification from the econometric point of view because it nests the alternative specifications presented in the previous sections. Specifically, the SDM specification can be expressed as a SEM or SAR model when certain restrictions are applied. When $\psi_2 = 0$ and $\beta_2 = 0$ we have the SAR model and if $-\rho\psi_1 = \psi_2$ and $-\rho\beta_1 = \beta_2$ the model is reduced to the SEM model where the spatial structure is modelled only in the error term (e.g. Anselin 1988; Ertur and Koch 2007; Bivand and Brunstad 2006).

The nesting characteristics of the SDM is particularly useful when we want to compare this specification to the SEM model. Analogous to the time series case, the common factor approach is based on the equivalence of two model specifications, one expressed with a spatially autoregressive error term (SEM) and the other in terms of spatially lagged model (SDM), according to the restrictions described above. Reversing the flow of reasoning, if we multiply both sides of the expression for the SEM model with the inverse term that contains the error structure, we obtain the SDM model.[70] Following Anselin (1988), the test on the coefficient constraint can be carried out by means of the Likelihood Ratio test under the null of the common factor hypothesis.[71] As Fingleton and López-Bazo (2006) highlight, studies of convergence using European data have often ended up with a spatial error

[70] See Appendix 5.9.5 for derivation details.
[71] The LR test that can be used to test for the common factor are described in the Appendix 5.9.4.2.

specification, despite the fact that this indicates that spatial externalities are not a substantive phenomena, but rather random shocks diffusing through space. Therefore, the common factor hypothesis will also be useful to provide some indication on how the spatial dimension operates between Brazilian micro-regions.

Furthermore, LeSage and Fisher (2008) and LeSage and Pace (2009) provide a discussion about the motivations and advantages of the SDM specification for growth models. They show that the use of a SDM specification rests on the plausibility of two circumstances that are likely to arise in applied regional spatial growth regressions, the spatial dependence in the disturbances of an OLS regression and endogeneity in the form of the existence of an omitted explanatory variable (that follows a spatial autoregressive process) that exhibits non-zero covariance with the variables in the model. Therefore, these plausible circumstances observed in applied spatial growth regressions make the SDM model a natural econometric choice over competing alternatives. Elhorst and Fréret (2009) also prefer an unconstrained SDM model to test whether there is evidence of political yardstick competition in France.

5.5 Cross-Section Results

This section reports the results for the spatial econometric specifications described in Section 5.4 for a cross-section set up. Firstly, we present the OLS results together with diagnostic tests designed to formally detect spatial dependence in the regressions. Secondly, we present the regression results for the spatial models described earlier. Finally, we repeat the same procedure for two regional sets of micro-regions to check whether the heterogeneity of the spatial patterns will provide different interpretations for the role of the SME sector in the process of economic growth.

5.5.1 Cross-Section Results for Brazil

In the previous sections we showed that all variables used in our model exhibit positive spatial autocorrelation that will probably make the OLS estimation unreliable. As shown later in this section, the spatial dependence is confirmed in the OLS estimates by diagnostic tests; therefore, spatial econometrics is employed to provide more consistent results. In this section, we present the OLS and spatial regression results for the model that includes only one aspect of the SME sector expressed by the size of this sector in terms of the employment share (*SMER*). All estimations reported in this section were performed in the software ***R***.[72]

The OLS estimates in Table 5.2 (column 1) show a negative sign of the coefficient on initial GDP per capita (indicating convergence) and on population growth. The coefficient on human capital is positive, confirming that this aspect is significant for economic growth in Brazil. Furthermore, the coefficient of the size of the SME sector (*SMER*) is negative and suggests that the size of this sector is not important for growth. These results are similar to those presented in Chapters 4. However, the OLS regressions are not reliable in the case of spatial dependence. To assess formally the presence of spatial dependence in the OLS regression we first report the Moran's *I* test statistic for the regression residuals. The Moran's *I* has a positive value of 0.3012, indicating the presence of spatial dependence in the residuals.

After detecting the spatial autocorrelation in the residual, we use the Lagrange Multiplier tests to test the OLS model against the alternative SEM and SAR models under the null of no spatial dependence. In relation to the spatial error model as the alternative, the LM_{ERR} and its robust version (LMR_{ERR}) are reported, whereas for the spatial lag model the LM_{LAG} and its robust version (LMR_{LAG}) are reported. If LM_{LAG} is more significant than LM_{ERR}, then the SAR model is the most appropriate model. Alternatively, in case both non-robust LM tests (LM_{LAG} and LM_{ERR}) are signiticants and LMR_{LAG} is more significant than LMR_{ERR}, the SAR model is the most appropriate model. Otherwise, following the same reasoning, the best choice is the SEM model. Turning to the results in Table 5.2, this decision rule suggests that the spatial dependence is better modelled by the SEM specification. More details on the LM tests are provided in the Appendix 5.9.4.1.

[72] For more information on this software follow this link, http://www.r-project.org/. We used the package *spdep* maintained by Roger Bivand (http://cran.r-project.org/web/packages/spdep/spdep.pdf) to perform the spatial analysis.

The results for the SAR and SEM specifications using the spatial weight Queen 1 are reported in Table 5.2 in columns 2 and 3, respectively.[73] Both models lead to the same qualitative analysis. We observe a significant convergence process and a negative and not significant sign of the population growth coefficient. The coefficient on human capital is positive and significant. Turning to the SMEs' aspects, the size of the SME sector is negatively related with economic growth, reinforcing the idea that the size of the SME sector is not positively related to economic growth.

Table 5.2 Cross-Section Results: Non-Spatial versus Spatial Models

	OLS (1)	**SAR** (2)	**SEM** (3)	**SDM** (4)
$\ln GDP_{t-1}$	-0.0229***	-0.0195***	-0.0263***	-0.0282***
	(-10.9350)	(-9.8073)	(-12.9269)	(-13.6028)
$\ln(n+d+g)$	-0.0089**	-0.0036	0.0014	0.0058
	(-2.2690)	(-1.0103)	(0.2967)	(1.1914)
$\ln SCHOOL$	0.0313***	0.0262***	0.0209***	0.0190***
	(7.9570)	(7.3355)	(5.7182)	(5.1382)
$\ln SMER$	-0.0028*	-0.0026*	-0.0045***	-0.0048***
	(-1.6690)	(-1.7247)	(-2.8127)	(-3.049)
ρ (SAR) / λ(SEM)		0.47293***	0.6674***	0.58361***
		(9.4544)	(15.476)	(11.965)
(Intercept)	-0.0275**	-0.0155	0.0167	-0.0147
	(-2.3810)	(-1.4868)	(1.1718)	(-1.1181)
W* $\ln GDP_{t-1}$				0.0166***
				(4.4065)
W* $\ln(n+d+g)$				-0.0083
				(-1.2221)
W* $\ln SCHOOL$				0.0010
				(0.147)
W* $\ln SMER$				0.0033
				(1.2569)
Observations	508	508	508	508
Log likelihood (LIK)	1333.32	1371.017	1394.615	1405.844
AIC	-2656.63	-2728	-2775.2	-2789.7
LR test		75.404(0.0000)	122.6(0.0000)	107.84(0.0000)
Wald test		89.386(0.0000)	239.52(0.0000)	143.17(0.0000)
LR COMFAC				22.4579(0.0002)
Moran's *I*	0.3012(0.0000)			
LM_{ERR}	131.085(0.0000)			
LMR_{ERR}	37.1108(0.0000)			
LM_{LAG}	95.8206(0.0000)			
LMR_{LAG}	1.8465(0.1742)			

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics and for the diagnostic tests are the levels of significance.

[73] Throughout this section, we only report the estimations using the Queen 1 spatial matrix but results are qualitatively similar when we use of alternative spatial weights. Regressions using alternative spatial weights are reported in the Appendix 5.9.6 and obtain similar qualitative results.

In the SAR specification the scalar spatial autoregressive parameter or the coefficient of the spatially lagged dependent variable is positive and significant, indicating the presence and magnitude of spatial autocorrelation in the dependent variable. It indicates that economic growth rates change 0.47% in reaction to a one percentage point increase in the growth rates of neighbouring regions.

The Likelihood Ratio (LR) and Wald tests compare the classic OLS regression against the alternative SAR model. Both tests reject the null pointing out to the model including the spatial lag.[74] These results are similar to those of the LM tests indicating the need to use of spatial econometrics.

In the SEM specification the spatial dependence is modelled as a random shock introduced into a specific region that will not only affect the growth rate in that region but will impact the growth rates of other regions. Similarly, the LR and Wald tests are employed to test the OLS specification against the alternative SEM model. Both tests reject the null and are in line with the LM tests that indicate the use of the SEM specification.

However, as discussed in Section 5.4.3, the spatial dependence in the disturbances of an OLS regression and endogeneity in the form of an omitted explanatory variable (that follows a spatial autoregressive process) that exhibits non-zero covariance with the variables in the model make the SDM model a natural econometric choice over competing alternatives (LeSage and Fisher 2008). These arguments are particularly useful in supporting the choice of the SDM over the SEM model once these models are intrinsically linked through the spatial common factor described earlier. However, to formally compare the two specifications we rely on the results of the Likelihood Ratio test for the common factor (LR COMFAC). A significant result for the LR COMFAC suggests that the SDM does perform better than the SEM and that the common factor constraints should be rejected. A test on the common factor hypothesis is a test on the constraints that each coefficient of a spatially lagged explanatory variable (e.g. W**SCHOOL*) equals the negative of the product of the spatial autoregressive coefficient (p) and the matching regression coefficient of its non-lagged variable (*SCHOOL*), as explained in Section 5.4.3.[75] For

[74] For more details on the LR and Wald tests please refer to the Appendix 5.9.4.
[75] In spdep, this can be implemented by means of the LR.sarlm function.

more details on the LR COMFAC test see the Appendix 5.9.5. The results for the LR COMFAC suggest that the common factor hypothesis is rejected with a p-value of 0.0002. This supports the view of LeSage and Fisher (2008) that recommends the SDM as the best specification. As in their arguments, the likely presence of omitted variables in our study (for instance, regional information on physical capital is not available) leads to a better fit of the SDM over the SEM specification.

Results for the SDM estimates are reported in Table 5.2 (column 4), and the initial level of GDP per capita has a coefficient with opposite sign from that associated with the spatial lags of this variable, a result also found by LeSage and Fisher (2008). Therefore, the positive coefficient on the spatial lag of initial GDP per capita level indicates that higher levels of GDP per capita in neighbouring regions positively impact economic growth in the home region. The coefficients related to the population growth also exhibit opposite signs, although without statistical significance. The coefficient on population growth is positive, a result contrary to the Solow growth model prediction but in line with the literature about Brazil (e.g. Nakabashi and Salvato 2007). As discussed in previous chapters this result might be related to the endogeneity between economic growth and population growth. The negative coefficient on the spatial lag of population growth indicates that higher population growth in neighbouring regions is negatively related to economic growth in the home region. This result also supports the idea that this variable is endogenous with respect to growth. If a given region is doing better than its neighbours, the population from the neighbouring regions migrate to this region; this pattern is reflected in the negative sign of the spatially lagged coefficient for the population growth. The average years of schooling variable is positively correlated with growth, indicating that education is important for growth. However, the spatially lagged coefficient of human capital is insignificant, indicating lack of positive spatial spillovers (from one region to the other) of human capital on economic growth. This result was also found for the overall level of human capital in LeSage and Fisher (2008) and Resende (2009). Higher levels of human capital are positively related with growth and the negative sign of the spatial lag of this variable indicates that a given region does not benefit from a higher level of human capital in the neighbouring region.[76] With respect to the size of the SME sector represented by *SMER*, the results show that it is negatively correlated with growth and the spatially lagged coefficient of this variable is positive, although insignificant.

We also present statistics to compare the fit of the OLS estimations with their spatial counterparts. These measures are the log likelihood, and the Akaike information criterion (AIC). The higher the value of the log-likelihood (LIK) functions, the better the fit of the model. For the remaining information criteria, the lower the measure, the better is the fit. The LIK and AIC measures confirm that models considering the spatial effects achieve a better fit than the OLS estimation. In particular the SDM performs better and this is an additional support to the LR COMFAC test that indicated the use of the SDM model.

5.5.1.2 Cross-Section Results: Extension with Human Capital

As in Chapter 4, we also extend the spatial analysis to include human capital of the SME sector, *SMEH*. The results are reported in Table 5.3. When we introduce *SMEH*, the Moran's *I* test statistic for residual in the OLS regression has a positive value of 0.3044, indicating the presence of spatial autocorrelation in the residual. In addition, the Lagrange Multiplier tests suggest that the spatial dependence is better modelled by the SEM specification, as in the case of the estimates reported in Table 5.2. Overall, the inclusion of *SMEH* does not change the conclusions drawn from the Table 5.2. Again, we observe a significant convergence process and a insignificant coefficient on the population growth variable. The coefficient on human capital is positive and significant and the size of the SME sector is negatively related with economic growth. Therefore, the effect of these variables on growth is robust to the inclusion of the human capital of SMEs.

Turning our attention to the additional variable *SMEH*, we observe that this variable is positive and significant in the SEM specification, indicating that this aspect of SMEs is more important than the size of this sector.

---

[76] This result might also be related with the migration process in Brazil, qualified workers seem to be attracted to one location and the concentration of more qualified people in this region does not translate into human capital spillovers to neighbouring regions.

When we test for the spatial common factor, testing the SDM against the SEM specification, the LR COMFAC test rejects the null and suggests the use of the SDM. This is in line with Table 5.2 and with the argumentation of LeSage and Fisher (2008) that this specification is the most appropriate one for regional growth regression studies. The results in column 4 also imply robustness of the results from the model without *SMEH* in the sense that we draw a similar qualitative interpretation. We can note, for instance, that the size of the SME sector is negatively correlated with growth and the interaction effects of this aspect of the SME sector across space is not statistically significant, although it presents a positive coefficient.

Table 5.3 – Cross-Section Results: Non-Spatial versus spatial models

	OLS (1)	**SAR** (2)	**SEM** (3)	**SDM** (4)
$\ln GDP_{t-1}$	-0.0227***	-0.0193***	-0.0262***	-0.0282***
	(-10.7750)	(-9.6642)	(-12.9887)	(-13.7729)
$\ln(n+d+g)$	-0.0079***	-0.0025	0.0043	0.0084*
	(-1.9470)	(-0.681)	(0.9214)	(1.7178)
ln*SCHOOL*	0.0291***	0.0237***	0.0143***	0.0123***
	(6.3020)	(5.6955)	(3.3852)	(2.897)
ln*SMER*	-0.0027	-0.0025	-0.0044***	-0.0048***
	(-1.5960)	(-1.6399)	(-2.8022)	(-3.1021)
ln*SMEH*	0.0046	0.0051	0.0125***	0.0127***
	(0.9090)	(1.1112)	(2.8364)	(2.8896)
ρ (SAR) / λ(SEM)		0.4739***	0.6882***	0.5946***
		(9.4795)	(16.58)	(12.416)
(Intercept)	-0.0304**	-0.0193	0.0109	-0.0083
	(-2.5360)	(-9.6642)	(0.7592)	(-0.6111)
W* $\ln GDP_{t-1}$				0.0167***
				(4.5078)
W* $\ln(n+d+g)$				-0.0137*
				(-1.9313)
W* ln*SCHOOL*				0.0115
				(1.4815)
W* ln*SMER*				0.0032
				(1.2264)
W* ln*SMEH*				-0.0236***
				(-2.8198)
Observations	508	508	508	508
Log likelihood (LIK)	1333.73	1371.635	1398.5	1412.139
AIC	-2655.47	-2727.3	-2781.0	-2798.3
LR test		75.804(0.0000)	129.53(0.0000)	113.37(0.0000)
Wald test		89.861(0.0000)	274.9(0.0000)	154.15(0.0000)
LR COMFAC				27.2778(0.0000)
Moran's *I*	0.3045(0.0000)			
LM_{ERR}	133.9354(0.0000)			
LMR_{ERR}	39.8905(0.0000)			
LM_{LAG}	96.4793(0.0000)			
LMR_{LAG}	2.4343(0.1187)			

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics and for the diagnostic tests are the levels of significance.

In addition, human capital of SMEs has a coefficient with opposite sign of those associated with the spatial lags of this variable. Higher levels of human capital in SMEs are positively related with growth, and the negative (and insignificant) sign of the spatial lag of this variable indicates that regions do not benefit from a higher level of human capital in neighbouring regions. This result was also found for the overall level of human capital in LeSage and Fisher (2008) for the European NUTS II and in Resende (2009) for the case of the Brazilian micro-regions. Our results might also be related with the migration process in Brazil, qualified workers seem to be attracted to one location and the concentration of more qualified people in this region does not translate into human capital spillovers to neighbouring regions. As a result, SMEs in regions with higher level of human capital only attract more skilled labour from neighbouring regions and do not generate any spillover for those regions. Therefore, the presence of SMEs does not generate economic growth but the level of human capital in the SMEs seems to be important for regional economic growth in Brazil, although there are no positive human capital externalities in the SME sector. The presence of more qualified workers is likely to be related to higher productivity and higher wages.[77] The estimation of Mincer equation for Brazil supports the idea that human capital and wages are positively related (e.g. Moura, 2008).

The results presented in this chapter are similar to those from the non-spatial estimations in Chapter 4, where we found indication that human capital of the SME sector is more important than the size of this sector. Nevertheless, as discussed in Chapter 4, Brazil is a country with considerable cross-regional asymmetries. Therefore, pooling all micro-regions together could provide difficulties in drawing inferences for region-specific public policies. The next section presents results for two different sets of regions with different level of development to observe whether there are different dynamics in the way the SME sector affects economic growth.

[77] The link between productivity and wages is in line with the neoclassical model presented in Chapter 2, which assumes that labour and capital factors are the major inputs into production - the derivative of the production function with respect to labour is equal to real wage, and as firms maximize profits, this should be equal to labour productivity (Solow, 1956).

5.5.1.3 Cross-section Results for Two Spatial Regimes

In Chapter 4 we used the distribution of the GDP per capita to identify different dynamics across regions. We found two different regimes stemming from a bimodal distribution of the GDP per capita. A lower club formed mainly by municipalities of Northeast and North regions and a higher club formed mainly by municipalities of South, Southeast and Centre-west regions. However, as in all results generated in Chapter 4, we assume that there is no spatial dependence that affects the coincidence of value similarity with location similarity. In other words, we assume that the value of the GDP per capita in one region is independent of its location.

However, as indicated in this chapter, micro-regions are likely to be more similar to the geographically closest micro-regions than to those further away. Therefore, considering the two clubs found in Chapter 4 for the regional analysis would ignore the presence of spatial dependence. To assess the existence of space heterogeneity and different clubs taking into account the spatial dependence we again use the LISA statistic (introduced in Section 5.3) as in Rey and Montouri (1999) and Le Gallo and Ertur (2003). A positive value for I_i indicates spatial clustering of similar values and negative values indicate cluster of dissimilar values. A cluster is classified as such when the value at a location is more similar to its neighbours than would be the case under spatial randomness. Figure 5.3 in Section 5.3 suggests the existence of two strong regional clusters. The first is the presence of the cluster of high-high values in the southern part of Brazil and the second cluster of low-low values dominates the spatial pattern in the northern part of the territory.

There seems to be an intrinsic polarization in the layout of the spatial regimes resulting in a clear north-south polarization. This finds support in the findings of Silveira-Neto and Azzoni (2006) for Brazilian states. This pattern is observed regardless of the spatial weights we use[78]. Despite the fact that each matrix provides a different LISA map, we always observe the same north-south polarization pattern. Therefore, we consider two alternative samples for the regional analysis. The first regional sample comprises 207 contiguous micro-regions in the northern part of

[78] Alternative LISA maps for the GDP per capita in 1980 and 2004 using different spatial weights are reported in the Appendix 5.9.1.

Brazil (Northeast and North) and the second sample comprises 301 contiguous micro-regions in the southern part (South, Southeast and Centre-west) of the country. This is in line with LeGallo and Dall'erba (2006) that estimated spatial growth regressions for two spatial regimes in the European Union NUTS II.

The same analysis is performed for these two spatial clusters and results for the northern sample is reported below in Table 5.4. As the results for the model without *SMEH* are robust to the inclusion of this measure, we only comment on results of the full model from Columns 5 to 8. As in the case of the country as a whole, the LM tests and the Moran`s *I* for the OLS residual autocorrelation indicate the need to use models that include the spatial dimension, particularly the SEM specification. Overall, the SEM model presents results qualitatively similar to those found for Brazil as a whole. We observe a significant convergence process, a positive sign of the population growth coefficient, and a positive and significant sign of the coefficient on human capital. When we turn to the SME sector, the size of *SMER* is negatively but not significantly related with economic growth, indicating that this factor is not important for this process. Additionally, the coefficient on *SMEH* is positive, indicating that this aspect of SMEs is more important than the size of this sector.

We also use the common factor hypothesis to test the SDM against the SEM specification. The LR COMFAC test does not reject the null of the common factor constraint in this case. The diagnostic test suggests the use of the SEM model but it is known from LeSage and Fisher (2008) that the SDM specification is the most appropriate for regional growth regression studies and is the only method amongst the three presented in this chapter that is suitable to capture externalities and spillovers working through the independent variables included in the model. It is interesting to note that either the SEM or SDM models provide similar results for the top five coefficients of the results in Table 5.4. For instance, results reported in columns 7 and 8 of the non-spatially lagged variables (top five coefficients) present similar results.

Analysing the results from the SDM specification we observe that as in the case of the country as a whole, the human capital of SMEs has coefficients with an opposite sign from those associated with the spatial lags of this variable. Higher levels of human capital in SMEs are positively related with growth and the negative sign of the spatial lag of this variable indicates that a given region does not benefit from higher level of human capital in the neighbouring regions. This result indicates that SMEs in regions with higher human capital only attract more skilled labour from neighbouring regions and do not generate any spillover for those regions. Both the SME sector size and the spatial interaction effect of this variable are negative and not significant, indicating that there is no aspect of the size of the SME sector improving the economic performance.

Table 5.4 – Cross-Section Results: Non-Spatial versus spatial models (Northern Regions)

	OLS (1)	SAR (2)	SEM (3)	SDM (4)	OLS (5)	SAR (6)	SEM (7)	SDM (8)
lnGDP_{t-1}	-0.0277***	-0.0234***	-0.0261***	-0.0255***	-0.0272***	-0.0229***	-0.0257***	-0.0259***
	(-8.4450)	(-7.5351)	(-8.1776)	(-7.8069)	(-8.2380)	(-7.3536)	(-8.1416)	(-8.0333)
ln(n+d+g)	0.0102	0.0157**	0.0134*	0.0141*	0.0122	0.0177**	0.0169**	0.0173**
	(1.3020)	(2.2079)	(1.7444)	(1.7671)	(1.5300)	(2.4647)	(2.1943)	(2.198)
ln*SCHOOL*	0.0211***	0.0159***	0.0165***	0.0163***	0.0175***	0.0123**	0.0102*	0.0105**
	(4.2530)	(3.5034)	(3.5706)	(3.4899)	(3.0720)	(2.3772)	(1.9585)	(2.0296)
ln*SMER*	-0.0028	-0.0021	-0.0015	-0.0009	-0.0027	-0.0019	-0.0014	-0.0011
	(-1.5000)	(-1.2149)	(-0.7966)	(-0.4654)	(-1.4250)	(-1.1335)	(-0.7402)	(-0.5724)
ln*SMEH*					0.0069	0.0070	0.0115**	0.0108**
					(1.2690)	(1.4241)	(0.01503)	(2.254)
ρ (SAR) / λ(SEM)		0.43046***	0.5175***	0.4995***		0.4320***	0.5477***	0.5245***
		(5.5936)	(6.3854)	(6.0645)		(5.6321)	(7.0114)	(6.5873)
(Intercept)	0.0329	0.0458**	0.0456**	-0.0035	0.0304	0.0433**	0.0422*	-0.0019
	(1.4600)	(2.2428)	(2.0582)	(-0.0934)	(1.3450)	(2.1227)	(1.9235)	(-0.0504)
W* lnGDP_{t-1}				0.0085				0.0098
				(1.3245)				(1.5505)
W* ln(n+d+g)				-0.0120				-0.0199
				(-0.8517)				(-1.3896)
W* ln*SCHOOL*				0.0079				0.0161
				(0.8203)				(1.5712)
W* ln*SMER*				-0.0025				-0.0021
				(-0.7816)				(-0.6651)
W* ln*SMEH*								-0.02125**
								(-2.0632)
Observations		207	207	207		207	207	207
Log likelihood (LIK)		570.0719	572.7184	574.745		571.0871	575.5597	578.7681
AIC		-1126.1	-1131.4	-1127.5		-1126.2	-1135.1	-1131.5
LR test		27.925(0.000000)	33.218(0.0000)	31.667(0.0000)		28.303(0.0000)	37.248(0.0000)	35.49(0.0000)
Wald		31.288(0.0000)	40.773(0.0000)	36.778(0.0000)		31.72(0.0000)	49.16(0.0000)	43.393(0.0000)
LR COMFAC				4.0532(0.3989)				6.417(0.2677)
Moran's *I*	0.2753(0.0000)				0.2871(0.0000)			
LM_{ERR}	42.6284(0.0000)				46.3411(0.0000)			
LMR_{ERR}	7.7125(0.0055)				10.5363(0.0012)			
LM_{LAG}	35.1643(0.0000)				35.818(0.0000)			
LMR_{LAG}	0.2485(0.6182)				0.0132(0.9086)			

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics and for the diagnostic tests are the levels of significance.

The results for the southern part of the country are presented in Table 5.5. The first thing to notice is that the results for the model without *SMEH* are robust to the inclusion of this measure, therefore, we only comment on the results obtained by the full model. In columns 5 we observe that the LM tests and the Moran`s *I* for the OLS residual autocorrelation indicate the use of the SEM specification. This specification presents results that are qualitatively similar to those found in Table 5.3 for Brazil and 5.4 for the northern sample. We observe a significant convergence process, a positive sign of the population growth coefficient, and the coefficient on human capital is positive and significant.

Regarding the SME sector we find that the size of the SME sector is negatively related with economic growth. For the augmented model with SMEs' human capital, the coefficient on *SMEH* is also positive but not significant. This indicates that this aspect of SMEs is not as important for growth as it is for other samples but we find the same indication that the quality of SMEs expressed by its human capital is more important than the size of the SME sector in the process of economic growth.

Turning to the common factor hypothesis, the LR COMFAC test does not reject the null of common factor again. The diagnostic tests for both regional sub-samples suggest the use of the SEM specification. However, the SEM and SDM model provide similar results for the non-spatially lagged variables. Nevertheless, we follow LeSage and Fisher (2008) and stick to the SDM specification arguing that it is the most appropriate method for regional growth regression studies and focus on the interpretation of results presented in column 8. There is no significant indication that higher levels of human capital in SMEs are positively related with growth and the negative and insignificant sign of the spatial lag of this variable indicates that a given region does not benefit from higher level of human capital in the neighbourhood.

Therefore, overall, the regional results are in line with the analysis for the country as a whole in the sense that the size of the SME sector does not seem to be positively correlated with economic growth in all samples. On the other hand, this sector seems to affect growth via the human capital embodied in SMEs, evidenced by the positive sign on the coefficient of this variable in the SEM and SDM specifications. However, SDM results show a negative sign on the spatial lag of *SMEH* in all samples and indicate that a given region does not benefit from higher level of human capital in the neighbouring regions, with the results for the southern regions being insignificant. This result again suggests that *SMEH* has no spatial externality effect on economic growth, regardless of the sample analysed.

Nevertheless, although the general conclusion is similar, some differences emerge in the regional analysis. For example, in terms of the level of significance and magnitudes of the coefficients related to the SME sector and different externalities coming from the size of the SME sector. Audretsch and Keilbach (2007), for instance, argue that entrepreneurship is linked to cultural phenomena that are clustered in space. Using regional data for Germany and spatial econometrics, they provide evidence that an increase in entrepreneurial activity in the adjacent regions increase entrepreneurship in a given region. Furthermore, cultural aspects of entrepreneurship are diffused slowly (if diffused) to regions that do not share those cultural characteristics. Therefore, differences in the regional empirical results are in line with the idea that entrepreneurship and SMEs are locally driven by cultural differences that impact entrepreneurial activities differently. Besides, as argued in Baumol (1990) and Dias and MacDermott (2006), better institutions provide positive incentives to productive entrepreneurship, in this sense, different regional results might also be related with differences in regional institutions.

However, the introductory results presented in this section are only based on cross-section estimations, limited binary spatial weights, and the manufacturing sector. It is well known that cross-section is less informative than panel data that contain more variation and less collinearity among the variables in the model. Additionally, as suggested by Beck et al. (2005a), the analysis of the commerce and services sectors would add to the analysis. Therefore, the next section is designed to provide a better and more complete analysis by providing estimations of spatial panel regressions (using a distance based spatial weight) for manufacturing and the extended SME sector including commerce and services.

Table 5.5 – Cross-Section Reuslts: Non-Spatial versus spatial models (Southern Regions)

	OLS (1)	SAR (2)	SEM (3)	SDM (4)	OLS (5)	SAR (6)	SEM (7)	SDM (8)
ln*GDP*$_{t-1}$	-0.0302***	-0.0285***	-0.0318***	-0.0318***	-0.0302***	-0.0285***	-0.0318***	-0.0317***
	(-9.7460)	(-10.0618)	(-11.8436)	(-11.6728)	(-9.7370)	(-10.0598)	(-11.8269)	(-11.6556)
ln(*n*+*d*+*g*)	-0.0013	0.0017	0.0001	0.0004	-0.0020	0.0017	0.0009	0.0015
	(-0.2610)	(0.378)	(0.009)	(0.0708)	(-0.3730)	(0.3492)	(0.1596)	(0.2302)
ln*SCHOOL*	0.0278***	0.0249***	0.0198***	0.0185***	0.0295***	0.0250***	0.0169**	0.0144*
	(4.2430)	(4.2754)	(3.0613)	(2.7483)	(3.7010)	(3.5308)	(2.116)	(1.7409)
ln*SMER*	-0.0105***	-0.0122***	-0.0144***	-0.0141***	-0.0107***	-0.0122***	-0.0142***	-0.0139***
	(-3.1060)	(-4.052)	(-4.9777)	(-4.8562)	(-3.1200)	(-3.9975)	(-4.8733)	(-4.7498)
ln*SMEH*					-0.0043	-0.0001	0.0070	0.0104
					(-0.3710)	(-0.0142)	(0.6032)	(0.8666)
ρ (SAR) / λ(SEM)		0.4963***	0.6296***	0.6042***		0.49624***	0.6322***	0.60087***
		(8.3643)	(10.606)	(9.8163)		(8.3632)	(10.69)	(9.7209)
(Intercept)	0.0088	0.0126	0.0249	0.0019	0.0122	0.0127	0.0186	0.0060
	(0.5010)	(0.8102)	(1.2125)	(0.0966)	(0.6150)	(0.7232)	(0.805)	(0.2714)
W* ln*GDP*$_{t-1}$				0.0203***				0.0203***
				(3.6906)				(3.6733)
W* ln(*n*+*d*+*g*)				-0.0001				-0.0026
				(-0.0119)				(-0.278)
W* ln*SCHOOL*				-0.0037				0.0030
				(-0.3362)				(0.2343)
W* ln*SMER*				0.0156**				0.0149**
				(2.5554)				(2.383)
W* ln*SMEH*								-0.0169
								(-0.9517)
Observations		301	301	301		301	301	301
Log likelihood (LIK)		824.082	836.56	837.9788		824.0822	836.741	838.5058
AIC		-1634.2	-1659.1	-1654		-1632.2	-1657.5	-1651.0
LR test		50.956(0.0000)	75.912(0.0000)	68.574(0.0000)		50.816(0.0000)	76.134(0.0000)	67.344(0.0000)
Wald		69.962(0.0000)	112.48(0.0000)	96.36(0.0000)		69.943(0.0000)	114.29(0.0000)	94.496(0.0000)
LR COMFAC				2.8376(0.5854)				3.5298(0.6189)
Moran's *I*	0.3323(0.0000)				0.3304(0.0000)			
LM_{ERR}	93.9258(0.0000)				92.8932	0.0000		
LMR_{ERR}	34.6757(0.0000)				33.8461	0.0000		
LM_{LAG}	61.2949(0.0000)				60.923	0.0000		
LMR_{LAG}	2.0447(0.1527)				1.8758	0.1708		

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics and for the diagnostic tests are the levels of significance.

5.6 Model Specification with Spatial Dependence for Panel Data

The empirical results reported and discussed in the previous sections are based only on cross-section estimates. However, panel data models have many advantages over the cross-section, as pointed out in the case of the non-spatial panel analysis in Chapter 4. Panel data are generally more informative because they contain more information, variation and less collinearity among the variables in the model. In the context of growth models, Islam (1995) argues that panel data estimations allow for time specific and individual specific levels of technology, the same argument is valid in the spatial context. A panel data approach would presume that spatial heterogeneity is a feature of the data (as indicated by the local Moran's *I*) and attempts to model that heterogeneity. The need to account for spatial heterogeneity comes from the fact that spatial units have individual characteristics that are space-specific and time invariant (Elhorst 2005). Cross-section estimates cannot control for spatial and time fixed effects, which means the results might be biased (Elhorst and Fréret 2009).

Spatial panel models only recently found application in regional convergence studies, and the application of this setting is at a very early stage (Arbia et al. 2008). Nonetheless, developments from Elhorst (2003, 2005 and 2010) provided the derivation of the maximum likelihood functions for spatial panel data. He shows that estimations of spatial models with fixed effects can be carried out with standard techniques developed by Anselin (1988) and Anselin and Hudak (1992) after demeaning the variables in the model. The fixed effect is considered by taking each variable included in the regression in deviation from its mean as in the within estimator (equivalent to the LSDV), a process called demeaning. Elhorst (2010) presents the maximum likelihood functions used to estimate spatial panel models along with a link to Matlab codes to estimate spatial panel models associated to this work on his website.[79] Our results for the SEM, SAR and SDM specifications for panel data are based on Elhorst's codes for Matlab.

[79] Elhorst's codes are available at <www.regroningen.nl/elhorst/software.shtml>. To be able to run his codes we also needed to download LeSage's routines to estimates spatial models at <www.spatial-econometrics.com> (LeSage 1999).

The spatial lag model (SAR) for panel data where the dependent variable depends on the dependent variable observed in neighbouring regions and on a set of control variables takes the following form:

$$gr_{it} = \rho \sum_{j=1}^{N} w_{ij} gr_{jt} - \ln(y_{t-1,i})\beta + \psi X_{it} + \alpha_i + \mu_t + \nu_{it} \qquad (5.7)$$

where ρ is the spatial autoregressive parameter, w_{ij} contains information about the spatial structure and connectivity between regions *i* and *j*, *gr* denotes the *GDP* annual per capita growth, ln y_{t-1} is the initial GDP per capita, β the convergence coefficient, and *X* represents a vector of control variables as described in Equation 5.4. Additionally, the terms α_i and μ_i denote the spatial and time specific fixed effects, respectively.

The spatial error model, on the other hand, considers that the dependent variable depends on a set of control variables and on the random error terms that are correlated across space. The version of Equation 5.4 for a panel data setting is as follows:

$$gr_{it} = -\ln(y_{t-1,i})\beta + X\psi_{it} + \alpha_i + \mu_t + \varepsilon_{it} \qquad (5.8)$$

$$\varepsilon_{it} = \lambda \sum_{j=1}^{N} w_{ij} \varepsilon_{jt} + \nu_{it}$$

where variables are defined as in Equation 5.7, λ is a scalar spatial error coefficient, ε is the error term, and $\upsilon \sim N(0, \sigma^2 I_n)$.

Finally, the unconstrained spatial Durbin model with spatial fixed effects takes the following form:

$$gr_{it} = \rho \sum_{j=1}^{N} w_{ij} gr_{jt} - \ln(y_{t-1,i})\beta_1 + \sum_{j=1}^{N} w_{ij} \ln(y_{t-1,j})\beta_2 + X_{it}\psi_1 + \sum_{j=1}^{N} w_{ij} X_{jt}\psi_2 + \alpha_i + \nu_{it} \qquad (5.9)$$

where everything is defined as in Equation 5.7, with the inclusion of the spatially lagged values of all conditioning variables. In all estimations we eliminate the spatial specific and time specific effects by demeaning all variables in the model.[80]

Another important difference in the panel data section is the use of a more sophisticated spatial weight matrix. Thus far, we only modelled space using binary weight matrices. However, these matrices are only a simple representation of the spatial structure and do not represent the real dimension of distance. To overcome these limitations, we use a weight matrix based on road distances from each pair of our 508 micro-regions.[81] This spatial weight represents real physical quantitative linkages between any given pair of regions. The weight matrix is expressed as an inverse distance matrix to account for the intuition of the first law of geography that states that closer regions are more related than further regions and is defined as:

$$W = \begin{cases} w_{ij} = 0 \text{ if } i = j \\ w_{ij} = \dfrac{1}{d_{ij}} 0 \text{ if otherwise} \end{cases}$$

where d_{ij} denotes the geographical distance between regions i and j.[82]

5.7 Panel Data Results

This section reports the alternative results for the spatial panel estimations for Brazilian micro-regions. Firstly, we present the results for the country as a whole analysing the SME sector in manufacturing and in manufacturing, commerce and

[80] See Appendix 5.9.3.
[81] We would like to thank Prof. Eduardo Haddad from the University of São Paulo (USP) for providing the road distances for each pair of Brazilian micro-regions. The alternative results for binary weights are reported in the Appendix 5.9.7.
[82] The weight matrix W is standardized by its largest characteristic root (or eigenvalue) to comply with the stationarity requirements without losing the original quantitative information on distances, since the row standardization process would cause the loss of the economic interpretation based on inverse distance (see Anselin 1988, pp. 23-24 and Elhorst 2010, pp. 379-380). The inverse distance row standardized version (Inv.Dist.Row-std) is presented in the Appendix 5.9.7 together with the binary ones and provides similar results.

services. Secondly, we perform the same analysis considering the northern and southern set of regions independently as in the cross-section analysis in Section 5.5.

5.7.1 Results for Brazil

Table 5.6 reports the results for Brazil considering only the SMEs in the manufacturing sector. We report the model without SMEs' human capital from Columns 1 to 3 and the augmented model with *SMEH* from Columns 4 to 6. As in the case of the cross-section estimates, the results are robust to the inclusion of *SMEH*. For this reason, although we report the results for the model without *SMEH*, we will comment only on the set of regressions that include the human capital aspect of SMEs throughout this section.

It is well known in the literature that panel models with fixed effects produce stronger convergence than the cross-section regression or the pooled panel estimations, as we can observe in the results from Chapter 4 (e.g. Islam 1995, Arbia et al. 2008). In the case of spatial growth models, Arbia et al. (2008) found that panel models produce faster convergence than cross-section estimates in the case of regional growth in Europe. We observe the same pattern in our estimates, with stronger convergence (captured by the coefficient of lagged GDP per capita) observed in our panel estimates.

Estimations reported from Columns 4 to 6 present some results that are in line with those from the cross-section estimates. We observe a significant convergence process and a population growth coefficient that is not negatively significant as predicted by the Solow model. The coefficient on human capital is positive and significant, confirming the importance of investments in education at micro-regional level. Turning the attention to the SME sector, the results show that the size of the SME sector is negatively related with economic growth. Furthermore, the coefficient on *SMEH* is positive and significant, indicating that this aspect of SMEs is more important than the size of this sector. The scalar spatial error coefficient, λ, or the spatial autocorrelation parameters, ρ, are positive and significant, confirming the importance of spatial effects in our panel.

Table 5.6 – Panel Results for Brazil

	SEM (1)	SAR (2)	SDM (3)	SEM (4)	SAR (5)	SDM (6)
lnGDP_{t-1}	-0.1415***	-0.1366***	-0.1427***	-0.1416***	-0.1367***	-0.1430***
	(-31.770)	(-31.2797)	(-31.2982)	(-31.8737)	(-31.3718)	(-31.4456)
ln*SCHOOL*	0.0915***	0.0625***	0.1184***	0.0832***	0.0532***	0.1108***
	(6.3951)	(5.6321)	(7.2650)	(5.7299)	(4.6361)	(6.7592)
ln($n+d+g$)	-0.0100	-0.0079	-0.0124	-0.0096	-0.0075	-0.0119
	(-1.2816)	(-1.0180)	(-1.5316)	(-1.2322)	(-0.9647)	(-1.4873)
ln*SMER*	-0.0088***	-0.0084***	-0.0092***	-0.0089***	-0.0086***	-0.0093***
	(-2.8486)	(-2.7312)	(-2.9702)	(-2.8871)	(-2.7767)	(-3.0072)
ln*SMEH*				0.0199***	0.0186***	0.0209***
				(3.3125)	(3.0848)	(3.4829)
W* lnGDP_{t-1}			0.1372***			0.1412***
			(3.1392)			(3.2482)
W* ln*SCHOOL*			-0.1999***			-0.1742**
			(-4.7305)			(-2.3812)
W* ln($n+d+g$)			0.0307			0.0306
			(0.2430)			(0.2426)
W* ln*SMER*			0.0778*			0.0788*
			(1.8025)			(1.8325)
W* ln*SMEH*						-0.0568
						(-0.6095)
ρ (SAR) / λ(SEM)	0.9210***	0.8939***	0.9220***	0.9350***	0.8920***	0.9290***
	(32.4425)	(20.0275)	(30.5760)	(42.0976)	(19.4252)	(36.4897)
R^2	0.5879	0.6128	0.6194	0.5895	0.6146	0.6218
log-likelihood (LIK)	3444.799	3434.3895	3450.3988	3450.3647	3439.1397	3456.4932
No Observations	2032	2032	2032	2032	2032	2032
LR_{FE}	1095.8824	1148.0944	1093.3590	1104.6548	1157.3870	1099.1166
LR_{FE} (*p*-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the t-statistics.

Table 5.6 also presents the likelihood ratio test to test for the existence of time and spatial specific effect (LR_{FE}). It tests the existence of spatial fixed effects only against the specification with both, time and spatial fixed effects. The results throughout this section always suggest the consideration of both fixed effects. More importantly, we observe that the fit of the SDM specification expressed by the R^2 and log-likelihood is the best among the three spatial alternatives.[83] This is also in line with the argumentation of LeSage and Fisher (2008) that this specification is the most appropriate for regional growth regression studies. As discussed in Section 5.4.3, LeSage and Fisher (2008) and LeSage and Pace (2009) suggest the use of a SDM specification based on the plausibility of two circumstances that are likely to arise in applied regional spatial growth regressions, the spatial dependence in the disturbances of an OLS regression and endogeneity in the form of the existence of an omitted explanatory variable (that follows a spatial autoregressive process) that exhibits non-zero covariance with the variables in the model. Therefore, these

[83] This result is observed in all estimations presented in this section.

plausible circumstances observed in applied spatial growth regressions make the SDM model a natural econometric choice over competing alternatives.

In the SDM estimates, the size of the SME sector is negatively correlated with growth but the interaction of this aspect of the SME sector across space is positive and statistically significant for growth, suggesting that the interaction effects of the size of the SME sector with the size of this sector in neighbouring regions contribute to economic growth. The spatial panel regressions seem to suggest that the fact that there are more SMEs next door incites productive entrepreneurship and new economic possibilities that affect economic performance positively. This is in line with Audretsch and Keilbach (2007) who argue that entrepreneurial activities have a more positive influence on neighbouring regions than on distant regions due to cultural aspects, for instance. In addition, human capital of SMEs has coefficients with the opposite sign from those associated with the spatial lags of this variable. Higher levels of human capital in SMEs are positively related with growth and the negative sign of the spatial lag of this variable indicates that a given region does not benefit from a higher level of human capital in the neighbouring region. This result was also found for the overall level of human capital in LeSage and Fisher (2008) and Resende (2009). Therefore, the presence of SMEs does not generate economic growth directly but through spatial interactions. Small businesses in a given region seem to benefit from SMEs in neighbouring regions. Additionally, the level of human capital in SMEs seems to be important for regional economic growth in Brazil, although there is no human capital spillover in the SME sector.

Thus far, the results presented in this chapter have been based only on the small business manufacturing sector. However, Beck et al. (2005a) suggested that it would be informative to have information on SME employment beyond manufacturing. Therefore, as in Chapter 4 we replicate the analysis completed for the manufacturing sector reported in Table 5.6 for an extended small business sector including commerce and services. As in Chapter 4, the new variables of interest are the employment share in manufacturing, commerce and services (*SMER2*) and the human capital of this sector (*SMEH2*).

Table 5.7 – Panel Results for Brazil – All Sectors

	SEM	SAR	SDM	SEM	SAR	SDM
ln*GDP*$_{t-1}$	-0.1425***	-0.1377***	-0.1435***	-0.1426***	-0.1378***	-0.1449***
	(-32.1026)	(-31.6182)	(-31.5733)	(-32.1109)	(-31.6292)	(-31.7586)
ln*SCHOOL*	0.0865***	0.0587***	0.1114***	0.0867***	0.0596***	0.1137***
	(6.0514)	(5.2917)	(6.8497)	(6.0685)	(5.3412)	(6.9950)
ln($n+d+g$)	-0.0118	-0.0098	-0.0137*	-0.0115	-0.0095	-0.0136*
	(-1.5089)	(-1.2720)	(-1.7029)	(-1.4639)	(-1.2269)	(-1.6890)
ln*SMER2*	-0.0287***	-0.0286***	-0.0286***	-0.0285***	-0.0284***	-0.0283***
	(-5.1599)	(-5.1505)	(-5.1536)	(-5.1318)	(-5.1203)	(-5.1045)
ln*SMEH2*				0.0095	0.0101	0.0116
				(0.6798)	(0.7363)	(0.8190)
W* ln*GDP*$_{t-1}$			0.13917***			0.1808***
			(3.1568)			(3.9962)
W* ln*SCHOOL*			-0.1871***			-0.2896***
			(-4.4273)			(-5.1566)
W* ln($n+d+g$)			0.0461			0.0093
			(0.3637)			(0.0733)
W* ln*SMER2*			0.1179			0.1429*
			(1.5992)			(1.9236)
W* ln*SMEH2*						-0.5205***
						(-2.7932)
ρ (SAR) / λ(SEM)	0.9219***	0.8890***	0.9250***	0.9220***	0.8870***	0.9320***
	(33.0247)	(18.6883)	(23.5260)	(33.0270)	(18.2889)	(52.3488)
R^2	0.5917	0.6163	0.6226	0.5918	0.6164	0.6242
log-likelihood (LIK)	3453.9855	3443.8461	3458.7499	3454.2176	3444.1163	3462.6524
Observations	2032	2032	2032	2032	2032	2032
LR_{FE}	1106.1970	1165.6934	1100.2601	1098.3175	1143.4475	1102.4475
LR_{FE} (*p*-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.7 reports the results for the model with the expanded SME sector. The empirical results are similar to those presented in Table 5.6, especially for the coefficients that are not related to the SME sector. In the SDM specification, for instance, the inclusion of the extended SME sector confirms the results in terms of sign, significance and quantitative effect of the variables other than SMEs' variables on growth. Regarding the SMEs' measures, we find the same qualitative results, that is, the negative effect stemming from the size of the SME sector, and a positive (weak) effect of the SME's human capital on growth. The difference is that the negative coefficient of the SME sector size and growth is twice as large when compared with results in Table 5.6, and the effect of SMEH2 is positive but not significant when the extended SME sector is considered. These results are similar to those found in the non-spatial estimations of Chapter 4. They indicate that human capital in SMEs is more important for the SME manufacturing sector than for the

SME sector including commerce and services. This is in line with the results presented in Chapter 4 and as discussed in that chapter, it is not surprising that the extended SME sector that includes commerce and services sectors consistently suggest that small businesses are negatively related to economic growth. This is because most of the commerce and services sectors are made of small high street shops that are likely not to be more productive than the manufacturing sectors. Thus, the results suggest that this pattern is robust and is observed in both spatial and a-spatial specifications.

5.7.2 Regional Results

The results from the local Moran index of spatial autocorrelation reported in Section 5.3 reveals the existence of spatial clusters in the northern and southern parts of the Brazilian territory. This confirms the presence of spatial heterogeneity that might cover some distinctive patterns specific for those groups of regions that are not captured by the estimates for the country as a whole. Therefore, we divide our sample into two sub-samples (southern and northern) and perform identical analyses separately for each set of regions.

Table 5.8 reports the results for the northern sample that comprises micro-regions of the north and northeast regions of Brazil. Looking at the SDM specification, that presents the best fit according to the log-likelihood, we observe that the size of the SME sector is negatively but not significantly related with economic growth, indicating that this factor is not important for this process. The coefficient on *SMEH* is positive, indicating that this aspect of SMEs is more important than the size of this sector.

The proxies related to the SME sector have coefficients with the opposite sign from those associated with the spatial lags of these variables. Higher levels of human capital in SMEs are positively related with growth and the negative sign of the spatial lag of this variable indicates that a given region does not benefit from higher level of human capital in the neighbouring regions, while the size of the SME sector is negative and the positive sign of the spatial lag of this variable indicates that a given region benefits from entrepreneurial activities in neighbouring regions.

Table 5.8 – Panel Results for the Northern Sample - Manufacturing

	SEM	SAR	SDM	SEM	SAR	SDM
ln*GDP*$_{t-1}$	-0.1402***	-0.1356***	-0.1444***	-0.1405***	-0.1359***	-0.1450***
	(-20.2876)	(-19.8703)	(-20.3797)	(-20.4666)	(-20.0464)	(-20.6646)
ln*SCHOOL*	0.1031***	0.1087***	0.0957***	0.0950***	0.1000***	0.0928***
	(5.2068)	(5.5930)	(4.6366)	(4.7910)	(5.1242)	(4.5081)
ln(*n*+*d*+*g*)	-0.0104	-0.0108	-0.0074	-0.0105	-0.0109	-0.0066
	(-1.0896)	(-1.1495)	(-0.7678)	(-1.1129)	(-1.1725)	(-0.6945)
ln*SMER*	-0.0049	-0.0033	-0.0053	-0.0050	-0.0034	-0.0049
	(-1.2240)	(-0.8029)	(-1.3145)	(-1.2585)	(-0.8415)	(-1.2241)
ln*SMEH*				0.0217***	0.0212***	0.0196***
				(3.2936)	(3.1855)	(2.9895)
W* ln*GDP*$_{t-1}$			0.1365**			0.1329**
			(2.0959)			(2.0570)
W* ln*SCHOOL*			0.0901			0.3281
			(0.5099)			(1.6400)
W* ln(*n*+*d*+*g*)			-0.0900			-0.1011
			(-0.8626)			(-0.9779)
W* ln*SMER*			0.1310***			0.14802***
			(2.6685)			(3.0107)
W* ln*SMEH*						-0.2704***
						(-2.9076)
ρ (SAR) / λ(SEM)	0.7440***	0.6109***	0.7249***	0.7499***	0.6350***	0.7050***
	(8.8498)	(5.9268)	(8.4257)	(9.1189)	(6.5036)	(7.7106)
R^2	0.6527	0.6598	0.6694	0.6565	0.6641	0.6761
log-likelihood (LIK)	1394.4878	1390.1122	1400.6039	1399.8741	1395.1222	1409.3048
Observations	828	828	828,	828	828	828
LR_{FE}	444.5653	446.3480	456.3466	453.2225	455.8539	467.5599
LR_{FE} (*p*-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Compared with the panel data results presented in Table 5.6, we reach the similar qualitative conclusions: the size of the SME sector is negatively related with economic growth and the coefficient on *SMEH* is positive and significant, indicating that this aspect of SMEs is more important than the size of this sector. However, we can observe some differences in terms of magnitude in the results for the northern regions. The SME sector is negatively but not significantly related with economic growth, indicating that the negative impact of this factor on growth is weaker than in the case of the country as a whole, and the spatial lagged coefficient for the SMEs' human capital is strongly significant. Finally, we can also note that the spillover effect from the dependent variable and the autocorrelation in the error term are weaker than for the country as whole (and weaker than for the southern regions described later in this section).

Table 5.9 – Panel Results for the Northern Sample – All Sectors

	SEM-All	SAR	SDM	SEM-All	SAR	SDM
lnGDP_{t-1}	-0.1407***	-0.1358***	-0.1446***	-0.1406***	-0.1356***	-0.1467***
	(-20.5400)	(-20.0483)	(-20.5839)	(-20.5060)	(-20.0101)	(-20.8722)
ln*SCHOOL*	0.0934***	0.0987***	0.0876***	0.0932***	0.0985***	0.0866***
	(4.7223)	(5.0694)	(4.2430)	(4.7128)	(5.0573)	(4.2190)
ln($n+d+g$)	-0.0125	-0.0131	-0.0096	-0.0126	-0.0133	-0.0086
	(-1.3288)	(-1.3950)	(-1.0005)	(-1.3334)	(-1.4174)	(-0.9068)
ln*SMER2*	-0.0266***	-0.0245***	-0.0259***	-0.0268***	-0.0247***	-0.0254***
	(-4.1541)	(-3.7803)	(-4.0214)	(-4.1576)	(-3.8001)	(-3.9515)
ln*SMEH2*				-0.0025	-0.0061	0.0019
				(-0.1520)	(-0.3737)	(0.1206)
W* lnGDP_{t-1}			0.1461**			0.1633**
			(2.1867)			(2.4373)
W* ln*SCHOOL*			0.1807			0.2443
			(0.9425)			(1.2773)
W* ln($n+d+g$)			-0.0575			-0.1582
			(-0.5352)			(-1.4216)
W* ln*SMER2*			0.2017**			0.2377***
			(2.4553)			(2.8864)
W* ln*SMEH2*						-0.6409***
						(-3.2637)
ρ (SAR) / λ(SEM)	0.7510***	0.6180***	0.7419***	0.7500***	0.6289***	0.7050***
	(9.1657)	(6.1770)	(9.1033)	(9.1194)	(6.4468)	(7.6842)
R^2	0.6581	0.6653	0.6752	0.6581	0.6655	0.6789
log-likelihood	1402.2671	1396.8685	1407.5458	1402.2783	1396.9294	1412.9032
No Observations	828	828	828	828	828	828
LR_{FE}	444.3886	447.0690	454.6217	438.9956	443.3526	458.0846
LR_{FE} (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

We also present the results for the extended SME sector (commerce, services, and manufacturing) in Table 5.9. The estimation of the SDM model shows that the coefficient on the size of the SME sector is negative and strongly significant, indicating that *SMER* does not foster economic growth directly. In addition, *SMEH* is positive but not significantly related to economic growth. This evidence provides additional support to previous results that suggest that the presence of SMEs in the manufacturing sector is less harmful to economic growth than the presence of small businesses in the commerce and services sectors. The spatially lagged variables present similar qualitative results when comparing to the Table 5.8.

The alternative results for the southern part of the country considering the SME manufacturing sector are presented in Table 5.10. The results for the model including *SMEH* are presented from columns 4 to 6. Compared to the northern regions and Brazil (as a whole), the strongest similarity is that we found the same indication that human capital in SMEs is more important than the size of the SME sector. The coefficient on *SMEH* is positive, although not significant, and the *SMER* is negatively related with economic growth. Additionally, proxies related to the SME sector have insignificant coefficients associated with the spatial lags of these variables, indicating that a given region does not benefit from higher level of human capital in neighbouring regions or from higher levels of entrepreneurial activities in SMEs in neighbouring regions. The diagnostic tests indicate that the SDM provides the best fit and that the model with time and spatial fixed effects is the best option. It is also worth noting that all spatial specifications provide the same interpretation in terms of non-lagged variables and the presence of spatial autocorrelation. The alternative results for the extended SME sector that includes commerce and services are presented in Table 5.11. The negative coefficient of the SME sector size increases (doubles) when comparing with results in Table 10, a pattern also observed in the results for the northern regions and Brazil. Therefore, the presence of SMEs seems to be less negatively associated with economic growth in the manufacturing sector.

Table 5.10 – Panel Results for the Southern Sample - Manufacturing

	SEM	SAR	SDM	SEM	SAR	SDM
lnGDP_{t-1}	-0.1461***	-0.1421***	-0.1498***	-0.1464***	-0.1423***	-0.1500***
	(-24.9920)	(-24.6219)	(-25.1782)	(-25.0417)	(-24.6716)	(-25.1922)
ln*SCHOOL*	0.1362***	0.1098***	0.1912***	0.1289***	0.1037***	0.1844***
	(5.7832)	(5.1247)	(6.6941)	(5.3631)	(4.7626)	(6.3098)
ln($n+d+g$)	-0.0092	-0.0058	-0.0184	-0.0075	-0.0039	-0.0173
	(-0.6541)	(-0.4202)	(-1.2431)	(-0.5314)	(-0.2796)	(-1.1687)
ln*SMER*	-0.0143***	-0.0148***	-0.0148***	-0.0144***	-0.0148***	-0.0150***
	(-2.9619)	(-3.0848)	(-3.0641)	(-2.9982)	(-3.0932)	(-3.0938)
ln*SMEH*				0.0214	0.0226	0.0169
				(1.4174)	(1.5243)	(1.1078)
W* lnGDP_{t-1}			0.2576***			0.2638***
			(3.6625)			(3.7098)
W* ln*SCHOOL*			-0.7322***			-0.7103***
			(-4.5612)			(-4.2071)
W* ln($n+d+g$)			0.3354*			0.3416*
			(1.7984)			(1.8262)
W* ln*SMER*			-0.1043*			-0.0992
			(-1.9280)			(-1.5821)
W* ln*SMEH*						-0.0240
						(-0.1253)
ρ (SAR) / λ(SEM)	0.922***	0.8920***	0.887***	0.9220***	0.8879***	0.8999***
	(34.1084)	(24.7908)	(23.5134)	(34.0951)	(23.8239)	(26.9078)
R^2	0.5515	0.583	0.5952	0.5528	0.5837	0.596
log-likelihood (LIK)	2071.07	2063.98	2082.14	2072.07	2065.14	2082.7701
Observations	1204	1204	1204,	1204	1204	1204
LR_{FE}	647.4885	648.1578	667.3786	648.3322	650.449,	667.8711
LR_{FE} (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

One interesting point is that southern regions seem to be more directly connected economically with neighbours as evidenced by the greater magnitude of the spatial autoregressive parameter and the coefficient of the spatial lag of the initial GDP per capita, when comparing with northern regions. This result might also be influenced by some externalities associated with the new economic geography models. For instance, Krugman (1999) argues that Brazil's south is a more attractive place to produce due to better transport infrastructure and higher market potential. Thus, southern Brazil attracts firms and investments that want to exploit economies of scale and access to consumers.[84] Additionally, the trade liberalization that occurred during the 1990's, due to the unilateral trade liberalization and the MERCOSUL regional trade agreement might have reinforced the advantages of

[84] Thus, this result can also be interpreted as a result of some spillovers that could have been caused by pecuniary externalities present in models of new economic geography (e.g. Krugman 1991).

Southern regions.[85] For instance, Martincus (2010) provides empirical evidence of an accentuated tendency during that decade for industries with a high degree of sectoral openness to locate in (Southern) Brazilian states that are closer to Argentina, the biggest trading partner in MERCOSUL.

Table 5.11 – Panel Results for the Southern Sample – All Sectors

	SEM	SAR	SDM	SEM	SAR	SDM
lnGDP_{t-1}	-0.1473***	-0.1435***	-0.1506***	-0.1473***	-0.1435***	-0.1506***
	(-25.0076)	(-24.6702)	(-25.1579)	(-25.0132)	(-24.6624)	(-25.1126)
ln*SCHOOL*	0.1387***	0.1127***	0.1932***	0.1371***	0.1123***	0.1896***
	(5.9108)	(5.2921)	(6.7627)	(5.8268)	(5.2703)	(6.5507)
ln($n+d+g$)	-0.0100	-0.0068	-0.019	-0.0094	-0.0062	-0.0182
	(-0.7133)	(-0.4929)	(-1.2822)	(-0.6646)	(-0.4497)	(-1.2285)
ln*SMER2*	-0.0319***	-0.0345***	-0.0304***	-0.0329***	-0.0354***	-0.0310***
	(-2.8534)	(-3.1261)	(-2.7011)	(-2.9267)	(-3.1877)	(-2.7400)
ln*SMEH2*				0.0210	0.0188	0.0143
				(0.7553)	(0.6968)	(0.4876)
W* lnGDP_{t-1}			0.2363***			0.2601***
			(3.3108)			(3.3154)
W* ln*SCHOOL*			-0.6723***			-0.6131***
			(-4.1843)			(-3.4093)
W* ln($n+d+g$)			0.2699			0.2887
			(1.4718)			(1.5614)
W* ln*SMER2*			-0.1669			-0.1713
			(-1.5739)			(-1.6134)
W* ln*SMEH2*						0.1601
						(0.4956)
ρ (SAR) / λ(SEM)	0.9219***	0.8880***	0.8999***	0.9209***	0.887***	0.902***
	(34.0822)	(23.8679)	(26.9607)	(33.6252)	(23.6074)	(27.5154)
R^2	0.5515	0.5829	0.5946	0.5519	0.5831	0.5948
log-likelihood (LIK)	2070.7564	2064.107	2080.6283	2071.0439	2064.3487	2080.9839
No Observations	1204	1204	1204	1204	1204	1204
LR_{FE}	658.2030	659.2349	676.5644	658.7671	659.6287	674.4791
LR_{FE} (*p*-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

[85] MERCOSUL (Mercado Comun do Sul) is translated to English as Southern Common Market. It is an common market and political agreement between Argentina, Brazil, Paraguay and Uruguay. Other Latin American States, such as Venezuela, Mexico, Chile, Bolivia, Equador, Colombia and Peru, are associated to MERCOSUL and might become members in the future.

5.8 Conclusion

This chapter has provided evidence on the importance of SMEs for economic growth in Brazil when considering the spatial dimension. In the exploratory spatial data analysis, measures of global and local spatial autocorrelation provide strong evidence accounting for the existence of spatial autocorrelation and spatial heterogeneity, suggesting the consideration of spatial econometrics techniques that consider the presence of spatial effects.

The spatial regressions that explicitly consider the spatial structure were estimated first in a cross-section setting. The empirical results for Brazil show that the spatial dynamics should be considered and reveal strong spatial dependence in the data. The empirical results also show that the SME sector size has a negative or neutral effect on growth and that SMEs' human capital seems to have a positive effect in this process, although there is no indication of positive spatial spillovers from this aspect of SMEs. The regional analysis provide similar general results in the sense that the presence of SMEs does not generate economic growth but the level of human capital in the SMEs seems to be important for regional economic growth in Brazil. Some differences emerge in terms of the level of significance and magnitudes of the coefficients related with the SME sector. Differences in the regional empirical results are probably related with the idea that southern regions are more interconnected, this is likely to be related to pecuniary externalities present in models of new economic geography (e.g. Krugman 1991) discussed earlier in this chapter. Also, differences in regional results might be related to the fact that entrepreneurship and SMEs are locally influenced by specific cultural characteristics, as discussed in Chapter 4, the literature suggests important local institutional differences in Brazil (e.g. Naritomi et al., 2009; Menezes-Filho et al., 2006) that might affect entrepreneurship differently as suggested by Baumol (1990) and Dias and McDermott (2006).

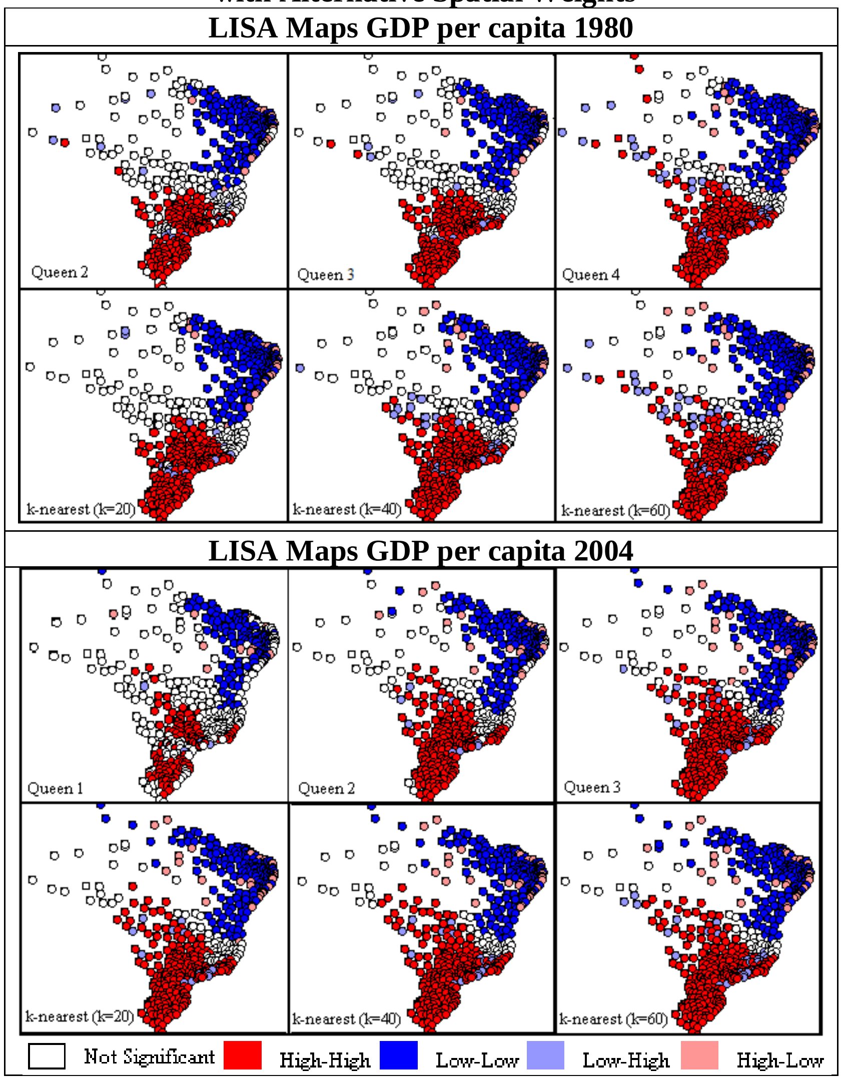
The panel analysis contains more variation and less collinearity among variables and provides more informative results. Overall, panel data results are in line with the cross-section estimates. They suggest a negative impact of the size of the SME sector on economic growth and the opposite impact stemming from the human capital of this sector. In addition, they also confirm the result that there is no indication of positive spatial spillovers derived from *SMEH* on economic growth. However, panel regressions suggest that the spatial interaction of entrepreneurial activities creates positive externalities for economic growth in Brazil and northern regions and suggest that SMEs' policies should be coordinated with a broader regional focus in order to explore spatial externalities. Thus, from the public policy point of view, results do not support the view that the size of the SME sector should be locally increased per se to improve economic performance. Instead, policy makers should better understand the spatial interactions of directly supporting SMEs in a given region to promote growth in its neighbourhood through the promotion of entrepreneurship.

Additionally, the panel data analysis also considered an extended SME sector with the inclusion of commerce and services. As in the case of non-spatial estimates (Chapter 4), the inclusion of commerce and services weakens the direct effect of the *SMEH*, and reinforces the negative effect of SME on growth, suggesting that SMEs in manufacturing are more important for growth; if the public focus is on increasing the productivity of the SME sector, the results suggest that public policy would be more succesfull if focused on manufacturing SMEs.

5.9 Appendix

5.9.1 Lisa Maps

Figure 5.4. LISA Maps for GDP per capita with Alternative Spatial Weights



5.9.2 Log-likelihood functions

5.9.2.1 Log-likelihood functions for Cross-Section Estimates

The maximum likelihood estimates are obtained using log-likelihood functions as described in Anselin (1988, p. 181) and LeSage and Pace (2009). The log-likelihood function for the SAR and SDM specifications is:

$$\ln L = -(N/2)\ln(\pi) - (N/2)\ln(\sigma^2) + \ln|I - \rho W| - \frac{[Y(I - \rho W) - X\beta]'[Y(I - \rho W) - X\beta]}{2\sigma^2}$$

where the vector Y represents the dependent variable of a model, X is the vector of independent variables, ρ is the autoregressive parameter, and W is a spatial weights matrix describing the connectivity between regions i and j.

The maximum likelihood estimates for the SEM specification is based on the log-likelihood function given:

$$\ln L = -(N/2)\ln(\pi) - (N/2)\ln(\sigma^2) + \ln|I - \lambda W| - \frac{(Y - X\beta)'(I - \lambda W)'(Y - X\beta)(I - \lambda W)}{2\sigma^2}$$

where the notation is the same as in the case of the SAR log-likelihood function, with the difference being the term λ that stands for the spatial error parameter. For detailed computational procedures see also Anselin (1988) and LeSage and Pace (2009).

5.9.2.2 Log-likelihood functions for Panel Data Estimates

Elhorst (2010) have shown that the spatial models can be estimated using the maximum likelihood procedure proposed by Anselin (1988) using the demeaning process described in Appendix 5.9.3. This approach is used to remove the individual and time specific effects.

The log-likelihood function corresponding to the demeaned specification for panel data for the SAR (and SDM) specification is presented in Elhorst (2010, p. 390) and is given by:

$$\ln L = -\frac{NT}{2}\ln(2\pi\sigma^2) + T\ln\left|I - \rho W\right| - \frac{1}{2\sigma^2}\sum_{i=t}^{N}\sum_{t=1}^{T}(y_{it}^* - \rho[\sum_{j=1}^{N} w_{ij} y_{jt}]^* - x_{it}^*\beta)^2$$

where y represents the dependent variable of a model, x represents the independent variables, w_{ij} is an element of a spatial weights matrix W describing the connectivity between regions i and j, and the asterisk denotes the demeaning procedure to extract the individual and/or time specific effect.

The log-likelihood function for a panel SEM specification if the spatial effects are assumed to be fixed is presented in Elhorst (2010, p. 393) and takes the following form:

$$\begin{aligned}\ln L = &-\frac{NT}{2}\ln(2\pi\sigma^2) + T\ln\left|I - \lambda W\right| \\ &- \frac{1}{2\sigma^2}\sum_{i=t}^{N}\sum_{t=1}^{T}\left\{y_{it}^* - \lambda[\sum_{j=1}^{N} w_{ij} y_{jt}]^* - (x_{it}^* - \lambda[\sum_{j=1}^{N} w_{ij} x_{jt})]^*)\beta\right\}^2\end{aligned}$$

where the variables are defined as in the SAR log-likelihood function described above.

5.9.3 Demeaning Process

The individual fixed effect is eliminated by demeaning the variables of the model, where variables are taken in deviation from their average over time. This transformation takes the following form for a given variable y:

$$y_{it}^* = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$$

A similar procedure is used to remove the time fixed effect through the following transformation:

$$y_{it}^{*} = y_{it} - \frac{1}{N}\sum_{i=1}^{N} y_{it}$$

5.9.4 Diagnostic Tests

The inference on the spatial dependence parameter can be based on the asymptotically equivalent methods of the likelihood ratio (LR), Lagrange multiplier (LM), and Wald test, with the choice usually down to computational convenience. The diagnostic tests used in this chapter are described below.

5.9.4.1 Lagrange Multiplier Tests for Spatial Dependence: Spatial *versus* Non-spatial Models

Four LM test statistics for spatial dependence are reported in the results in cross-section and refer to the spatial model as the alternative model.[86] Two tests, LM-Lag and Robust LM-Lag, consider the spatial lag model as the alternative and the remaining two, LM-ERR and Robust LM-ERR, considered the spatial error model as the alternative model.

The LM test for the spatial error model as the alternative one takes the form:

$$LM_{ERR} = \frac{[\tilde{\varepsilon}' W \tilde{\varepsilon} / \tilde{\sigma}^2]^2}{T}$$

where T = tr[$(W' + W)W$], with tr being the trace of matrix and $\tilde{\varepsilon}$ and $\tilde{\sigma}^2$ being the estimates of those parameter associated with the null hypothesis. As shown in

[86] See Anselin and Hudak (1992) and Anselin et al. (1996) for a detailed discussion on these tests. These tests have the advantage that they can be implemented estimating only the OLS regressions.

Burridge (1980), this statistic is distributed as a chi-squared statistic with one degree of freedom. The LM test version for the spatial lag model takes the following form:

$$LM_{LAG} = \frac{[\tilde{\varepsilon}' Wy / \tilde{\sigma}^2]^2}{T_1}$$

where $T_1 = [(WX\tilde{\beta})(I - X(X'X)^{-1}X')(WX\tilde{\beta} + (tr[(W' + W)W])\tilde{\sigma}^2] / \tilde{\sigma}^2$, and the rest of notation is as above.

The robust versions of these tests proposed by Anselin et al. (1996) are presented below. It consists of a test that is robust to local misspecification. The robust version for the LM_{ERR} test is:

$$LMR_{ERR} = \frac{[(\tilde{\varepsilon}' W\tilde{\varepsilon} / \tilde{\sigma}^2) - T(T_1)^{-1}(\tilde{\varepsilon}' W\tilde{\varepsilon} / \tilde{\sigma}^2)]^2}{T[1 - T(T_1)]^{-1}}$$

Likewise, the robust version for the LM_{LAG} is:

$$LMR_{LAG} = \frac{[\tilde{\varepsilon}' Wy / \tilde{\sigma}^2 - \tilde{\varepsilon}' W\tilde{\varepsilon} / \tilde{\sigma}^2]^2}{T_1 - T}$$

It is important to consider the robust options only when the standard versions are significant. When they are not significant, the properties of the robust version may no longer hold. The rejection of the standard forms requires the consideration of the robust forms. Summarizing the decision rule, if neither of the tests is significant, consider the model without spatial dependence. If only one test reject the null, consider the estimation of this type of spatial model. Finally, if both standard versions are significant, consider the most significant robust version as the appropriate model.

5.9.4.2 The Likelihood Ratio Test (LR test)

This test is based on the idea that if one restriction is valid, the log-likelihood function should not lead to a large reduction in the log-likelihood function. If L_R and L_U are the log-likelihood function for the restricted model and unrestricted models, respectively, the likelihood ratio takes the following form:

$$\lambda = -2\ln\frac{L_R}{L_U} = -2[\ln L_R - \ln L_U]$$

where this ratio is positive and between 0 and 1. This test statistic has a chi-squared distribution with degrees of freedom equal to the number of restrictions imposed, under the null that the restricted model is preferred.

5.9.4.3 The Wald test

To calculate the Wald test we need to estimate only the constrained model. The test statistic is given by the expression below:

$$W = [c(\hat{\theta}) - q]'(Var[c(\hat{\theta}) - q])^{-1}[c(\hat{\theta}) - q]$$

where we impose a restriction that $c(\theta) = q$. This test has a chi-squared distribution with degrees of freedom equal to the number of restrictions imposed, under the null

5.9.5 The Common Factor Test

This test allows a comparison of the likelihood function values from SDM and SEM models following the principles of the likelihood ratio test described earlier in Section 5.9.4.2. Considering a simple case with only one conditioning variable the structure of this test can be described as follows:[87]

[87] This description is based on Anselin (1988, p.227).

$$gr_t = \beta \ln(y_0) + \varepsilon_t$$
$$\varepsilon_t = \lambda W \varepsilon_t + \upsilon_t$$

or the equivalent expression of the first specification (SEM) in matrix form:

$$gr = \beta \ln(y_0) + (\mathrm{I} - \lambda W)^{-1} \upsilon$$

After multiplying both sides by (I-λW) we obtain:

$$gr = \lambda W gr + \beta \ln(y_0) - \lambda W \ln(y_0) \beta + \upsilon$$

or the equivalent form of the second specification (SDM)

$$gr = \lambda W gr + \beta \ln(y_0) + \gamma W \ln(y_0) + \upsilon$$

In order for the two specifications to be equivalent, the product of the coefficients $\lambda\beta$ should equal the negative of the coefficient γ. Therefore, the test on the coefficient constraints is implemented under the following null hypothesis: H_0: $\lambda\beta + \gamma = 0$. If we reject the null the LR test suggests the use of the SDM model.

5.9.6 Additional Results for Cross-Section Estimates

Table 5.12 – Cross-Section Results with Alternative Weight Matrices for *SMER* (Brazil)

	Queen2 SAR	Queen2 SEM	Queen2 SDM	Queen3 SAR	Queen3 SEM	Queen3 SDM	Queen4 SAR	Queen4 SEM	Queen4 SDM	K-20th SAR	K-20th SEM	K-20th SDM	K-40th SAR	K-40th SEM	K-40th SDM	K-60th SAR	K-60th SEM	K-60th SDM
$\ln GDP_{t-1}$	-0.0206***	-0.0286***	-0.0298***	-0.0213***	-0.028820***	-0.0297***	-0.021983***	-0.02951***	-0.0300***	-0.0204***	-0.027562***	-0.0286***	-0.021161***	-0.028412***	-0.0293***	-0.021517***	-0.028834***	-0.0293***
	(-9.7278)	(-13.3082)	(-13.6961)	(-9.76)	(-12.8779)	(-13.1866)	(-9.9082)	(-13.0433)	(-13.2572)	(-9.6237)	(-12.6102)	(-12.9192)	(-9.6238)	(-12.5413)	(-12.9151)	(-9.66)	(-12.7904)	(-13.0958)
$\ln(n+d+g)$	-0.0043	0.0040	0.0066	-0.0055	0.005137	0.0072	-0.0069286*	0.0058299	0.0084*	-0.0038	0.0072573	0.0109**	-0.005439	0.0088828*	0.0133***	-0.0065902*	0.0092457**	0.0140***
	(-1.1327)	(0.8451)	(1.3423)	(-1.4275)	(1.0872)	(1.4872)	(-1.7818)	(1.2664)	(1.7842)	(-1.0294)	(1.5142)	(2.2228)	(-1.4178)	(1.8503)	(2.7124)	(-1.6952)	(1.9681)	(2.9011)
$\ln SCHOOL$	0.0282***	0.0198***	0.0190***	0.0298***	0.022393***	0.02166***	0.0309839***	0.023705***	0.0227***	0.0277***	0.019292***	0.0180***	0.029643***	0.0211253***	0.0196***	0.0306077***	0.0206141***	0.0192***
	(7.5553)	(5.212)	(4.9741)	(7.7949)	(5.746)	(5.5441)	(7.9794)	(6.0422)	(5.7986)	(7.4411)	(4.9541)	(4.6208)	(7.742)	(5.3384)	(4.9707)	(7.8855)	(5.2341)	(4.907)
$\ln SMER$	-0.0026	-0.0053***	-0.0055***	-0.0021	-0.003680**	-0.0037**	-0.0022173	-0.003894**	-0.00385**	-0.0021	-0.004191**	-0.0046***	-0.002202	-0.0038554**	-0.0039**	-0.0021625	-0.0042424**	-0.0041**
	(-1.6125)	(-3.2821)	(-3.418)	(-1.2802)	(-2.1349)	(-2.1525)	(-1.3278)	(-2.2728)	(-2.2365)	(-1.3313)	(-2.5733)	(-2.8455)	(-1.344)	(-2.2996)	(-2.3656)	(-1.297)	(-2.5426)	(-2.4731)
ρ or λ	0.4895***	0.8127***	0.6533***	0.44647***	0.84948***	0.6351***	0.36654***	0.89673***	0.6163***	0.5284***	0.78783***	0.614***	0.4801***	0.82908***	0.5169***	0.43154***	0.88385***	0.52413***
	(6.7733)	(17.076)	(9.4065)	(4.5813)	(15.333)	(6.4425)	(2.8651)	(17.482)	(4.8105)	(7.5917)	(15.431)	(7.9801)	(4.9586)	(14.2)	(4.1907)	(3.6225)	(18.416)	(3.5842)
(Intercept)	-0.0189*	0.0276*	-0.0125	-0.0229**	0.0280077*	-0.02843	-0.0267165**	0.0284174*	-0.0604**	-0.0177	0.035918**	-0.0223	-0.023281**	0.0378652**	-0.0549**	-0.0267378**	0.0384404**	-0.0757***
	(-1.7229)	(1.8035)	(-0.7823)	(-2.0208)	(1.8011)	(-1.4076)	(-2.2949)	(1.7632)	(-2.2283)	(-1.6114)	(2.3492)	(-1.3855)	(-2.027)	(2.4714)	(-2.5681)	(-2.2497)	(2.4915)	(-2.8438)
W* $\ln GDP_{t-1}$			0.0163***			0.0175**			0.0218**			0.0118**			0.0105			0.0051
			(2.8837)			(2.4115)			(2.3796)			(1.9626)			(1.2866)			(0.5164)
W* $\ln(n+d+g)$			-0.0062			-0.0127			-0.0269**			-0.0135*			-0.0258***			-0.0303***
			(-0.7591)			(-1.2894)			(-2.1666)			(-1.6934)			(-2.598)			(-2.589)
W* $\ln SCHOOL$			0.0071			0.0014			-0.0066			0.0122			0.0147			0.02758
			(0.6631)			(0.0985)			(-0.3669)			(1.0827)			(0.9401)			(1.4151)
W* $\ln SMER$			0.0042			-0.0033			-0.00423			0.0026			-0.0052			-0.0063
			(1.1601)			(-0.6468)			(-0.675)			(0.6307)			(-0.9883)			(-0.9547)
Observations	508	508	508	508	508	508	508	508	508	508	508	508	508	508	508	508	508	508
Log likelihood	1352.26	1379.05	1387.97	1341.88	1363.509	1370.506	1336.34	1356.974	1364.388	1353.489	1374.541	1382.22	1342.402	1362.123	1372.131	1337.865	1360.353	1371.22
Akaike Criterion	-2690.5	-2744.1	-2753.9	-2669.8	-2713	-2719	-2658.7	-2699.9	-2706.8	-2693	-2735.1	-2742.4	-2670.8	-2710.2	-2722.3	-2661.7	-2706.7	-2720.4
LR test	37.881	91.48	60.294	17.145	60.387	28.601	6.0504	47.317	14.3	40.348	82.452	45.192	18.173	57.615	16.735	9.0987	54.075	10.361
LR test (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Wald	45.877	291.59	88.482	20.988	235.09	41.505	8.2088	305.62	23.141	57.634	238.11	63.681	24.588	201.63	17.562	13.123	339.15	12.846
Wald (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
COMFAC			17.8354			13.995			14.8296			15.3574			20.0171			21.7343
COMFAC (p-value)			0.001329			0.007311			0.005068			0.004015			0.0005			0.0002

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.13 – Cross-Section Results with Alternative Weight Matrices for SMEH (Brazil)

	Queen2C SAR	Queen2C SEM	Queen2C SDM	Queen3C SAR	Queen3C SEM	Queen3C SDM	Queen4C SAR	Queen4C SEM	Queen4C SDM	K-20th SAR	K-20th SEM	K-20th SDM	K-40th SAR	K-40th SEM	K-40th SDM	K-60th SAR	K-60th SEM	K-60th SDM
$\ln GDP_{t-1}$	-0.02039***	-0.02843***	-0.029582***	-0.02111***	-0.02859***	-0.02947***	-0.02179***	-0.02929***	-0.02982**	-0.020138***	-0.027294***	-0.028452***	-0.02087**	-0.02813**	-0.02911***	-0.02122***	-0.02856***	-0.02897***
	(-9.607)	(-13.2294)	(-13.6135)	(-9.6429)	(-12.7664)	(-13.0633)	(-9.794)	(-12.9279)	(-13.1398)	(-9.4545)	(-12.5078)	(-12.8682)	(-9.4532)	(-12.4014)	(-12.7967)	(-9.4858)	(-12.6528)	(-12.9115)
$\ln(n+d+g)$	-0.00331	0.006077	0.008556*	-0.004538	0.006905	0.009127*	-0.006012	0.007173	0.00987*	-0.002465	0.0094767*	0.01295***	-0.004113	0.01076*	0.01521**	-0.005293	0.010886**	0.015529***
	(-0.8515)	(1.2573)	(1.7221)	(-1.1402)	(1.4253)	(1.8344)	(-1.4924)	(1.5171)	(2.0337)	(-0.6372)	(1.9477)	(2.5951)	(-1.0356)	(2.1918)	(3.0333)	(-1.3148)	(2.2616)	(3.155)
$\ln SCHOOL$	0.026126***	0.014992***	0.014012***	0.02772***	0.01858***	0.017585***	0.02891***	0.02073**	0.019513**	0.024735***	0.014077***	0.012926***	0.02678**	0.01705**	0.01549***	0.027828***	0.017084***	0.015812***
	(5.9655)	(3.3536)	(3.1226)	(6.1714)	(4.0389)	(3.8093)	(6.343)	(4.4769)	(4.2134)	(5.6632)	(3.1178)	(2.8499)	(5.9676)	(3.6828)	(3.3495)	(6.123)	(3.7181)	(3.4596)
$\ln SMER$	-0.002455	-0.00524***	-0.005445***	-0.00198	-0.003555**	-0.003623**	-0.002114	-0.00376*	-0.003701**	-0.0019517	-0.004086**	-0.00458***	-0.002047	-0.00371**	-0.003840**	-0.002004	-0.004077**	-0.003957**
	(-1.5416)	(-3.2348)	(-3.3676)	(-1.2119)	(-2.0644)	(-2.0794)	(-1.2643)	(-2.1919)	(-2.1459)	(-1.2296)	(-2.5196)	(-2.824)	(-1.2479)	(-2.2183)	(-2.3007)	(-1.2007)	(-2.4428)	(-2.3832)
$\ln SMEH$	0.004464	0.009236**	0.009670**	0.004493	0.007488	0.007880	0.004382	0.00588	0.006372	0.0062898	0.010437**	0.010557**	0.005986	0.00805	0.0081625*	0.005819	0.007087	0.007011
	(0.9231)	(1.9636)	(2.0618)	(0.906)	(1.5357)	(1.6205)	(0.8722)	(1.2042)	(1.3088)	(1.3051)	(2.188)	(2.2191)	(1.2094)	(1.6607)	(1.6979)	(1.1633)	(1.4704)	(1.4712)
ρ or λ	0.48936***	0.82053***	0.65852***	0.44605***	0.8538***	0.63601***	0.36425***	0.89767***	0.61324***	0.53375***	0.79762***	0.61608***	0.48805***	0.83422***	0.50926***	0.44341***	0.88624***	0.54423***
	(6.773)	(17.717)	(9.5776)	(4.58)	(15.718)	(6.4647)	(2.8436)	(17.622)	(4.7618)	(7.7047)	(16.161)	(8.0558)	(5.0638)	(14.631)	(4.0963)	(3.7566)	(18.806)	(3.8429)
(Intercept)	-0.020390	0.022953	-0.0084889	-0.0211106	0.024299	-0.024958	-0.0217958	0.02534	-0.06035*	-0.0201382	0.029938*	-0.016641	-0.020868	0.033658**	-0.048548**	-0.0212212	0.034621**	-0.080603**
	(-9.607)	(1.4756)	(-0.4889)	(-9.6429)	(1.5394)	(-1.1318)	(-9.794)	(1.5517)	(-2.0195)	(-9.4545)	(1.9205)	(-0.9696)	(-9.4532)	(2.1626)	(-2.0047)	(-9.4858)	(2.2097)	(-2.5456)
W* $\ln GDP_{t-1}$			0.015875***			0.017129**			0.021149**			0.011938**			0.010136			0.0046302
			(2.825)			(2.3607)			(2.3096)			(1.9992)			(1.2362)			(0.4678)
W* $\ln(n+d+g)$			-0.0089333			-0.016188			-0.02787**			-0.019212**			-0.030313***			-0.029121**
			(-1.034)			(-1.4813)			(-2.0509)			(-2.1702)			(-2.7367)			(-2.4015)
W*$\ln SCHOOL$			0.0141499			0.008119			-0.00275			0.0217456*			0.023555			0.0279303
			(1.1824)			(0.5221)			(-0.1383)			(1.7464)			(1.3787)			(1.3132)
W*$\ln SMER$			0.0038979			-0.003625			-0.004594			0.002427			-0.005323			-0.006049
			(1.0696)			(-0.7091)			(-0.7283)			(0.5939)			(-1.0068)			(-0.9092)
W*$\ln SMEH$			-0.0146435			-0.01426			-0.005876			-0.022950*			-0.018968			0.002393
			(-1.1534)			(-0.818)			(-0.2681)			(-1.6696)			(-1.0072)			(0.1076)
Observations	508	508	508	508	508	508	508	508	508	508	508	508	508	508	508	508	508	508
LIK	1352.682	1380.964	1390.243	1342.298	1364.683	1371.889	1336.721	1357.697	1365.243	1354.338	1376.906	1385.219	1343.129	1363.494	1373.768	1338.537	1361.431	1372.363
AIC	-2689.4	-2745.9	-2754.5	-2668.6	-2713.4	-2717.8	-2657.4	-2699.4	-2704.5	-2692.7	-2737.8	-2744.4	-2670.3	-2711	-2721.5,	-2661.1	-2706.9	-2718.7
LR test	37.898	94.461	61.687	17.13	61.899	28.785	5.9752	47.928	13.996	41.209	86.345	45.373	18.792	59.522	15.726	9.6068	55.395	10.975
LR test (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0145	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0019	0.0000	0.0009
Wald	45.873	313.89	91.73	20.976	247.06	41.792	8.086	310.52	22.674	59.362	261.18	64.896	25.642	214.08	16.779	14.112	353.66	14.768
Wald (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.000001	0.0000	0.0000	0.0000	0.000000	0.0000	0.000041	0.000172	0.0000	0.000121
COMFAC			18.5591,			14.4125,			15.0921,			16.626,			20.5478,			21.8655,
COMFAC(p-value)			0.002321			0.01319			0.009976			0.005266			0.0009859			0.000555

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.14 – Cross-Section Results with Alternative Weight Matrices for *SMER* (Northern)

	Queen2C SAR	Queen2C SEM	Queen2C SDM	Queen3C SAR	Queen3C SEM	Queen3C SDM	Queen4C SAR	Queen4C SEM	Queen4C SDM	K-20th SAR	K-20th SEM	K-20th SDM	K-40th SAR	K-40th SEM	K-40th SDM	K-60th SAR	K-60th SEM	K-60th SDM
lnGDP_{t-1}	-0.024401***	-0.026957***	-0.027230***	-0.025341***	-0.02708***	-0.027177***	-0.026458***	-0.0278027**	-0.026376***	-0.023945***	-0.026131***	-0.025916***	-0.025150***	-0.026604***	-0.027033***	-0.0259428***	-0.0270219***	-0.023877***
	(-7.6051)	(-8.3404)	(-8.3352)	(-7.7933	(-8.2564	(-8.1714)	(-8.1198)	(-8.5511)	(-8.0799)	(-7.5108)	(-8.0092)	(-7.7887)	(-7.8146)	(-8.1614)	(-8.2183)	(-8.1479)	(-8.5098)	(-7.0966)
ln(*n*+*d*+*g*)	0.016719**	0.0128743*	0.0137487*	0.0172107**	0.016219**	0.0176817**	0.017601**	0.0203673**	0.0213967***	0.017917**	0.01519876*	0.015880**	0.019019**	0.020012**	0.021467***	0.0191057**	0.0216926***	0.019406**
	(2.253)	(1.6475)	(1.7365)	(2.255	(2.0109	(2.1659)	(2.2641)	(2.5311)	(2.6631)	(2.4127)	(1.9139)	(1.9646)	(2.5082)	(2.4977)	(2.6534)	(2.5329)	(2.7285)	(2.4173)
ln*SCHOOL*	0.017678***	0.018177***	0.018519***	0.019283***	0.020361***	0.0207847***	0.0197556***	0.0201129***	0.0186824***	0.017378***	0.0193234***	0.019225***	0.01917***	0.019872***	0.01986***	0.0197529***	0.0195767***	0.015793***
	(3.7705)	(3.885)	(3.9559)	(4.0362	(4.3183	(4.3568)	(4.0877)	(4.2142)	(3.9427)	(3.6923)	(4.0749)	(4.0126)	(4.008)	(4.1704)	(4.147)	(4.1313)	(4.1405)	(3.2367)
ln*SMER*	-0.001980	-0.001717	-0.001489	-0.0016989	-0.0005416	0.000616	-0.0018067	-0.0006814	0.0003027	-0.001166	-0.0008405	-0.000734	-0.001365	-0.000457	-0.0000576	-0.0013991	-0.0006612	-0.000439
	(-1.13)	(-0.9053)	(-0.7807)	(-0.9484	(-0.2746	(0.3007)	(-0.9932)	(-0.3503)	(0.1545)	(-0.6644)	(-0.4434)	(-0.3799)	(-0.7607)	(-0.2393)	(-0.0299)	(-0.7765)	(-0.3506)	(-0.235)
ρ or λ	0.48147***	0.63428***	0.61697***	0.49393***	0.66468***	0.54384***	0.47588***	0.68945***	0.14631	0.53244***	0.64636***	0.62256***	0.54467***	0.69909***	0.27204	0.62403***	0.77703***	-0.24163
	(4.5572)	(5.9659)	(5.7411)	(3.6952	(4.9953	(3.3382)	(2.8818)	(4.4958)	(0.49492)	(4.9109)	(5.748)	(5.2825)	(3.6284)	(5.0463)	(1.0747)	(3.6402)	(6.1462)	(-0.56228)
(Intercept)	0.046965**	0.042998*	-0.012085	0.047004**	0.0499461*	0.0177687	0.0481848**	0.0614772**	-0.328182**	0.0502432**	0.0494241**	0.0154492	0.051381**	0.060928***	-0.22223*	0.049817**	0.0649721***	-0.741513***
	(2.2108)	(1.8859)	(-0.2202)	(2.161	(2.1259	(0.2133)	(2.1841)	(2.6148)	(-2.1883)	(2.3626)	(2.1425)	(0.2451)	(2.3707)	(2.6287)	(-1.7322)	(2.2949)	(2.8305)	(-3.2764)
W*lnGDP_{t-1}			0.0076481			0.00903301			0.0557434**			0.0138644			0.040779			0.1226840***
			(0.7492)			(0.5657)			(2.3349)			(1.1432)			(1.6218)			(2.919)
W*ln(*n*+*d*+*g*)			-0.008961			-0.0088293			-0.1450559***			-0.010593			-0.10253**			-0.300166***
			(-0.4513)			(-0.2849)			(-2.633)			(-0.4539)			(-2.0508)			(-3.431)
W*ln*SCHOOL*			0.0230364			-0.0024159			-0.0326025			-0.009164			-0.019194			-0.0724345
			(1.2444)			(-0.0896)			(-0.9677)			(-0.4688)			(-0.5706)			(-1.529)
W*ln*SMER*			-0.003169			-0.0114936			0.0038865			-0.000870			-0.002318			0.0260305*
			(-0.7000)			(-1.6127)			(0.3577)			(-0.1701)			(-0.2600)			(1.6453)
Observations	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207
LIK	565.3434	568.3236	570.4919	561.8654	563.2308	565.5857	559.5202	560.5682	566.9682	565.0326	566.5707	566.6997	561.1943	562.817	566.4083	560.5216	562.4047	569.6342
AIC	-1116.7	-1122.6	-1119	-1109.7	-1112.5	-1109.2	-1105	-1107.1	-1111.9	-1116.1	-1119.1	-1111.4	-1108.4	-1111.6	-1110.8	-1107	-1110.8	-1117.3
LR test	18.468	24.428	21.584	11.512	14.243	7.6131	6.8215	8.9176	0.24417	17.846	20.923	15.493	10.17	13.415	0.99049	8.8244	12.591	0.27791
LR test (p-value)	20.768	35.592	0.00000	0.0007	0.0002	0.0057	0.0090	0.0028	0.62121	0.0000	0.00000	0.0000	0.0014	0.0002	0.3196	0.0029	0.00038	0.59807
Wald	0.000000	0.000000	32.96	13.654	24.953	11.144	8.3046	20.212	0.24495	24.117	33.039	27.905	13.165	25.465	1.1551	13.251	37.776	0.31615
Wald (p-value)	0.046012		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
COMFAC			4.3365,			4.7098,			12.7999,			0.258,			7.1826,			14.459,
COMFAC(p-value)			0.3624			0.3184			0.0123			0.9924			0.1265			0.005965

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.15 – Cross-Section Results with Alternative Weight Matrices for SMEH (Northern)

	Queen2C SAR	Queen2C SEM	Queen2C SDM	Queen3C SAR	Queen3C SEM	Queen3C SDM	Queen4C SAR	Queen4C SEM	Queen4C SDM	K-20th SAR	K-20th SEM	K-20th SDM	K-40th SAR	K-40th SEM	K-40th SDM	K-60th SAR	K-60th SEM	K-60th SDM
ln*GDP*$_{t-1}$	-0.024013***	-0.026532***	-0.026799***	-0.024950***	-0.026677***	-0.0269639***	-0.026039***	-0.0273753**	-0.026292***	-0.02347***	-0.025698***	-0.025723***	-0.024670***	-0.026138***	-0.02656***	-0.025454***	-0.0265388***	-0.022981***
	(-7.4642)	(-8.2078)	(-8.2273)	(-7.6499)	(-8.113)	(-8.0867)	(-7.9643)	(-8.3826)	(-8.109)	(-7.3333)	(-7.8897)	(-7.7201)	(-7.6376)	(-8.0106)	(-8.0322)	(-7.9688)	(-8.3356)	(-6.9049)
ln(*n*+*d*+*g*)	0.0183101**	0.014643*	0.015769**	0.018775**	0.0179138**	0.0191803**	0.0191744**	0.0219344**	0.023029***	0.019796***	0.0173226**	0.0176565**	0.020903***	0.022041***	0.023465***	0.02101***	0.0237286***	0.022518***
	(2.4314)	(1.8566)	(1.9797)	(2.4245)	(2.1917)	(2.323)	(2.4302)	(2.6843)	(2.8616)	(2.6272)	(2.1596)	(2.1655)	(2.7155)	(2.717)	(2.8626)	(2.7426)	(2.9398)	(2.8183)
ln*SCHOOL*	0.014750***	0.014461***	0.014475***	0.016306***	0.0171666**	0.0182122***	0.0166585***	0.0170371**	0.016801***	0.013999***	0.0153001***	0.0153576***	0.015754***	0.016096***	0.016275***	0.0163006***	0.0159970***	0.011422**
	(2.7578)	(2.6841)	(2.7039)	(2.9873)	(3.1499)	(3.3088)	(3.013)	(3.0735)	(3.0991)	(2.6077)	(2.8277)	(2.8188)	(2.8779)	(2.9348)	(2.9713)	(2.9752)	(2.9414)	(2.0814)
ln*SMER*	-0.001874	-0.001642	-0.001366	-0.0016011	-0.000457	0.0007138	-0.0017093	-0.00056	0.00004156	-0.001039	-0.0007736	-0.0007894	-0.001239	-0.000357	0.0000408	-0.00127	-0.0005068	-0.000971
	(-1.0706)	(-0.8693)	(-0.7191)	(-0.895)	(-0.2323)	(0.3494)	(-0.9414)	(-0.2914)	(0.0214)	(-0.5933)	(-0.4102)	(-0.4099)	(-0.6915)	(-0.1873)	(0.0212)	(-0.7064)	(-0.2694)	(-0.5238)
ln*SMEH*	0.005677	0.006844	0.007886	0.0057577	0.0059556	0.0060252	0.0059893	0.0056362	0.0087122*	0.0064881	0.00763948	0.0074150	0.0065677	0.0070165	0.006998	0.0066175	0.0067190	0.010631**
	(1.1229)	(0.171631)	(1.5798)	(1.1162)	(0.246363)	(1.1773)	(1.1465)	(0.276808)	(1.6848)	(1.2823)	(0.130311)	(1.463)	(1.2701)	(0.170319)	(1.3304)	(1.2752)	(0.18783)	(2.1187)
ρ or λ	0.47606***	0.63772***	0.61564***	0.48531***	0.65972***	0.50909***	0.46424***	0.67899***	-0.088145	0.53088***	0.65093***	0.63267***	0.54214***	0.69896***	0.27273	0.62158***	0.77363***	-0.47567
	(4.4905)	(6.0358)	(5.7204)	(3.5938)	(4.9075)	(2.98)	(2.7654)	(4.3204)	(-0.26538)	(4.8953)	(5.8435)	(5.4801)	(3.6057)	(5.0437)	(1.0769)	(3.5996)	(6.0382)	(-1.0299)
(Intercept)	0.044732**	0.040039*	-0.021191	0.0446543**	0.0477434*	-0.012971	0.0456215**	0.0593784**	-0.4705***	0.047821**	0.04634098**	0.02410774	0.0488954**	0.058243**	-0.22596*	0.0473324**	0.0627084**	-0.9593826***
	(2.1032	(1.7556)	(-0.3674)	(2.0512)	(2.0335)	(-0.1461)	(2.0673)	(2.5299)	(-2.9556)	(2.2486)	(2.011)	(0.3718)	(2.2566)	(2.5153)	(-1.6575)	(2.18)	(2.737)	(-4.0897)
W*ln*GDP*$_{t-1}$			0.0073479			0.0074352			0.042613*			0.0136392			0.037855			0.11474799***
			(0.7205)			(0.4676)			(1.7709)			(1.1306)			(1.4239)			(2.7477)
W*ln(*n*+*d*+*g*)			-0.008389			-0.0068144			-0.13251**			-0.0142005			-0.099158*			-0.2879891***
			(-0.4227)			(-0.2168)			(-2.3743)			(-0.6029)			(-1.811)			(-3.2558)
W*ln*SCHOOL*			0.0258615			0.0018278			-0.011087			-0.0036973			-0.011146			-0.06798151
			(1.3624)			(0.0679)			(-0.326)			(-0.1854)			(-0.3018)			(-1.4629)
W*ln*SMER*			-0.003206			-0.012154*			0.0006618			-0.0005103			-0.002928			0.02491593
			(-0.7122)			(-1.7074)			(0.0614)			(-0.1003)			(-0.3172)			(1.5978)
W*ln*SMEH*			0.0028445			0.0186382			0.097528**			-0.0168481			0.0025802			0.15387744**
			(0.1646)			(0.6488)			(2.1646)			(-0.8477)			(0.0418)			(2.3274)
Observations	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207	207
LIK	565.9727	569.2533	571.8051	562.4855	563.9	566.5301	560.1734	561.1561	570.1403	565.8524	567.708	567.9857	561.9977	563.7528	567.3145	561.3314	563.2681	573.8056
AIC	-1115.9	-1122.5	-1117.6	-1109	-1111.8	-1107.1	-1104.3	-1106.3	-1114.3	-1115.7	-1119.4	-1110	-1108	-1111.5	-1108.6	-1106.7	-1110.5	-1121.6
LR test	18.074	24.635	21.211	11.1	13.929	6.1456	6.4756	8.4409	0.075277	17.834	21.545	16.197	10.124	13.634	0.9975	8.7915	12.665	0.99728
LR test (p-value)	0.0000212	0.000000	0.0000041	0.00086342	0.0001898	0.013174	0.010937	0.0036688	0.7838	0.0000241	0.00000345	0.00005709	0.0014635	0.000222	0.31792	0.0030264	0.00037259	0.31797
Wald	20.164	36.431	32.723	12.915	24.084	8.8806	7.6472	18.666	0.070425	23.964	34.146	30.032	13.001	25.438	1.1596	12.957	36.46	1.0608
Wald (p-value)	0.0000071	0.000000	0.0000000	0.0003259	0.0000009	0.0028822	0.005686	0.00001557	0.79072	0.000000	0.000000005	0.00000004	0.00031132	0.0000004	0.28154	0.00031872	0.00000000	0.30304
COMFAC			5.1037,			5.2602,			17.9685,			0.5553,			7.1235,			21.0751,
COMFAC(p-value)			0.4034			0.385			0.002986			0.99			0.2116			0.000784

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.16 – Cross-Section Results with Alternative Weight Matrices for *SMER* (Southern)

	Queen2C SAR	Queen2C SEM	Queen2C SDM	Queen3C SAR	Queen3C SEM	Queen3C SDM	Queen4C SAR	Queen4C SEM	Queen4C SDM	K-20th SAR	K-20th SEM	K-20th SDM	K-40th SAR	K-40th SEM	K-40th SDM	K-60th SAR	K-60th SEM	K-60th SDM
lnGDP_{t-1}	-0.028957***	-0.032032***	-0.032148***	-0.029334***	-0.032265***	-0.032688***	-0.030083***	-0.033377***	-0.033855***	-0.02865***	-0.03076***	-0.0309***	-0.02922***	-0.03149***	-0.03188***	-0.02950***	-0.03206***	-0.03272***
	(-9.7328)	(-11.2363)	(-11.2925)	(-9.5627)	(-10.7037)	(-10.8939)	(-9.7231)	(-10.9893)	(-11.22)	(-9.5663)	(-10.3848)	(-10.5445)	(-9.5283)	(-10.2793)	(-10.4539)	(-9.5723)	(-10.419)	(-10.7484)
ln(n+d+g)	0.0013222	0.0031282	0.0043507	-0.000272	0.0025274	0.004323	-0.000668	0.0026523	0.0020112	0.001485	0.005857	0.010084	0.001675	0.006044	0.010889*	0.000881	0.005239	0.010042*
	(0.2803)	(0.5293)	(0.6972)	(-0.0558)	(0.4315)	(0.6865)	(-0.1348)	(0.4638)	(0.3213)	(0.313)	(1.0135)	(1.6357)	(0.3432)	(1.0742)	(1.7803)	(0.1784)	(0.9588)	(1.6552)
ln*SCHOOL*	0.025527***	0.0150808**	0.0127512*	0.027127***	0.018973***	0.016066**	0.0284029***	0.0226318***	0.020922***	0.024695***	0.013551**	0.008336	0.02587***	0.018181***	0.011214	0.02760***	0.02119***	0.013332*
	(4.1604)	(2.1984)	(1.8064)	(4.2648)	(2.7459)	(2.2389)	(4.3952)	(3.3244)	(2.9605)	(4.0004)	(1.9773)	(1.1615)	(4.0766)	(2.6563)	(1.5294)	(4.3024)	(3.1889)	(1.8642)
ln*SMER*	-0.012165***	-0.014998***	-0.014249***	-0.011218***	-0.013567***	-0.013069***	-0.0111018***	-0.013094***	-0.012403***	-0.01221***	-0.01401***	-0.01390***	-0.01158***	-0.01307***	-0.01342***	-0.01131***	-0.01296***	-0.01298***
	(-3.8508)	(-5.005)	(-4.7376)	(-3.4365)	(-4.293)	(-4.1461)	(-3.3605)	(-4.0968)	(-3.8995)	(-3.8464)	(-4.5387)	(-4.521)	(-3.5561)	(-4.0944)	(-4.2149)	(-3.4429)	(-4.0528)	(-4.0632)
ρ or λ	0.55798***	0.74501***	0.65629***	0.51312***	0.72511***	0.56443***	0.49157***	0.7665***	0.46208**	0.55999***	0.70638***	0.57111***	0.54168***	0.70397***	0.53335***	0.5637***	0.75523***	0.52197***
	(6.9853)	(10.072)	(7.4225)	(4.5878)	(6.929)	(4.0776)	(3.4259)	(6.657)	(2.3837)	(6.8602)	(8.4017)	(5.3271)	(4.7819)	(6.1568)	(3.4215)	(4.2198)	(6.5044)	(2.755)
(Intercept)	0.0106697	0.0407179*	-0.0093492	0.0053677	0.0343498	-0.018464	0.00390491	0.0314645	-0.000942	0.011657	0.047996**	-0.046835	0.011986	0.043725**	-0.08305**	0.007740	0.038502**	-0.15044**
	(0.645)	(1.8868)	(-0.3824)	(0.3096)	(1.6139)	(-0.571)	(0.219)	(1.5142)	(-0.0212)	(0.6974)	(2.3143)	(-1.6044)	(0.6897)	(2.1801)	(-2.0548)	(0.4367)	(1.9903)	(-2.4349)
W*lnGDP_{t-1}			0.0206602**			0.0208168*			0.038279***			0.008636			0.006730			0.00551
			(2.5671)			(1.9228)			(2.7356)			(0.9377)			(0.5397)			(0.3533)
W*ln(n+d+g)			-0.0032283			-0.004955			-0.000896			-0.015478			-0.022942*			-0.032945*
			(-0.316)			(-0.3964)			(-0.0568)			(-1.3899)			(-1.6461)			(-1.8358)
W*ln*SCHOOL*			0.0151892			0.0198689			-0.001191			0.044173**			0.057642**			0.088052**
			(0.9384)			(0.887)			(-0.0392)			(2.363)			(2.1053)			(2.2994)
W*ln*SMER*			0.030050***			0.04254***			0.072850***			0.027919**			0.030427			0.0468453*
			(2.9562)			(2.687)			(3.0496)			(2.2075)			(1.3675)			(1.7426)
Observations	301	301	301	301	301	301	301	301	301	301	301	301	301	301	301	301	301	301
LIK	813.145	824.9054	828.9127	805.4696	813.8103	817.753	802.3125	810.3658	815.4377	812.3482	818.6394	823.9866	806.1867	811.0466	815.4659	803.952	809.4743	815.7313
AIC	-1612.3	-1635.8	-1635.8	-1596.9	-1613.6	-1613.5	-1590.6	-1606.7	-1608.9	-1610.7	-1623.3	-1626	-1598.4	-1608.1	-1608.9	-1593.9	-1604.9	-1609.5
LR test	29.082	52.603	37.319	13.731	30.413	14.358	7.4172	23.524	4.994	27.489	40.071	21.553	15.165	24.885	8.7362	10.696	21.741	5.0196
LR test (p-value)	0.00000	0.0000000	0.0000000	0.00021	0.0000	0.00015111	0.006460	0.00000	0.025436	0.0000001	0.0000000	0.000003	0.000098	0.0000	0.00311	0.0010	0.000003	0.025062
Wald	48.795	101.45	55.093	21.048	48.011	16.627	11.737	44.315	5.6822	47.063	70.589	28.378	22.866	37.906	11.707	17.807	42.307	7.5903
Wald (p-value)	0.0000000	0.0000000	0.0000000	0.00000	0.00000	0.0000455	0.0006127	0.000000	0.017138	0.000000	0.0000000	0.0000000	0.0000017	0.00000	0.00062	0.00002	0.0000000	0.0058684
COMFAC			8.0146			7.8852			10.1438			10.6945			8.8387			12.514
COMFAC (p-value)			0.09105			0.09588			0.0000			0.0000			0.06526			0.01391

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.17 – Cross-Section Results with Alternative Weight Matrices for SMEH (Southern)

	Queen2C SAR	Queen2C SEM	Queen2C SDM	Queen3C SAR	Queen3C SEM	Queen3C SDM	Queen4C SAR	Queen4C SEM	Queen4C SDM	K-20th SAR	K-20th SEM	K-20th SDM	K-40th SAR	K-40th SEM	K-40th SDM	K-60th SAR	K-60th SEM	K-60th SDM
lnGDP_{t-1}	-0.02892***	-0.03184***	-0.03191***	-0.029310***	-0.03221***	-0.03254***	-0.030082***	-0.03336***	-0.03382***	-0.028575***	-0.030508***	-0.03065***	-0.02917***	-0.03135***	-0.03158***	-0.02947***	-0.032022***	-0.03279***
	(-9.7248)	(-11.1996)	(-11.2393)	(-9.5557)	(-10.7224)	(-10.9038)	(-9.7226)	(-11.0007)	(-11.2105)	(-9.5503)	(-10.337)	(-10.485)	(-9.5142)	(-10.2459)	(-10.3838)	(-9.5622)	(-10.413)	(-10.8125)
ln(n+d+g)	0.002006	0.005547	0.006918	0.000176	0.00556	0.007972	-0.000647	0.004730	0.005007	0.0028009	0.009076	0.01323**	0.002506	0.00888	0.01424**	0.001427	0.007321	0.012451*
	(0.3987)	(0.8983)	(1.07)	(0.0339)	(0.8902)	(1.2001)	(-0.1221)	(0.7684)	(0.7495)	(0.5537)	(1.4915)	(2.0655)	(0.4814)	(1.4611)	(2.1859)	(0.2707)	(1.2243)	(1.906)
ln*SCHOOL*	0.02392***	0.008156	0.004769	0.026071***	0.011516	0.006624	0.028353***	0.0178412**	0.014041	0.0216318***	0.004634	-0.00159	0.02398***	0.01163	0.00234	0.026372***	0.016776**	0.007958
	(3.2126)	(0.951)	(0.5446)	(3.372)	(1.3264)	(0.7383)	(3.6061)	(2.0782)	(1.5693)	(2.893)	(0.5336)	(-0.1774)	(3.1209)	(1.3412)	(0.2544)	(3.3918)	(1.9932)	(0.8896)
ln*SMER*	-0.01197***	-0.01442***	-0.01353***	-0.01109***	-0.01301***	-0.01232***	-0.011095***	-0.012726***	-0.01179***	-0.011869***	-0.013335***	-0.0130***	-0.01137***	-0.01257***	-0.01274***	-0.01117***	-0.012645***	-0.01180***
	(-3.7404)	(-4.7719)	(-4.461)	(-3.3524)	(-4.0921)	(-3.8875)	(-3.3133)	(-3.9514)	(-3.6733)	(-3.6906)	(-4.2902)	(-4.2016)	(-3.4423)	(-3.9055)	(-3.9693)	(-3.3521)	(-3.9254)	(-3.642)
ln*SMEH*	0.004049	0.015968	0.019255	0.002681	0.016777	0.02262*	0.0001262	0.0108007	0.015199	0.007643	0.0196774	0.02316*	0.004725	0.014130	0.019336	0.003141	0.009785	0.013160
	(0.3765)	(1.3052)	(1.5455)	(0.241)	(1.3594)	(1.7837)	(0.0112)	(0.8971)	(1.2353)	(0.7075)	(1.5762)	(1.8206)	(0.4262)	(1.1522)	(1.5481)	(0.2808)	(0.8198)	(1.1027)
ρ or λ	0.5626***	0.75664***	0.65367***	0.51801***	0.7503***	0.55296***	0.49183***	0.78049***	0.46616**	0.57295***	0.73057***	0.57689***	0.55221***	0.73046***	0.5506***	0.57195***	0.77244***	0.45637**
	(7.1036)	(10.542)	(7.3457)	(4.6675)	(7.6503)	(3.8988)	(3.4204)	(7.0891)	(2.3512)	(7.1475)	(9.2306)	(5.4147)	(4.9478)	(6.8529)	(3.6226)	(4.3367)	(7.0502)	(2.1984)
(Intercept)	0.007435	0.028132	-0.00158	0.003183	0.022978	0.00500	0.0038010	0.0245145	-0.010523	0.0055906	0.033692	-0.03878	0.008257	0.035151	-0.05868	0.00520	0.032804	-0.25404***
	(0.3993)	(1.1821)	(-0.0528)	(0.1641	(0.9908	(0.1152	(0.1916	(1.0952	(-0.1578	(0.2968	(1.4652	(-1.0953	(0.4219	(1.611	(-1.0689	(0.262	(1.5744	(-2.7729
W*lnGDP_{t-1}			0.020046**			0.0200008*			0.036804***			0.00898			0.006463			0.002249
			(2.4933)			(1.8625)			(2.6157)			(0.9835)			(0.5212)			(0.1455)
W*ln(n+d+g)			-0.006608			-0.011365			-0.002911			-0.01913*			-0.02624*			-0.03805**
			(-0.6144)			(-0.8626)			(-0.1767)			(-1.6465)			(-1.8551)			(-2.1246)
W*ln*SCHOOL*			0.0259379			0.0364783			0.009412			0.05517***			0.07110**			0.098789**
			(1.4477)			(1.4937)			(0.2886)			(2.7771)			(2.4455)			(2.565)
W*ln*SMER*			0.028142***			0.037642**			0.077537***			0.02719**			0.02599			0.078374**
			(2.6483)			(2.1826)			(2.6981)			(2.08)			(1.0769)			(2.4037)
W*ln*SMEH*			-0.026845			-0.045571			-0.009053			-0.02965			-0.03714			0.045563
			(-1.119)			(-1.3953)			(-0.1993)			(-1.1768)			(-1.1383)			(1.0588)
Observations	301	301	301	301	301	301	301	301	301	301	301	301	301	301	301	301	301	301
LIK	813.2148	825.7436	830.1507	805.498	814.7021	819.5173	802.3126	810.761	816.2334	812.5894	819.8378	825.6569	806.2741	811.6819	816.7957	803.9901	809.8007	817.2988
AIC	-1610.4	-1635.5	-1634.3	-1595	-1613.4	-1613	-1588.6	-1605.5	-1606.5	-1609.2	-1623.7	-1625.3	-1596.5	-1607.4	-1607.6	-1592	-1603.6	-1608.6
LR test	29.081	54.139	35.983	13.648	32.056	13.23	7.277	24.174	4.9933	27.831	42.327	21.342	15.2	26.016	9.423	10.632	22.253	3.2058
LR test (p-value)	0.000000	0.000000	0.0000000	0.00022	0.000000	0.000275	0.006984	0.000000	0.025446	0.000000	0.00000	0.000003	0.0000	0.00000	0.002142	0.0011115	0.000002	0.073377
Wald	50.461	111.14	53.96	21.786	58.527	15.201	11.699	50.255	5.5279	51.087	85.205	29.319	24.48	46.962	13.123	18.807	49.706	4.8327
Wald (p-value)	0.0000000	0.000000	0.0000000	0.00000	0.0000000	0.0000966	0.0006252	0.00000	0.018715	0.0000000	0.00000000	0.0000	0.00000	0.00000000	0.0002916	0.000014	0.00000	0.027924
COMFAC			8.8143			9.6304			10.9448			11.6381			10.2277			14.9964
COMFAC(p-value)			0.1167			0.08641			0.05248			0.0401			0.06904			0.01038

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

5.9.7 Additional Results for Panel Estimations

Table 5.18 Alternative Panel Results for Brazil (SAR Model)

	Alternative Panel Results for Brazil – Manufacturing							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
lnGDP_{t-1}	-0.139914***	-0.131006***	-0.134449***	-0.136305***	-0.133528***	-0.131654***	-0.13399***	-0.136502***
	(-31.681382)	(-28.806176)	(-29.887617)	(-30.490168)	(-31.335374)	(-30.609336)	(-29.608494)	(-30.793405)
ln*SCHOOL*	0.054383***	0.049916***	0.052557***	0.053735***	0.045548***	0.045555***	0.047586***	0.049438***
	(4.691483)	(4.43105)	(4.581708)	(4.64967)	(4.085353)	(4.065226)	(4.153558)	(4.266588)
ln($n+d+g$)	-0.010883	-0.003833	-0.004651	-0.006259	-0.000295	-0.002082	-0.005135	-0.00681
	(-1.392759)	(-0.503704	(-0.600154)	(-0.802986)	(-0.039153)	(-0.275558)	(-0.665888)	(-0.872301)
ln*SMER*	-0.008935***	-0.008224***	-0.007505**	-0.007469**	-0.008722***	-0.008986***	-0.007834**	-0.007987**
	(-2.871781)	(-2.720742)	(-2.431474)	(-2.401444)	(-2.91805)	(-2.990984)	(-2.551514)	(-2.568231)
ln*SMEH*	0.019276***	0.019716***	0.018101***	0.017648***	0.019922***	0.020059***	0.019237***	0.018543***
	(3.174595)	(3.342806)	(3.012103)	(2.914122)	(3.414489)	(3.420336)	(3.215186)	(3.056831)
W*dep.var	0.738969***	0.512991***	0.51799***	0.511999***	0.361994***	0.507943***	0.52798***	0.489981***
	(8.71534)	(7.192823)	(6.845849)	(6.768926)	(14.581694)	(14.823013)	(6.337721)	(9.229249)
R^2	0.6072	0.6297	0.6153	0.6095	0.6374	0.6337	0.6187	0.608
log-likelihood	3434.8545	3473.8184	3442.2709	3429.598	3484.3919	3480.7689	3448.5597	3425.2884
Observations	2032	2032	2032	2032	2032	2032	2032	2032
LR-test	1189.0335	1151.6014	1157.2755	1169.3269	1232.7281	1172.5099	1168.3890	1174.4354
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inverse Distance Row-standardised**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.18 Alternative Panel Results for Brazil (SAR Model) - *continued*

	Alternative Panel Results for Brazil – All Sectors							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lnGDP_{t-1}	-0.140977***	-0.132075***	-0.135397***	-0.137445***	-0.134467***	-0.132785***	-0.135089***	-0.137592***
	(-31.787362)	(-29.090367)	(-30.152212)	(-31.096129)	(-31.527296)	(-30.861575)	(-29.892626)	(-31.055501)
ln*SCHOOL*	0.060858***	0.056324***	0.058241***	0.059494***	0.052171***	0.052291***	0.053988***	0.055594***
	(5.406362)	(5.150218)	(5.230837)	(5.302207)	(4.814294)	(4.80252)	(4.854101)	(4.940856)
ln($n+d+g$)	-0.013083*	-0.006019	-0.006633	-0.008221	-0.002536	-0.004401	-0.007201	-0.008842
	(-1.67371)	(-0.79088)	(-0.856167)	(-1.054742)	(-0.336507)	(-0.581904)	(-0.933449)	(-1.132243)
ln*SMER*	-0.029376***	-0.029008***	-0.028167***	-0.027547***	-0.027374***	-0.028793***	-0.028336***	-0.028148***
	(-5.245032)	(-5.335129)	(-5.083509)	(-4.931832)	(-5.080934)	(-5.319838)	(-5.134894)	(-5.030362)
ln*SMEH*	0.008643	0.007161	0.00875	0.010339	0.002288	0.004798	0.009421	0.009554
	(0.621804)	(0.530907)	(0.636667)	(0.746214)	(0.171176)	(0.357343)	(0.688071)	(0.688261)
W*dep.var	0.749982***	0.513978***	0.527989***	0.507991***	0.361971***	0.502983***	0.52798***	0.489974***
	(6.594564)	(7.26751)	(7.048641)	(9.643986)	(14.598953)	(14.711891)	(6.386791)	(9.24917)
R^2	0.6092	0.6316	0.6176	0.6115	0.6385	0.635	0.6207	0.6099
log-likelihood	3440.2058	3479.0723	3448.0805	3435.0772	3487.3692	3484.8422	3453.7448	3430.399
Observations	2032	2032	2032	2032	2032	2032	2032	2032
LR-test	1179.9978	1140.7306	1145.5252	1157.4488	1222.7544	1162.1570	1156.7704	1161.7980
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist. Row-std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.19 Alternative Panel Results for Brazil (SEM Model)

Alternative Panel Results for Brazil – Manufacturing								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lnGDP_{t-1}	-0.141444***	-0.147074***	-0.144064***	-0.143568***	-0.155999***	-0.147517***	-0.14352***	-0.141965***
	(-31.880378)	(-32.90528)	(-32.273621)	(-32.115374)	(-35.281389)	(-33.051849)	(-32.278302)	(-31.835314)
ln*SCHOOL*	0.076607***	0.099954***	0.100741***	0.096319***	0.078791***	0.088662***	0.091267***	0.088224***
	(5.365387)	(6.242636)	(6.28699)	(6.045707)	(5.521624)	(5.542686)	(5.761791)	(5.406492)
ln(n+d+g)	-0.009247	-0.006859	-0.011977	-0.013366*	-0.000239	-0.000854	-0.009886	-0.012821
	(-1.182272)	(-0.896857)	(-1.520782)	(-1.687259)	(-0.031987)	(-0.11057)	(-1.256426)	(-1.6171)
ln*SMER*	-0.008771***	-0.009603***	-0.008511***	-0.007648**	-0.011908***	-0.010605***	-0.007998***	-0.008811***
	(-2.835177)	(-3.259949)	(-2.794463)	(-2.474887)	(-4.204978)	(-3.598576)	(-2.627378)	(-2.840637)
ln*SMEH*	0.019852***	0.023839***	0.022524***	0.020118***	0.022561***	0.023512***	0.02386***	0.021418***
	(3.294974)	(4.1468)	(3.794309)	(3.360388)	(4.059404)	(4.097334)	(4.041644)	(3.558653)
λ(SEM)	0.921981***	0.705984***	0.731946***	0.736***	0.534996***	0.66697***	0.72097***	0.711998***
	(32.975531)	(12.930876)	(17.25376)	(14.882821)	(21.631685)	(9.197765)	(17.289575)	(7.953586)
R^2	0.5901	0.5871	0.5872	0.588	0.588	0.5884	0.5884	0.5889
log-likelihood	3445.9226	3529.1748	3478.6166	3456.3382	3564.1907	3533.0836	3483.1445	3447.3113
Observations	2032	2032	2032	2032	2032	2032	2032	2032
LR-test	1149.5399	1155.8631	1136.6677	1123.8283	1267.3717	1147.4002	1125.5978	1111.3346
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv.Dist.Row-std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.19 Alternative Panel Results for Brazil (SEM Model) - *continued*

	Alternative Panel Results for Brazil – All Sectors							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lnGDP_{t-1}	-0.142535***	-0.147657***	-0.144651***	-0.144273***	-0.156436***	-0.148189***	-0.144275***	-0.142831***
	(-32.123496)	(-33.050954)	(-32.389633)	(-32.286874)	(-36.751905)	(-33.109235)	(-32.402745)	(-32.027695)
ln*SCHOOL*	0.080162***	0.104421***	0.103678***	0.098856***	0.084373***	0.093635***	0.095352***	0.091223***
	(5.694772)	(6.672251)	(6.535472)	(6.288054)	(6.135296)	(5.914319)	(5.966745)	(5.678843)
ln($n+d+g$)	-0.011	-0.007948	-0.013124*	-0.014496*	-0.001828	-0.001883	-0.010787	-0.014285*
	(-1.40532)	(-1.036084)	(-1.663672)	(-1.827971)	(-0.243606)	(-0.242739)	(-1.368472)	(-1.799212)
ln*SMER*	-0.028651***	-0.027277***	-0.026791***	-0.025811***	-0.026948***	-0.027436***	-0.026671***	-0.028665***
	(-5.143245)	(-5.134539)	(-4.891743)	(-4.635072)	(-5.279287)	(-5.208794)	(-4.85819)	(-5.128034)
ln*SMEH*	0.011696	0.0147	0.013968	0.012282	0.007682	0.010035	0.016249	0.009363
	(0.837238)	(1.087073)	(1.012545)	(0.881567)	(0.580589)	(0.738251)	(1.168592)	(0.666953)
λ(SEM)	0.921986***	0.695992***	0.725993***	0.724965***	0.523995	0.660995***	0.710943***	0.703995***
	(32.98236)	(20.525434)	(16.960461)	(14.462502)	(0.001)	(8.843582)	(9.459148)	(7.73323)
R^2	0.5922	0.5898	0.5899	0.5905	0.5905	0.5909	0.591	0.5913
log-likelihood	3450.2754	3529.5951	3480.3803	3459.003	3561.6643	3532.3242	3484.5471	3450.5848
Observations	2032	2032	2032	2032	2032	2032	2032	2032
LR-test	1144.5250	1149.8534	1127.1371	1115.8246	1262.9035	1144.3203	1117.3736	1104.7352
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist. Row-std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.20. Alternative Panel Results for Brazil (SDM Model)

Alternative Panel Results for Brazil – Manufacturing								
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
lnGDP_{t-1}	-0.142351***	-0.148098***	-0.145116***	-0.14512***	-0.158453***	-0.148676***	-0.144148***	-0.142755***
	(-31.734252)	(-32.990768)	(-32.399526)	(-32.301267)	(-36.219154)	(-33.329493)	(-32.337816)	(-31.983953)
ln*SCHOOL*	0.053416***	0.107584***	0.110945***	0.106837***	0.097896***	0.102087***	0.102885***	0.101675***
	(4.128094)	(6.473066)	(6.670219)	(6.508448)	(6.216638)	(6.172103)	(6.236746)	(6.142149)
ln($n+d+g$)	-0.006876	-0.007704	-0.013732*	-0.015425*	-0.00285	-0.001377	-0.010662	-0.014151*
	(-0.862668)	(-1.005202)	(-1.738268)	(-1.936633)	(-0.381474)	(-0.178134)	(-1.350278)	(-1.785411)
ln*SMER*	-0.009513***	-0.009354***	-0.008341***	-0.007544**	-0.011642***	-0.010084***	-0.008078***	-0.008919***
	(-3.035746)	(-3.170152)	(-2.742072)	(-2.443855)	(-4.104348)	(-3.426379)	(-2.656571)	(-2.878225)
ln*SMEH*	0.019978***	0.023624***	0.022967***	0.020525***	0.022443***	0.023785***	0.02363***	0.021791***
	(3.316209)	(4.098267)	(3.872098)	(3.427967)	(4.03669)	(4.153028)	(4.002983)	(3.629338)
W*lnGDP_{t-1}	0.143754***	0.120429***	0.114558***	0.119115***	0.11268***	0.11714***	0.113453***	0.110143***
	(4.163505)	(9.905121)	(5.993189)	(4.929175)	(29.212193)	(8.060883)	(6.160182)	(5.35925)
W*lnschool	0.082926	-0.0789***	-0.068781**	-0.078878**	-0.085307***	-0.09904***	-0.086498***	-0.088648***
	(1.250885)	(-3.197608)	(-2.434458)	(-2.41696)	(-4.251582)	(-4.111428)	(-3.132725)	(-2.801112)
W* ln($n+d+g$)	-0.162717	-0.022381	0.024696	0.05312	-0.023984	-0.044036**	0.011517	0.048305
	(-1.436401)	(-1.064703)	(0.829355)	(1.382129)	(-1.63442)	(-2.136907)	(0.424256)	(1.385492)
W* ln*SMER*	0.088272*	0.013653	0.019061	0.003028	0.016724***	0.019498**	-0.007354	-0.000622
	(1.892891)	(1.33772)	(1.377896)	(0.182538)	(2.855866)	(2.085097)	(-0.573144)	(-0.038962)
W*l ln*SMEH*	-0.186488**	-0.046648**	-0.079181***	-0.071196*	-0.018865	-0.027918	-0.056957**	-0.048679
	(-2.377877)	(-2.32361)	(-2.76872)	(-1.903531)	(-1.616554)	(-1.486953)	(-2.20345)	(-1.432915)
W*dep.var.	0.901999***	0.674968***	0.708974***	0.706989***	0.527967***	0.66495***	0.704984***	0.696981***
	(21.529722)	(19.39233)	(9.68196)	(8.550204)	(10.910312)	(9.196802)	(9.230489)	(16.359246)
R^2	0.6138	0.6558	0.6343	0.624	0.6798	0.6602	0.637	0.6207
log-likelihood	3423.7883	3533.5497	3484.3651	3461.6948	3577.6234	3539.9268	3487.6394	3451.5683
Observation	2032	2032	2032	2032	2032	2032	2032	2032
LR-test	1095.5135	1155.2962	1141.6921	1128.575	1262.12	1146.9836	1124.7989	1112.0610
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist. Row-std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.20. Alternative Panel Results for Brazil (SDM Model) - *continued*

	Alternative Panel Results for Brazil – All Sectors							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
lnGDP_{t-1}	-0.143491***	-0.148898***	-0.145813***	-0.14607***	-0.158728***	-0.149426***	-0.145172***	-0.143745***
	(-31.993127)	(-33.115207)	(-32.473632)	(-32.528008)	(-36.113682)	(-33.34891)	(-32.48553)	(-32.177711)
ln*SCHOOL*	0.056584***	0.11136***	0.115557***	0.112162***	0.101076***	0.10512***	0.105456***	0.10482***
	(3.786819)	(6.738036)	(6.970775)	(6.879506)	(6.437504)	(6.374508)	(6.427849)	(6.377111)
ln($n+d+g$)	-0.009225	-0.008614	-0.015161*	-0.01686**	-0.004006	-0.001961	-0.011704	-0.015618**
	(-1.152277)	(-1.121038)	(-1.913844)	(-2.116083)	(-0.5335)	(-0.252449)	(-1.47915)	(-1.966321)
ln*SMER*	-0.03095***	-0.027278***	-0.026226***	-0.025381***	-0.027213***	-0.027681***	-0.026135***	-0.028248***
	(-5.514211)	(-5.132617)	(-4.786197)	(-4.567554)	(-5.31571)	(-5.249635)	(-4.764425)	(-5.058118)
ln*SMEH*	0.012784	0.015286	0.013862	0.013156	0.008277	0.009467	0.017315	0.00924
	(0.915543)	(1.129383)	(1.003272)	(0.945207)	(0.626709)	(0.695608)	(1.240474)	(0.655393)
W*lnGDP_{t-1}	0.144865***	0.118656***	0.108003***	0.119537***	0.112563***	0.116259***	0.107762***	0.10752***
	(4.203065)	(9.711712)	(6.408888)	(5.675058)	(28.893256)	(7.726726)	(6.848959)	(4.963092)
W*lnschool	-0.030851	-0.113054***	-0.131587***	-0.161564***	-0.089424***	-0.108595***	-0.127579***	-0.127375***
	(-1.035018)	(-5.047303)	(-5.188886)	(-5.61533)	(-4.638145)	(-4.861941)	(-5.236224)	(-4.563714)
W* ln($n+d+g$)	-0.120236	-0.025575	0.025668	0.02847	-0.022707	-0.04463**	-0.000229	0.039559
	(-1.077768)	(-1.222262)	(0.880682)	(0.758499)	(-1.540872)	(-2.173588)	(-0.008609)	(1.158124)
W* ln*SMER*	0.189359**	0.009573	0.027752	0.017782	0.012799	0.013129	-0.000463	0.024388
	(2.479233)	(0.518044)	(1.118255)	(0.57339)	(1.187999)	(0.756557)	(-0.020044)	(0.848189)
W* ln*SMEH*	-0.099507*	-0.085915**	-0.135728**	-0.226227***	0.002633	-0.007729	-0.11395**	-0.112632
	(-1.662536)	(-2.043812)	(-2.33733)	(-2.891418)	(0.104642)	(-0.19727)	(-2.104895)	(-1.563385)
W*dep.var.	0.895982***	0.678953***	0.68295***	0.687964***	0.527951***	0.653998***	0.68499***	0.683952***
	(19.898376)	(19.505566)	(14.550408)	(12.553161)	(10.961235)	(8.742679)	(15.364538)	(12.079103)
R^2	0.6152	0.6561	0.634	0.6254	0.6783	0.6591	0.637	0.6217
log-likelihood	3429.9041	3534.2163	3485.1655	3466.419	3572.8919	3538.1747	3489.0774	3454.8303
Observations	2032	2032	2032	2032	2032	2032	2032	2032
LR-test	1097.0524	1148.367	1127.6187	1120.5716	1259.6419	1140.2913	1117.0471	1103.3535
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist. Row-std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.21 Alternative Panel Results for Northern Regions (SAR Model)

	Alternative Panel Results for Northern Regions – Manufacturing							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln*GDP*$_{t-1}$	-0.138334***	-0.131845***	-0.134464***	-0.135626***	-0.132551***	-0.131977***	-0.13359***	-0.136481***
	(-20.406983)	(-19.024836)	(-19.452448)	(-19.77126)	(-19.298401)	(-19.165207)	(-19.353695)	(-19.818623)
ln*SCHOOL*	0.098981***	0.098711***	0.101576***	0.102303***	0.091511***	0.092852***	0.096851***	0.098764***
	(5.098862)	(5.087133)	(5.167354)	(5.202339)	(4.787422)	(4.823287)	(4.961268)	(5.007474)
ln(*n*+*d*+*g*)	-0.013246	-0.010077	-0.010869	-0.010758	-0.006977	-0.007538	-0.009985	-0.01074
	(-1.420497)	(-1.08234)	(-1.156899)	(-1.144357)	(-0.7597)	(-0.81591)	(-1.068284)	(-1.138467)
ln*SMER*	-0.003634	-0.003615	-0.002923	-0.002968	-0.003603	-0.004174	-0.003494	-0.003383
	(-0.906628)	(-0.903723)	(-0.724162)	(-0.735258)	(-0.913449)	(-1.052406)	(-0.870515)	(-0.834537)
ln*SMEH*	0.021352***	0.022456***	0.022174***	0.021543***	0.021315***	0.022179***	0.022088***	0.021206***
	(3.226835)	(3.400737)	(3.32838)	(3.229945)	(3.271519)	(3.38313)	(3.330847)	(3.168142)
W*dep.var.	0.415988***	0.367987***	0.368985***	0.425991***	0.274961***	0.369954***	0.371978***	0.268988***
	(3.028963)	(5.820169)	(4.253199)	(4.441568)	(6.47491)	(6.281635)	(4.852005)	(2.57955)
R^2	0.6672	0.6686	0.6628	0.6621	0.6773	0.6733	0.6657	0.6594
log-likelihood	1392.2064	1400.7358	1394.9511	1394.0657	1408.5027	1405.321	1397.6904	1391.4895
Observations	828	828	828	828	828	828	828	828
LR-test	489.6081	458.6072	465.6886	467.3306	511.6631	468.6279	466.4659	473.1064
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist. Row-std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.21 Alternative Panel Results for Northern Regions (SAR Model) - *continued*

	Alternative Panel Results for Northern Regions – All Sectors							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln*GDP*$_{t-1}$	-0.137997***	-0.131562***	-0.134384***	-0.135461***	-0.132255***	-0.131786***	-0.133454***	-0.136234***
	(-20.376346)	(-18.982415)	(-19.426265)	(-19.740402)	(-19.243171)	(-19.120964)	(-19.325778)	(-19.782504)
ln*SCHOOL*	0.097323***	0.097324***	0.100271***	0.100809***	0.090316***	0.091775***	0.095642***	0.097364***
	(5.026123)	(5.023785)	(5.09823)	(5.123783)	(4.732761)	(4.774683)	(4.906779)	(4.944099)
ln($n+d+g$)	-0.015607*	-0.012486	-0.013224	-0.013108	-0.009396	-0.01008	-0.012397	-0.013095
	(-1.670144)	(-1.337892)	(-1.403427)	(-1.390152)	(-1.020161)	(-1.087661)	(-1.322462)	(-1.384288)
ln*SMER*	-0.02544***	-0.025794***	-0.02439***	-0.024211***	-0.024403***	-0.02533***	-0.025072***	-0.024427***
	(-3.93415)	(-3.999642)	(-3.746321)	(-3.716743)	(-3.833787)	(-3.955062)	(-3.869775)	(-3.733391)
ln*SMEH*	-0.005443	-0.005626	-0.005512	-0.005692	-0.007872	-0.00818	-0.005689	-0.00622
	(-0.334699)	(-0.346414)	(-0.336218)	(-0.347045)	(-0.49078)	(-0.506723)	(-0.348632)	(-0.377899)
W*dep.var.	0.417965***	0.363995***	0.343991***	0.398986***	0.270963***	0.358999***	0.355986***	0.255982**
	(3.131089)	(5.812596)	(3.936514)	(4.108788)	(6.377254)	(6.059106)	(4.61961)	(2.468889)
R^2	0.6689	0.67	0.6636	0.6631	0.6784	0.6744	0.6668	0.6607
log-likelihood	1394.0536	1402.5587	1396.2045	1395.5352	1410.1158	1406.9279	1399.3015	1393.1311
Observation	828,	828,	828,	828,	828,	828,	828,	828,
LR-test	478.4175,	446.1124,	453.3827,	456.2589,	498.0178,	456.6047,	454.4141,	461.6917,
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist. Row-std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.22 Alternative Panel Results for Northern Regions (SEM Model)

Alternative Panel Results for Northern Regions – Manufacturing	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lnGDP_{t-1}	-0.140595***	-0.141005***	-0.140972***	-0.140121***	-0.146777***	-0.140781***	-0.139269***	-0.139172***
	(-20.484719)	(-20.13399)	(-20.304157)	(-20.319082)	(-21.348823)	(-20.300936)	(-20.084572)	(-20.206278)
ln*SCHOOL*	0.10013***	0.080229***	0.078205***	0.08384***	0.081919***	0.077239***	0.080235***	0.086435***
	(5.087403)	(3.938828)	(3.865917)	(4.180006)	(4.132833)	(3.811622)	(3.956066)	(4.278031)
ln(n+d+g)	-0.01107	-0.00859	-0.010529	-0.010198	-0.004226	-0.004519	-0.009798	-0.010545
	(-1.169635)	(-0.922649)	(-1.121993)	(-1.079666)	(-0.458115)	(-0.48569)	(-1.036039)	(-1.115509)
ln*SMER*	-0.00495	-0.00568	-0.005022	-0.004232	-0.00702*	-0.005791	-0.004488	-0.004749
	(-1.229735)	(-1.44451)	(-1.265416)	(-1.056908)	(-1.839559)	(-1.470686)	(-1.125337)	(-1.175618)
ln*SMEH*	0.021675***	0.023949***	0.023547***	0.02149***	0.021981***	0.023223***	0.024344***	0.022925***
	(3.269513)	(3.675256)	(3.569383)	(3.248554)	(3.474478)	(3.578473)	(3.702149)	(3.450252)
λ(SEM)	0.690984***	0.53798***	0.61396***	0.604983***	0.417999***	0.493964***	0.51197***	0.446958***
	(6.664285)	(7.886299)	(7.35076)	(5.917642)	(9.306575)	(7.526092)	(5.940827)	(3.847846)
R^2	0.6565	0.6559	0.6559	0.6563	0.6551	0.6557	0.6561	0.6563
log-likelihood	1396.5918	1410.432	1403.0382	1398.3542	1425.467	1413.2635	1403.5068	1394.9064
Observation	828,	828,	828,	828,	828,	828,	828,	828,
LR-test	481.3069,	460.9811,	469.1964,	464.2575,	522.4770,	459.8765,	462.0811,	469.1986,
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist. Row-std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.22 Alternative Panel Results for Northern Regions (SEM Model) - *continued*

	Alternative Panel Results for Northern Regions – All Sectors							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln*GDP*$_{t-1}$	-0.140652***	-0.141234***	-0.141071***	-0.140083***	-0.147087***	-0.141105***	-0.139314***	-0.139144***
	(-20.526801)	(-20.176085)	(-20.312371)	(-20.329077)	(-21.383071)	(-20.311413)	(-20.087906)	(-20.217155)
ln*SCHOOL*	0.098643***	0.079818***	0.077568***	0.081788***	0.081627***	0.075486***	0.080309***	0.086157***
	(5.026874)	(3.92981)	(3.83678)	(4.080008)	(4.132668)	(3.726822)	(3.969518)	(4.276911)
ln($n+d+g$)	-0.013245	-0.010195	-0.011941	-0.011938	-0.006111	-0.005494	-0.011625	-0.012513
	(-1.39779)	(-1.093532)	(-1.269588)	(-1.262034)	(-0.662164)	(-0.590249)	(-1.2258)	(-1.320293)
ln*SMER*	-0.026818***	-0.027584***	-0.02559***	-0.0251***	-0.026854***	-0.026754***	-0.026244***	-0.025874***
	(-4.137659)	(-4.34248)	(-3.982635)	(-3.875195)	(-4.391757)	(-4.25027)	(-4.059224)	(-3.965746)
ln*SMEH*	-0.002887	0.007076	0.00406	-0.001623	0.002939	-0.000063	0.003599	-0.002164
	(-0.175725)	(0.438656)	(0.250214)	(-0.099378)	(0.185921)	(-0.003923)	(0.219977)	(-0.13162)
λ(SEM)	0.692995***	0.540971***	0.594995***	0.595949***	0.417977***	0.522995***	0.495967***	0.417976***
	(6.72333)	(7.967196)	(6.874197)	(5.724829)	(9.305876)	(8.275209)	(5.630583)	(3.472687)
R^2	0.6581	0.6574	0.6575	0.6579	0.6568	0.6573	0.6577	0.6581
log-likelihood	1399.0474	1412.4753	1404.0266	1400.0435	1427.5831	1414.6942	1404.546	1396.2182
Observation	828	828	828	828	828	828	828	828
LR-test	468.2681	450.6210	457.4338	452.6266	517.4567	450.8766	448.7839	455.7533
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist. Row-std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.23. Alternative Panel Results for Northern Regions (SDM Model)

Alternative Panel Results for Northern Regions – Manufacturing								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lnGDP_{t-1}	-0.143566***	-0.142755***	-0.141881***	-0.143195***	-0.150191***	-0.141213***	-0.139941***	-0.141808***
	(-20.895736)	(-20.174347)	(-20.501544)	(-20.928394)	(-21.41593)	(-20.097915)	(-20.028712)	(-20.665008)
ln*SCHOOL*	0.098422***	0.074672***	0.075787***	0.077041***	0.079375***	0.071717***	0.076179***	0.079439***
	(4.94165)	(3.61198)	(3.70212)	(3.787061)	(3.957376)	(3.455209)	(3.702112)	(3.889585)
ln($n+d+g$)	-0.008384	-0.007425	-0.010089	-0.008632	-0.004252	-0.003449	-0.00859	-0.009549
	(-0.881276)	(-0.79934)	(-1.086375)	(-0.928129)	(-0.460478)	(-0.370266)	(-0.908916)	(-1.02335)
ln*SMER*	-0.005562	-0.006098	-0.004839	-0.004244	-0.006722*	-0.005668	-0.004961	-0.005091
	(-1.39276)	(-1.548542)	(-1.224263)	(-1.07298)	(-1.752762)	(-1.4409)	(-1.250107)	(-1.274785)
ln*SMEH*	0.020513***	0.02223***	0.021827***	0.019348***	0.021121***	0.022113***	0.022474***	0.019332***
	(3.132826)	(3.404839)	(3.34029)	(2.964929)	(3.33492)	(3.401263)	(3.424208)	(2.930358)
W* lnGDP_{t-1}	0.055431*	0.079088***	0.072354**	0.040249	0.083057***	0.064974***	0.050309*	0.018713
	(1.738592)	(3.623866)	(2.340483)	(1.015195)	(6.02232)	(3.164633)	(1.824473)	(0.503896)
W**SCHOOL*	0.144612**	0.106125*	0.264269***	0.578337***	-0.005564	0.052094	0.175213**	0.323593***
	(2.236093)	(1.851035)	(3.241639)	(4.685852)	(-0.152521)	(0.977414)	(2.415508)	(3.377195)
W* ln($n+d+g$)	-0.180806	-0.031506	-0.024279	-0.016375	-0.015294	-0.033594	0.001021	0.00741
	(-1.43775)	(-1.10873)	(-0.552265)	(-0.290002)	(-0.770827)	(-1.147116)	(0.02761)	(0.168343)
W*ln*SMER*	0.094657**	0.006477	0.047235*	0.078158**	0.01871**	0.017919	0.009566	0.063899**
	(2.1327)	(0.386728)	(1.898963)	(2.106753)	(2.326222)	(1.285742)	(0.483761)	(2.438572)
W* ln*SMEH*	-0.195695**	-0.068846***	-0.165416***	-0.271482***	-0.018188	-0.038506*	-0.131294***	-0.244267***
	(-2.440846)	(-2.632096)	(-3.980821)	(-4.22254)	(-1.324676)	(-1.670506)	(-3.774833)	(-4.737672)
W*dep.var.	0.52897***	0.482952***	0.491996***	0.499985***	0.393981***	0.447984***	0.404978***	0.215991
	(4.149489)	(6.630306)	(4.981028)	(4.64219)	(8.616676)	(6.491945)	(4.184338)	(1.521248)
R^2	0.6754	0.6824	0.6802	0.6791	0.6996	0.6838	0.6765	0.673
log-likelihood	1395.5278	1415.9955	1415.6425	1414.861	1431.1846	1416.7101	1410.7677	1408.6104
Observations	828	828	828	828	828	828	828	828
LR-test	473.0087	467.4763	491.1883	495.5561	523.1667	463.2095	471.6229	493.2492
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist. Row-std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.23. Alternative Panel Results for Northern Regions (SDM Model) - *continued*

	Alternative Panel Results for Northern Regions – All Sectors							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln*GDP*$_{t-1}$	-0.143928***	-0.14493***	-0.143496***	-0.143388***	-0.150427***	-0.142048***	-0.141878***	-0.142614***
	(-21.07318)	(-20.603801)	(-20.791176)	(-21.090043)	(-21.429139)	(-20.169109)	(-20.36313)	(-20.716666)
ln*SCHOOL*	0.088761***	0.075572***	0.077144***	0.080882***	0.078587***	0.069635***	0.072946***	0.077299***
	(4.440238)	(3.700752)	(3.785588)	(3.997778)	(3.924064)	(3.363194)	(3.564346)	(3.779309)
ln($n+d+g$)	-0.009697	-0.010181	-0.011739	-0.012268	-0.006705	-0.004461	-0.009184	-0.010346
	(-1.02055)	(-1.102338)	(-1.265688)	(-1.324363)	(-0.725456)	(-0.477461)	(-0.972694)	(-1.101921)
ln*SMER*	-0.027861***	-0.025456***	-0.022836***	-0.023839***	-0.02559***	-0.025937***	-0.023873***	-0.023079***
	(-4.359487)	(-4.040311)	(-3.598724)	(-3.764403)	(-4.148422)	(-4.095539)	(-3.720528)	(-3.560683)
ln*SMEH*	-0.004823	0.006254	-0.000373	0.000077	0.004587	-0.000637	0.004193	-0.003402
	(-0.300274)	(0.391282)	(-0.023285)	(0.004848)	(0.290272)	(-0.039394)	(0.257135)	(-0.208314)
W*ln*GDP*$_{t-1}$	0.088467***	0.059134***	0.043396	0.004341	0.084622***	0.063445***	0.022886	-0.012758
	(2.780112)	(2.670318)	(1.332628)	(0.099801)	(6.147712)	(3.082472)	(0.799928)	(-0.329017)
W**SCHOOL*	0.074301**	0.12698**	0.253815***	0.55515***	-0.004777	0.055649	0.170906**	0.276897***
	(2.548112)	(2.232075)	(3.148497)	(4.54492)	(-0.132608)	(1.060129)	(2.428578)	(2.949712)
W* ln($n+d+g$)	-0.258823**	-0.046278	-0.079815*	-0.081744	-0.015402	-0.042107	-0.020467	-0.01729
	(-1.962559)	(-1.588808)	(-1.64961)	(-1.1914)	(-0.766712)	(-1.418072)	(-0.533713)	(-0.361365)
W*ln*SMER*	0.139146**	0.05564**	0.091973**	0.189254***	0.027628**	0.034034	0.071816**	0.102741**
	(2.10771)	(2.061938)	(2.317865)	(3.113527)	(2.055746)	(1.520197)	(2.230156)	(2.451894)
W* ln*SMEH*	-0.347174***	-0.289563***	-0.478566***	-0.759139***	-0.041002	-0.072334	-0.363831***	-0.47297***
	(-3.193859)	(-4.450481)	(-4.650289)	(-4.802337)	(-1.304954)	(-1.377891)	(-4.305008)	(-4.006492)
W*dep.var.	0.609941***	0.414991***	0.399976***	0.395983***	0.393999***	0.44198***	0.355989***	0.200951
	(5.926775)	(5.311783)	(3.642772)	(3.232644)	(8.613556)	(6.3672)	(3.508235)	(1.387527)
R^2	0.6792	0.6889	0.6837	0.6838	0.7009	0.6851	0.6802	0.6726
log-likelihood	1392.7947	1426.1646	1421.1869	1421.7797	1432.8714	1418.5812	1416.2767	1408.0907
Observations	828	828	828	828	828	828	828	828
LR-test	454.8014	473.1601	486.2076	490.6113	523.4576	453.9118	468.0581	472.3553
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist. Row-std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.24 Alternative Panel Results for Southern Regions (SAR Model)

	Alternative Panel Results for Southern Regions – Manufacturing							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln*GDP*$_{t-1}$	-0.142583***	-0.13692***	-0.141778***	-0.143162***	-0.137877***	-0.137078***	-0.140023***	-0.142865***
	(-24.203417)	(-23.760452)	(-24.448245)	(-24.473755)	(-23.71762)	(-23.853399)	(-24.309671)	(-24.559396)
ln*SCHOOL*	0.100478***	0.104194***	0.10667***	0.105862***	0.086035***	0.087777***	0.095703***	0.102226***
	(4.523063)	(4.845805)	(4.87098)	(4.806238)	(4.066502)	(4.11062)	(4.373687)	(4.594548)
ln($n+d+g$)	-0.012086	-0.003043	-0.005319	-0.005865	0.006039	0.003602	0.001147	-0.003667
	(-0.855782)	(-0.227)	(-0.386241)	(-0.419588)	(0.452446)	(0.270026)	(0.083969)	(-0.262963)
ln*SMER*	-0.016136***	-0.013458***	-0.013332***	-0.012904***	-0.015323***	-0.014863***	-0.012505***	-0.012451***
	(-3.311618)	(-2.909696)	(-2.80586)	(-2.673905)	(-3.326992)	(-3.228506)	(-2.654667)	(-2.589036)
ln*SMEH*	0.024154	0.017592	0.014783	0.016123	0.026146*	0.021244	0.015111	0.015703
	(1.595572)	(1.222693)	(1.000259)	(1.074855)	(1.827127)	(1.484494)	(1.0322)	(1.051697)
W*dep.var.	0.702992***	0.593974***	0.629981***	0.59899***	0.399995***	0.581975***	0.642999***	0.623983***
	(7.449043)	(14.834633)	(12.163939)	(8.784362)	(12.449702)	(14.413665)	(12.409246)	(8.961644)
R^2	0.5677	0.6107	0.5893	0.5765	0.6141	0.6145	0.5966	0.5799
log-likelihood	2059.1136	2100.5016	2073.3616	2057.9616	2099.5383	2103.4482	2081.3234	2061.2705
Observations	1204	1204	1204	1204	1204	1204	1204	1204
LR-test	673.4735	664.8494	666.7019	664.6059	683.5969	669.0898	670.8461	667.8138
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist.Row std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.24 Alternative Panel Results for Southern Regions (SAR Model) - *continued*

	Alternative Panel Results for Southern Regions – All Sectors							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lnGDP_{t-1}	-0.143808***	-0.13798***	-0.142936***	-0.144316***	-0.138907***	-0.138186***	-0.141122***	-0.143952***
	(-24.196474)	(-23.805753)	(-24.470859)	(-24.486273)	(-23.692303)	(-23.871437)	(-24.315223)	(-24.550626)
ln*SCHOOL*	0.109919***	0.111639***	0.113019***	0.112058***	0.096976***	0.096588***	0.102102***	0.108502***
	(5.0545)	(5.282355)	(5.26486)	(5.200576)	(4.667102)	(4.602023)	(4.751944)	(4.975453)
ln(n+d+g)	-0.014609	-0.005514	-0.007259	-0.007561	0.00264	0.000869	-0.000883	-0.005553
	(-1.036434)	(-0.412501)	(-0.52868)	(-0.542594)	(0.197902)	(0.065285)	(-0.064849)	(-0.399431)
ln*SMER*	-0.037538***	-0.031349***	-0.032073***	-0.032455***	-0.031935***	-0.033405***	-0.029647***	-0.030342***
	(-3.317583)	(-2.922127)	(-2.90995)	(-2.899893)	(-2.982332)	(-3.125763)	(-2.712909)	(-2.720162)
ln*SMEH*	0.019884	-0.002271	0.007267	0.017535	0.00007	0.00235	0.003124	0.00879
	(0.723161)	(-0.086803)	(0.270384)	(0.643523)	(0.002697)	(0.090308)	(0.117562)	(0.324606)
W*dep.var.	0.700945***	0.601967***	0.636947***	0.60598***	0.396992***	0.584967***	0.642997***	0.631996***
	(7.383335)	(15.073078)	(12.398392)	(8.980709)	(12.24148)	(14.416446)	(12.35018)	(9.178282)
R^2	0.5668	0.6106	0.5893	0.5767	0.6122	0.6138	0.5963	0.5799
log-likelihood	2058.1682	2099.9335	2073.1899	2058.0865	2096.871	2102.089	2080.9748	2061.1105
Observations	1204	1204	1204	1204	1204	1204	1204	1204
LR-test	683.7240	671.5450	674.2132	672.8828	689.1384	675.5625	677.2036	674.9723
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist.Row.std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.25 Alternative Panel Results for Southern Regions (SEM Model)

	Alternative Panel Results for Southern Regions – Manufacturing							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln*GDP*$_{t-1}$	-0.146902***	-0.152268***	-0.148373***	-0.146975***	-0.161245***	-0.155162***	-0.148449***	-0.146003***
	(-25.059927)	(-26.689511)	(-25.733493)	(-25.270182)	(-28.721531)	(-27.514246)	(-25.873956)	(-25.272191)
ln*SCHOOL*	0.118801***	0.190099***	0.181051***	0.149964***	0.149336***	0.174557***	0.171074***	0.167185***
	(5.031679)	(6.919226)	(6.686483)	(5.746245)	(5.757369)	(6.343311)	(6.237529)	(6.24261)
ln(n+d+g)	-0.006217	-0.008281	-0.017057	-0.01695	0.00852	0.005132	-0.007866	-0.015694
	(-0.442773)	(-0.589395)	(-1.181195)	(-1.174567)	(0.624763)	(0.369321)	(-0.53004)	(-1.056838)
ln*SMER*	-0.014294***	-0.014296***	-0.012859***	-0.011165**	-0.017121***	-0.017668***	-0.011953**	-0.011204**
	(-2.965457)	(-3.144244)	(-2.685778)	(-2.290243)	(-3.943414)	(-3.983714)	(-2.533166)	(-2.315622)
ln*SMEH*	0.02317	0.024374*	0.019001	0.014312	0.029958**	0.024051*	0.015869	0.011172
	(1.533165)	(1.688658)	(1.262415)	(0.937355)	(2.137606)	(1.694803)	(1.065142)	(0.735645)
λ(SEM)	0.904987***	0.738974***	0.75296***	0.734974***	0.565988***	0.759948***	0.766978***	0.757955***
	(27.040824)	(19.09484)	(15.086679)	(11.88765)	(17.406117)	(21.364284)	(16.391189)	(12.828254)
R^2	0.5532	0.547	0.5481	0.5509	0.5493	0.5484	0.5495	0.5495
log-likelihood	2068.8968	2136.4584	2090.9771	2069.9257	2156.0158	2149.6491	2102.4703	2072.5474
Observations	1204	1204	1204	1204	1204	1204	1204	1204
LR-test	664.0487	712.4636	686.4722	669.7392	739.9670	713.8774	684.1647	664.9278
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist.Row.std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.25 Alternative Panel Results for Southern Regions (SEM Model) - *continued*

	Alternative Panel Results for Southern Regions – All Sectors							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lnGDP_{t-1}	-0.147867***	-0.152558***	-0.149329***	-0.148007***	-0.160983***	-0.155714***	-0.149119***	-0.147026***
	(-25.039493)	(-26.546199)	(-25.716431)	(-25.284315)	(-28.298653)	(-27.364806)	(-25.812997)	(-25.261874)
ln*SCHOOL*	0.127333***	0.200398***	0.187338***	0.153777***	0.16343***	0.186736***	0.178384***	0.172115***
	(5.510088)	(7.371631)	(7.03822)	(6.010465)	(6.345234)	(6.873725)	(6.605719)	(6.538556)
ln($n+d+g$)	-0.00817	-0.009221	-0.01808	-0.017569	0.006247	0.004208	-0.008679	-0.016891
	(-0.582927)	(-0.654736)	(-1.250829)	(-1.21729)	(0.455521)	(0.301939)	(-0.58386)	(-1.136257)
ln*SMER*	-0.033138***	-0.027123**	-0.030367***	-0.028433**	-0.025442**	-0.033656***	-0.025387**	-0.027516**
	(-2.944343)	(-2.536771)	(-2.699211)	(-2.488672)	(-2.472945)	(-3.221346)	(-2.29907)	(-2.421738)
ln*SMEH*	0.027574	0.006348	0.018958	0.027971	-0.007834	0.003774	0.002939	0.001574
	(0.988472)	(0.229929)	(0.674695)	(0.993494)	(-0.291437)	(0.13808)	(0.105234)	(0.055832)
λ(SEM)	0.907986***	0.737954***	0.74897***	0.737973***	0.567987***	0.753949***	0.772982***	0.756984***
	(28.003776)	(19.016143)	(14.82629)	(12.043935)	(17.513295)	(20.86058)	(16.815298)	(12.774959)
R^2	0.5523	0.5457	0.5475	0.5506	0.5474	0.5471	0.5484	0.5489
log-likelihood	2067.9641	2133.5815	2090.4206	2070.2912	2149.4492	2145.7525	2101.4828	2072.6093
Observations	1204	1204	1204	1204	1204	1204	1204	1204
LR-test	675.1416	723.4399	696.0438	679.6545	749.2894	726.8070	695.0347	674.4220
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist. Row.std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.26. Alternative Panel Results for Southern Regions (SDM Model)

	Alternative Panel Results for Southern Regions – Manufacturing							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lnGDP_{t-1}	-0.150268***	-0.153564***	-0.149828***	-0.149501***	-0.16277***	-0.156799***	-0.149655***	-0.147262***
	(-25.375253)	(-26.803058)	(-25.963601)	(-25.866262)	(-29.033825)	(-27.831561)	(-25.997967)	(-25.473398)
ln*SCHOOL*	0.084503***	0.20468***	0.216563***	0.198995***	0.169882***	0.193858***	0.190778***	0.195232***
	(3.365642)	(7.0033)	(7.311046)	(6.931902)	(6.128221)	(6.753251)	(6.627982)	(6.836183)
ln($n+d+g$)	-0.003857	-0.009957	-0.027324*	-0.031158**	0.005608	0.002985	-0.009996	-0.020369
	(-0.261405)	(-0.694856)	(-1.821232)	(-2.087829)	(0.408565)	(0.213774)	(-0.659716)	(-1.338576)
ln*SMER*	-0.014265***	-0.014564***	-0.01465***	-0.013412***	-0.017258***	-0.01686***	-0.012467***	-0.01281***
	(-2.926182)	(-3.181412)	(-3.042962)	(-2.763643)	(-3.963275)	(-3.786686)	(-2.633158)	(-2.63602)
ln*SMEH*	0.027827*	0.02465*	0.019901	0.01831	0.0298**	0.022727	0.014561	0.009636
	(1.831857)	(1.705802)	(1.31467)	(1.202828)	(2.140229)	(1.608359)	(0.974496)	(0.633377)
W* lnGDP_{t-1}	0.301368***	0.13532***	0.095769***	0.11512***	0.126634***	0.156274***	0.128775***	0.115729***
	(5.062278)	(7.809498)	(3.631917)	(3.473574)	(11.663708)	(9.715895)	(5.155102)	(3.163942)
W*lnschool	-0.091744	-0.19884***	-0.312831***	-0.391857***	-0.15547***	-0.204389***	-0.247606***	-0.319016***
	(-0.960941)	(-4.206515)	(-5.235633)	(-5.27147)	(-4.254763)	(-4.647239)	(-4.753146)	(-4.935089)
W* ln($n+d+g$)	0.370025**	-0.026859	0.128058**	0.409135***	-0.040913	-0.066153**	0.009448	0.102638*
	(2.236812)	(-0.678841)	(2.060786)	(4.775156)	(-1.610338)	(-1.974477)	(0.212162)	(1.731979)
W* ln*SMER*	0.020591	0.012712	-0.039931*	-0.113311***	0.015299*	0.036925***	-0.021371	-0.074184***
	(0.344724)	(0.824159)	(-1.823805)	(-3.989)	(1.655734)	(2.593622)	(-1.04355)	(-2.696019)
W*ln*SMEH*	-0.029108	-0.054406	-0.090004	-0.139399*	-0.016232	-0.018747	-0.033312	-0.036908
	(-0.263662)	(-1.194311)	(-1.425254)	(-1.724679)	(-0.608083)	(-0.466925)	(-0.587632)	(-0.47854)
W*dep.var.	0.85198***	0.731976***	0.667979***	0.529969***	0.567967***	0.750962***	0.733998***	0.634966***
	(19.393052)	(18.832235)	(11.051004)	(5.960019)	(17.581946)	(20.809851)	(14.434418)	(8.182043)
R^2	0.5844	0.6405	0.6064	0.5969	0.6667	0.6557	0.6156	0.5939
log-likelihood	2069.9506	2139.1217	2097.6648	2088.7692	2165.1898	2157.1257	2106.1929	2081.4381
Observations	1204	1204	1204	1204	1204	1204	1204	1204
LR-test	664.7269	714.5724	695.7395	704.5312	750.4950	724.3710	690.5047	680.9804
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist.Row std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Table 5.26. Alternative Panel Results for Southern Regions (SDM Model) - *continued*

	Alternative Panel Results for Southern Regions – All Sectors							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lnGDP_{t-1}	-0.151129***	-0.15365***	-0.15067***	-0.150791***	-0.16139***	-0.157194***	-0.150049***	-0.147674***
	(-25.367834)	(-26.605392)	(-25.872819)	(-25.785261)	(-28.355364)	(-27.616467)	(-25.880289)	(-25.265487)
ln*SCHOOL*	0.086704***	0.216701***	0.223427***	0.2014***	0.185136***	0.205633***	0.195578***	0.196949***
	(3.667262)	(7.470222)	(7.622499)	(7.079956)	(6.674631)	(7.213934)	(6.883492)	(6.974352)
ln($n+d+g$)	-0.003481	-0.009479	-0.027211*	-0.03147**	0.004364	0.004354	-0.009473	-0.02114
	(-0.23649)	(-0.661046)	(-1.811604)	(-2.10024)	(0.315955)	(0.310342)	(-0.624068)	(-1.385921)
ln*SMER*	-0.031196***	-0.027363**	-0.031925***	-0.029521***	-0.028394***	-0.033193***	-0.025703**	-0.028414**
	(-2.73337)	(-2.551963)	(-2.832739)	(-2.5932)	(-2.765977)	(-3.180548)	(-2.328092)	(-2.499507)
ln*SMEH*	0.055003**	0.006506	0.021739	0.039571	-0.008511	-0.000549	0.002128	0.001151
	(1.970769)	(0.233503)	(0.764189)	(1.396924)	(-0.314789)	(-0.020011)	(0.075446)	(0.040365)
W* lnGDP_{t-1}	0.282068***	0.133403***	0.096456***	0.099092***	0.120495***	0.157704***	0.133283***	0.126929***
	(4.441942)	(7.128891)	(3.356242)	(2.770488)	(10.66071)	(9.326481)	(5.098811)	(3.37313)
W*ln*SCHOOL*	-0.042992	-0.21579***	-0.305005***	-0.377581***	-0.169757***	-0.222382***	-0.228477***	-0.260733***
	(-0.522247)	(-4.552131)	(-4.937537)	(-4.807483)	(-4.668495)	(-5.052507)	(-4.355103)	(-3.9304)
W* ln($n+d+g$)	0.297929*	-0.014381	0.09433	0.330028***	-0.02919	-0.055729*	0.005383	0.059867
	(1.702003)	(-0.377186)	(1.571365)	(4.039707)	(-1.158803)	(-1.71959)	(0.126072)	(1.044684)
W* ln*SMER*	0.028048	-0.000122	-0.03466	-0.12102**	-0.017583	0.041636	-0.036705	-0.06638
	(0.237286)	(-0.003909)	(-0.869019)	(-2.513554)	(-0.840009)	(1.427273)	(-0.96796)	(-1.465121)
W*ln*SMEH*	-0.145878	0.016778	-0.079516	-0.301109**	0.102012**	0.057738	0.06138	0.047358
	(-1.142274)	(0.225942)	(-0.749635)	(-2.153993)	(2.154796)	(0.861775)	(0.672031)	(0.389575)
W*dep.var.	0.855971***	0.718955***	0.687998***	0.604961***	0.539981***	0.739972***	0.731978***	0.684992***
	(20.077196)	(17.910099)	(11.923157)	(7.696882)	(16.155188)	(19.896554)	(14.346685)	(9.712627)
R^2	0.5844	0.6381	0.6054	0.5952	0.6622	0.6525	0.615	0.5925
log-likelihood	2068.2425	2136.1149	2095.3985	2085.0349	2161.6888	2152.6601	2105.3471	2077.9969
Observations	1204	1204	1204	1204	1204	1204	1204	1204
LR-test	672.4435	724.0318	701.2071	704.3686	759.2383	734.7601	699.1313	680.1825
LR-test(p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sp. Weight	**Inv. Dist.Row.std**	**K-20th**	**K-40th**	**K-60th**	**Queen1**	**Queen2**	**Queen3**	**Queen4**

Note: * p-value<0.10, ** p-value<0.05, *** p-value<0.01. Numbers in brackets for the coefficients are the *t*-statistics.

Appendix 5.9.8. Additional Results: a Spatial Filtering Approach

In this section, we provide a robustness check on the importance of space for economic growth in Brazil, however, instead of considering the spatial structure explicitly, we analyse the impact of the explanatory variables on growth after removing the spatial dependence from the data. Therefore, the aim of this chapter is to re-examine the importance of the SME sector and other independent variables for economic growth after spatially filtering the data, thus removing spatial effects. We will therefore examine whether there is any impact of the independent variables on growth once we remove the spatial effect. We draw on Badinger et al. (2004) and Battisti and Di Vaio (2008), which analyse the convergence process in Europe after removing the spatial autocorrelation from the data. The conclusions drawn from this chapter provides additional evidence of the important of spatial structure for the process of economic growth in Brazil.

5.9.8.1 Model Specification without Spatial Dependence

The previous chapter provides evidence of the existence of spatial autocorrelation and spatial heterogeneity that can lead to misleading estimations when spatial econometrics is not considered. The common practice is to explicitly incorporate the spatial information in the regression specification as we did in the previous chapter. Alternatively, in this chapter, instead of correcting for the spatial autocorrelation by including the spatial structure into the model, we use a different approach and want to correct for the spatial dependence by removing it from the data.

The specification used in this chapter to study SMEs and economic growth using Brazilian micro-regional data is the same general growth Equation 4.1 (presented in Chapter 4) using data that explicitly removes the spatial effect:

$$gr^{*}_{it} = a_i - b\ln y^{*}_{i,t-1} + \beta \ln SMER^{*}_{it} + \alpha \ln SMEH^{*}_{it} + \psi \ln X^{*}_{it} + v_{it} \qquad (5.10)$$

where everything is defined as in equation 4.1 with the difference being the asterisk, *, denoting the spatially filtered variable. The subscript i denotes each individual micro-region, t represents each period of time considered, gr^* denotes the spatially filtered annual GDP per capita growth in each cross-section, $\ln y_{t-1}$ is the initial GDP per capita, a_i represents the time invariant economy specific effect, and v is the error term. Furthermore, $SMER^*$ and $SMEH^*$ are the spatially filtered versions of the relative size and human capital of the SME sector. The vector X^* encompasses other growth determinants as discussed in Chapter 4, after removing the spatial dependence.

5.9.8.2 Estimation Issues

In this section we follow Badinger et al. (2004) and present the procedure that allows for the estimation of the growth equation for a panel data setting that eliminates the spatial dependence and controls for endogeneity. Badinger et al. (2004) propose a two-step procedure that consists of filtering the data to separate the spatial effect from the data and subsequently applying the GMM estimators for dynamic panels used in Chapter 4 to treat the endogeneity problem. This procedure is described in detail in the following section.

5.9.8.2.1 The First Step: The Spatial Filtering

As argued in Badinger et al. (2004) and Battisti and Di Vaio (2008), the aim of the spatial filtering technique is to separate the spatial regional interdependencies from the data allowing for the use of conventional estimators that are based on the assumption of spatially uncorrelated errors.

As described in Getis and Griffith (2002), Badinger et al. (2004), Ferstl (2007) and Battisti and Di Vaio (2008), the spatial filter is based on the local statistic of spatial dependence G_i developed by Getis and Ord (1992) and Ord and Getis (1995). The G_i statistic is the defining element of a filter device and also reveals local spatial dependencies as LISA statistics presented in Chapter 4. It is given by the following expression:

$$G_i(\delta) = \frac{\sum_j w_{ij}(\delta) x_j}{\sum_j x_j}, \qquad i \neq j$$

where w_{ij} is the weight matrix, δ is the distance decay parameter, and X_j represents observations of a given variable. Furthermore, the expected value of the G_i statistic is given by:

$$\mathrm{E}[G_i] = \frac{\sum_j w_{ij}(\delta)}{(n-1)}$$

This expression represents the realization at a location i when no autocorrelation occurs. Dividing this expression (for a given location) by the observed value of the local autocorrelation given by G_i results in a ration that represents the spatially uncorrelated part of the data (e.g. Badinger et al. 2004 and Ferstl 2007). Therefore, the expression below is used to filter the data in every location:

$$x_i^* = x_i \left(\frac{\dfrac{\sum_j w_{ij}(\delta)}{(n-1)}}{\dfrac{\sum_j w_{ij}(\delta) x_j}{\sum_j x_j}} \right)$$

where the difference between $x_i - x_i^*$ represents the spatial component of a vector variable X at location i and consequently x_i^* represents the spatially filtered or spaceless variable.

In this chapter we use a distance based specification of the spatial structure that has a negative exponential function to model the distance decay function as in Badinger et al. (2004) and Ferstl (2007):

$$w_{ij} = \begin{cases} \exp(-\delta d_{ij}) \text{ if } i \neq j \\ 0 \quad \text{if otherwise} \end{cases}$$

where d_{ij} denotes the geographical distance by road between two micro-regions as defined in Chapter 4.

As argued in Badinger et al. (2004), the choice of δ is crucial and the objective is to minimize the remaining spatial autocorrelation in the filtered variable by varying the distance decay parameter δ. The aim is to filter the data and vary the parameter δ to check what value of this parameter provides a filtered variable with minimum spatial autocorrelation.[88] The spatial filtering removes the spatial autocorrelation for a given variable only for one period of time at a time. Therefore, we apply this filter for each variable in each time (t) period available.

5.9.8.2.2 The Second Step: Estimation of Standard Dynamic Panel Data Models

In the first step we eliminate the spatial dependencies in the data in order to use the filtered variables in conventional estimations that are based on the assumption that the error term is not spatially correlated. Therefore, our second step consists of applying standard panel data estimators to our spatially filtered data once we certify that we do not have spatially autocorrelated errors. As in Chapter 4, we estimate the OLS, LSDV and the GMM differenced GMM Arellano and Bond (1991) estimator (GMM-DIFF) and the system GMM Blundell and Bond (1998) estimator (GMM-SYS) to control for endogeneity by extracting the exogenous component from the independent variables and check the effect of the conditioning variables, including our SME variables on growth. We exploit the same data used in Chapters 4 and 5, therefore, the final panel is a balanced panel data from 1985 to 2004 with 508

[88] We follow Ferstl (2007) and minimize the global autocorrelation given by the Moran's I in the filtered variable (X^*) using the following objective function: $\delta_{opt} = \underbrace{\min}_{\delta} |z_I(X^*)|$, where Z_I is the z-transformed Moran's I. The goal of this objective function is to minimize the remaining spatial autocorrelation in the filtered variable by varying the distance decay parameter δ. In other words, the software searches for the value of δ that provides the filtered data with the minimum spatial autocorrelation.

Brazilian micro-regions organised in intervals as close as possible to five years to avoid business cycle influences.

5.9.8.3 Empirical Results and discussion

The existing literature provides evidence of convergence for Brazilian regions and suggests that convergence is conditioned to structural factors such as population growth, physical capital and human capital using either standard or spatial econometrics. In particular, recent studies recognize the importance of spatial spillovers to the process of economic growth (e.g. Mossi et al. 2003; Silveira-Neto and Azzoni 2006; Resende 2009). One interesting point is made by Silveira-Neto and Azzoni (2006) who suggest that growth determinants reflect the regional patterns of Brasil. Resende (2009) also indicates that values of conditioning variables might reflect the spatial structure of the country. The spatial filtering procedure used in this chapter offers another way of looking at this phenomenon. We want to see what will be the importance of growth determinants after removing the spatial autocorrelation from the data. In other words, what is left behind after removing the influence of space?

Firstly, we want to check if our filtering procedure removed the autocorrelation from the error term in the regression specification and test the OLS model against specifications with spatial correction. As in Chapter 4, we assess formally the presence of spatial dependence in the OLS regression using the Moran's *I* test statistic for the regression residuals and the Lagrange Multiplier tests to test OLS against the alternative SEM and SAR models under the null hypothesis of no spatial dependence. To perform these tests, for the cross-section specification, we use the weight matrix based on the inverse road distances from each pair of our 508 micro-regions to provide a direct comparison with the results using the original unfiltered variables reported in Table 5.3.

After presenting the spatial dependence diagnostic tests in column 1, Table 5.27 mimics Table 4.4 in construction and uses the spaceless version of the variables to analyse the effect of the conditioning variables and the SME sector on growth rates once we remove the spatial autocorrelation. Overall, we can observe that growth rates are affected mainly by the initial value of the GDP per capita, confirming that micro-regions are converging in Brazil, even after controlling for the spatial influence. Also, the remaining conditioning variables become insignificant, a clear indication that their values are intrinsically related to the spatial structure of the country.

Importantly, the Moran's *I* has a value of 0.0002 and is not statistically significant, indicating a lack of spatial dependence in the residuals. Furthermore, to consider a model with spatial dependence as the alternative to the OLS regression, the LM_{ERR} test and its robust version (LMR_{ERR}), and the LM_{LAG} test and its robust version (LMR_{LAG}) presented in Chapter 5 are reported at the bottom of column 1. All tests fail to reject the null of no spatial dependence, suggesting that regression residuals are not spatially autocorrelated. Therefore, these diagnostic tests suggest that the filtering procedure successfully removed the spatial autocorrelation from the data and from the errors of an OLS regression.

Now that we have removed the autocorrelation from the data we can rely on estimations that assume non-autocorrelation in the residuals using standard panel data techniques. The OLS estimation for panel data in column 2 confirms the presence of convergence among Brazilian micro-regions and indicates that the initial GDP is the only aspect that affects growth after removing the spatial autocorrelation. Human capital and population growth, as well as the proxies related to the SME sector became insignificant, a clear indication that their values are intrinsically related to the spatial structure of the country. When we incorporate the micro-regional specific effect using the LSDV to control for micro-regional specific time invariant heterogeneity, the results remain the same, only the initial GDP per capita is important for economic growth in the Brazilian micro-regions.

Nevertheless, OLS and LSDV estimations are subject to endogeneity bias and as in Chapter 4, to consider the endogeneity problems we apply the differenced GMM Arellano and Bond (1991) estimator (GMM-DIFF) and the system GMM Blundell and Bond (1998) estimator (GMM-SYS). In all GMM estimations, the autocorrelation test AR2 indicates validity of the instruments and the Hansen test for joint validity of the instruments also always fail to reject the null that instruments are valid.[89] Therefore, there are no problems with the validity of instruments in this set of regressions. All conditioning variables but the initial GDP became insignificant, a clear indication that their values are intrinsically related to the spatial structure of the country. The exception is the GMM-SYS estimation, which suggests a weak positive effect of human capital embodied in SMEs on growth. In a different manner, this is an additional support for the idea that human capital in the SME sector is more important that its size.

Therefore, because the values of the variables are related to the spatial structure, even if there is a non-spatial effect of a given variable on growth, it is difficult to disentangle this effect from the spatial one. The empirical results are robust to a change in the cut-off used to define SMEs and this can be confirmed in Table 5.27 from columns 8 to 13 where we use 500 employees instead of 250 to classify an establishment as small. Alternatively, we also present the results for the extended SME sector with commerce and services in Table 5.28 and estimations provide similar qualitative results.

The results reported in Table 5.27 support the view of Silveira-Neto and Azzoni (2006) and Resende (2009) who argue that values of growth determinants carry spatial information within them. Silveira-Neto and Azzoni (2006), for instance, find that there is no spatial dependence when considering spatial specifications for conditional growth regressions for Brazil using the conditional variables that have very strong regional or geographic patterns across Brazilian states. The effect of the conditioning variables on growth seems to be related to its geographical location.

Therefore, in Table 5.28, we provided additional evidence on the importance of space for the process of economic growth in Brazil addressing the spatial dependence and endogeneity. We estimated the growth equation for 508 Brazilian micro-regions by a two-step procedure consisting of eliminating the spatial dependence in the data and using the GMM estimators to treat endogeneity.

The diagnostic tests in column 1 of Table 5.27 indicate that the spatial autocorrelation was removed from the growth equation that uses spatially filtered data. Therefore, standard methods of panel estimators were applied and we found

[89] As explained in Chapter 4, the Hansen statistic tests the exogeneity of the instrument set under the null that instruments are valid.

two main interesting results. First, the empirical results from the filtered regressions confirm the existence of convergence even after removing the spatial dependence. Second, the filtered conditioning variables, including the aspects of the SME sector, became statistically insignificant for the economic growth process.

Therefore, the use of the spatial filter shows the importance of space by removing it from the estimations. The values of the conditioning variables across regions seem to be intrinsically linked with geographical location, supporting the view that conditioning variables carry strong spatial information with them. It seems that the inherent characteristics of the conditioning variables that affect growth are correlated across space and that information is removed by the spatial filter. Hence, estimations using filtered variables confirm the importance of space for economic growth in Brazil, since the spatial pattern intrinsic in the conditional variables is the component of these variables that seems to affect growth. However, these estimates cannot say anything about how space affects regional growth.

Table 5.27. Industry Employment Share, *SMEH* and Growth for Spatially Filtered Variables

	SME250							SME500					
	(1) OLS Cross-section	(2) OLS	(3) LSDV	(4) DiffGMM 1-step	(5) DiffGMM 2-step	(6) SysGMM 1-step	(7) SysGMM 2-step	(8) OLS	(9) LSDV	(10) DiffGMM 1-step	(11) DiffGMM 2-step	(12) SysGMM 1-step	(13) SysGMM 2-step
lnGDP_{t-1}	-0.0417***	-0.200***	-0.200***	-0.284***	-0.197***	-0.198***	-0.199***	-0.200***	-0.200***	-0.258***	-0.199***	-0.199***	-0.199***
	(-191.87)	(-299.39)	(-259.43)	(-3.08)	(-12.54)	(-74.96)	(-180.58)	(-299.41)	(-259.39)	(-3.98)	(-27.76)	(-76.88)	(-158.20)
ln*School*	0.0006	0.0003	0.00004	132.9	-11.64	-2.279	-1.705	0.0004	0.00006	91.80	-30.02	-1.915	-1.320
	(0.173)	(0.12)	(0.01)	(1.00)	(-0.12)	(-0.51)	(-0.71)	(0.12)	(0.02)	(1.00)	(-1.07)	(-0.46)	(-0.55)
ln$(n+g+d)$	-2.1658	-1.687	1.896	-11.21	-3.394	-1.010	-1.176	-1.668	1.899	-7.341	-3.122	0.0474	-1.666
	(-1.4880)	(-0.94)	(0.64)	(-0.99)	(-0.30)	(-0.14)	(-0.33)	(-0.94)	(0.64)	(-1.07)	(-0.38)	(0.01)	(-0.43)
ln*SMER*	0.0588	-0.0520	-0.662	2.257	-3.555	-6.094	-3.229	-0.0776	-0.559	0.192	-3.054	-6.343	-2.330
	(0.124)	(-0.08)	(-0.57)	(0.40)	(-1.42)	(-0.91)	(-1.07)	(-0.11)	(-0.46)	(0.05)	(-1.04)	(-0.89)	(-0.47)
ln*SMEH*	-2.8188**	-1.795	1.955	-32.14	1.301	35.89	15.92*	-1.837	1.684	-6.659	-1.054	36.27	11.52
	(-2.394)	(-1.30)	(0.92)	(-0.81)	(0.05)	(0.93)	(1.81)	(-1.35)	(0.80)	(-0.46)	(-0.30)	(0.95)	(0.86)
*Dummy*1995		-0.0127	-0.764	1459.9	-7.630	-33.39	-13.83	-0.0194	-0.686	1004.4	120.6	-29.09	-11.38
		(-0.01)	(-0.64)	(0.75)	(-0.03)	(-0.71)	(-0.90)	(-0.02)	(-0.58)	(0.72)	(0.77)	(-0.64)	(-0.60)
*Dummy*2000		0.960	-0.404	1443.2	-4.659	-37.77	-14.98	0.979	-0.313	990.7	127.0	-34.03	-12.21
		(0.79)	(-0.29)	(0.74)	(-0.02)	(-0.74)	(-0.92)	(0.80)	(-0.23)	(0.72)	(0.81)	(-0.69)	(-0.61)
*Dummy*2005		-0.531	-2.395	1426.2	-3.748	-46.03	-18.57	-0.500	-2.254	975.6	131.5	-42.51	-15.28
		(-0.39)	(-1.45)	(0.74)	(-0.01)	(-0.80)	(-1.08)	(-0.37)	(-1.37)	(0.71)	(0.82)	(-0.76)	(-0.71)
Observations	508	2032	2032	1524	1524	2032	2032	2032	2032	1524	1524	2032	2032
Adjusted R^2	0.9864	0.978	0.971					0.978	0.971				
Instruments				14	14	19	19			14	14	19	19
AR2				0.464	0.859	0.652	0.702			0.486	0.285	0.663	0.838
Hansen				0.755	0.755	0.273	0.273			0.163	0.163	0.238	0.238
Moran's *I*	0.0002(0.56)												
LM_{ERR}	0.0013(0.97)												
LMR_{ERR}	0.0998 (0.75)												
LM_{LAG}	0.998 (0.32)												
LMR_{LAG}	1.0965(0.29)												

Note: *t* statistics in parentheses and * p-value<0.10, ** p-value<0.05, *** p-value<0.01. The filtering procedure was implemented using MATLAB codes as described in Ferstl (2007). The cross-section OLS regression in column 1 was estimated in *R*, and the panel data estimations were implemented in Stata.

Table 5.28. Industry, Commerce and Services Employment Share, *SMEH* and Growth for Spatially Filtered Variables

	SME250						SME500					
	(1) OLS	(2) LSDV	(3) DiffGMM 1-step	(4) DiffGMM 2-step	(5) SysGMM 1-step	(6) SysGMM 2-step	(7) OLS	(8) LSDV	(9) DiffGMM 1-step	(10) DiffGMM 2-step	(11) SysGMM 1-step	(12) SysGMM 2-step
lnGDP_{t-1}	-0.200***	-0.200***	-0.262***	-0.202***	-0.204***	-0.200***	-0.200***	-0.200***	-0.257***	-0.205***	-0.202***	-0.200***
	(-299.53)	(-258.93)	(-3.97)	(-23.98)	(-45.56)	(-142.60)	(-299.53)	(-259.06)	(-4.23)	(-29.68)	(-157.56)	(-267.51)
ln*School*	0.0002	0.0001	97.82	-20.22	6.942	1.448	0.0002	0.0001	89.26	-2.780	2.505	0.584
	(0.07)	(0.04)	(1.03)	(-0.73)	(0.74)	(0.43)	(0.08)	(0.04)	(1.02)	(-0.11)	(0.84)	(0.36)
ln*(n+g+d)*	-1.719	1.970	-10.00	-2.288	9.223	-3.038	-1.783	2.013	-8.607	-3.759	5.976	-1.153
	(-0.96)	(0.66)	(-1.06)	(-0.29)	(0.83)	(-0.50)	(-1.00)	(0.68)	(-1.04)	(-0.60)	(0.84)	(-0.66)
ln*SMER*	0.862	-0.124	-6.212	0.410	-8.564	-4.216	0.785	0.315	-10.02	2.451	-4.018	-2.676
	(0.70)	(-0.06)	(-0.71)	(0.14)	(-1.07)	(-1.60)	(0.56)	(0.14)	(-0.88)	(0.54)	(-1.04)	(-0.99)
ln*SMEH*	-5.364*	1.308	12.96	-5.804	-11.32	3.104	-5.023*	1.919	18.86	-5.718	2.283	3.349
	(-1.84)	(0.21)	(0.58)	(-0.63)	(-0.80)	(0.37)	(-1.80)	(0.31)	(0.74)	(-0.91)	(0.35)	(0.73)
*Dummy*1995	0.318	-0.525	1067.0	176.6	77.60	2.517	0.339	-0.656	972.5	161.1	26.81	1.600
	(0.27)	(-0.37)	(0.75)	(1.08)	(0.98)	(0.10)	(0.29)	(-0.45)	(0.75)	(1.45)	(1.23)	(0.19)
*Dummy*2000	0.957	0.110	1051.4	180.8	77.22	2.618	0.989	-0.0341	958.3	162.3	27.05	1.789
	(0.82)	(0.08)	(0.75)	(1.10)	(0.99)	(0.11)	(0.84)	(-0.02)	(0.74)	(1.47)	(1.23)	(0.21)
*Dummy*2005	-0.126	-1.647	1032.1	184.6	76.25	1.399	-0.125	-1.866	939.6	163.2	24.91	0.846
	(-0.09)	(-0.80)	(0.74)	(1.13)	(0.98)	(0.05)	(-0.09)	(-0.88)	(0.73)	(1.48)	(1.15)	(0.10)
Observations	2032	2032	1524	1524	2032	2032	2032	2032	1524	1524	2032	2032
Adjusted R^2	0.978	0.971					0.978	0.971				
Instruments			14	14	19	19			14	14	19	19
AR2			0.480	0.328	0.613	0.938			0.483	0.226	0.695	0.963
Hansen			0.431	0.431	0.0153	0.0153			0.617	0.617	0.0000	0.0000

Note: t statistics in parentheses and * p-value<0.10, ** p-value<0.05, *** p-value<0.01"

Chapter 6

Are Small Establishments More Cyclically Sensitive in Brazil?

6.1 Introduction

The previous chapters analysed the importance of SMEs for the economic growth process. However, the output of economies does not grow smoothly, and this chapter examines the characteristics of employment fluctuations in small businesses during the business cycle. Little is known about the behaviour of the level of employment in SMEs during business cycles, and further studies could provide valuable information about the best policy to be followed in order to dampen employment and economic fluctuations.

The paucity of evidence on the behaviour of SMEs during business cycles is surprising given that the SME sector employs the majority of the labour force, both in developed and in developing countries.[90] To date, the literature on the sensitivity of SMEs to cycles is still scarce. In a rare and influential paper on this subject, Gertler and Gilchrist (1994) provide evidence indicating that small firms are more sensitive to cyclical conditions. They show that small firms contract substantially more than large enterprises (LEs) after tight money events and account for a significantly disproportionate amount of the resultant decline in manufacturing. On the other hand, recent papers by Moscarini and Postel-Vinay (2009, 2010) appear to contradict this view. For instance, using employment series, Moscarini and Postel-Vinay (2009) present new empirical evidence for a set of countries (U.S., Canada, Denmark, UK and Brazil) and extensive evidence at regional and sectoral level for

[90] See Ayyagary et al. (2007) and Beck et al. (2005a).

the U.S., that suggests that large firms are more sensitive than small ones to business cycle conditions.

The limited attention paid to the analysis of the behaviour of SMEs during the cycles is also expressed by the scant evidence for developing countries. Hence, more attention to the documentation of the sensitivity of SMEs to cycles, particularly in terms of employment, is needed in these countries.[91] The evidence drawn from a developing country's data might also be important to construct theoretical models that take into account features of the SME sector that are particularly important for the business cycle in developing countries. In addition, this evidence might also be important to devise social public policies designed to deal with higher levels of unemployment during a downturn.

The aim of this chapter is to address this gap in the literature by providing an analysis of the behaviour of SMEs' employment during business cycles in the context of a developing country using Brazilian data. Firstly, the analysis follows Moscarini and Postel-Vinay (2009, 2010) in analysing the behaviour of the difference in the growth rates between large and small firms during business cycles in Brazil. Secondly, following Moscarini and Postel-Vinay (2010) and Gertler and Gilchrist (1994), this chapter uses Vector Autoregressions (VARs) to analyse the response of small and large businesses to changes in unemployment, monetary policy and credit constraint. This approach sheds more light on the variables that determine the cyclical behaviour of firms of different sizes. Therefore, this chapter intends to document the relative behaviour of the employment series of establishments in different size bins in Brazil and analyse some aspects that may determine this behaviour.

The remainder of the chapter is as follows. Section 6.2 presents the basic framework to check how sensitive small businesses are relatively to larger ones and describes the unique dataset used in this chapter. Section 6.3 discusses the results at the national, regional and sectoral levels. Section 6.4 analyses the response of SMEs

[91] For instance, Gertler and Gilchrist (1994) uses U.S. data and Moscarini and Postel-Vinay also provide detailed evidence for the U.S. only. They also provide additional initial evidence for a set of countries from which Brazil is the only developing country.

and LEs to changes in unemployment, monetary policy and credit constraint. Lastly, the final section concludes.

6.2 Basic Framework and Data

6.2.1 Basic Framework

This section presents the framework used to shed some light on the sensitivity of small and large firms to business cycle fluctuations in Brazil. We draw on the work of Moscarini and Postel-Vinay (2009, 2010), who analyse the correlations between measures of relative performance by size class and business cycle conditions. For instance, Moscarini and Postel-Vinay (2009) suggest the use of the difference in employment growth rates between large and small firms as a measure of relative firm performance, this can be expressed as:

$$GD_t^j = \left[\frac{(L_t^j - L_{t-1}^j)}{L_{t-1}^j}\right] - \left[\left(\frac{(S_t^j - S_{t-1}^j)}{S_{t-1}^j}\right)\right] \qquad (6.1)$$

where L is the employment level in LEs , S the employment level in SMEs, j is an index for each region or sector, and t denotes time.[92] Consequently, GD is the difference in employment growth rates between LEs and SMEs. As in Moscarini and Postel-Vinay (2009), the aim is to observe how the deviation from the trend of this difference correlates with business cycle conditions in Brazil as a whole, in its states, and across sectors.

The main measure of business cycle conditions used in this chapter is based on the detrended real wholesale revenue index. The main reason for using this indicator is that it is available for each Brazilian state, which allows us to perform a more comparable regional analysis using a local measure of the business cycle.

[92] The index j is used for sector and region, therefore, one GD series is calculated for each sector and each region.

Alternative measures of business cycle conditions will be employed for the case of the country as a whole, since we have a broader range of measures available.

As will be discussed in the next sub-section, we use the band-pass filter developed by Christiano and Fitzgerald (2003), and the Hodrik-Prescott (HP) filter to extract the cyclical component of the series.

6.2.2 Detrending the Series

In order to analyse the relationship between the deviations from the trend of the difference in employment growth rates between large and small firms and the business cycle conditions, the first task is to detrend the series.

Hodrick and Prescott (1981, 1997), for instance, propose a procedure for representing a time series as the sum of a smoothly varying trend component and a cyclical component. Their procedure is based on the prior knowledge that the growth component varies smoothly over time and that a time series y_t is the sum of a trend component τ_t and a cyclical component c_t. For more details on the HP filter see Appendix 6.6.2.

One important issue related with the HP filter is the selection of the smoothing parameter, λ. Hodrick and Prescott suggest λ = 1,600 as a value for the smoothing parameter using quarterly data. While the value of λ = 1,600 seems to be the consensus in the literature, there is less agreement when we move to other frequencies. Despite this dispute, Ravn and Uhlig (2002) consider that it is likely that the HP filter will remain one of the standard methods for detrending series. They study how the Hodrick-Prescott filter should be adjusted when changing the frequency of observations, and suggest that the smoothing parameter should be adjusted accordingly to the fourth power of a change in the frequency of observations. Therefore for $\lambda = 1600$ using quarterly data, this implies that for annual data $\lambda = 1600/4^4 = 6.25$, and for monthly data $\lambda = 1600\ (3^4) = 129600$.

However, the HP filter has some limitations. Baxter and King (1999) argue that the detrending and smoothing techniques to carry out trend-cycle decomposition in the spirit of the HP filter require only that the detrending procedure produces a stationary business-cycle component but does not explicitly specify the statistical characteristics of business cycles. In this sense, Baxter and King (1999) argue that the HP filter ignores the definition of a business cycle. Additionally, they also argue that the HP filter performs badly near the end of the samples. In the same line, Pollock (2000) argues that whereas the HP filter is an excellent device for representing the broad trend of a time series, it often fails in the task of generating a detrended series.

As an alternative, Christiano and Fitzgerald (2003) suggest the use of a band pass filter, which performs better than the HP filter, particularly when annual or monthly data is used. Besides, the band pass filter can explicitly specify the statistical characteristics of business cycles. We follow the usual definition of business cycles as being the cyclical components of no less than six quarters (1.5 years) in duration, and that typically last fewer than 32 quarters (eight years) (e.g. Baxter and King 1999; Christiano and Fitzgerald 2003; Rua and Nunes 2005). Therefore, the band pass filter will pass through components of the time series with periodic fluctuations between six and 32 quarters, while removing components at higher and lower frequencies[93]. Christiano and Fitzgerald (2003) show that the HP filter performance near the end points is relatively poor compared with their band pass filter, and that it also outperforms the HP filter outside the tail areas. Therefore, Christiano and Fitzgerald (2003)'s band pass filter is our preferred device to extract the cyclical component of the series used here. Nevertheless, results for the HP filter are provided in the Appendix 6.6.5 and provide similar qualitative results.

6.2.3 Data

We consider a dataset with the aggregate results of the employment level of individual establishments at different size bins at national, regional, and sectoral

[93] For more details on the Christiano and Fitzgerald (2003)'s band-pass filter see Appendix 7.6.2.2 and Christiano and Fitzgerald's (2003) and Rua and Nunes (2005). In our period of analysis, Brazil went through three recessions. The average duration of the complete business cycle (associated with these recessions) is 43 months (around 11 quarters) according to the Brazilian Economic Cycle Dating Committee (CODACE) of the Fundação Getúlio Vargas (FGV). We tested the stationarity of the series before detrending them (Appendix 7.6.3). This is because when applying the Christiano and Fitzgerald (2003)'s filter, the stochastic or deterministic trends of a data series must be removed.

levels. The employment series are constructed using the CAGED (*Cadastro Geral de Empregados e Desempregados*) database. This is a comprehensive administrative census dataset collected monthly by the Ministry of Labour. It covers the main type of formal employment link, "*celetistas*", covering approximately 32.5 million workers in the formal sector on 31/07/2009. Every month, establishments are required to report any employment variation to the Ministry of Labour, in this sense CAGED is a dataset of job flows.[94] If an establishment fails to report this variation, it faces automatic fines proportional to the length of the delay and the number of declarations omitted.[95] However, because the severance payment is based on CAGED records, employers and workers have a strong incentive to fulfil CAGED records.

This database covers the formal sector of the economy providing information pertaining to employment variation that can be retrieved according to various regional and sectoral levels. The series are constructed by means of a two step procedure used by the Ministry of Labour. At the beginning of each year, the stock of workers across sectors and regions is provided by the Ministry of Labour. Subsequently, we use CAGED database to retrieve the monthly net employment variation across sectors and regions backwards to construct the series.[96] We used this procedure to construct the employment series for SMEs and LEs nationally, regionally, and sectorally.[97] The initial employment stock is based on its values on 01/01/2009, thus, we also use net employment variation forward until 31/07/2009 to construct the series until the most recent data available at the time of the data collection. The software SGT micro, provided under request to the Ministry of

---

[94] CAGED stands for General Register of Employed and Unemployed and it is similar to the American Business Employment Dynamics (BED) dataset in the sense that it is a dataset of job flows.

[95] The employer that submit the RAIS/CAGED declaration late has to pay a fine of R$ 425,64 (US$ 266,00 using the exchange rate of 12/08/2010), plus R$ 106,40 (US$ 67,00) for each additional two months of delay. An additional percentage is added to the fine according to the size of the firm: 0% to 4% - firms between 0 and 25 employees; 5% to 8,0% - firms between 26 and 50 employees; 9% to 12%- firms between 51 and 100 employees; 13% to 16,0% - firms between 101 and 500 employees; 17% to 20,0% - firms between with more than 500 employees. If an employer omits declarations, a fine of also applies in the same terms explained above. An additional fine of R$ 26,60 (US$ 17,00) per employee omitted also applies. For more details see: http://www.rais.gov.br/informar.asp#penalidade

[96] Net employment variation equals the number of jobs created less the number of jobs destroyed in a given month.

[97] To obtain the stock of workers (not the series) in specific size bins a special request should be made to the Ministry of Labour.

Labour, is used to retrieve the net employment variation from CAGED. The monthly employment data is split into 27 federal units and classified according to the 2-digit sectoral classification of the IBGE (*Instituto Brasileiro de Geografia e Estatística*).[98] We are able to construct monthly employment data across states and sectors with information disaggregated by size bin from the first point available in time 1:2000 to 7:2009, which includes the recent recession. Therefore, we have 115 observations for each state and for the 2-digit IBGE's sectoral classification.[99] The series constructed from the information retrieved from CAGED are not seasonally adjusted, therefore, all series were seasonally adjusted before extracting the cyclical components.[100]

It is worth noting that the classification of SMEs varies across and within countries. Ayyagari et al. (2007), for instance, provide a thorough discussion about the difficulties of collecting SMEs data and of finding a common measure of SMEs. They show that the most common criterion used to classify SMEs is based on employment information, using a cut-off to define SMEs that generally varies between 100 and 500 employees, with a large number of sources using 250 employees as a cut-off. For instance, the European Union and Beck et al. (2005a) adopt 250 employees as a cut-off to classify SMEs. Therefore, the cut-off of 250 employees seems to be a reasonable choice based on existing classifications and is also in line with the literature on small businesses. Alternatively, we also follow Moscarini and Postel-Vinay (2009) who suggest cut-offs of less than 50 and more than 500 employees to classify SMEs and LEs, respectively. Thus, we have two series for the difference in employment growth rates between large and small firms. The first is the difference in employment growth rates between employment growth rates in establishments with more than 250 and those with less than 250 employees, and the second is the same difference but in establishments with more than 500 and less than 50 employees.

[98] Brazilian Institute of Geography and Statistics.

[99] According to the Ministry of Labour, for construction, public administration and agriculture sectors there are some problems with data quality stemming from inaccurate responses in small businesses and small municipalities. Please refer to Appendix 7.6.1 for the definition of the 2-digit sectoral classification and denomination of Brazilian states.

[100] All data used in the chapter is correspondingly seasonally adjusted using the standard ARIMA X-12 procedure.

To check the behaviour of the differential employment growth during the economic cycles, various measures regarded as coinciding with business cycles were collected, at monthly and quarterly frequency. For the variables available at monthly frequency, the seasonally adjusted real wholesale revenue index for Brazil and its states are retrieved from IBGE. These series are only available from January 2000 onwards and match the availability of the employment series. The industrial production in the manufacturing index is also obtained from IBGE. Finally, we have also used lagged unemployment as a measure of business cycle conditions as in Moscarini and Postel-Vinay (2009). Unfortunately, due to methodology changes, the official series of unemployment for Brazil before and after the year of 2002 are not compatible. As an alternative, we use the unemployment rate in the most important metropolitan regions of Brazil calculated by Dieese (*Departamento Intersindical de Estatística e Estudos Socioeconômicos*).[101] A limitation of the monthly data is the paucity of information on GDP per capita at this frequency. To provide a robustness check, we collecte the GDP data at quarterly frequency and deflate it by the official domestic price index (IPCA), both are obtained from IBGE. The real GDP series is then seasonally adjusted before extracting its cyclical component.

One important element of any data specification is the choice of the data frequency. The analysis performed here is mainly based on results using monthly data. The use of this frequency offers some advantages over lower frequency data. It offers the computational advantage that more degrees of freedom are available, especially in our context of a limited time span of 10 years. Finally, given that the Brazilian business cycle is characterised by a high degree of volatility (e.g. Kanczuc 2004; Ellery et al. 2002), working with monthly data also has the advantage of reducing the risk of major structural changes in the series when compared with quarterly data.

[101] The unemployment rate of metropolitan areas of Belo Horizonte, Porto Alegre, Recife, Salvador, São Paulo, and Distrito Federal. Dieese stands for Inter-Union Department of Statistics and Socio-Economic Studies.

6.3 Stylised Facts

This section presents the results for the correlations of the detrended difference in employment growth series based on equation (6.1) and detrended measures of business cycles. Using monthly data, this section will document the behaviour of small and large firms during cycles at national, regional and sectoral level. We also comment on the results using quarterly data.

6.3.1 Business Cycles and Business Size in Brazil

In order to provide a preliminary insight into the behaviour of the employment series of establishments of different sizes, Figure 6.1 presents the detrended employment growth rates (normalized) for SMEs and LEs plotted against a measure of business cycles represented by the detrended real wholesale series, from January 2000 to July 2009. The figure also presents the shaded regions to identify the periods of contraction experienced by the Brazilian economy dated by the Brazilian Economic Cycle Dating Committee (CODACE) of the Fundação Getúlio Vargas (FGV). This dating of peaks and troughs is similar to that found by Issler et al. (2009) for the monthly coincident index of the Brazilian economy for the overlapping period of analysis.[102] Both business cycles dates of the Brazilian economy are reported in Table 6.1.

Table 6.1 - Brazilian Business Cycles Dating

CODACE Monthly Data		Issler et al. (2009) Monthly Data	
Peaks	Troughs	Peaks	Troughs
1980:M10	1983:M02	1980:M10	1981:M09
		1982:M07	1983:M02
1987:M02	1988:M10	1987:M02	1988:M10
1989:M06	1991:M12	1989:M06	1990:M04
		1991:M07	1991:M12
1994:M12	1995:M09	1994:M12	1995:M07
1997:M10	1999:M02	1997:M10	1999:M02
2000:M12	***2001:M09***	***2000:M12***	***2001:M09***
2002:M10	***2003:M06***	***2002:M10***	***2003:M06***
2008:M07	***2009:M01***	***NA***	***NA***

Note: In italics and bold are the dates within the time period of our analysis

[102] The monthly coincident index is an average of the four coincident series production, income, sales and employment. The index treats the fluctuations of all four series equally in computing the index.

As we can see in Figure 6.1, it is noteworthy that our main measure for cycles (real wholesale) presents a sharp decline at about the beginning of the shaded area and shows the improvement of business cycle conditions after the recessions. This is an indication that the real wholesale index performs well as our indicator of business cycle conditions.

Turning to the pattern of the employment series, the first thing to note is that, overall, both series seem to display a pro-cyclical pattern. This is in line with the well documented fact that employment series are pro-cyclical and that when an economy is in a downward (upward) phase of the business cycle, changes in employment are negative (positive) (e.g. Liu and Spector, 2005). However, from this picture, it is difficult to observe a clear pattern of the sensitivity of the employment series in SMEs and LEs. The first recession of the decade, which started in 2001, reflected a sum of factors such as the collapse of the Argentinean Peso, electricity shortages in July, and the September 11th episode. In this period, the employment growth rates in LEs seem to have been hit harder.[103] It is also important to bear in mind that the 2001 recession imposed higher constraints on LEs that are more electricity intensive. On the other hand, in the subsequent recession in 2003, SMEs seem to suffer much more than LEs, and their employment growth rates present a steeper relative decline during this recession. Finally, the last recession of the period of analysis shows that both SMEs and LEs were hit hard by the financial crisis.

In order to better visualise the relative behaviour of LEs and SMEs during the business cycle, we calculate the difference in employment growth rates between LEs and SMEs as in equation (6.1) and detrend these series. To contrast the pattern of these series with the behaviour of the business cycle conditions we plot it against the detrended series of the real wholesale index. If the difference in employment growth rates is counter-cyclical, then SMEs are more sensitive to business cycles than LEs. Figure 6.2 shows the difference in growth rates using the two alternative cut-offs (< 250 and <50 for SMEs, and >250 and >500 for LEs) against the detrended real

[103] In 2001, Brazil experienced an energy shortage that led to rationing for 9 months, from June 2001 to March 2002 caused by lower levels of investment in the energy sector together with adverse climate conditions that led to the 2001 drought that provided less rain to the Brazilian Dams (see Carvalho, 2006).

wholesale series. Using both cut-off criteria, the difference in growth rates seems to be counter-cyclical, indicating that SMEs shed proportionally more jobs in recessions and gain more in booms. Table 6.2 shows that the correlations between the differential growth series and real wholesale are negative and confirm the visual impression we get from this figure.

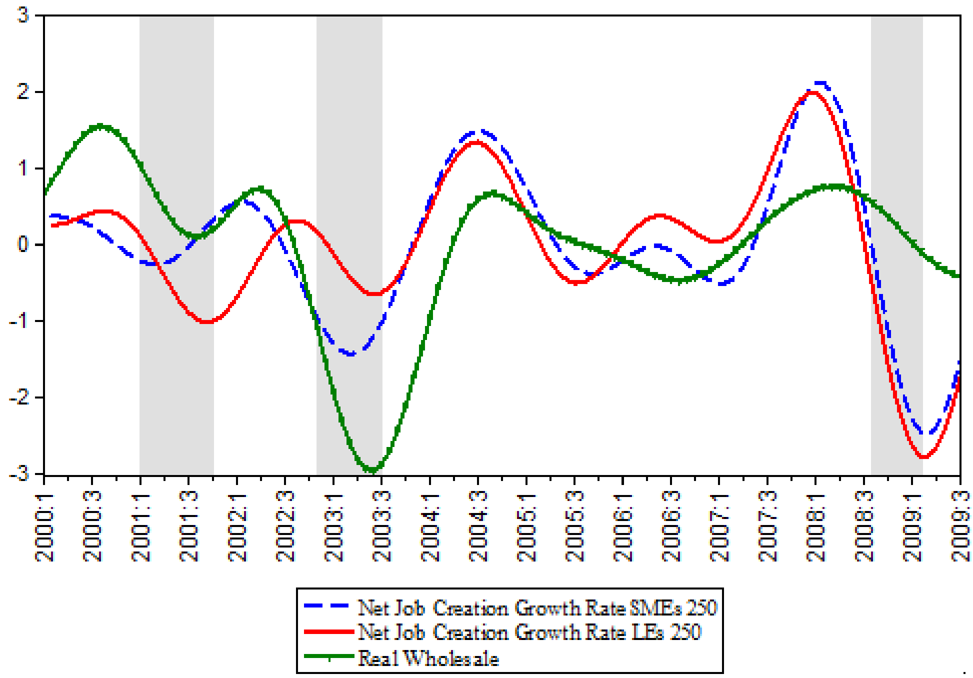


Figure 6.1 – Employment Growth Rates in LEs and SMEs (detrended)

This evidence contradicts the results provided by Moscarini and Postel-Vinay (2009), who state that large firms are more sensitive to cycles than small ones. It also seems to contradict the predictions of Moscarini and Postel-Vinay (2008)'s theoretical model based on heterogeneous firms competing for workers in the labour markets. Given that the level of productivity is positively associated with the size of the firms, their basic idea is that in periods of recessions small firms hire cheaply from unemployed workers. As the reservoir of unemployment dries out, more productive LEs find it profitable to start raising wages to raid workers from less productive SMEs competitors. Wages rise as workers upgrade by quitting to higher paying employers. Workers quit mostly from small, low paying firms to large, high paying firms. As a result, LEs can keep growing their employment through that channel and SMEs have their employment growth restricted in relative terms during expansions. Therefore, SMEs grow in size faster than LEs when the labour market is slack and vice versa when the labour market turns tight.

Besides, Moscarini and Postel-Vinay (2009) also suggest that the result that LEs are more sensitive to cycles is a regularity that is true for the U.S. but also for countries in different stages of development such as Denmark, Brazil, Canada, and United Kingdom. It is noteworthy that Brazil is the only developing country in their analysis and still presents results in line with its developed counterparts. Our stylised facts, however, do not support this evidence and are instead in line with the view of Gertler and Gilchrist (1994) that suggests that SMEs account for a significantly disproportionate share of the manufacturing decline during recessions. They argue that the difference in terms of behaviour across firm size bins is mainly due to financial liquidity constraints on small firms. After tight monetary policy and at the onset of recessions, credit flows to small firms contract relatively more compared to larger firms. The fact that small firms are financially constrained is well documented, and Beck and Demirgüç-Kunt (2006) provide a rich literature review about this fact[104]. In addition, Beck et al. (2005a) and Beck et al. (2005b, 2006) argue that financial constraints on small businesses are worse in developing countries. Hence, the results in Moscarini and Postel-Vinay (2009) that large firms are more sensitive to cycles, even in a developing country context, deserves a more detailed analysis.

[104] It is interesting to note that Moscarini and Postel-Vinay (2009) show that SMEs were hit harder in the U.S in the last recession which originated in the financial sector. This is one exception in their results and indicates the importance of financial constraints for firms' performance.

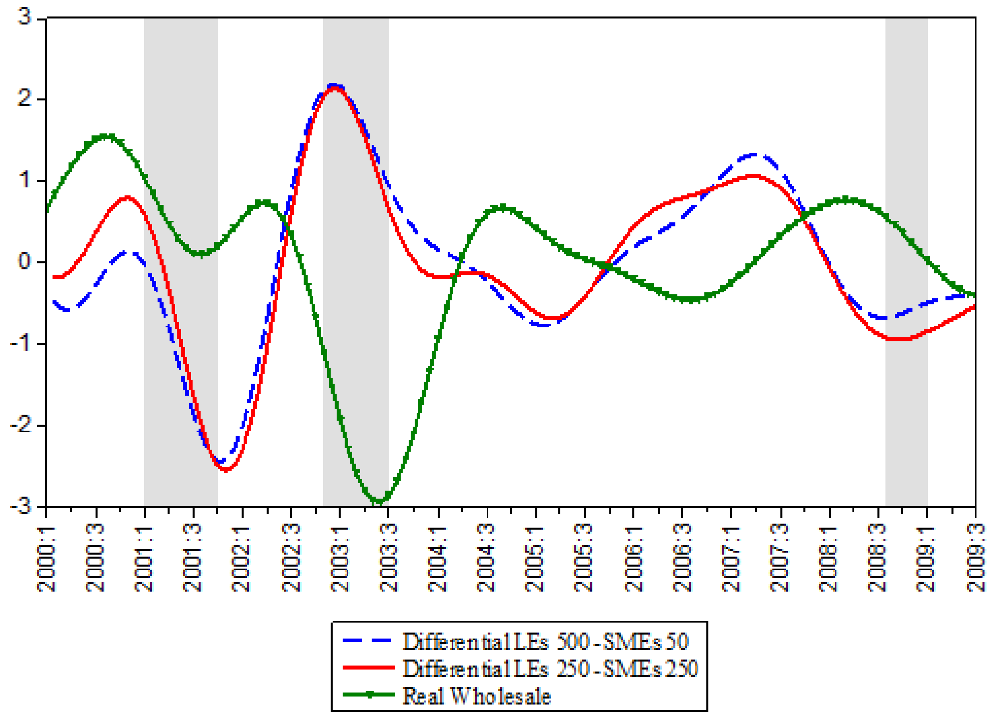


Figure 6.2 – Differential Employment Growth Rates and Cycles (detrended)

The difference between our results and the evidence provided by Moscarini and Postel-Vinay (2009) for Brazil might be related to several issues. First, Moscarini and Postel-Vinay (2009) use annual information from RAIS. Instead, we use CAGED that allowed us to construct employment series at monthly frequency for a different period of analysis that also encompasses the recent financial crises. Hence, the use of our monthly information might represent an improvement in terms of data quality because it allows for more degrees of freedom as opposed to the annual RAIS used by them. Second, the longitudinal annual panel based on RAIS used by Moscarini and Postel-Vinay (2009) is subject to an important type of bias. They allocated individual firms that existed in 1995 to their respective size bins and tracked the growth rate of employment in the two groups over the subsequent 10 years. There are two main problems with this procedure using RAIS that will be discussed below.

The first problem is related to the fact that during the nineties RAIS benefited from important advances regarding data quality. A governmental campaign and improvements in the data collection process expanded RAIS coverage and quality.

One main improvement occurred in 1997 (after the initial point in time of their data for Brazil), when the data started to be collected via the internet, improving its quality and coverage. As noted in Saboia (2000), the RAIS database improved substantially during the 1990s, however, it still presented patchy coverage in many regions and sectors. Second, in the early 2000s, the mortality rate of SMEs was around 60% after four years from the creation of the firms (e.g. SEBRAE, 2007). This can create an important bias towards a bigger presence of LEs as the SMEs that existed in 1995 die and new small businesses created after that date are not considered in Moscarini and Postel-Vinay (2009)'s data.

Therefore, the panel used by Moscarini and Postel-Vinay (2009) to analyse the Brazilian economy lost a significant number of SMEs (as they had a high mortality rate) and might also be biased towards more structured (probably larger) firms that followed all administrative regulations at early times of RAIS. As a result, their data is likely not to be representative of the formal economy as a whole. For instance, in 2006 their data include 16 million workers as opposed to 28 million in the data used in this chapter. In addition, the distribution of employment by size bins supports our claim that our data is more representative of the Brazilian economy. Using the cut-off of 50 employees to classify SMEs, in the data used in this chapter, 46% of the workers are employed by small firms. In contrast, in their data this share falls to 31%, indicating a bias towards larger firms.[105] This may well explain why they found opposite results.

Also, they use the one-year lagged unemployment rate as an indicator of business cycle conditions, while our results are mainly based on the use of the real wholesale index. The use of an unemployment rate series is problematic because unemployment may not be an appropriate measure of business cycle conditions for a country with rigid labour market regulations and where the cost of hiring and firing is high (e.g. World Bank 2002; De Barros and Corseuil 2001).

---

[105] We thank Carlos Corseuil from IPEA that provided us with the same data produced by IPEA for Moscarini and Postel-Vinay (2009). Also, the RAIS data considering the whole universe of establishments with "celetistas" in 2006 also shows a higher share of SMEs, around 48% of the workers are employed by small firms, very similar to our CAGED data. When we consider "celetistas" e "estatutarios", the share of SMEs is also larger than in Moscarini and Postel-Vinay (2009), around 40% of the workers are employed by small firms.

Nevertheless, to check whether our results are sensitive to different business cycle measures we perform the same analysis using data on the industrial production in manufacturing, and also the 1-year lagged unemployment as in Moscarini and Postel-Vinay (2009). Table 6.2 summarises the cross-correlations between all variables capturing business cycle conditions and the differential growth rates[106]. For instance, the differential growth rate (500-50) presents a strong negative correlation with the real wholesale, a negative correlation with manufacturing production, and a weak positive correlation with unemployment, perhaps as a result of the rigidity of the labour market. In addition, the table of correlations suggests that the lagged unemployment is not counter-cyclical to the wholesale and manufacturing production. Also it shows that even if we use the lagged unemployment as our measure of business cycle conditions our results do not lend support to the argument that LEs are more sensitive to cycles in Brazil.

Table 6.2 - Correlations between Cycle Measures and Differential Firm Growth

	Growth Differential (GD) (500-50)	Growth Differential (GD) (250-250)	Real Wholesale	Unemployment (*t*-12)	Manufacturing Production
GD (500- 50)	1.00	0.96	-0.52	0.06	-0.18
GD (250- 250)	0.96	1.00	-0.38	0.20	-0.02
Real Wholesale	-0.52	-0.38	1.00	0.24	0.72
Unemployment (*t*-12)	0.06	0.20	0.24	1.00	0.24
Manufacturing Production	-0.18	-0.02	0.72	0.24	1.00

6.3.2 State Level Analysis

The evidence for the country as a whole summarized in Table 6.2 might cover regional and sectoral specificities in a country marked by its regional asymmetries (e.g. Ferreira 2000; Laurine et al. 2005; Silveira-Neto and Azzoni, 2006). These regional differences might well be related with the evidence reported in Cunha and Moreira (2006) and Martincus and Molinari (2007) that the Brazilian

[106] Figure 7.5 in the Appendix 7.6.5 mimics Figure 7.2 but considers all measures of business cycles considered in the Table 7.2.

regional business cycles are less synchronised than regional cycles in developed countries.

Table 6.3. Correlations between Real Wholesale and Differential Firm Growth

States	Differential (250-250)	States	Differential (500-50)
Sao Paulo	-0.50	Rondonia	-0.65
Goiás	-0.43	Bahia	-0.59
Brazil	**-0.38**	**Brazil**	**-0.52**
Bahia	-0.37	Sao Paulo	-0.50
Rondonia	-0.29	Rio de Janeiro	-0.40
Mato Grosso	-0.29	Goiás	-0.30
Paraíba	-0.28	Ceará	-0.28
Rio Grande do Sul	-0.18	Rio Grande do Sul	-0.25
Ceará	-0.17	Pernambuco	-0.21
Rio de Janeiro	-0.14	Acre	-0.21
Mato Grosso do Sul	-0.13	Mato Grosso	-0.18
Alagoas	-0.09	Paraíba	-0.16
Sergipe	-0.07	Piauí	-0.15
Espirito Santo	-0.05	SantaCatarina	-0.14
Pernambuco	-0.04	Espirito Santo	-0.09
Tocantins	-0.02	Alagoas	-0.07
Rio Grande do Norte	0.00	Tocantins	-0.07
Acre	0.02	Sergipe	-0.06
SantaCatarina	0.02	Paraná	-0.04
Paraná	0.03	Amazonas	0.00
Amapá	0.10	Minas Gerais	0.05
Amazonas	0.11	Mato Grosso do Sul	0.08
Roraima	0.13	Distrito Federal	0.17
Minas Gerais	0.17	Amapá	0.17
Distrito Federal	0.18	Maranhao	0.19
Piauí	0.18	Rio Grande do Norte	0.28
Maranhao	0.24	Roraima	0.39
Pará	0.81	Pará	0.77

This section presents the results for 27 Brazilian states, correlating the differential employment growth rates with the local state wholesale index.[107] Table 6.3 reports the results for the two differential growth rates based on our two different size definitions. The first thing to note is that, overall, the difference in the growth rates becomes more counter-cyclical or more negatively correlated with the business cycle when we use the second criterion (differential growth (500-50)). For instance, for Brazil as a whole, the correlation between the real wholesale index and

[107] Brazil is divided into 27 Federal Units, 26 states and the Federal District of Brasília, for the sake of simplicity we use the word states instead of Federal Units.

differential growth (250-250) is -0.38 and increases to -0.52 when differential growth (500-50) is used instead. Secondly, the counter-cyclical behaviour of the differential growth rate is also observed in most of the Brazilian states. For instance, using the differential growth (500-50), two thirds of the Brazilian states support the argument that SMEs are more cyclically sensitive to business cycles. These states represent around 81% of the formal employment in Brazil in 2009. Nonetheless, there is a substantial variation in the way the difference growth rates correlate with business cycles. For instance, the most important state, São Paulo, presents clearly the same pattern as the country as a whole, indicating that SMEs are more sensitive to cycles than LEs. On the other hand, the state of Pará presents a clear pro-cyclical behaviour of the growth differences, suggesting that LEs are more sensitive to cycles.

Hence, the higher cyclical sensitivity of small businesses is also observed within states. Only few states, representing less than 20% of formal employment in Brazil, present a different dynamic. This different pattern might be related to the industry composition in each state and to other specific characteristics of the states' economies, and the sectoral analysis can shed extra light on the regional results.

6.3.3 Sectoral Level Analysis

In order to analyse the sectoral behaviour of the employment series of establishments in different size bins we use data from eight broad sectors that represent the 2-digit sectoral level according to IBGE's classification. Table 6.4 presents the results for the differential growth rates in each sector. In all sectors, but commerce, the small businesses are more sensitive to business cycles than large one. Clearly, the difference in growth rates in the manufacturing sector presents the most counter-cyclical behaviour, suggesting that the SMEs in this sector are highly sensitive to business cycle fluctuations. On the other hand, the commerce sector presents the opposite result, indicating that LEs are more sensitive to cycles in this sector.

This difference might suggest that the industry composition in each state might be partially responsible for the heterogeneous results found in Table 6.3. If the heterogeneous results for the Brazilian states are generated by sectoral differences, it would be expected that states with a higher share of employment in the commerce sector would tend to present a result where LEs are more sensitive than SMEs. Conversely, a higher share of employment in the manufacturing sector would be associated with a more counter-cyclical behaviour of the differential growth rates. However, the employment distribution of commerce across states does not vary much (representing about 20% to 30% of the states' formal employment) to explain the heterogeneity of the results found in Table 6.3 (see Table 6.7 in the Appendix 6.6.5). Additionally, states with important shares of formal employment in manufacturing, such as AM and PR, still present a counter-cyclical pattern. Hence, differences in the results of the Brazilian states can be possibly related to states' individual characteristics rather than to the sectoral composition alone. These individual characteristics might be related to the regional institutional differences and wealth disparities within Brazil and discussed in Chapters 4 and 5. These regional differences can also be observed, for instance, in human capital, and in different degree of enforcement of the same federal laws (e.g. Almeida and Carneiro, 2010).

Table 6.4. Correlations between Cycles and Differential Firm Growth (Sectors)

Panel A		
Differential (250-250)	**Real Wholesale**	**Manufacturing Production**
Manufacturing (MAN)	-0.77	-0.48
Extraction (EXT)	-0.33	-0.23
Construction (COT)	-0.29	-0.16
Services (SER)	-0.14	0.01
Agriculture – (AGR)	-0.13	-0.03
Public Services - (IPS)	-0.10	-0.05
Public Administration - (PAD)	-0.07	-0.22
Commerce - (COM)	0.30	0.36
Panel B		
Differential (500-50)	**Real Wholesale**	**Manufacturing Production**
Manufacturing (MAN)	-0.73	-0.54
Extraction (EXT)	-0.31	-0.01
Services (SER)	-0.28	-0.12
Construction (COT)	-0.12	-0.03
Public Services - (IPS)	-0.10	-0.13
Agriculture - (AGR)	-0.08	-0.11
Public Administration - (PAD)	-0.05	-0.31
Commerce - (COM)	0.28	0.36

Notes: The sectoral series for small establishments with fewer than 50 employees may present lower quality. Accordingly to the Ministry of Labour, problems related to the omission of declarations and misleading information are more frequent in small municipalities, which usually have smaller firms. These problems are more frequent in construction, agriculture and public administration related activities. See, for instance, the Ministry of Labour technical note MTE 079/2009 available at http://www.mte.gov.br/pdet/ajuda/notas_comunic/nt07909.asp.

It is a noticeable surprise that the differential growth rates of the commerce sector exhibt the most pro-cyclical behaviour, with large firms being more sensitive to cycles than small ones. It is important to bear in mind that generally speaking, the small-scale, informal, often low productivity, frequently family-based enterprises employ between 30% and 70% of the urban work force in Latin America (Maloney 2004). Specifically for the Brazilian case, about 50% of the workers in Brazil resort to informality according to IPEA (Instituto de Pesquisa Econômica Aplicada).[108] Also, Henley et al. (2009) provide evidence showing that the informal sector accounts for a significant share of employment in Brazil. Using data from Pesquisa Nacional por Amostra de Domicílios (PNAD), which is a large scale annual household survey, they suggest three different measures of informality and found that it ranges from 50% to 55% according to the definition used.[109] In this context of significant informality, the commerce sector might represent a significant share of the informal sector and therefore any variation in the level of employment in this sector is not captured by CAGED. For instance, a detailed survey undertaken by IBGE and SEBRAE (Economia Informal Urbana 2003) indicates that 13.86 million workers in urban areas were employed by informal enterprises in 2003, and the commerce sector is the largest informal sector, responsible for 34.9% of urban informal employment.

This substantial share of informal workers, particularly in the commerce sector, might be related with the fact that these workers are not productive enough to match a formal job vacancy and that severance and income tax might be too high relative to the productivity level of the commerce sector. Albrecht et al. (2009), for instance, build a search and matching model with an informal sector that resembles the Latin America labour market. They show that a given level of workers productivity is necessary to match a formal job vacancy and that higher severance and income taxes reduce the rate at which workers find formal sector jobs.[110]

[108] Institute of Applied Economic Research.

[109] It is interesting to note that, if one considers the overlaps between all measures of informality, 63% of economically active workers are classified as informal in Brazil.

[110] Information for Brazil provided by La Porta and Shleifer (2008) and De Paula and Scheinkman

Additionally, businesses in the commerce sector are smaller than those in other sectors. This sector presents a higher share of formal employment in the first size bin of our data, and has 73.10% of its workforce employed by establishments in the size class with less than 50 employees. This might also discourage redundancy because any job shed is a threat to the existence of the establishment given the small numbers of workers.[111]

Therefore, the empirical evidence supports the view that SMEs are more cyclically sensitive than LEs and the sectoral and state level evidence reinforce this view. Moreover, there is a substantial variation in the regional patterns but this heterogeneity does not seem to be determined by the sectoral composition. The marked difference is the pro-cyclical behaviour of the differential growth rate of the commerce sector. However, this is likely to be affected by the fact that an important part of this sector resorts to informal workers (that is not taken into account in our data), possibly because its workers are not productive enough to match a formal sector job vacancy in a environment of rigid labour markets and high levels of severance and income taxes.

A drawback of the analysis when using monthly data is the paucity of GDP data. Lucas (1977), for instance, defines business cycles as movements about trend in GDP. In addition, as argued in Stock and Watson (1998), although the business cycle is technically defined as co-movements across many sectors and series, fluctuations in aggregate output are at the core of a business cycle. Therefore, the cyclical component of real GDP is a useful proxy for the overall business cycle and is a useful benchmark for comparisons. The GDP series for Brazil are available at quarterly frequency and the use of this data provides a good robustness check for the relationship between the differential growth of employment and business cycles. Alternative results using quarterly information detrended using the band pass filter and HP filter are reported in the Appendix 6.6.5 (Table 6.6) and provide similar qualitative results, suggesting that small businesses are more cyclically sensitive than

(2009) support the view that formal firms are more productive than informal ones.

[111] Unfortunately, we are not able to disaggregate the size bins to check for the share of employment in smaller establishments because we were not provided with the stock of employment for smaller size bins. For instance, it will be useful to have information for establishments with less than 9 (or even 4) workers (See Table 7.7 in the Appendix 7.6.5).

large ones in Brazil. Unfortunately, quarterly GDP data is not available at state level and we could not perform the same analysis using the detrended local GDP to proxy for the local business cycle.

6.4 The Response of Small and Large Firms to Unemployment, Monetary Policy and Credit Constraint.

In the previous section, the descriptive analysis documented that small firms are more sensitive to cycles than large ones. Our evidence contradicts the results found in Moscarini and Postel-Vinay (2009, 2010) that says otherwise (for a set of countries that includes Brazil) and is more in line with the view presented in Gertler and Gilchrist (1994). Therefore, we now complement our analysis using Structural Vector Autoregressions (SVAR) to check the impact of monetary policy and credit constraints on the performance of businesses of different sizes.

6.4.1 The SVAR Impulse Response Analysis

We use Structural Vector Autoregressions (SVAR) set up to analyse how shocks in macroeconomic variables affect employment in small and large businesses[112]. Our main objective is to know the response of the differences in employment growth rates between LEs and SMEs to an impulse in another macroeconomic variable.

Our empirical strategy here is to specify a model and then focus on the structural parameters and resulting impulse responses. The basic VAR model of order p can be represented as:

$$y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + u_t \tag{6.2}$$

where y_t is a ($Kx1$) vector of observable time series variables, A_j(j=1,...., p) are (K x K) coefficient matrices and u_t is K-dimensional residual with zero mean. This model

[112] The description of the SVAR impulse response is based on Lutkepohl (2005)

is commonly represented using a lag operator L defined as $Ly_t = y_{t-1}$. Then we can also represent equation (6.2) as $A(L)y_t = u_t$, where $A(L) = I_k - A_1 L - ... A_p L^p$.

Importantly, the VAR is sensitive to the lag order p chosen. As shown in Lutkepohl (2005), choosing p that is unnecessarily large will reduce the forecast precision of the corresponding VAR. Also, the precision of the impulse response will be affected. Therefore, we use the Akaike Information Criteria (AIC) and the Schwartz Criteria (SC) for choosing the adequate VAR order. These tests are described in Appendix 6.6.4.

After estimating the model and choosing the appropriate lag order, the aim is to know the response of one variable to an impulse in another variable in the system. An example provided in Lutkepohl (2005) illustrates this strategy. Suppose the effect of an innovation (or shock) in a system with 3 time series y_1, y_2 and y_3. Then, the variable y_1 increases by one unit in period t = 0, that means, $u_{1,0}$ = 1. It is now possible to follow what happens to the system during the subsequent periods if no further shocks occur. Hence, the vector y (y_1, y, y_3) with t= 0 can be represented as illustrated below to trace the shocks:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix} \tag{6.3}$$

where the first matrix on the right-hand side, A, is a matrix with the coefficients. We trace a unit shock in the first variable in period t = 0 as follows:

$$y_0 = \begin{bmatrix} y_{1,0} \\ y_{2,0} \\ y_{3,0} \end{bmatrix} = \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad y_1 = \begin{bmatrix} y_{1,1} \\ y_{2,1} \\ y_{3,1} \end{bmatrix} = A_1 y_0 = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix},$$

$$y_2 = \begin{bmatrix} y_{1,2} \\ y_{2,2} \\ y_{3,2} \end{bmatrix} = A_1 y_1 = A_1^2 y_0 \text{ , and so on.}$$

Therefore, we can observe the effect of one innovation in one variable on the others in the system. However, the impulse responses may not reflect the relations between variables properly because the components of u_t (equation 6.2) might be correlated, $\Sigma_u = E(u_t u_t')$ may not be diagonal. As a result, we need to orthogonalise the impulses, that is, find a model with instantaneously uncorrelated residuals.

A conventional approach to find a model with uncorrelated residuals is to model the instantaneous relations between the observable variables directly by considering a structural model of the form:

$$\mathrm{A} y_t = A_1^* y_{t-1} + \ldots + A_p^* y_{t-p} + \varepsilon_t \tag{7.4}$$

where $A_j^* = \mathrm{A} A_j \, (j = 1, \ldots.. p)$ and $\varepsilon_t = \mathrm{A} u_t \sim (0, \Sigma_\varepsilon = \mathrm{A} \Sigma_u \mathrm{A}')$. Therefore, an appropriate choice of A will provide a diagonal covariance matrix for ε_t.

Alternatively, it is possible to identify the structural innovations ε_t directly from the forecast errors u_t. In this case we have that $\mathrm{B}\varepsilon_t = u_t$, with $(\mathrm{B}\Sigma_\varepsilon \mathrm{B}' = \Sigma_u)$, and assuming $\varepsilon_t =\sim (0, I_k)$ results that $\Sigma_u = \mathrm{B}\mathrm{B}'$. Combining both restrictions together gives rise to the following expression:

$$\mathrm{A} u = \mathrm{B} \varepsilon_t , \quad \text{where} \quad \varepsilon_t \sim (0, I_k) \tag{7.5}$$

where the assumption of uncorrelated innovations imposes the following restriction on A and B:

$$\mathrm{A} \Sigma_u \mathrm{A}' = \mathrm{B}\mathrm{B}' \tag{7.6}$$

while the two matrices A and B have K^2 elements each, we need $2k^2 - \frac{1}{2}k(k+1)$ additional restrictions to identify all $2K^2$ elements of A and B. As in this example K=3, 12 restrictions on A and B would be necessary to identify this model. In this chapter, to restrict A, we will use a lower triangular matrix with ones on the main diagonal and B will be a diagonal matrix.[113]

We will use this strategy to observe orthogonised impulse responses to check the effect of shocks in monetary policy and credit constraints on the performance of businesses of different sizes. This will complement the initial analysis provided in previous sections and will enable us to investigate how different aspects of the economy impact on employment series in different size bins and determine the sensitivity of this series to business cycles.

6.4.2 The Impulse Response Analysis

The use of the SVAR is particularly useful to contrast the two views about the performance of firms of different sizes during the cycles. As discussed earlier, Gertler and Gilchrist (1994) suggest that small businesses suffer more than large ones during recessions because credit flows contract relatively more for small enterprises. On the other hand, Moscarini and Postel-Vinay (2008, 2009) suggest that SMEs are less cyclically sensitive because they hire proportionally more from unemployment, this is complemented by the argument put forward in Moscarini and Postel-Vinay (2010) that advocates that the interest rate rises during booms tame small firms' growth and this effect is lessened during recessions (due to lower interest rates).

We use the SVAR as in Moscarini and Postel-Vinay (2010) to entertain the hypotheses outlined above. As in their work, we estimate a SVAR with the SELIC rate (SELIC)[114], inflation rate (INF), unemployment (UNP) and the relative measure of the performance of small and large businesses given by the employment growth

---

[113] Therefore, there will be six restrictions on A and six on B. This is similar to the use of a Choleski decomposition.

[114] The SELIC rate is the Brazilian Central Bank reference interest rate. It is similar to the U.S. Federal Reserve's federal funds rate.

differential (GD) as in equation 6.1.[115] It is also worth noting that Gertler and Gilchrist (1994) used a similar approach using VARs to analyse the effect of credit channels on the performance of firms of different sizes, the difference is that they used GNP growth instead of unemployment as a measure of the cyclical conditions of the economy. This system encompasses the main target variables (unemployment and inflation) of a monetary policy conducted using the principles of the Taylor rule (Taylor, 1993), that is in line with conducting a monetary policy with inflation target used in Brazil.[116] Therefore, the SVAR allows us to further examine the effects that monetary policy and cyclical conditions might have on the relative performance of businesses of different sizes through credit channels.

The SVAR is estimated with the variables in their stationary levels and it is a VAR of order two estimated on monthly data over the period 1:2000 to 7:2009.[117] To analyse the impact of shocks in monetary policy and in the cyclical conditions of the economy on the relative performance of small and large businesses we report a set of impulse response functions with their confidence band in Figure 6.3. We observe that the monetary policy virtually does not respond to unemployment but does respond to inflation. This is in line with Moura and Carvalho (2010) who argue that Brazil follows Taylor's principles and is strongly concerned with the achievement of the inflation target via a strong monetary policy response.

More important, however, are the responses of the differential growth rate to monetary policy and cyclical conditions. Contrary to the credit channel suggested by Moscarini and Postel-Vinay (2010), it seems that the relative growth rate of employment tends to decline when the SELIC increases (although this effect is not

[115] The variables are ordered in the VAR as follows: UNP, INF, GD and SELIC. The GD series is given by the difference between employment growth rates in businesses with more than 500 employees and less than 50 employees .The alternative cut-off using 250 employees to define small and large businesses provide similar qualitative results and are reported in the Appendix 6.6.6.2 (Figure 6.10). The funds rate (SELIC) is placed last to capture the idea that monetary policy may adjust to current events but its effects operate on other variables only in the following month. The results presented are not sensitive to the ordering. Data about the official inflation (IPCA) is retrieved from IBGE and SELIC rates were retrieved from IPEADATA (Institute of Applied Economic Research database).

[116] In short, the Taylor rule means that the central bank moves interest rates to achieve the targets.

[117] The SVAR of lag order 2 is selected according to the standard AIC and SC criteria described in the Appendix 7.6.4. The former tends to overestimate the true lag order and suggests a SVAR of order 3 and the latter tends to underestimate the true order and suggests a SVAR of 1 (see Lutkepohl, 2005). According to the PP and ADF unit root tests reported in the Appendix 7.6.3, inflation and the employment differential growth rates are $I(0)$ and the remaining variables are $I(1)$.

significant), indicating that the credit channel proposed by them does not seem to work in the Brazilian case, a higher interest rate does not tame SMEs growth. This evidence might be related with the fact that the interest rate does not constrain SMEs from getting loans when it is at a higher level if they have the credit availability that might help them increase their productivity. Yet, the fact that we observe an insignificant effect of an innovation to SELIC on the differential growth rate, as evidenced by the confidence band, can be related with the argument of Gertler and Gilchrist (1994) that the financial propagation mechanism is likely to be asymmetric over the cycle, more potent in downturns than in booms. Thus, the response of the differential growth rate to a shock to SELIC might be reflecting more the effect such shock has in recessions but is mixed up with a weaker effect of this relationship during booms. Hence, an unambiguous measure of credit constraints is needed to provide direct evidence on the importance of credit constraints to the performance of businesses of different sizes.[118]

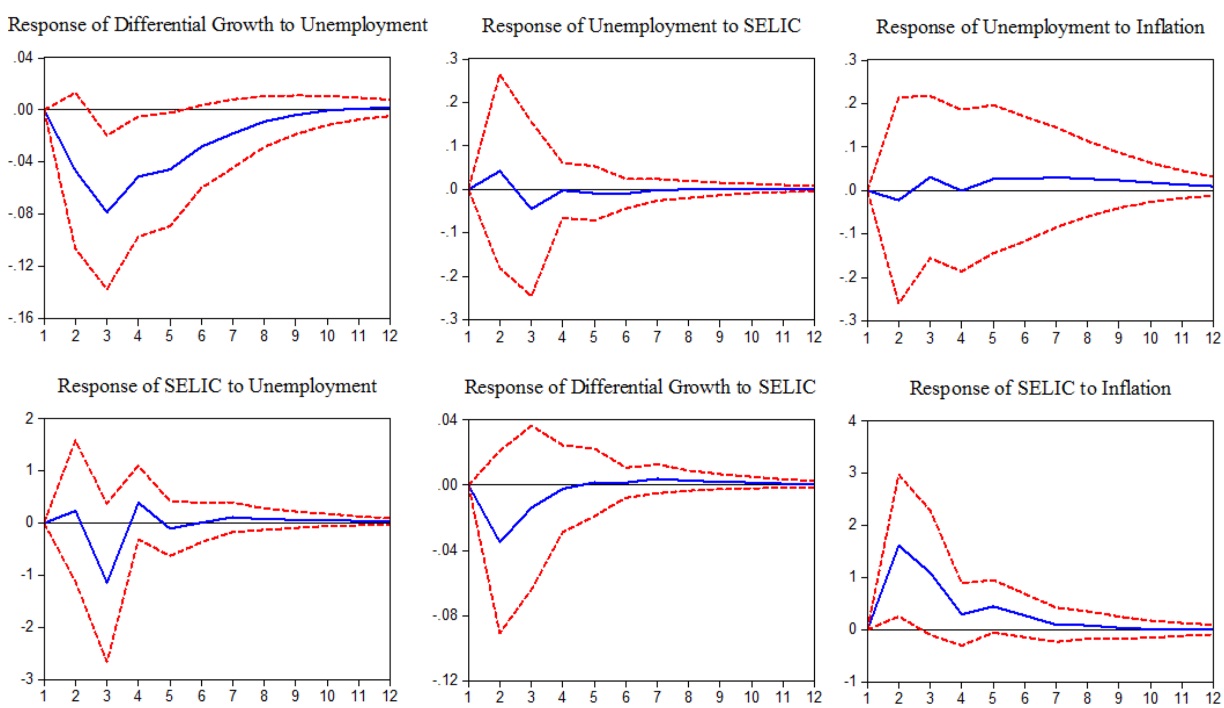


Figure 6.3 –Impulse Response Functions for a SVAR of order 2 with Unemployment, Inflation, Differential Growth and SELIC.

[118] For instance, Gertler and Gilchrist (1994) use the so-called Romer episodes (see Romer and Romer 1990) to proxy for periods of tight money and credit constraint.

On the other hand, the response of the differential growth rate to cyclical conditions provides support to the argument of Moscarini and Postel-Vinay (2008, 2009, 2010) that small businesses hire proportionally more from unemployment. A shock to unemployment leads to a decline in the differential growth rate, meaning that the less productive small firms, that offer low paid jobs, hire proportionally more than more productive large firms. Therefore, the SVAR exercise identified the effect of small firms hiring proportionally more from unemployment as suggested theoretically by Moscarini and Postel-Vinay (2008). However, this is not the same to say that the performance of small firms is less sensitive to cycles because this effect is only one aspect of business cycle conditions and SELIC is not an ideal credit constraint proxy due to its asymmetrical behaviour as argued in Gertler and Gilchrist (1994).

Nevertheless, to further investigate what kind of innovation impacts on the differential growth rates, we now introduce a more explicit credit constraint variable in the SVAR. This variable is provided by the Brazilian Central Bank and is the percentage of credit operations with non-earmarked funds in arrears from 15 to 90 days in financial institutions in Brazil, hereinafter, CC. We expected that this proxy will be able to better capture credit availability, as defaults can be seen as an indication of lack of credit; thus constituting a more direct proxy of credit constraint. It is reasonable to think that the percentage of defaults capture credit constraint regardless of the level of interest rate. It is a better proxy for credit constraint than the interest rate that can have an asymmetrical behaviour over the cycle as argued earlier in this chapter. The new set of impulse responses are reported in Figure 6.4. A similar credit constraint proxy is used, for instance, in Aghion et al. (2008).[119]

[119] Banks in Brazil have a legal obligation to report any default of firms and individuals, thus these numbers refer to the whole of the financial system in Brazil. A similar credit constraint proxy is used, for instance, in Aghion et al. (2008). The length of the sample reduces six months because CC is only available from July 2000 onwards. Again, the SVAR of order 2 is selected according to the AIC and SC criteria. We use unemployment in our baseline estimates but the results are qualitatively the same when we use other measures that capture the cyclical conditions of the economy. The Appendix 7.6.6.1 reports the impulse response functions using manufacturing production and real wholesale instead of unemployment and provides similar qualitative interpretation. Also, the alternative cut-off using 250 employees to define small and large businesses provide similar qualitative results and are

The analysis from the impulse responses is similar to that drawn from Figure 6.3. The important difference is that one innovation to the credit constraint variable (CC) increases the differential growth rate between small and large firms. More credit constraint hits small firms harder. This is consistent with the stylised facts presented in the descriptive sections of this chapter and supports the view that SMEs are more credit constrained. This can be a driving force in the performance of small businesses during the cycles, mainly during a downturn. Following the inclusion of CC, the differential growth rate still responds negatively to an innovation in unemployment, a sign that there is an effect of small businesses hiring proportionally more in recessions than in booms. However, SMEs might shed more jobs in downturns due to other factors at work, one of them being the credit constraint as argued in Gertler and Gilchrist (1994). Therefore, Moscarini and Postel-Vinay's ideas might represent an important advance in the way we think about the cyclically sensitivity of small and large firms. However, for the Brazilian case, the effect of hiring cheap from unemployment does not guarantee that large firms are more cyclically sensitive. Other reasons, such as the credit constraint effect, may well affect small businesses to the extent that they suffer more during recessions when the credit flows contract relatively more for them.

---

reported in the Appendix 7.6.6.2 (Figure 7.11).

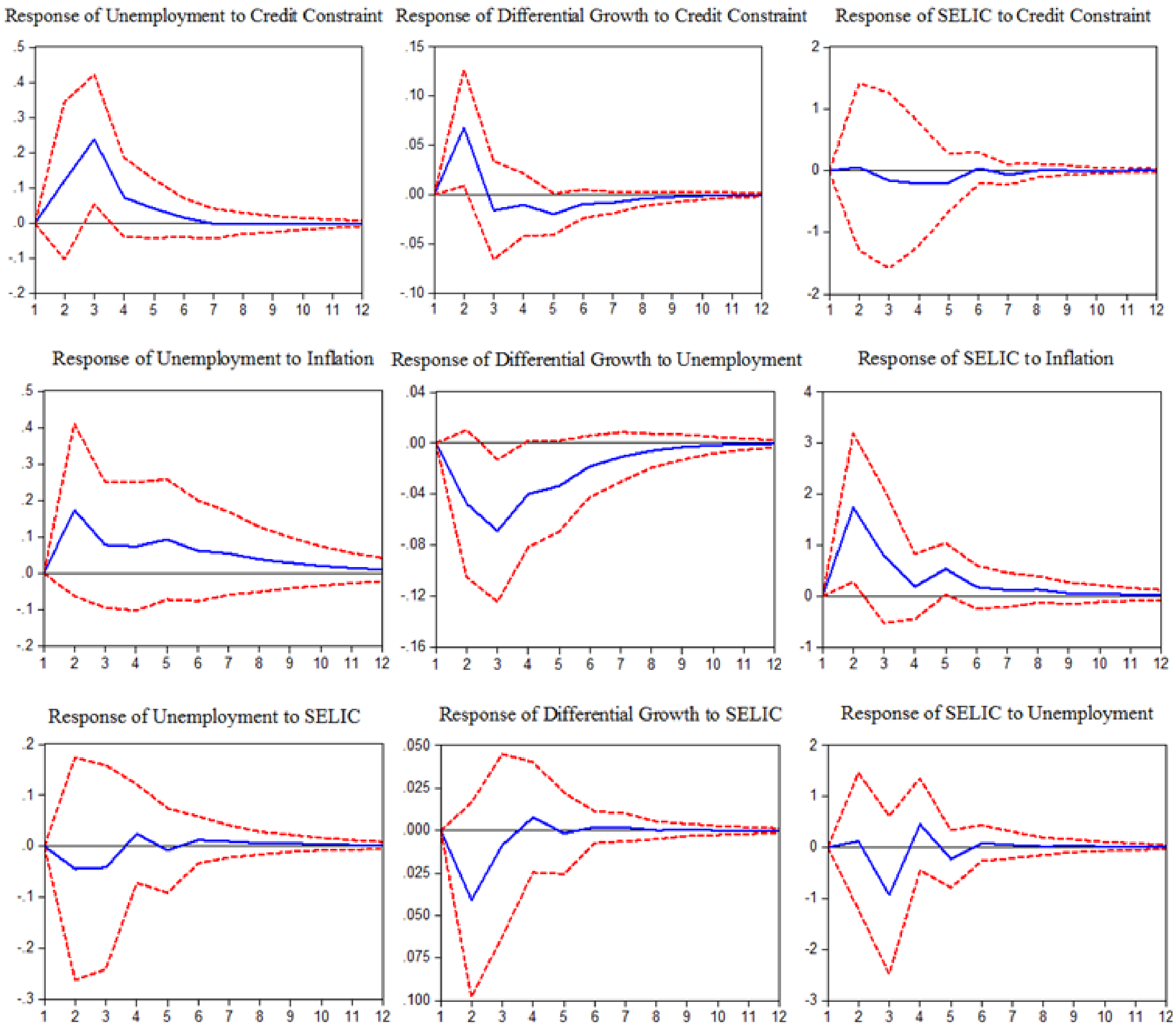


Figure 6.4 – Impulse Response Functions for a SVAR of order 2 with Unemployment, Inflation, Credit Constraint, Differential Growth and SELIC.

6.5 Conclusion

This chapter has presented some stylised facts for a developing country showing that smaller employers are more cyclically sensitive than larger ones. Using Brazilian data, we showed that small businesses shed proportionally more jobs in recessions and gain more in booms. This pattern is robust to various business cycle measures and different data frequencies, and the pattern observed for the country as a whole is also observed at regional and sectoral level.

The impulse response analysis provided interesting insights regarding the labour market dynamics. It points to the existence of the effect of small firms hiring cheaply from unemployment proportionally more than large ones during recessions. On the other hand, innovations to credit constraint hit small firms harder and might help to explain the empirical regularity that small firms are more cyclically sensitive than large ones for the Brazilian case.

Therefore, from the public policy perspective, our results provide evidence and support the view that SMEs are more sensitive to cycles in the context of a developing country. Consequently, policies designed to dampen employment shocks during business cycles and protect employment during recessions should aim at easing the financial constraints to small businesses in Brazil. Temporary tax reduction on the cost of employment for small firms during downturns could prevent some layoffs. Additionally, the design of a mechanism that provides SME lending during recessions might be a promising solution to dampen employment fluctuations in downturns.

6.6 Appendix

6.6.1 Industries and States

Industries

Extraction (Mining and Quarrying) (EXT); Manufacturing (MAN); Industries Public Services – Electricity, Gas, Water Supply and others (IPS); Construction (COT); Commerce (COM); Services (SER); Public Administration (PAD); Agriculture (AGR). [120]

Federal Units (States)

Distrito Federal (DF); Mato Grosso (MG); Mato Grosso do Sul (MS); Goiás (GO); Rio Grande do Sul (RS); Santa Catarina (SC); Paraná (PR); Rio de Janeiro (RJ); São Paulo (SP); Minas Gerais (MG); Espirito Santo (ES); Sergipe (SE); Bahia (BA); Pernambuco (PE); Rio Grande do Norte (RN); Ceará (CE); Paraiba (PB); Alagoas (AL); Piaui (PI); Maranhão (MA); Amazonas (AM); Amapá (AP); Rondônia (RO); Acre (AC); Roraima (RR); Pará (PA); Tocantins (TO)

6.6.2 HP filter and Christiano Fitzgerald Band Pass Filter

6.6.2.1 The HP Filter

The HP Filter is based on the prior knowledge that the growth component varies smoothly over time, and that a time series y_t is the sum of a trend component τ_t and a cyclical component c_t:

$$y_t = \tau_t + c_t \qquad \text{for } t = 1, \ldots\ldots\ldots, T.$$

[120] Importantly, CAGED includes only the main formal type of contract, the so-called “celetistas”. Therefore, CAGED only includes in the public administration sector (PAD) the employees of state owned enterprises or private enterprises working in the public administration sector. The public servants are not included in CAGED because they have a different type of formal contract and are known as “estatutários”.

In this conceptual framework the average of c_t over long time periods is assumed to be near zero. Therefore, the problem for determining the growth components is given as follows:

$$\min_{\tau_t} \sum_{t=1}^{T} ((y_t - \tau_t)^2 + \lambda((\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}))^2)$$

where the parameter λ is a positive number which penalizes variability in the growth component series. The larger the value of λ, the smoother is the solution for the cyclical component series. As argued in Section 3, we use λ = 1,600 as a value for the smoothing parameter using quarterly data, and $\lambda = 1600\ (3^4) = 129600$ for monthly data.[121]

6.6.2.2 The Christiano-Fitzgerald Band-Pass Filter

If we want to isolate fluctuations of a given periodicity in a series y_t, the filtered series should be given by:

$$y_t^* = B(L) y_t$$

where the ideal band-pass filter $B(L)$ has the following structure:

$$B(L) = \sum_{j=-\infty}^{\infty} B_j L^j \text{ , with } L^l y_t \equiv y_{t-l}$$

and the B_j's are given by:

$$B_j = \frac{\sin(jb) - \sin(ja)}{\pi j}, \quad j \geq 1$$

$$B_0 = \frac{b-a}{\pi}, \text{ where } a = \frac{2\pi}{p_u},\ b = \frac{2\pi}{p_l}$$

[121] We detrended our series through the HP filter using EViews.

However, to compute y_t^* using *B(L)* requires an infinite number of observations on y_t. Hence, Christiano and Fitzgerald (2003) suggest the estimation of y_t^* using $\hat{y}_t^*$, a linear function of the available data and therefore:

$$\hat{y}_t = \hat{B}^{p,f}(L) y_t \text{, with } f = T - t \text{ and } p = t \text{ -1}$$

where the ideal band-pass filter $\hat{B}^{p,f}(L)$ using a linear approximation to generate an infinite series has the following structure:

$$\hat{B}_j^{p,f}(L) = \sum_{j=-f}^{p} \hat{B}_j^{p,f} L^j \text{, where } L^h y_t \equiv y_{t-h}$$

and selecting the filter weights $\hat{B}_j^{p,f}$ solving:

$$\min_{\hat{B}_j^{p,f}, j=-f,\ldots,p} \int_{-\pi}^{\pi} \left| B(e^{-i\omega}) - \hat{B}^{p,f}(e^{-i\omega}) \right|^2 f_y(\omega) \mathrm{d}\omega$$

where the $f_y(\omega)$ is the spectral density of y_t at frequency ω, which measures the contribution of each frequency component to the overall variance of y_t.[122]

[122] We tested extensively for the stationarity of the series used in our analysis before detrending them. The series of the difference in employment growth rates between SMEs and LEs, and unemployment were treated as (*I*(0)). Alternatively, the remaining series were treated as *I*(1) with drift adjustment. This is because when applying Christiano and Fitzgerald filter, the stochastic or deterministic trends of a data series must be removed. Christiano and Fitzgerald (2003)' filter was applied to our data using EViews.

6.6.3 Unit Root Results

To test the stationarity of the series we use the augmented Dicker-Fuller (ADF) and the Phillips-Perron (PP) tests. The augmented Dicker-Fuller (ADF) takes the following form:

$$\Delta y_t = \alpha_1 + \alpha_2 t + \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t$$

This test for a unit root under the null hypothesis that the series contain a unit root (H: γ = 0) against the alternative hypothesis that the series is stationary. The test is given by the normal *t*-statistic on the coefficient of the lagged dependent variable.

As an alternative, the Phillips-Perron (PP) test is also provided to test the stationary of the series. This test is given by:

$$\Delta y_{t-1} = \alpha_1 + \alpha_2 t + \gamma y_{t-1} + u_t$$

where PP test makes a correction to the *t*-statistic of the coefficient γ to account for the residual autocorrelation. As in the ADF test, the null hypothesis that the series contain a unit root (H: γ = 0) is tested against the alternative hypothesis that the series is stationary. The results of these tests for all time series used in this chapter are reported below.

Table 6.5 Unit Root Test Results

	Test	**Differential (500-50)**	**Differential (250-250)**	**Real Wholesale**	**Unemploy-ment**	**Manufactu-ring**	**SELIC**	**Credit Constraint**	**Inflation**
Brazil	**ADF Levels**	**-5.87**	**-8.91**	**-1.20**	**-1.84**	**-2.20**	**-2.51**	**-2.96**	**-4.35**
Brazil	**ADF 1st Diff**			**-10.44**	**-4.22**	**-10.03**	**-5.16**	**-9.39**	
Brazil	**PP Levels**	**-9.30**	**-9.08**	**-1.18**	**-1.38**	**-2.20**	**-2.65**	**-4.38**	**-4.41**
Brazil	**PP 1st Diff**			**-10.45**	**-7.19**	**-10.01**	**-15.06**	**-12.83**	
State/ Sector	**Test**	**Differential (500-50)**	**Differential (250-250)**	**Real Wholesale**	**State/ Sector**	**Test**	**Differential (500-50)**	**Differential (250-250)**	**Real Wholesale**
AC	ADF Levels	-0.81	-5.59	-2.04	RJ	ADF Levels	-9.77	-10.84	-1.03
AC	ADF 1st Diff			-14.29	RJ	ADF 1st Diff			-10.63

AC	PP Levels	-15.60	-11.20	-2.16	RJ	PP Levels	-9.74	-10.84	-1.47
AC	PP 1st Diff			-14.32	RJ	PP 1st Diff			-16.29
AL	ADF Levels	-7.29	-7.38	-1.45	RN	ADF Levels	-12.39	-4.61	-2.34
AL	ADF 1st Diff			-13.18	RN	ADF 1st Diff			-2.10
AL	PP Levels	-14.85	-18.43	-1.47	RN	PP Levels	-12.39	-10.84	-2.01
AL	PP 1st Diff			-13.09	RN	PP 1st Diff			-11.74
AP	ADF Levels	-11.34	-3.62	-2.19	RS	ADF Levels	-3.34	-3.69	-1.31
AP	ADF 1st Diff			-16.57	RS	ADF 1st Diff			-3.39
AP	PP Levels	-11.32	-10.05	-2.83	RS	PP Levels	-8.98	-9.09	-2.64
AP	PP 1st Diff			-16.74	RS	PP 1st Diff			-17.08
AM	ADF Levels	-3.89	-5.44	-1.65	RO	ADF Levels	-4.43	-3.08	-3.70
AM	ADF 1st Diff			-13.03	RO	ADF 1st Diff			-8.00
AM	PP Levels	-7.40	-5.54	-1.64	RO	PP Levels	-8.04	-8.80	-3.64
AM	PP 1st Diff			-13.03	RO	PP 1st Diff			-12.77
BA	ADF Levels	-10.62	-9.83	-1.12	RR	ADF Levels	-5.48	-9.18	-1.93
BA	ADF 1st Diff			-7.03	RR	ADF 1st Diff			-2.87
BA	PP Levels	-10.62	-9.86	-1.24	RR	PP Levels	-9.39	-9.19	-2.11
BA	PP 1st Diff			-12.82	RR	PP 1st Diff			-15.43
CE	ADF Levels	-9.08	-8.18	-1.39	SC	ADF Levels	-4.25	-4.55	-2.23
CE	ADF 1st Diff			-14.20	SC	ADF 1st Diff			-8.15
CE	PP Levels	-7.11	-7.19	-1.61	SC	PP Levels	-10.89	-11.20	-2.23
CE	PP 1st Diff			-13.84	SC	PP 1st Diff			-11.34
DF	ADF Levels	-10.60	-10.43	-1.94	SP	ADF Levels	-4.29	-3.87	-0.98
DF	ADF 1st Diff			-15.73	SP	ADF 1st Diff			-12.40
DF	PP Levels	-10.62	-10.44	-2.05	SP	PP Levels	-9.46	-8.73	-0.87
DF	PP 1st Diff			-16.25	SP	PP 1st Diff			-12.61
ES	ADF Levels	-5.98	-6.09	-2.45	SE	ADF Levels	-6.66	-7.39	-1.87
ES	ADF 1st Diff			-3.51	SE	ADF 1st Diff			-15.94
ES	PP Levels	-8.85	-9.80	-1.84	SE	PP Levels	-7.62	-8.41	-2.23
ES	PP 1st Diff			-11.50	SE	PP 1st Diff			-15.95
GO	ADF Levels	-10.99	-10.65	-2.10	TO	ADF Levels	-5.21	-5.26	-1.97
GO	ADF 1st Diff			-11.24	TO	ADF 1st Diff			-3.92
GO	PP Levels	-10.99	-10.65	-1.87	TO	PP Levels	-7.43	-8.71	-3.07
GO	PP 1st Diff			-11.87	TO	PP 1st Diff			-18.02
MA	ADF Levels	-9.08	-8.23	-2.63	AGR	ADF Levels	-8.67	-9.02	
MA	ADF 1st Diff			-11.81	AGR	ADF 1st Diff			
MA	PP Levels	-9.20	-8.35	-2.64	AGR	PP Levels	-8.57	-9.01	
MA	PP 1st Diff			-11.90	AGR	PP 1st Diff			
MT	ADF Levels	-3.66	-5.31	-2.76	COM	ADF Levels	-8.55	-9.63	
MT	ADF 1st Diff			-14.01	COM	ADF 1st Diff			
MT	PP Levels	-7.66	-7.76	-2.51	COM	PP Levels	-9.95	-9.58	
MT	PP 1st Diff			-13.74	COM	PP 1st Diff			
MS	ADF Levels	-7.45	-9.67	-2.34	COT	ADF Levels	-3.10	-5.25	
MS	ADF 1st Diff			-13.16	COT	ADF 1st Diff			
MS	PP Levels	-7.43	-9.69	-2.33	COT	PP Levels	-25075.97	-8.18	
MS	PP 1st Diff			-13.52	COT	PP 1st Diff			
MG	ADF Levels	-4.63	-9.20	-2.10	EXT	ADF Levels	-5.94	-8.10	
MG	ADF 1st Diff			-15.11	EXT	ADF 1st Diff			

MG	PP Levels	-9.67	-9.29	-2.14	EXT	PP Levels	-8.79	-8.06
MG	PP 1^{st} Diff			-15.48	EXT	PP 1^{st} Diff		
PA	ADF Levels	-8.31	-8.38	-2.60	MAN	ADF Levels	-3.80	-4.81
PA	ADF 1^{st} Diff			-4.61	MAN	ADF 1^{st} Diff		
PA	PP Levels	-8.61	-8.73	-3.03	MAN	PP Levels	-10.57	-10.58
PA	PP 1^{st} Diff			-14.87	MAN	PP 1^{st} Diff		
PB	ADF Levels	-9.19	-9.49	-1.77	PAD	ADF Levels	-4.92	-6.34
PB	ADF 1^{st} Diff			-12.95	PAD	ADF 1^{st} Diff		
PB	PP Levels	-13.42	-16.98	-2.23	PAD	PP Levels	-8.82	-11.21
PB	PP 1^{st} Diff			-12.94	PAD	PP 1^{st} Diff		
PR	ADF Levels	-8.26	-3.52	-2.20	IPS	ADF Levels	-10.82	-10.66
PR	ADF 1^{st} Diff			-7.85	IPS	ADF 1^{st} Diff		
PR	PP Levels	-8.33	-9.61	-2.19	IPS	PP Levels	-10.96	-10.73
PR	PP 1^{st} Diff			-14.15	IPS	PP 1^{st} Diff		
PE	ADF Levels	-11.94	-11.75	-1.64	SER	ADF Levels	-3.27	-3.46
PE	ADF 1^{st} Diff			-14.58	SER	ADF 1^{st} Diff		
PE	PP Levels	-14.67	-13.59	-1.83	SER	PP Levels	-8.76	-8.95
PE	PP 1^{st} Diff			-14.58	SER	PP 1^{st} Diff		
PI	ADF Levels	-9.88	-11.23	-3.65				
PI	ADF 1^{st} Diff			-10.07				
PI	PP Levels	-9.85	-11.69	-3.45				
PI	PP 1^{st} Diff			-14.11				

Notes: The values reported in this table are the test values for the ADF and PP unit root tests. Series that show only the results in levels are stationary and series that also show the results for the variable in their first difference are stationary in their first difference $I(1)$. The ADF unit root test for the credit constraint indicates that this series has a unit root, however, this result is not clear from using the PP unit root test. For that reason, the SVAR analysis considering credit constraint as $I(0)$ is provided in the Appendix 6.6.6.3. Results are qualitatively similar to the ones reported in section 6.4.2.

6.6.4 Choosing the VAR order

The Akaike Information Criteria (AIC) for choosing a VAR order is given by:

$$AIC(m) = \ln\left|\tilde{\Sigma}_u(m)\right| + \frac{2mK^2}{T}$$

where m denotes the order of the VAR process fitted to the data, T is the sample size, K the dimension of the time series, and $\tilde{\Sigma}_u(m)$ is the maximum likelihood estimator of the residual covariance matrix obtained by fitting the VAR*(m*) model. The estimate $\hat{p}(\text{AIC})$ for p is chosen so this criterion is minimized. As shown in Lutkepohl (2005), in moderate sample sizes this test tends to choose the correct VAR order more often than the SC criterion reported below.

Alternatively, the Schwarz Criteria (SC) is given by:

$$SC(m) = \ln\left|\tilde{\Sigma}_u(m)\right| + \frac{\ln T}{T} mK^2$$

where the notation is the same as described for the AIC, and the estimate for p is chosen when this criterion is minimized. Contrary to the AIC, this test tends to underestimate the true order of a VAR.

6.6.5 Additional Results (Tables and Figures)

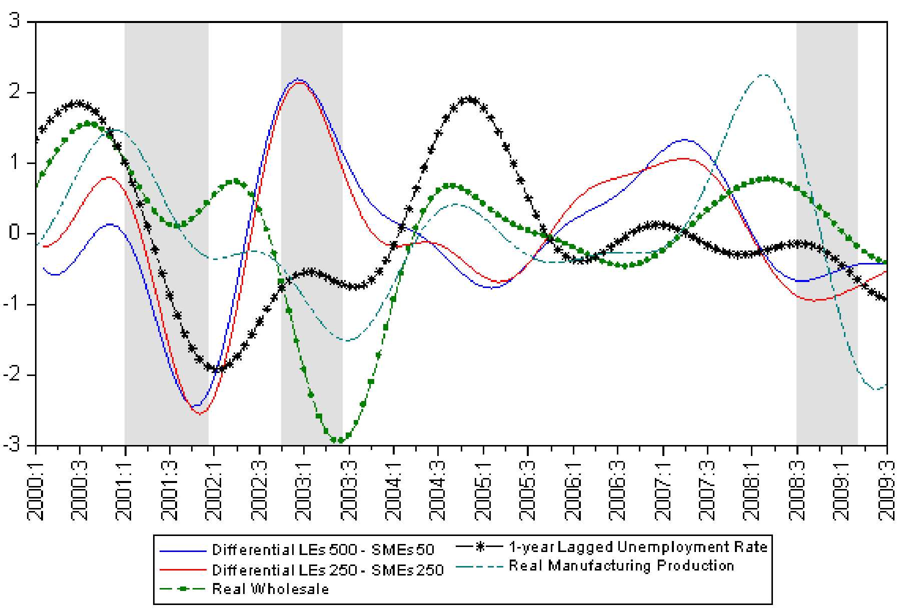


Figure 6.5 – Differential Employment Growth Rates and Cycles (detrended CF Filter)

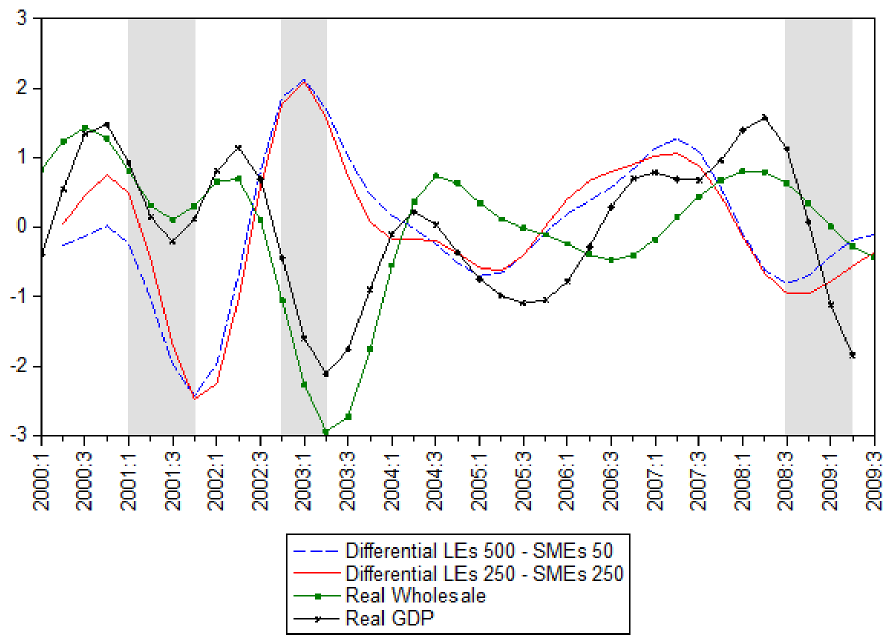


Figure 6.6 – Differential Employment Growth Rates and Cycles at Quarterly Frequency (detrended CF Filter)

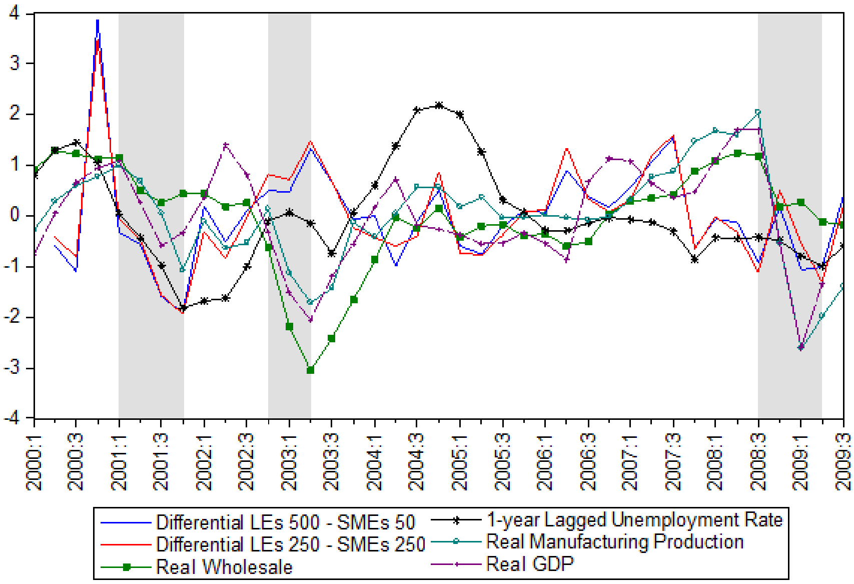


Figure 6.7 –Differential Employment Growth Rates and Cycles at Quarterly Frequency (detrended HP Filter)

Table 6.6. Correlations between Cycle Measures and Differential Firm Growth at Quarterly Frequency – HP Filter and CF Band pass Filter

	Growth Differential (500-50)		Growth Differential (250-250)		Real Wholesale		Unemployment (*t*-12)		Manufacturing Production		Real GDP	
	CF	HP	CF	HP	CF	HP	CF	HP	CF	HP	CF	HP
GD (500-50)	1.00	1.00	0.96	0.97	-0.57	-0.20	0.13	0.16	-0.17	0.02	-0.22	0.06
GD (250-250)	0.96	0.97	1.00	1.00	-0.46	-0.22	0.23	0.20	-0.03	-0.01	-0.14	-0.03
Real Wholesale	-0.57	-0.20	-0.46	-0.22	1.00	1.00	0.19	0.00	0.78	0.58	0.74	0.65
Unemployment (t-12)	0.13	0.16	0.23	0.20	0.19	0.00	1.00	1.00	0.22	0.27	-0.07	0.01
Manufacturing Production	-0.17	0.02	-0.03	-0.01	0.78	0.58	0.22	0.27	1.00	1.00	0.79	0.77
Real GDP	-0.22	0.06	-0.14	-0.03	0.74	0.65	-0.07	0.01	0.79	0.77	1.00	1.00

Note: CF stands for Christiano and Fitzgerald (2003)'s band pass filter and HP for the Hodrick-Prescott filter.

Table 6.7 - Total Employment Shares across Brazilian States and Sectors (2009)

	EXT	MAN	IPS	COT	COM	SER	PAD	AGR	TOTAL
Employment Shares Across Different across States and Sectors									
RO	0.44	18.50	1.53	6.94	34.27	29.30	3.28	5.74	100.00
AC	0.30	9.73	1.59	14.10	30.10	29.99	10.10	4.09	100.00
AM	0.23	31.55	1.58	7.68	19.28	37.75	0.99	0.94	100.00
RR	0.13	6.95	2.69	17.11	31.64	36.90	0.63	3.95	100.00
PA	1.79	16.51	1.38	9.16	27.08	33.56	2.93	7.59	100.00
AM	2.95	5.56	2.13	8.90	32.14	45.22	0.51	2.60	100.00
TO	0.77	11.87	2.65	12.29	30.38	27.11	0.94	13.99	100.00
MA	0.22	10.65	1.81	12.17	28.47	36.32	3.78	6.58	100.00
PI	0.33	12.38	2.10	10.11	28.82	36.41	6.38	3.46	100.00
CE	0.29	26.01	0.79	6.34	19.44	39.74	4.33	3.06	100.00
RN	2.12	20.13	1.52	8.63	24.54	36.35	1.73	4.98	100.00
PB	0.42	24.04	2.67	7.43	23.32	33.12	3.93	5.06	100.00
PE	0.21	20.86	1.58	7.07	22.45	40.60	1.83	5.40	100.00
AL	0.28	36.71	1.52	4.70	20.23	30.12	2.62	3.82	100.00
SE	1.36	15.86	1.36	8.91	21.10	41.78	3.80	5.82	100.00
BA	0.87	14.52	1.17	7.59	23.83	41.63	3.49	6.89	100.00
MG	1.41	22.16	0.91	7.88	21.51	36.22	1.75	8.15	100.00
ES	1.91	17.80	1.19	8.04	24.73	39.75	1.39	5.21	100.00
RJ	0.65	12.78	1.59	6.15	21.88	51.63	4.47	0.86	100.00
SP	0.15	25.33	0.85	4.88	20.63	41.70	2.50	3.96	100.00
PR	0.25	28.18	1.11	5.06	23.40	34.94	1.58	5.49	100.00
SC	0.50	37.60	1.09	4.32	21.01	31.12	1.31	3.04	100.00
RS	0.40	31.66	1.05	3.94	22.18	35.40	1.42	3.95	100.00
MS	0.53	18.81	0.74	5.67	23.97	32.67	1.05	16.55	100.00
MT	0.49	18.87	0.84	6.46	27.67	26.59	1.91	17.17	100.00
GO	0.78	22.24	1.07	6.96	24.33	34.42	1.26	8.94	100.00
DF	0.06	5.59	1.26	7.37	22.46	61.53	0.76	0.97	100.00
Total	0.54	23.11	1.11	6.04	22.12	39.79	2.43	4.87	100.00
Sdt	**0.72**	**8.84**	**0.56**	**3.05**	**4.21**	**7.43**	**2.08**	**4.18**	**0.00**
Employment Shares Across Different Size Bins									
Size Class	**EXT**	**MAN**	**IPS**	**COT**	**COM**	**SER**	**PAD**	**AGR**	**TOTAL**
0-50	33.69	31.60	14.21	31.85	73.10	42.47	5.26	60.64	45.71
50-250	20.12	25.75	19.85	30.02	19.57	20.90	15.40	18.41	22.01
250-500	7.87	11.97	11.05	12.69	4.77	9.22	13.21	5.96	9.03
+500	38.32	30.68	54.89	25.44	2.56	27.41	66.13	14.99	23.25

6.6.6 Robustness Checks

6.6.6.1 Alternative Results Using the Differential Growth using 500 and 50 employees as Cut-offs to Classify LEs and SMEs.

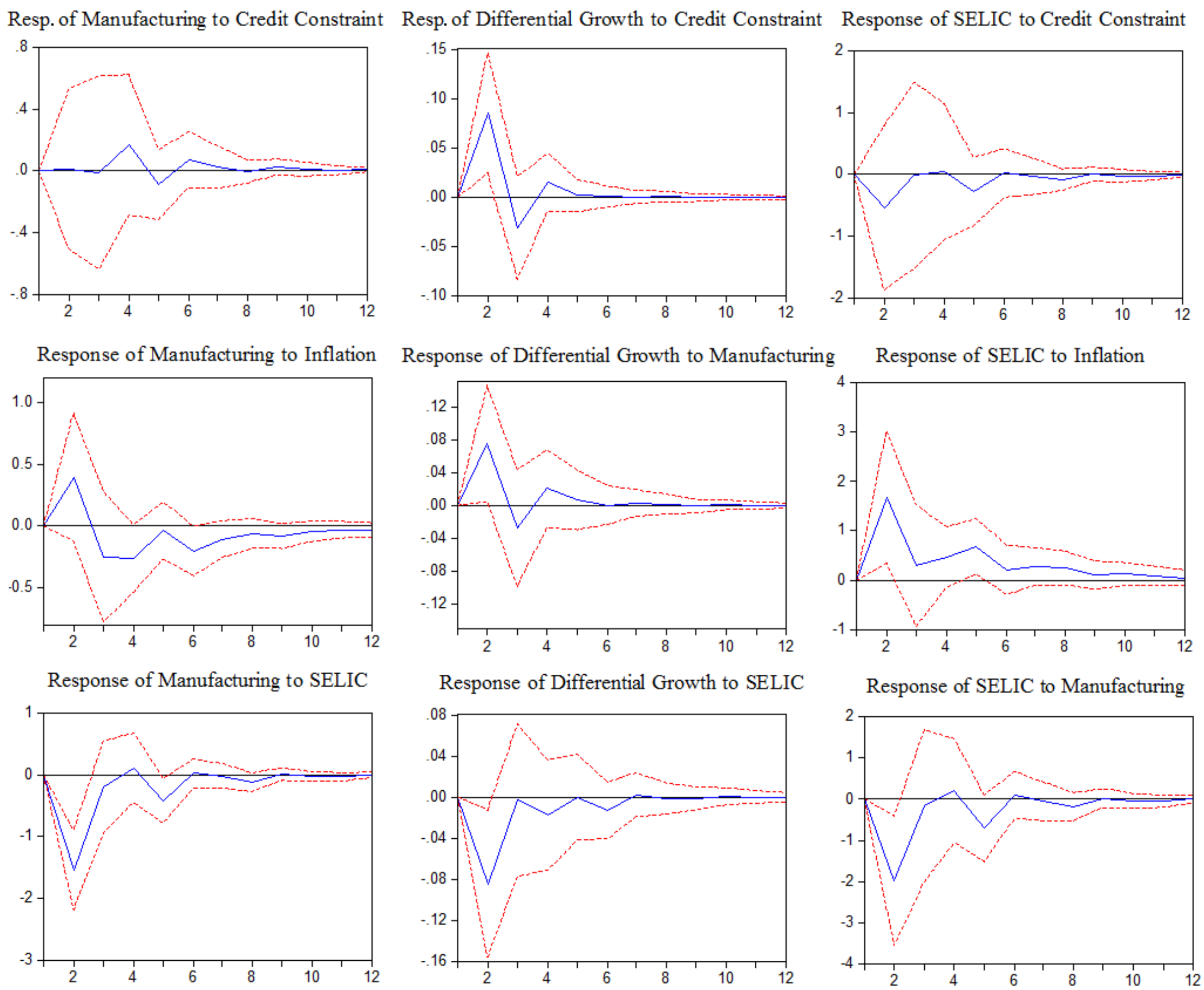


Figure 6.8 – Impulse response for a SVAR of order 2 with Manufacturing index, Inflation, Credit Constraint, Differential Growth and SELIC.

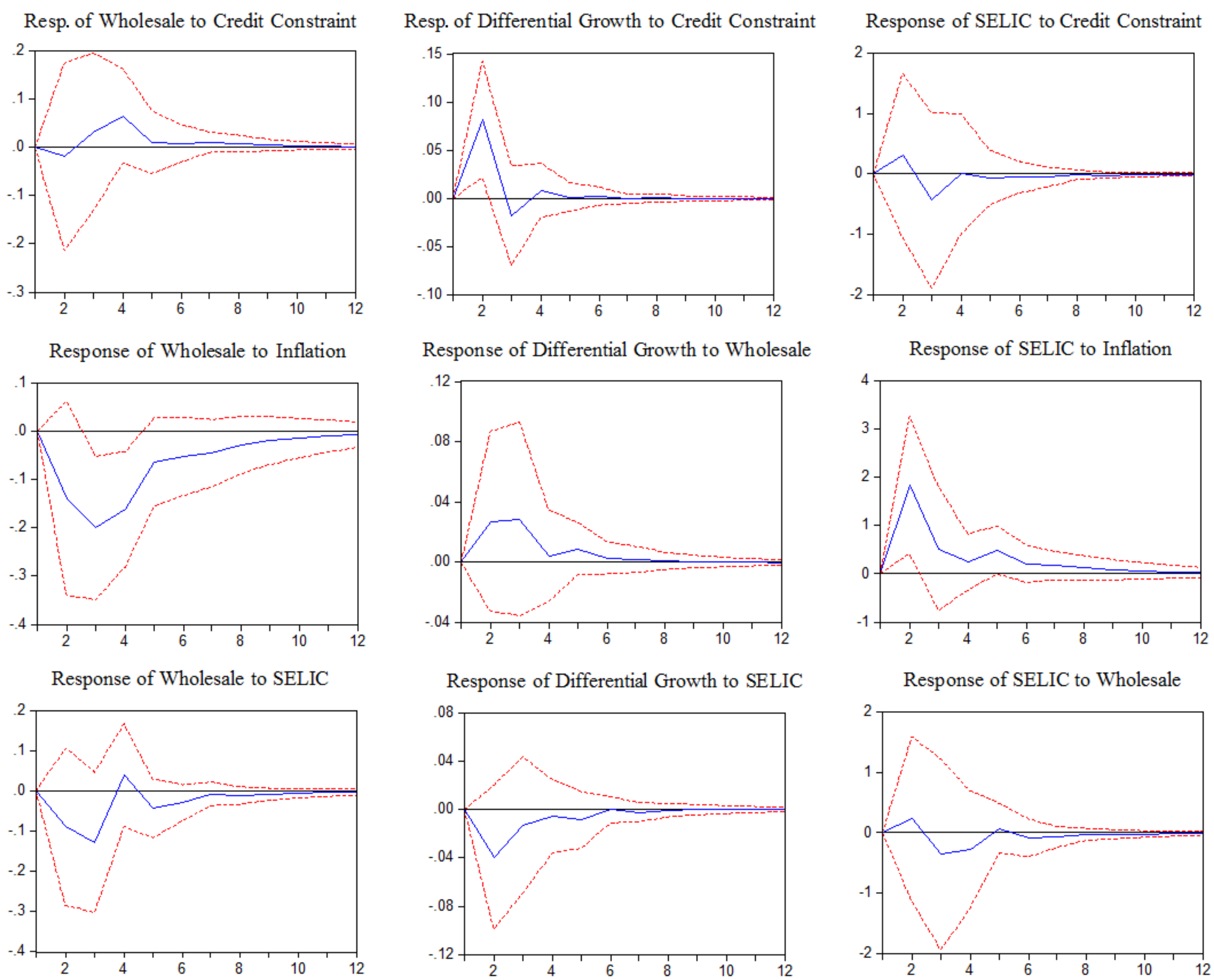


Figure 6.9 – Impulse response for a SVAR of order 2 with Real Wholesale index, Inflation, Credit Constraint, Differential Growth and SELIC.

6.6.6.2 Alternative Results Using the Differential Growth using 250 Employees as Cut-offs to Classify LEs and SMEs.

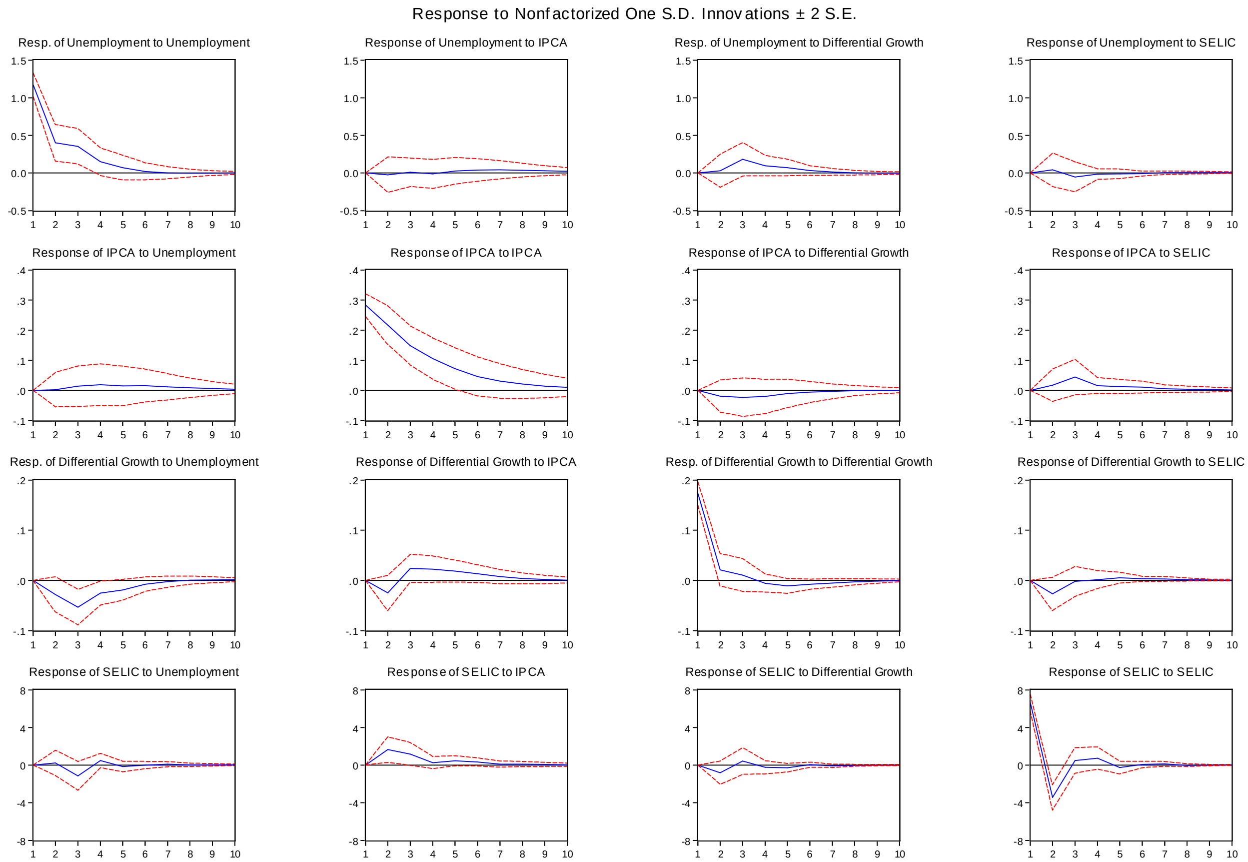


Figure 6.10 –Impulse Response Functions for a SVAR of order 2 with Unemployment, Inflation, Differential Growth (Using 250 Employees as Cut-off) and SELIC.

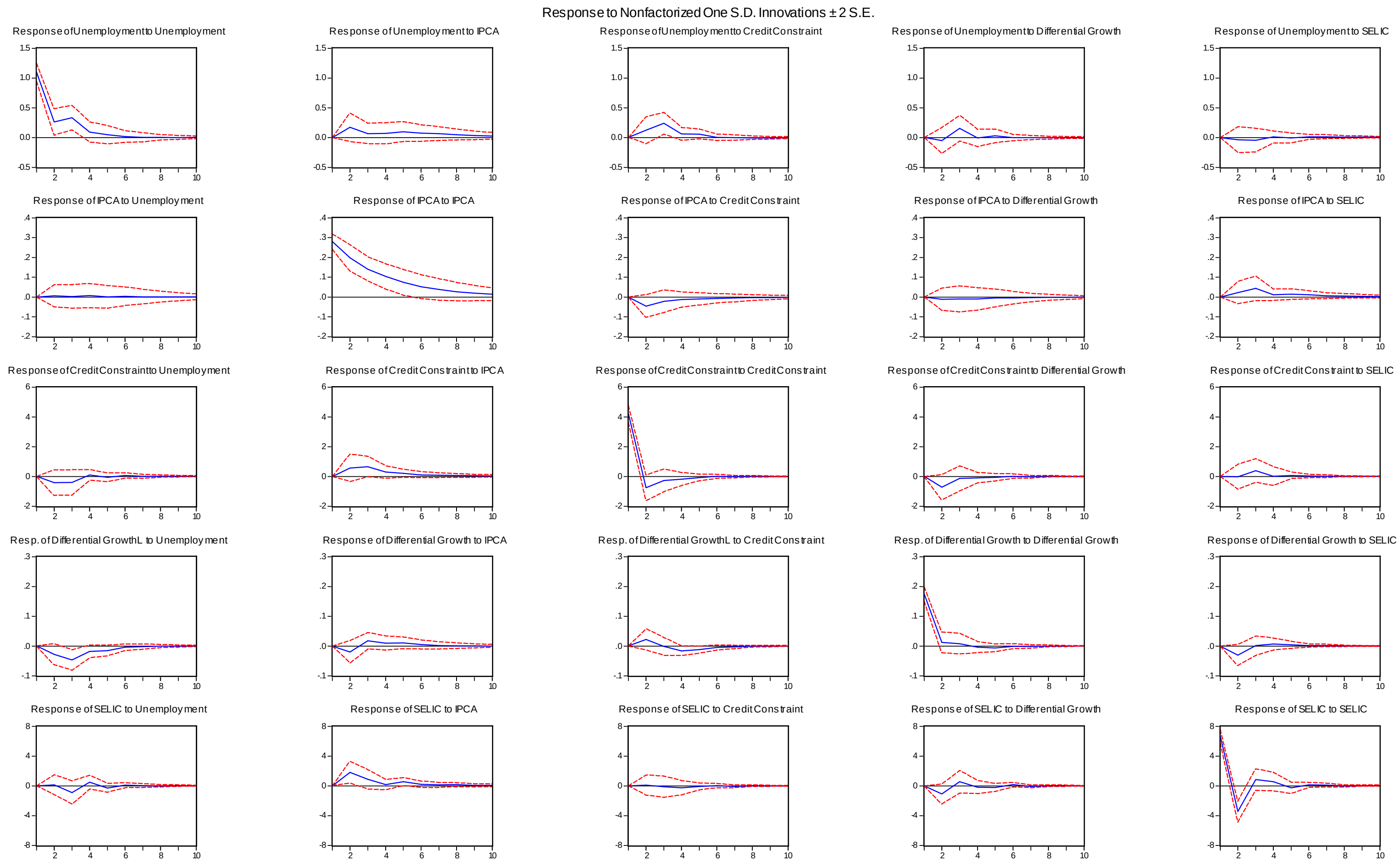


Figure 6.11 – Impulse Responses for a SVAR of order 2 with Unemployment, Inflation, Credit Constraint, Differential Growth (250-250) and SELIC.

6.6.6.3 Alternative Results Considering Credit Constraint as *I*(0).

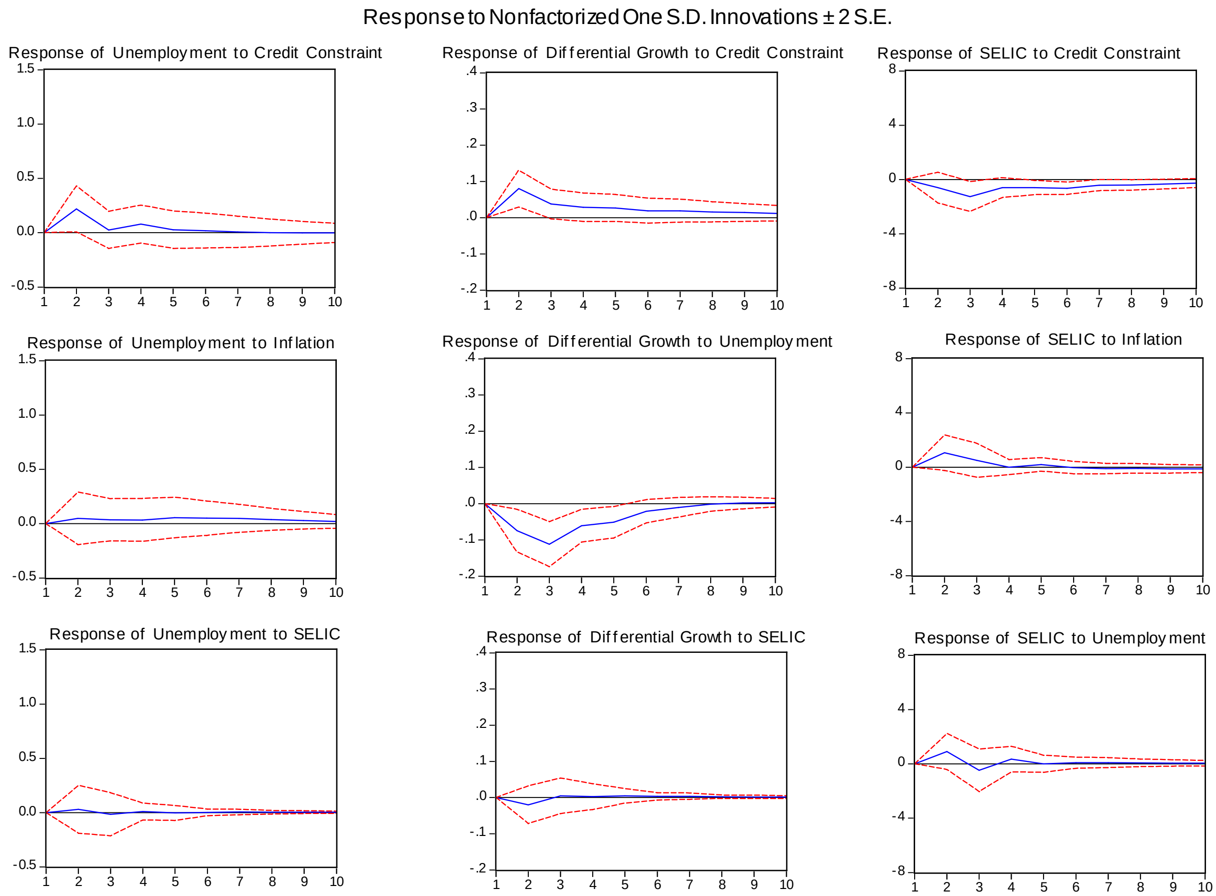


Figure 6.12 – Impulse response for a SVAR of order 2 with Unemployment, Inflation, Credit Constraint (*I*(0)), Differential Growth and SELIC.

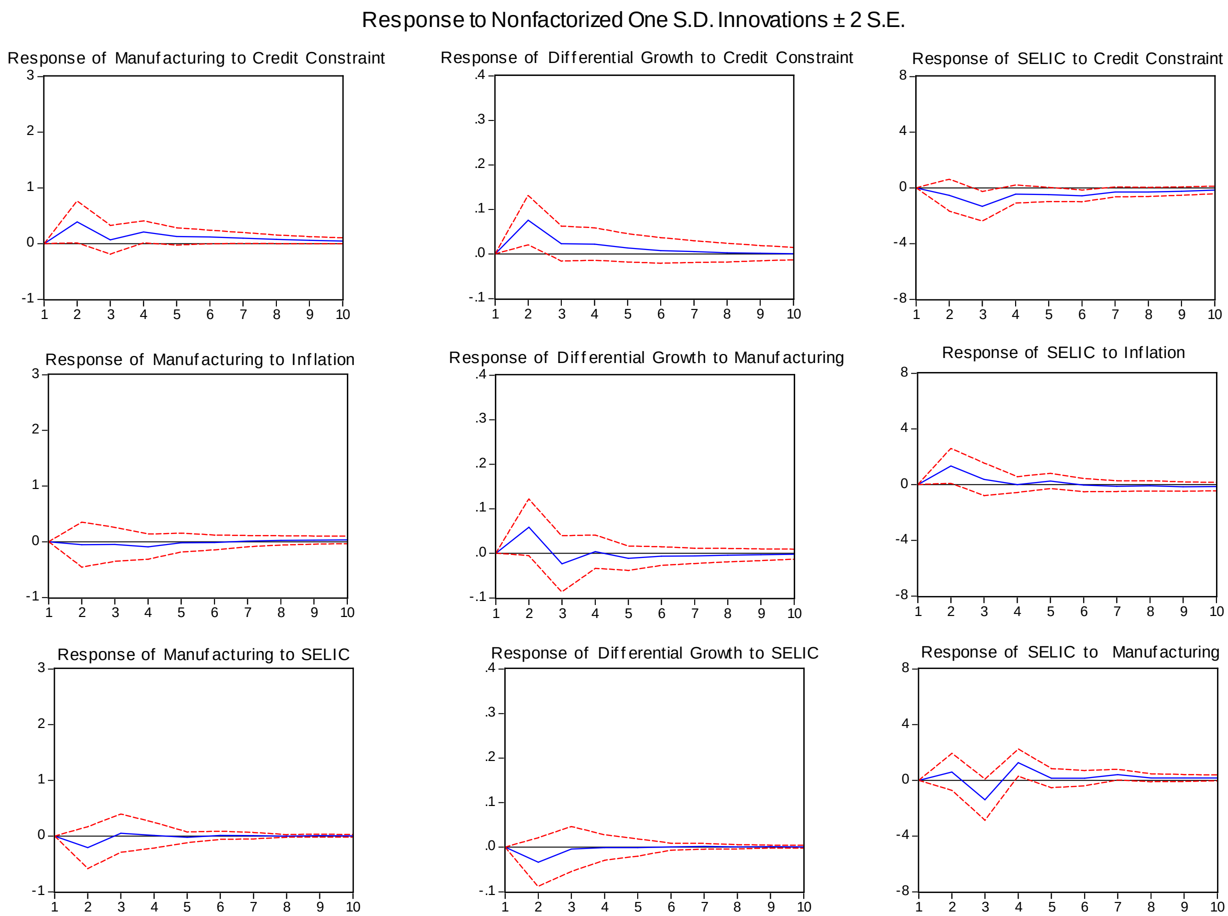


Figure 6.13 – Impulse response for a SVAR of order 2 with Manufacturing, Inflation, Credit Constraint (*I*(0)), Differential Growth and SELIC.

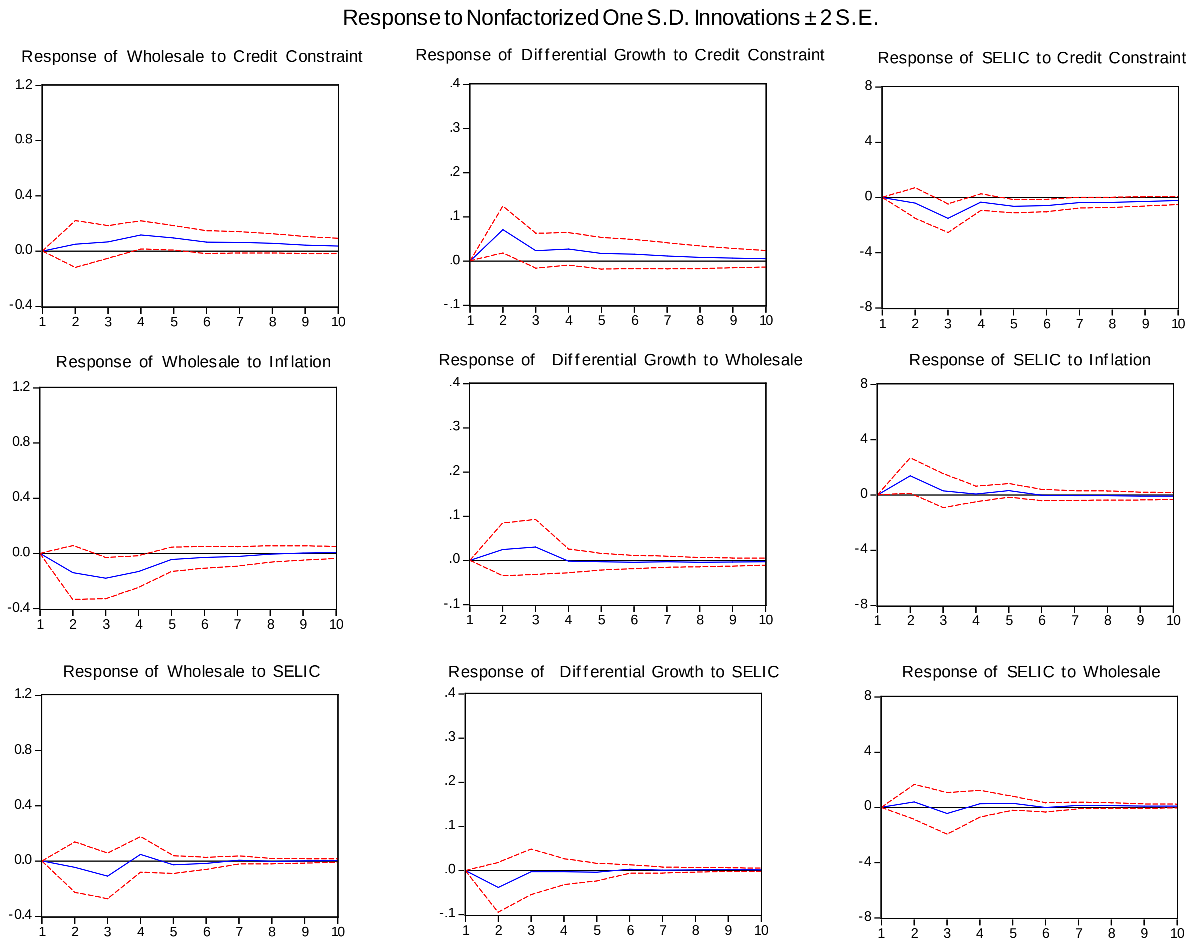


Figure 6.14 – Impulse response for a SVAR of order 2 with Wholesale, Inflation, Credit Constraint (*I*(0)), Differential Growth and SELIC.

Chapter 7

Summary and Conclusions

This thesis studies the importance of small and medium-sized businesses in the Brazilian economy. It provides a literature review that introduces the issues related to the importance of this sector for economic growth and its sensitivity to business cycles in Chapter 2. To explore those issues in the Brazilian economy, a unique dataset was constructed from the databases described in Chapter 3.

Chapter 4 examines the importance of small and medium-sized businesses for regional economic growth in Brazil. The chapter investigates how the size of this sector and its human capital impact on economic growth. It uses a growth regression as in Beck et al. (2005) and extends it to incorporate the human capital embodied in small and medium-sized businesses into the analysis. The panel data econometric approach includes the use of estimators developed by Arellano and Bond (1991) and Blundell and Bond (1998) that use instrumental variables to control for endogeneity. The chapter finds that a larger proportion of small and medium-sized businesses in the Brazilian regions do not promote regional economic growth and the human capital embodied in those businesses does not have a clear positive impact on growth. Regional inequalities observed in Brazil call for a regional analysis, and the empirical results confirm that small and medium-sized businesses do not promote regional growth, regardless of the level of GDP per capita of the regions. Nevertheless, the regional analysis indicates that human capital embodied in small businesses is important for regional economic growth in richer regions. The indication that SMEs' human capital is more important for rich regions reflects the fact that in this set of regions human capital can contribute for gains in productivity through their absorptive and innovation capabilities. Moreover, the results are in line with the idea that rich regions are likely to be associated with better institutions that provide the necessary conditions to incite more human capital formation and make effective use of this capital.

Conversely, a worse institutional quality that seems to be related to the group of poor regions is likely to incite more unproductive entrepreneurship and less human capital formation, which is associated with lower productivity levels. The findings, however, are potentially subject to endogeneity and do not consider the spatial interactions among regions which may lead to erroneous conclusions.

Chapter 5 buids on Chapter 4 and considers the existence of spatial regional interactions in the process of regional growth in Brazil using spatial econometrics. It was noted in the chapter that the literature has shown that spatial dependence is an important part of the economic growth process and so if each individual region is treated as an independent entity regression results may be biased. The chapter, therefore, uses spatial econometrics estimators developed by Anselin (1988) for cross-section estimations and modified by Elhorst (2003, 2005, 2010) for panel data estimations to consider how spatial interactions affect regional economic growth. In particular, the chapter contributes to the literature by providing for the first time an analysis on how spatial interactions in the small business sector might affect regional economic growth.

The empirical results confirm a strong spatial dependence in the process of regional economic growth in Brazil, supporting the arguments for using spatial econometrics techniques. The spatial panel regressions are in line with the results presented in Chapter 4 in the sense that a larger proportion of the small business sector is not positively related to the process of economic growth and human capital embodied in this sector is more important in process. Interestingly, the results suggest that a given region benefits from more small firms in neighbouring regions but there is no indication of positive spillovers stemming from human capital embodied in these firms. The chapter also examines the regional inequalities observed in Brazil using measures of global and local autocorrelation that suggested the presence of two clusters in Brazil: one encompassing poor regions in the northern part of the country and another one encompassing rich ones in the southern Brazil. In general, the regional analysis is in line with the results for the country as a whole. The presence of small and medium-sized firms is negatively related to growth but the human capital embodied in these firms seems to impact positively on regional economic growth. However, there are some differences in the results for the two spatial clusters in terms of the level of significance and magnitude of the coefficients estimated, which are consistent with the argument that the activities of small firms are influenced by local characteristics. Therefore, the empirical evidence presented in this chapter does not support the public policy view that the size of the small and medium-sized sector should be locally increased *per se* to improve economic performance. Instead, policy makers should better understand the spatial interactions of directly supporting this sector in a given region to promote growth on its neighbourhood; public policy should be coordinated with a broader regional focus in order to explore entrepreneurship externalities in Brazil. However, if the focus is on localised direct support to small firms, educational policies should be a focus of the public policy by enabling more human capital formation. The appendix of Chapter 5 complements the and provides a robustness check by emphasising the importance of the spatial dependence by removing it from the data as in Badinger et al. (2004). The results from the estimations with the data without the influence of spatial autocorrelation confirm the presence of a significant convergence in the GDP per capita. Besides, the results show that the conditioning variables (including the variables that capture aspects of the small and medium-sized firms) lose their statistical significance, suggesting that the effect of the conditioning variables that affect growth are intrinsically related with the spatial pattern in Brazil as argued in Silveira-Neto and Azzoni (2006). These results confirm the importance of the spatial dependence in the process of regional economic growth in Brazil.

While chapters 4 and 5 in this thesis analyse the importance of small businesses for regional economic growth, the last empirical chapter, Chapter 6, changes the focus and investigates the behaviour of small businesses during business cycles. The chapter follows Moscarini and Postel-Vinay (2009) and uses a measure of relative firm performance in terms of employment to check whether employment in small firms is more cyclically sensitive than in larger firms. The stylised facts for Brazil show that employment in small firms is relatively more sensitive to business cycle fluctuations than in larger ones. This pattern is also observed across Brazilian states and various sectors. Following the documentation that small businesses are more sensitive to business cycle fluctuations, Structural

Vector Autoregressions are used as in Gertler and Gilchrist (1994) and Moscarini and Postel-Vinay (2010) to investigate how shocks in different macroeconomic variables affect employment series in businesses of different sizes. The results suggest that employment series in small businesses seem to be more sensitive to cycles due to credit constraints. Therefore, policy makers should aim at easing financial constraints for small firms to protect employments in these firms during the downturns.

In concluding, this thesis has analysed the role of the small and medium-sized businesses related to two economic aspects in Brazil. Firstly, it considered the importance of these businesses in the process of regional growth in Brazil. Secondly, it provided evidence that suggests that small firms are relatively more sensitive to cyclical conditions due to credit constraints. Future research could build on the results provided in this thesis to further understand the role of small firms in the Brazilian economy. Specificaly, the use of firm level data could provide an insight into how specific characteristics of small firms are associated with economic growth and with the sensitivity of these firms to cyclical conditions. For instance, in the strand of research on regional growth and SMEs, firm level data could be used to provide micro evidence on the return of human capital at firm level in other to better understand the process of productivity increase and regional growth process. Firm level data could also be used to identify firms that export and try to understand how those firms contribute to employment generation and growth. In the strand of research on SMEs and cycles, firm level data could be used to provide a more detailed understanding on how the skills and occupational area of the SMEs employees are related to cycles, this would allow policy makers to provide more effective labour market policies to maintain jobs during crises.

References

Abreu M, de Groot H, Florax R (2005) Space and Growth: A Survey of Empirical Evidence and Methods. *Région et Développement* 21: 13-44.

Acemoglu D, Johnson S, Robinson JA (2001) The colonial origins of comparative development: An empirical investigation. *American Economic Review* 91(5): 1369-1401.

Acs Z, Amoros J (2008). Entrepreneurship and competitiveness dynamics in Latin America. *Small Business Economics* 31(3): 305–322.

Acs Z, Desai S, Hessels J (2008). Entrepreneurship, economic development and institutions. *Small Business Economics* 31(3): 219–234.

Aghion P, Askenazy P, Berman N, Cette G, Eymard L (2008) Credit Constraints and the Cyclicality of R&D Investment: Evidence from France. NER - R 198, Banque de France.

Aguayo-Tellez E, Muendler M, Poole J (2010). Globalization and Formal-Sector Migration in Brazil. *World Development* 38(6): 840-856.

Aguiar M, Gopinath G (2007) Emerging market business cycles: The cycle is the trend. *Journal of Political Economy* 115(1): 69-102.

Albrecht J, Navarro L, Vroman S (2009) The effects of labour market policies in an economy with an informal sector. *Economic Journal* 119(539): 1105-1129.

Almeida R, Carneiro P (2009). Enforcement of labor regulation and firm size. *Journal of Comparative Economics,* 37(1): 28-46.

Anselin L (1988) Spatial Econometrics: Methods and Models. Kluwer, Dordrecht.

Anselin L (1995) Local Indicators of Spatial Association—LISA. *Geographical Analysis* 27: 93–115.

Anselin L (1996) The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In Fischer M, Scholten H, Unwin D (eds) *Spatial Analytical Perspectives on GIS,* 111–25 Chapter 8. Taylor and Francis, London.

Anselin L, Bao S (1997) Exploratory spatial data analysis linking SpaceStat and ArcView. In: Fisher M, Getis A (eds) *Recent developments in spatial analysis,* 35-60 Chapter 3. Springer, Berlin Heidelberg, New York.

Anselin L, Bera AK, Florax R, Yoon MJ (1996) Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26(1): 77-104.

Anselin L, Hudak S (1992) Spatial econometrics in practice : A review of software options. *Regional Science and Urban Economics* 22(3):509-536.

Arbia G, Battisti M, Di Vaio G (2010) Institutions and geography: Empirical test of spatial growth models for European regions. *Economic Modelling* 27(1): 12-21.

Arbia G, Le Gallo J, Piras G (2008) Does Evidence on Regional Economic Convergence Depend on the Estimation Strategy? Outcomes from Analysis of a Set of NUTS2 EU Regions. *Spatial Economic Analysis* 3(2): 209-224.

Arellano M (2003) Panel Data Econometrics, Oxford University Press: Advanced Texts in Econometrics.

Arellano M, Bond S (1991) Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* 58: 277-297.

Arellano M, Bover O (1995) Another Look at the Instrumental Variable Estimation of Error-Components Models. *Journal of Econometrics* 68:29-51.

Audretsch D (2007) Entrepreneurship Capital and Economic Growth. *Oxford Review of Economic Policy*. 23(1): 63–78.

Audretsch D, Carree M, Van Stel A, Thurik R (2002) Impeded Industrial Restructuring: The Growth Penalty. *Kyklos* 55(1): 81-98.

Audretsch D, Fritsch M (2002) Growth Regimes over Time and Space. Regional *Studies* 36(2): 113-124.

Audretsch D, Keilbach M (2004) Entrepreneurship Capital and Economic Performance. *Reginal Studies* 38(8): 949-959.

Audretsch D, Keilbach M (2007) The Localisation of Entrepreneurship Capital: Evidence from Germany. *Papers in Regional Science* 86(3): 351-365.

Ayyagari M, Beck T, Demirguc-Kunt A (2003) Small and Medium Enterprises across the Globe: A New Database. World Bank mimeo.

Ayyagari M, Beck T, Demirguc-Kunt, A (2007) Small and medium enterprises across the globe. *Small Business Economics* 29(4): 415-434.

Azzoni CR (2001) Economic growth and regional income inequality in Brazil. *Annals of Regional Science* 35: 133–152

Badinger H, Muller W, Tondl G (2004) Regional Convergence in the European Union 1985-1999: A spatial dynamic panel analysis. *Regional Studies* 38(3): 241-253.

Barro R (1991) Economic Growth in a Cross Section of Countries. *Quarterly Journal of Economics* 106(2): 407-43.

Barro R (2001) Human capital: Growth, history, and policy-a session to honor Stanley Engerman. Human capital and growth. *American Economic Review* 91(2): 12–17.

Barro R, Sala-i-Martin X (1992) Convergence. *Journal of Political Economy* 100(2): 223-251.

Battisti M, Vaio G (2008) A spatially filtered mixture of β-convergence regressions for EU regions, 1980–2002. *Empirical Economics* 34(1): 105-121.

Baum C, Schaffer M, Stillman S (2003) Instrumental Variables and GMM: Estimation and Testing. *Stata Journal* 3(1): 1-31.

Baumol W (1990) Entrepreneurship: Productive, Unproductive, and Destructive. *Journal of Policy Modeling* 30: 575-589.

Baumol W (2008) Small Enterprises. Large Firms, Productivity Growth and Wages. *Journal of Political Economy*30:575-589.

Baxter M, King RG (1999) Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. *The Review of Economics and Statistics* 81(4): 575-593.

Beck T, Demirgüç-Kunt A (2006) Small and medium-size enterprises: Access to finance as a growth constraint. Journal of Banking and Finance 30(11): 2931-2943.

Beck T, Demirguc-Kunt A, Laeven L, Levine R (2008). Finance, Firm Size and Growth. Journal of Money, Credit and Banking 40:1379-1405.

Beck T, Demirguc-Kunt A, Levine R (2005a) SMEs, Growth, and Poverty: Cross-Country Evidence. *Journal of Economic Growth* 10(3): 199-229.

Beck T, Demirgüç-Kunt A, Maksimovic V (2005b) Financial and Legal Constraints to Growth: Does Firm Size Matter? *Journal of Finance* 60(1): 137-177

Beck T, Demirgüç-Kunt A, Maksimovic V (2006) The influence of financial and legal institutions on firm size. *Journal of Banking and Finance* 30(11): 2995-3015.

Bivand R, Brunstad R (2006) Regional Growth in Western Europe: Detecting Spatial Misspecification Using the R Environment. *Papers in Regional Science*. 85(2): 277-297.

Blundell R, Bond S (1998) Initial Conditions and moment restrictions in Dynamic Panel Data Models. *Journal of Econometrics* 87:111-143.

Bond S, Hoeffler A, Temple J (2001) GMM Estimation of Empirical Growth Models. Centre for Economic Performance (CEPR) Discussion Paper nº3048, London.

Bosker M (2007) Growth, agglomeration and convergence: A space-time analysis for European regions. *Spatial Economic Analysis* 2: 91–110

Brock W, Evans D (1989). Small Business Economics. *Small Business Economics* 1: 7-20.

Burns AM, Mitchell WC (1946) Measuring Business Cycles. National Bureau of Economic Research, New York.

Burridge P (1980) On the Cliff-Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society B* 42: 107-108.

Campbell JY, Mankiw NG (1987) Are output fluctuations transitory? *Quarterly Journal of Economics* 102(4): 857-880.

Carree M, Thurik A (1998) Small firms and economic growth in Europe. *Atlantic Economic Journal* 26: 137–146

Carvalho GO (2006) Environmental Resistance and the Politics of Energy Development in the Brazilian Amazon. *The Journal of Environment & Development* 15(3): 245-268

Caselli F, Esquivel G, Lefort F (1996) Reopening the Convergence Debate. *Journal of Economic Growth* 1(3): 363-389.

Christiano LJ, Fitzgerald TJ (2003). The Band Pass Filter. *International Economic Review* 44(2): 435-465.

Coelho RL, Figueiredo L (2007) Uma análise da hipótese da convergência para os municípios Brasileiros. *Revista Brasileira de Economia* 61: 331–352.

Cravo T (2010) SMEs and Economic Growth in the Brazilian Micro-Regions. *Papers in Regional Science* 89(4): 711-734.

Cravo T, Gourlay A, Becker B (2010) SMEs and Regional Economic Growth in Brazil. *Small Business Economics*, in Press, Accepted Manuscript, DOI 10.1007/s11187-010-9261-z.

Cunha A, Moreira LP (2006) Ciclos Econômicos Regionais no Brasil de 1985 a 2002: Uma Introdução. *Revista de Economia Contemporânea* 10(1): 115-138.

Dallerba S, LeGallo J (2008) Regional convergence and the impact of European structural funds over 1989–1999: A spatial econometric analysis. *Papers in Regional Science* 87: 219–244

De Barros RP, Corseuil CH (2001) The impact of regulations on Brazilian labor market performance. Research Network Working Paper No.R-427, Inter-American Development Bank.

De Paula A, Scheinkman J (2009) The Informal Sector: An Equilibrium Model and Some Empirical Evidence from Brazil. Penn Institute for Economic Research Working Paper 09-044.

Dias J, McDermott J (2006) Institutions, Education, and Development: The Role of Entrepreneurs. *Journal of Development Economics* 80: 299-328.

Du J, Lu Y, Tao Z (2008) Economic institutions and FDI location choice: Evidence from US multinationals in China. *Journal of Comparative Economics* 36(3): 412-429.

Durlauf S, Johnson P, Temple J (2005) Growth Econometrics. In Aghion P and Durlauf S (eds) *Handbook of Economic Growth* Chapter 8: Elservier.

Elhorst JP (2003) Specification and Estimation of Spatial Panel Data Models, *International Regional Science Review* 26(3): 244-268.

Elhorst JP (2005) Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis* 37(1): 85-106.

Elhorst JP (2010) Spatial Panel Data Models. In Fischer M, Getis A. (eds) Handbook of Applied Spatial Analysis. Springer, Berlin.

Elhorst JP, Freret S (2009) Evidence Of Political Yardstick Competition In France Using A Two-Regime Spatial Durbin Model With Fixed Effects. *Journal of Regional Science* 49(5):931-951.

Ellery R, Gomes V, Sachsida A (2002) Business Cycle Fluctuations in Brazil. *Revista Brasileira de Economia* 56(2): 269-308.

Ertur C, Koch W (2007) Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics* 22(6): 1033-1062.

Ertur C, Le Gallo J, Baumont C (2006) The European regional convergence process, 1980-1995: Do spatial regimes and spatial dependence matter? *International Regional Science Review* 29(1): 3-34.

Esteves L (2007) A Note on Gibrat's Law, Gibrat's Legacy and Firm Growth: Evidence from Brazilian Companies. *Economics Bulletin* 12(19) 1-7.

European Commission. The 2003 Observatory of European SMEs; SMEs in Europe 2003. Luxembourg, 2004.

Ferreira A (2000) Convergence in Brazil: Recent trends and long-run prospects. *Applied Economics* 32(4): 479-489.

Ferreira A, Diniz C (1995) Convergência entre as rendas per capita estaduais no Brasil. *Revista de Economia Política* 15(4): 38-56.

Ferstl R (2007) Spatial Filtering with EViews and MATLAB. *Austrian Journal of Statistics* 36(1): 17-26.

Fingleton B (2000) Spatial econometrics, economic geography, dynamics and equilibrium: a ‘third way’? *Environment and Planning A* 32: 1481–1498

Fingleton B, Lopez-Bazo E (2006) Empirical Growth Models with Spatial Effects. *Papers in Regional Science* 85(2): 177-198.

Figueiredo L,Garcia R (2003) Impactos de Diferentes Metodologias para o Cálculo da Taxa Líquida de Migração Sobre o Estudo do Crescimento Economico e da Distribuição da População dos Estados Brasileiros. In: Encontro Transdiciplinar Espaço e População. Abep/Nepo, Campinas.

Galor, O., (1996). Convergence? Inference from Theoretical Models. The Economic Jounal, 106, 1056-1069.

Gertler M, Gilchrist S (1994) Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms. *The Quarterly Journal of Economics* 109(2): 309-340.

Getis A, Griffith DA (2002) Comparative spatial filtering in regression analysis. *Geographical Analysis* 34(2): 130-140.

Getis A, Ord JK (1992) The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis* 24: 189-206.

Globerman S, Shapiro D (2002) Global foreign direct investment flows: The role of governance infrastructure. *World Development* 30(11): 1899-1919.

Griffith R, Redding S, Van Reenen J (2004) Mapping the two faces of R&D: Productivity growth in a panel of OECD industries. *The Review of Economics and Statistics* 86(4): 883-895.

Hansen L (1982) Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50(3): 1029-1054.

Henley A (2005) On Regional Growth Convergence in Great Britain. *Regional Studies* 39(9): 1245-1260.

Henley A, Arabsheibani GR, Carneiro FG (2009) On Defining and Measuring the Informal Sector: Evidence from Brazil. *World Development* 37(5): 992-1003.

Hodrick RJ, Prescott EC (1981) Post-war U.S. Business Cycles: An Empirical Investigation. Discussion Paper 451, Carnegie-Mellon University.

Hodrick RJ, Prescott EC (1997) Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking* 29(1): 1 – 16.

Hoeffler AE (2002) The augmented Solow model and the African growth debate. *Oxford Bulletin of Economics & Statistics* 64(2): 135.

IBGE/Sebrae (2005). Economia Informal Urbana 2003. IBGE, Rio de Janeiro, Brazil.

Islam N (1995). Growth Empirics: A Panel Data Approach. *Quarterly Journal of Economics* 110(4): 1127-1170.

Issler JV, Notini H, Rodrigues C (2009) Constructing Coincident and Leading Indices of Economic Activity for the Brazilian Economy. Ensaios Econômicos 694, Fundação Getúlio Vargas (FGV).

Johnson P, Takeyama L (2001) Initial conditions and economic growth in the US states. *European Economic Review* (4-6)919 – 927.

Kanczuk F (2004). Real interest rates and Brazilian business cycles. *Review of Economic Dynamics* 7(2): 436-455.

Krugman P (1999) The Role of Geography in Development. *International Regional Science Review* 22(2): 142-161.

Kydland F, Prescott E (1990) Business Cycles: Real Facts and a Monetary Myth. *Federal Reserve Bank of Minneapolis Quarterly Review* 14(2): 3-18.

La Porta R, Shleifer A (2008). The Unofficial Economy and Economic Development. *Brookings Papers on Economic Activity* 2: 275-363

Lau L J, Jamison DT, Lius S, Rivkin S (1993) Education and economic growth: some cross-sectional evidence from Brazil. *Journal of Development Economics* 41: 45-70.

Laurini M, Andrade E, Valls P (2005) Income convergence clubs for Brazilian Municipalities: a non-parametric analysis. *Applied Economics* 37(18): 2099-2118.

Le Gallo J. and Dall'erba S. (2006) Evaluating the Temporal and Spatial Heterogeneity of the European Convergence Process, 1980–1999. *Journal of Regional Science* 46(2): 269-288.

Le Gallo J, Ertur C (2003) Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995. *Papers in Regional Science* 82(2): 175-201.

LeSage JP (1999). The Theory and Practice of Spatial Econometrics, a manual to accompany the Spatial Econometrics Toolbox, available at: www.spatial-econometrics.com

LeSage JP, Fischer M (2007) Spatial Growth Regressions: Model Specification, Estimation and Interpretation, SSRN Working Paper.

LeSage JP, Fischer M (2008) Spatial Growth Regressions: Model Specification, Estimation and Interpretation. *Spatial Economic Analysis* 3(3): 275-304.

LeSage JP, Pace RK (2009) Introduction to Spatial Econometrics. CRC Press/Taylor & Francis Group, London.

Levine R, Renelt D (1992) A Sensitivity Analysis of Cross-Country Growth Regressions. *American Economic Review*, 82(4): 942-963.

Liu T, Spector L (2005) Dynamic employment adjustments over business cycles. *Empirical Economics* 30(1): 151-169.

Lopez-Bazo E, Vaya E, Artis M (2004) Regional Externalities and Growth: Evidence from European Regions. *Journal of Regional Science* 44(1): 43-73.

Lucas R (1988) On the Mechanics of Economic Development. *Journal of Monetary Economics* 22(1): 3-42.

Lucas R (1977) Understanding Business Cycles. In Carnegie-Rochester Conference Series on Public Policy 5, 7-29.

Lütkepohl H (2005) New Introduction to Multiple Time Series Analysis. Springer-Verlag, Berlin, Germany.

Maloney WF (2004) Informality Revisited. *World Development* 32(7): 1159-1178.

Mankiw G, Romer D, Weil D (1992) A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics* 107(2): 404-437.

Martincus C (2010) Spatial Effects of Trade Policy: Evidence from Brazil. *Journal of Regional Science* 50(2):541-569.

Martincus CV, Molinari A (2007) Regional business cycles and national economic borders: What are the effects of trade in developing countries? *Review of World Economics* 143(1): 140-178.

Maza A, Villaverde J (2009) Spatial Effects on Provincial Cconvergence and Income Distribution in Spain: 1985–2003. *Tijdschrift voor Economische en Sociale Geografie* 100(3): 316–331.

Menezes-Filho NA, Marcondes R, Pazello E, Scorzafave LG (2006) Instituições e Diferenças de Renda entre os Estados Brasileiros: Uma análise histórica. In: XXXIV Encontro Nacional de Economia. ANPEC, 2006. v. 1.

Menezes-Filho N, Muendler M, Ramey G (2008) The Structure of Worker Compensation in Brazil with a Comparison to France and the United States. *Review of Economics and Statistics* 90(2): 324-346.

Michelacci C, Silva O (2007) Why so many local entrepreneurs? *Review of Economics and Statistics* 89(4): 615-633.

Mohl P, Hagen T (2010) Do EU structural funds promote regional growth? New evidence from various panel data approaches. *Regional Science and Urban Economics* 40(5): 353-365.

Moore GH (1961) Leading and Confirming Indicators of General Business Changes. In Geoffrey H. Moore (ed). *Business Cycle Indicators*. Chapter 3, 45-109.

Moscarini G, Postel-Vinay F (2008) The Timing of Labor Market Expansions: New Facts and a New Hypothesis. In Acemoglu D, Rogoff K, Woodford M (eds) NBER *Macroeconomics Annual* 23: 1-51.

Moscarini G, Postel-Vinay F (2009). Large Employers Are More Cyclically Sensitive. NBER Working Papers 14704.

Moscarini G, Postel-Vinay F (2010). Unemployment and Small Cap Returns: The Nexus. *American Economic Review Papers and Proceedings*.

Mossi MB, Aroca P, Fernández IJ, Azzoni CR (2003) Growth Dynamics and Space in Brazil. *International Regional Science Review* 26(3): 393-418.

Moura ML, Carvalho A (2010) What can Taylor rules say about monetary policy in Latin America? *Journal of Macroeconomics* 32(1): 392-404.

Moura R (2008) Testando as Hipóteses do Modelo de Mincer para o Brasil. *Revista Brasileira de Economia*, 62(4): 407–449

Mueller P (2007) Exploiting Entrepreneurial Opportunities: The Impact of Entrepreneurship on Growth. *Small Business Economics* 28: 355-362.

Muendler M (2007) Trade and Workforce Changeover in Brazil. National Bureau of Economic Research (NBER), Working Papers nº12980, Cambridge, USA.

Naritomi J, Soares R, Assunção J (2009) Institutional Development and Colonial Heritage within Brazil. Pontifical Catholic University of Rio de Janeiro (PUC-Rio). Discussion Paper No. 7579.

Nakabashi L, Salvato M (2007) Human Capital Quality in the Brazilian. *Revista ANPEC* 8:211-229.

Nelson R, Phelps E (1966) Investment in humans, technological diffusion, and economic growth. *American Economic Review* 56(1/2): 65-75.

Neumeyer PA, Perri F (2005) Business cycles in emerging economies: the role of interest rates. *Journal of Monetary Economics* 52(2): 345-380.

Nickell S (1981) Biases in Dynamic-Models with Fixed Effects. *Econometrica* 49(6):1417-1426.

Nystrom K (2008) The institutions of economic freedom and entrepreneurship: Evidence from panel data. *Public Choice* 136: 269–282

Ord JK, Getis A (1995) Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis* 27: 286-306.

Parker S (2004) The Economics of Self-Employment and Entrepreneurship, Cambridge, UK.

Pinto E (2011) Firms' relative sensitivity to aggregate shocks and the dynamics of gross job flows. *Labour Economics* 18(1): 111-119.

Pollock D (2000) Trend estimation and de-trending via rational square-wave filters. *Journal of Econometrics* 99(2): 317-334.

Poole JP (2007) Multinational Spillovers through Worker Turnover. Department of Economics, University of California, Santa Cruz.

Quah D (1996) Empirics for economic growth and convergence. *European Economic Review* 40(6): 1353-1375.

Quah D (1997a) Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs. *Journal of Economic Growth* 2: 27-59.

Quah D (1997b) Twin peaks: Growth and convergence in models of distribution dynamics. *Economic Journal* 106:1045-1055.

Ramajo J, Márquez M, Hewings G, Salinas M (2008) Spatial heterogeneity and interregional spillovers in the European Union: Do cohesion policies encourage convergence across regions? European Economic Review 52(3): 551-567.

Rand J, Tarp F (2002) Business Cycles in Developing Countries: Are They Different? *World Development* 30(12): 2071-2088.

Ravn MO, Uhlig H (2002) On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations. *The Review of Economics and Statistics* 84(2): 371-376.

Resende G (2009) Multiple dimensions of regional economic growth: The Brazilian case, 1991-2000. Proceedings of the 49th congress of the European Regional Science Association (ERSA). Lodz, Poland.

Rey S, Montouri B (1999) US regional income convergence: A spatial Econometric perspective. *Regional Studies* 33: 143-156

Reynolds P, Storey D, Westhead P (1994) Cross-national Comparisons of the Variation in New Firm Formation Rates. *Regional Studies* 28(4): 443-456.

Rodrik D, Subramanian A, Trebbi F (2004) Institutions rule: The primacy of institutions over geography and integration in economic development. Journal of Economic Growth 9(2):131-165

Romer C, Romer D (1990) Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz. National Bureau of Economic Research (NBER) Working Papers 2966.

Romer P (1986) Increasing Returns and long Run Growth. *Journal of Political Economy* 94(5): 1002-1037.

Romer P (1990) Endogenous Technological Change. *Journal of Political Economy* 98: S71-S102.

Roodman D (2009a) A note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics* 71(1): 135-158.

Roodman D (2009b) How to do xtabond2: an introduction to difference and system GMM in Stata. *Stata Journal* 9(1): 86-136.

Rua A, Nunes LC (2005) Coincident and leading indicators for the euro area: A frequency band approach. *International Journal of Forecasting* 21(3): 503-523.

Sabóia J (2000) Desconcentração industrial no Brasil nos anos 90: um enfoque regional. *Pesquisa e Planejamento Econômico* 30(1): 69-116.

Sala-i-Martin X (1996) The Classical Approach to Convergence Analysis. *Economic Journal* 106: 1019-1036.

Sala-i-Martin X (2002) 15 Years of New Growth Economics: What Have We Learnt?" Universitat Pompeu Fabra Economics and Business Working Paper No. 620, June Available at SSRN: http://ssrn.com/abstract=320765

SEBRAE 2007. Fatores condicionantes e taxas de sobrevivencia e mortalidade das micro e pequenas empresas no Brasil. SEBRAE, Brasilia, Brazil.

Silveira-Neto R, Azzoni CR (2006) Location and regional income disparity dynamics: The Brazilian Case. *Papers in Regional Science* 85(4): 599-613.

Silverman BW (1986). Density Estimation. Chapman and Hall, London.

Solow R (1956) A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics* 70(1): 65-94.

Solow R (2007). The last 50 Years in Growth Theory and the Next 10. *Oxford Review of Economic Policy* 23(1): 3-14.

Stock JH, Watson MW (1998) Business cycle fluctuations in US macroeconomic time series. National Bureau of Economic Research (NBER), Working Papers 6528, Cambridge, USA.

Taylor J (1993) Discretion versus policy rules in practice. *Carnegie–Rochester Conference Series on Public Policy* 39: 195–214.

Temple J (1999). The New Growth Evidence. *Journal of Economic Literature* 37(1): 112-156.

Tobler WR (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46: 234-240

Van Stel A, Caree M, Thurik R (2005) The Effect of Entrepreneurial Activity on National Economic Growth. *Small Business Economics* 24: 311-321.

Vandenbussche J, Aghion P, Meghir C (2005) Growth, Distance to Frontier and Composition of Human Capital. *Journal of Economic Growth* 11: 97-127.

Veracierto M (2008) On the cyclical behavior of employment, unemployment and labor force participation. *Journal of Monetary Economics* 55(6): 1143-1157.

Wennekers S, Thurik R (1999) Linking entrepreneurship and economic growth. *Small Business Economics* 13: 27–55

Wennekers S, Van Stel A, Thurik R, Reynolds P (2005) Nascent entrepreneurship and the level of economic development. *Small Business Economics* 30(3): 325-325.

Windmeyer F (2005) A Finite Sample Correction for the Variance of Linear Efficient Two step GMM Estimators. *Journal of Econometrics* 126: 25-51.

World Bank (2002) Brazil Jobs Report. Washington, DC, Report No. 24408-BR.