Multicollinearity and Financial Constraint in Investment Decisions: A Bayesian Ridge Regression*

Aquiles Elie Guimarães Kalatzis ¹, Camila Fernanda Bassetto ², Carlos Roberto Azzoni ³

Abstract

This paper is concerned with the investment decisions considering the presence of financial constraints of 373 large Brazilian firms in the period from 1997 to 2004, using panel data. We have used a bayesian econometric model classified by the firm capital intensity, considering ridge regression for the problem of multicollinearity among the model variables. Priors are assumed for the parameters, classifying the model in random or fixed effects. Normal and Student t distributions have considered for the model’s error. It was used a recursive predictive density criterion for the investment model comparisons. Results suggest the presence of multicollinearity provides important changes in the parameters and evidences the differences in the model parameters estimation with or without ridge regression. The estimation indicate that financial constraints are more important for capital-intensive firms, probably due to their lower profitability indexes, higher fixed costs and higher degree of property diversification, increasing agency costs.

Key Words: Investment Decision, Financial Constraint, Bayesian Ridge Regression.

JEL: G3, G30, G31

1. Introduction

Explaining investment decisions is a hot topic in the literature of investment theory. Understanding the role of different factors is of great importance, especially in developing economies, such as Brazil, for improvements in financial markets can lead to reductions in transactions and information costs, thus affecting investment decisions. Investment decisions are not as frequent as other firm’s decisions, and decisions rules of thumb are hard to develop. The literature points out the role of information asymmetries in the credit market, and sources of conflicts between managers and capital owners, such as moral hazard, adverse selection and agency costs, all affecting the cost of debt.

On the empirical side, many studies have attempted to identify factors related to investment decisions, introducing liquidity variables in recognition of the important role of internal funds on investment. The use of micro panel data allows for the abandonment of the representative firm and the introduction of firm heterogeneity. Usually, firms are grouped according to size to control for heteroskedasticity. Investment is usually the dependent variable in regressions in which the explanatory variables are cash flow, debt, sales and others. These are evidently correlated, especially cash flow and sales, but few discussions are found in the literature on this problem. In this paper we deal with such multicollinearity by using a ridge regression model with a bayesian approach.

In the next section a brief review of the literature will be present. The data and the model will be presented in the following section. The variables will be then examined to indentify the presence of multicollinearity. Next, the bayesian procedure will be presented, including the posteriori densities for the parameters. Model selection is performed next, by using the ordinate predictive density criterion. Finally, results are presented and analyzed.

2. Investment Decisions and Financial Constraint

The bulk of the literature on investment decisions in the last decades has concentrated on the effects of financial constraints. A firm is considered constrained if the costs or accessibility to external sources is an obstacle to implement investments (Kaplan and Zingales, 1997); from another viewpoint, if an unexpected increase in internal resources leads to an increase in investment, the firm is considered as financially restricted (Bond and Reenen, 2003). Empirical studies have shown that restrictions play an important role in investment decisions and that the way firm heterogeneity is introduced is a key aspect. With their seminal study, Fazzari, Hubbard and Petersen, (1988) - FHP - found that the sensitivity of investment to the availability of internal resources increased with the degree of financial constraint. Kaplan and Zingales (1997) - KZ - argued that there is no strong theoretical reason for the sensitivity of investment to cash flow to be monotonic function either of internal fund availability or of problems due to asymmetric information in capital markets.

In financial markets there are strong information asymmetries, particularly when the borrowers know better about the project than the lenders who financed it and when information is not efficiently transferred. If, on the one hand, the entrepreneurs know the quality of their own investments, the lenders do not know the differences between them. The empirical literature has appointed that it is not simple to use the cash flow coefficient to indicate if a firm is financially constrained, because the liquidity can serve as a proxy for unobserved determinants of investment and to designate a potential of future profits. In this case, the cash flow was not an appropriate variable to capture if a firm can be financially constraint. A frequent explanation of models based on asymmetric information is that the investment of firms should respond in a specific manner conditioned on liquidity. Thus, it is essential to control for the investment opportunities of the firms. Several empirical studies about investment search to control for the investment opportunities. Hoshi, Kashyap and Scharfstein, (1991) use a classification of the firms by the long-term relationship in keiretsu form and compare them with independent firms. Firms that are members of the group keiretsu can transfer information about the quality of individual projects more efficiently, hence reducing the asymmetric information. Their results show that firms with a long-term relationship are less financially constraint, while independent firms without a long-term relationship are more financially constraint.

Bond and Meghir (1994), comparing the role of cash flow in the absence and presence of funding schemes, obtained as the main result the existence of significant differences in the decisions of the investment firms when classified according to its policy of financing. The companies classified in regimes with reduced payment of dividends and emissions shares were those that showed greater sensitivity of the financial variables in relation to investment. Chirinko and Schaller (1995) examined the importance of cash flow on investment of

---

* The research reported in this paper has been supported by a grant from FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo, SP-Brazil.

1 Associate Professor of Production Engineering, University of Sao Paulo, (EESC/USP) Brazil, aquiles@sc.usp.br

2 Production Engineering, University of Sao Paulo, (EESC/USP) Brazil, camila@sc.usp.br

3 Professor of Economics, University of Sao Paulo, (FEA/USP) Brazil, cazzoni@usp.br
Canadian firms. They ranked firms according to their conditions of transmitting information in their financial positions to investors, defined as: maturity of the firm, degree of concentration of ownership and participation of its members with inter-related groups. The results showed that firms with worse conditions to transmit information showed higher coefficients of liquidity. This classification would include in the analysis the problem on the asymmetry of information and transaction costs.

Using a structural model and estimating a version of the model of Abel and Blanchard (1986), Gelichrist and Himmelberg (1995) analyzed the behavior of 428 American firms in the period from 1979 to 1989. The firms were classified by level of payment of dividends, size, access to the bond market and classification of risk of the firm by the bond rating. The result was that the variable cash flow showed greater significance for firms classified a priori as restricted financially. Hsiao, Pesaran and Tahmiscioglu (1997) grouped 561 U.S companies based on capital intensity and concluded that ignoring differences in individual behavior of firms can lead to underestimation of the estimated coefficients. They found that more-capital intensive firms have greater sensitivity of the investment to cash flow.

In an attempt to explain the contradictory results found by FHP (1988) and KZ (1997), Povel and Raith (2001) and Cleary, Povel and Raith (2005) argue that the choice of different variables able to represent constraint may lead to financial effects different on investment of the firm. The relationship between financial constraint and sensitivity of the investment to cash flow will depend on which measure is used to classify firms. If they are classified using measures of information asymmetry such as proxy for financial constraint and are not in financial distress, the sensitivity of the investment to cash flow is high and greater will be the role of market imperfections. Moreover, if the measures used in the division of the sample make a strong relationship with the net value of the firm or with internal funds, there will be an inconclusive result in the identification of the financial constraint.

To contribute to the intense debate on the role of cash flow in the decisions of investment we consider the problem of multicollinearity between variables cash flow, sales and debt. This is done considering a regression model ridge within a bayesian approach and sorting firms by capital intensity. In our knowledge of the problem multicollinearity in the model of investment decisions has not yet been considered.

### 3. Data and Model

The sample was constructed with 373 of the 500 larger firms in the period dating from 1997 to 2004 in Brazil, the equivalent to the Fortune 500 to the Brazilian context. Firms from public utility sector, as well as the ones belonging to public administrations and inconsistent data (federal, state or local) were eliminated. Following Jorgenson and Siebert (1968), we have created two groups of firms, based in intensity capital, which is measured by the ratio of capital stock to sales.6

The presence of multicollinearity among the variables cash flow, sales and debt was assumed. We estimate the standard model:7

\[
\left( \frac{I}{K} \right)_{it} = \alpha_i + \omega_j + \beta_{i,j} \left( \frac{I}{K} \right)_{i,j-1} + \beta_{s,j} \left( \frac{I}{K} \right)_{i,j-1}^2 + \beta_{c,j} \left( \frac{CF}{K} \right)_{it} + \beta_{d,j} \left( \frac{S}{K} \right)_{it} + \beta_{d,j} \left( \frac{D}{K} \right)_{it} + \epsilon_{it}
\]

(1)

where \( i \) identifies the firm, \( i = 1,\ldots,N \); \( t \) identifies the year, \( t = 1,\ldots,T \); \( j \) is the group, \( j = 1,2 \); \( \alpha_i \) is the firm-specific effect, incorporating non-observable peculiarities such as managerial capability; \( \omega_j \) is the time component, introducing all factors that vary between years but that are common to all firms; \( K_{it} \) is capital stock; \( I_{it} = K_{it} - K_{i,t-1} \) is investment; \( CF_{it} \) is cash flow; \( S_{it} \) are sales; \( D_{it} \) is debt and \( \epsilon_{it} \) is the error term.

Since all variables are divided by \( K \), investment is expressed as a rate, and the explanatory variables are expressed as ratios to capital stock. The choice of the variables is made considering the wide literature on investment decision. These studies assume the existence of a known investment function, in which the heterogeneity of the firms can be considered by inclusion of a specific effect for each firm. The cash flow variable captures the effects of possible liquidity restriction on investment, but it also represents the potential for future yields. The introduction of indebtedness captures tax effects and the fact that leverage could raise the value of firm. Several studies insist that leverage is positively related to improvements in operational efficiency. The use of lagged values for dependent variable considers the dynamic aspect of investment behavior. The quadratic variable was introduced due to indications of a non-linear in the adjustment of capital stock, reflecting a quadratic form of cost adjustment.

Panel data models have been used to estimate the parameters of models such as (1), with either fixed or random effects. Although the use of micro data in panel form allows for more information on the data, it makes firm heterogeneity an issue. In this paper we introduce a specific effect for each firm to take into consideration non-observed aspects. Another effect is introduced for non-observed temporal aspects, which would influence all firms simultaneously, but differently for distinct years.

Collinearity among the explanatory variables in model (1) could compromise the inference on the parameters of interest. Although non-orthogonal variables are very common in all regression models, usually they are not a damaging problem. In some cases, however, depending on the degree of collinearity, all inferences based on a regression model could be compromised. If the problem is serious enough, estimated coefficients will change strongly with the introduction (or drop) of additional variables or observations; some theoretically important variables could present non-significant coefficients; large values for the estimated coefficient standard deviation. The collinearity can be detected through the Conditional Number (CN), of Tolerance (TOL) and the Variance Inflation Factor (VIF). Beyond these values, as the cash flow of the firm is derived from sales, to verify the presence of collinearity problems, it is interesting to observe also the existing correlation between these variables and the regression determination coefficients when the dependent variable is regressed on the other explicative variables.

---

4 This exogenous classification made by rating agencies considers the operational and financial performance under a historical approach of the company.

7 More details in Birkes and Dodge (1993), and Farrar and Glauber (1967).
In order to clarify, $R_1^2$, $R_2^2$ and $R_3^2$ are the determination coefficients when the dependent variable is regressed on the $(CF/K)$, $(S/K)$ and $(D/K)$, respectively, and $R^2$ is the determination coefficient of the general regression, that is, when the dependent variable $(I/K)$ is regressed on all the explicative variables $(CF/K)$, $(S/K)$, $(S/K)$.

In Table 1 are presented the values found for the Conditional Number, VIF, TOL and the regression determination coefficients $(R_1^2, R_2^2, R_3^2)$:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Correlation between variables</th>
<th>Coefficient of Determination</th>
<th>VIF</th>
<th>TOL = $1/VIF$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$CF/K$</td>
<td>1</td>
<td>$R_1^2 = 0.1784$</td>
<td>1.22</td>
<td>0.8216</td>
</tr>
<tr>
<td>$S/K$</td>
<td>0.42</td>
<td>$R_2^2 = 0.2786$</td>
<td>1.39</td>
<td>0.8214</td>
</tr>
<tr>
<td>$D/K$</td>
<td>0.13</td>
<td>$R_3^2 = 0.1369$</td>
<td>1.16</td>
<td>0.8631</td>
</tr>
</tbody>
</table>

Table 1: Diagnostic of the Variables

Mean Conditional Number 2.66

It has found low values for VIF and TOL. Then, at the first time, it does not have collinearity problems between the variables that compose the considered model, given in expression (1). Despite the correlations presented in Table 1 not to be great, the existing correlation mainly between the variable $(CF/K)$ and $(S/K)$ must be considered inside of an economic context, since cash flow is derived from sales.

Another rule practical to identify collinearity is suggested by Klein (1962), pointing that the multicollinearity would be a problem only if the $R^2$ of the auxiliary regressions will be greater than the $R^2$ of the general regression. The obtained general regression determination coefficient was $R^2 = 0.0676$. In this manner, as the values obtained for $R_1^2$, $R_2^2$ and $R_3^2$ have been much greater than $R^2$ of the general regression, it will be considered the correlation among the cash flow, sales and debt of the firm in the model with regression ridge.

4. Bayesian inference and ridge regressions

The estimates of the model parameters used in this work are gotten considering a bayesian approach. Moreover, the problem of existing correlation among the cash flow and debt variables is considered through the ridge regression. Then, in this section is done a brief presentation both the main characteristics of the bayesian theory and the basic concepts for using ridge regression. The estimation and inferences on the parameters in the bayesian approach come from two sources of information: one indicating the prior knowledge represented by the probability distribution, denoted by $\pi(\theta)$, and another one representing the distribution of the contained information in the data, denoted by $f(y/\theta)$, that is, the parameters likelihood function. Through the combination of both the information we get posterior density:

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{\int f(y|\theta)\pi(\theta)\,d\theta} \propto f(y|\theta)\pi(\theta)$$

(2)

where $\pi(\theta|y)$ is posterior density, and the first term of the right side shows the product of the prior $\pi(\theta)$ for the conditioned likelihood on the data. The formula given in (2) is known as Bayes’ formula and, from the posterior distribution, it is possible to make inference without loss of information. The term $\pi(\theta)$, omitted in the second side of (2), is considered as a normalization factor, which, does not depend on $\theta$ with fixed $y$, can be considered a constant to prevent that the posterior density is improper, that is, $\int \pi(\theta)\,d\theta \neq 1$.

The posterior density calculation in many occasions can be impracticable, due to a complex process in the integrals calculation. In these cases, when the integrals can not be analytically calculated, approximations using numerical methods must be used. The posterior densities for the parameters can be calculated using algorithms of simulation for method Monte Carlo Markov Chain (MCMC). When the random vector dimension will be very great, the integration to get the marginal densities of each parameter becomes impracticable. In
this in case, an alternative is the use of the method of Gibbs Sampling, which generates samples when the posterior conditional densities present a known form. In the absence of known distributions, the Metropolis-Hasting is used. This algorithm allows the posterior densities to reach a distribution of interest or balance through the construction of a Markov Chain. The generated sample convergence can formally be verified by the method considered for Gelman and Rubin (1952). The authors considered a method of simulation based on some generated Markov Chains from several initial points.

Ridge regression was introduced by Hoerl (1962), for the resolution of a problem regarding to chemical engineering. With his experience in regression analysis, Hoerl (1962) discovered that in the presence of linear approximately relations between the explanatory variables, the estimators gotten through the squared least did not make sensible when placed in the context of the process that generated the data due to the great variances presented for them. He proposed as alternative method to the squared least, the use of the ridge regression, which presented better estimators than that gotten ones with the method of the squared least.

Considering the existence of multicollinearity, we will use a Bayesian ridge regression model. Starting with a basic regression model such as:

\[ y = \alpha + X \beta + \varepsilon \]  

(3)

where \( \alpha \) is a unknown constant, \( X \) is a known matrix \((nxp)\) of observations about the explanatory variables, \( y \) is a observations vector \((nx1)\), \( \beta \) is a vector \((px1)\) of parameters that will be estimate and \( \varepsilon \) is the random vector \((nx1)\) the errors.

Using the squares least method, \( \alpha \) and \( \beta \) are estimated for:

\[
\begin{align*}
\hat{\alpha}_{LS} &= \bar{y} \\
\hat{\beta}_{LS} &= (X'X)^{-1}X'y
\end{align*}
\]  

(4)

For using the ridge regression, firstly the matrix of explanatory variables \( X \) is replaced by some matrix of standardized variables \( Z \). The most common standardization is referent to mean and standard deviation of the variable.

Standardizing the variable \( X \), we consider \( z_i = \frac{x_i - \bar{x}}{s_x} \), where \( \bar{x} \) is the mean and \( s_x \) is the standard deviation of \( x_i \)’s.

Considering standardized variables, the model (3) becomes:

\[ y = 1\mu + Z\gamma + \varepsilon \]  

(5)

where \( Z \) is a standardized variables matrix, and \( \mu \) and \( \gamma \) are the parameters.

The multicollinearity is softened by an increase of small positive amounts in the elements of the main diagonal line of the correlation matrix.

According to Hoerl and Kennard (1970), \( X'X \) represent the explanatory variables correlation matrix. Using standardized variables, the correlation matrix becomes \( Z'Z \).

For the model given in (5), considering the ridge regression, the estimators for \( \mu \) and \( \gamma \) are defined by:

\[
\begin{align*}
\hat{\mu} &= \bar{y} \\
\hat{\gamma} &= (Z'Z + kI)^{-1}Z'y, \quad k \geq 0
\end{align*}
\]  

(6)

where

\[ k = p \frac{\hat{\sigma}_{LS}^2}{\|\hat{\gamma}_{LS}\|^2} \]  

(7)

The least squares estimator of \( \sigma^2 \) is \( \hat{\sigma}_{LS}^2 \), the estimator of \( \hat{\gamma} \) is \( \hat{\gamma}_{LS} \), \( \|\hat{\gamma}_{LS}\| \) is the norm of this estimator, and \( p \) is the number of parameters. \( k \) is a small positive increment in the correlation matrix. \( \hat{\sigma}_{LS}^2 \) and \( \hat{\gamma}_{LS} \) are the least squares estimator of \( \sigma^2 \) and \( \gamma \), respectively, \( \|\hat{\gamma}_{LS}\| \) is the corresponding norm to \( \hat{\gamma}_{LS} \) and \( p \) is the model parameter number.

The justification for using the formula (6) is that ridge regression improves the precision of the squared least estimator. Some properties of the estimator \( \hat{\gamma} \) justify its consideration as alternative for the estimator of squared least. For example, Hoerl and Kennard (1970) showed that a series of values exists for which the average quadratic error for the ridge regression is lesser than squared least one. The ridge estimator can be considered a bayesian estimative if it is assumed that \( \gamma \) is a random vector. Conditional on a fixed value of \( \gamma \), we suppose that \( y \) has a multivariate normal distribution with mean vector \( 1\mu + Z\gamma \) and covariance matrix \( \sigma^2 I \), and suppose that \( \gamma \) is a random vector with a multivariate normal distribution with mean vector \( 0 \) and covariance matrix \( \tau^2 I \).

In the bayesian approach, the prior distribution of \( \gamma \) is combined with data to produce the posterior distribution of \( \gamma \), that is, its conditional distribution given \( y \). It can be shown that the posterior mean vector of \( \gamma \) is \( (Z'Z + k\tau I)^{-1}Z'y \), where \( k^* = (\sigma^2/\tau^2)^{10} \).

---

8 The origin of this name is result of articles of Metropolis et al. (1953) and Hasting (1970).
9 This formula to calculate \( k \) was suggested by Hoerl, Kennard and Baldwin (1975). Others one has been suggested as it can be seen in Drapper and Van Nostrand (1979).
10 Lindley and Smiths, 1972.
Lindley and Smith (1972) showed that if \( y \sim N \left( X \beta, \sigma^2 I \right) \), then \( \hat{\beta} = \left( k \right) \) is the Bayes estimator, with \( k = \sigma^2 / \sigma^2_p \). Besides that, \( \hat{\beta} = \left( k \right) \) is a restricted estimative of the LS estimator, since it minimizes the sum of residuals subject to the restriction about the length of the factor of estimated coefficients. However, these properties do not explicitly define an appropriate value for \( k \) in any specific application.

The series of values for which the ridge estimator dominates the LS estimator in terms of the mean quadratic error depends on the unknown values of \( \beta \) and \( \sigma^2 \). In the Bayesian interpretation, \( k \) is a proportion of unknown variances\(^{11}\).

In order to estimate the parameters of the model in a Bayesian model, it is necessary to specify the prior distributions. We take two distributions for the error: normal and Student \( \text{t} \), with 5 degrees of freedom\(^{12}\). Although the Student \( \text{t} \) distribution is consistent to heteroskedasticity we classified the firms by capital intensity. Software Winbugs was used (Spiegelhalter, 1995); convergence was verified by the Gelman-Rubin index (Gelman and Rubin, 1952). Both random and fixed effects models were estimated. In the latter case, uniform prior distributions are considered for firm-specific and time-specific effects. Therefore, the prior distributions for the parameters are specified in a single stage. In the random effect model it is assumed that these effects have normal distribution with unknown variances, thus demanding two stages for the specification of the priors, in a hierarchical structure (Koop, 2003). In order to facilitate notation in the expressions of the probability distributions, we use:

\[
\begin{align*}
\left( \frac{I}{K} \right)_{i.t} &= y_{i.t}, \quad \left( \frac{I}{K} \right)_{i,t-1} = x_{i,\text{it}}, \quad \left( \frac{I}{K} \right)^2_{i,t-1} = x_{2\text{it}}, \quad \left( \frac{CF}{K} \right)_{it} = x_{3\text{it}}, \quad \left( \frac{S}{K} \right)_{it} = x_{4\text{it}}, \quad \left( \frac{D}{K} \right)_{it} = x_{5\text{it}} \quad (8)
\end{align*}
\]

The variables and their respective coefficients, \( \beta_{3j} \), \( \beta_{4j} \), and \( \beta_{5j} \), are standardized:

\[
\begin{align*}
\zeta_{3ji} &= \left( \frac{x_{3\text{it}} - \overline{x}_{3\text{it}}} {sd(x_{3\text{it}})} \right), \quad \zeta_{2ji} = \left( \frac{x_{2\text{it}} - \overline{x}_{2\text{it}}} {sd(x_{2\text{it}})} \right), \quad \zeta_{3ji} = \left( \frac{x_{3\text{it}} - \overline{x}_{3\text{it}}} {sd(x_{3\text{it}})} \right) \\
\beta_{3ji} &= \frac{\beta_{3j}} {sd(x_{3\text{it}})}, \quad \beta_{4ji} = \frac{\beta_{4j}} {sd(x_{4\text{it}})}, \quad \beta_{5ji} = \frac{\beta_{5j}} {sd(x_{5\text{it}})}
\end{align*}
\]

The model then becomes:

\[
y_{i.t} = \alpha_j + \omega_j x_{i,\text{it}} + \beta_{3j} \zeta_{3ji} + \beta_{4j} \zeta_{2ji} + \beta_{5j} \zeta_{3ji} + \epsilon_{i.t} \quad (10)
\]

The posterior densities for the error, as well as the parameters are defined below, for both the fixed-effect and random-effect models.

### 4.1. Fixed-effect model

For a normal distribution for the errors, the prior distributions are specified in a single stage:

\[
\begin{align*}
\alpha_i &\sim U \left( e_1, f_1 \right); \quad e_1, f_1 \text{ known, } i = 1, 2, ..., 373; \\
\omega_t &\sim U \left( e_2, f_2 \right); \quad e_2, f_2 \text{ known, } t = 1, 2, ..., 8; \\
\beta_{lj} &\sim N \left( u_{lj}, v_{lj} \right); \quad u_{lj}, v_{lj} \text{ known, } l = 1, 2, ..., 5; j = 1, 2; \\
\sigma^2 &\sim IG \left( c, d \right); \quad c, d \text{ known}
\end{align*}
\]

where \( U \left( \delta, \phi \right) \) indicates a uniform distribution, \( N \left( \mu, \sigma^2 \right) \) a normal distribution, and \( IG \left( \xi, \psi \right) \) an inverse gamma distribution.

If the prior distributions given in (11) are independent, their combination leads to the joint prior distribution of the model parameters. Combining the joint prior distribution with the likelihood function, we have the joint posterior distribution, which can be seen in Appendix.

### 4.2. Random-effect model

For a normal distribution for the errors, the prior distribution has a hierarchical structure, since they are specified in two stages:

First stage:

\[
\begin{align*}
\alpha_i &\sim N \left( g_1, \sigma^2_\alpha \right); \quad g_1 \text{ known, } i = 1, 2, ..., 373; \\
\omega_t &\sim N \left( g_2, \sigma^2_\omega \right); \quad g_2 \text{ known, } t = 1, 2, ..., 8; \\
\beta_{lj} &\sim N \left( u_{lj}, v_{lj} \right); \quad u_{lj}, v_{lj} \text{ known, } l = 1, 2, ..., 5; j = 1, 2; \\
\sigma^2 &\sim IG \left( c_1, d_1 \right); \quad c_1, d_1 \text{ known}
\end{align*}
\]

Second stage:

\[\text{...}\]

---

\(^{11}\) Many algorithms for the biased parameters have been proposed in the literature (Gibbons, 1981).

\(^{12}\) "If a normal distribution is too restrictive, you can have a more flexible distribution by taking a weighted average of more the one Normal distribution” (Koop, 2003).
\[ \sigma^2_{\alpha} \sim IG(c_1, d_1); \quad c_1, d_1 \text{ known}; \]
\[ \sigma^2_{\omega} \sim IG(c_3, d_3); \quad c_3, d_3 \text{ known}. \] (13)

Taking the distribution given in (12) and (13) as independent and combining them, the joint posterior distribution of the parameters is obtained. Again, combining the joint prior distribution with the likelihood function, we have the joint posterior distribution, which can be also seen in Appendix.

For the normal and the Student \( t \) distributions, the choice between a fixed or random-effects model is based on the posterior Conditional Predictive Ordinate (CPO)\(^{13}\), implemented by the Monte Carlo Markov Chain simulation method (Gelfand and Dey, 1994).

The Table 2 shows the found OPC values for the models:

Table 2: Indicators for model choice. CPO = Conditional Predictive Ordinate

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Student ( t )</th>
<th>Ridge Regression</th>
<th>Student ( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPO</td>
<td>2.19e+08</td>
<td>1.51e+13</td>
<td>1.06e+17</td>
<td>7.35e+21</td>
</tr>
<tr>
<td>CPO</td>
<td>3.84e+10</td>
<td>5.73e+14</td>
<td>6.20e+21</td>
<td>1.17e+21</td>
</tr>
</tbody>
</table>

Table 2 indicates that the Student \( t \) distribution fixed-effect model is the preferred one. Therefore, the remaining of the paper is based on estimations using this model.

5. Estimation and Results

In order to provide robust evidence on the role of different factors on firms’ investment decision, we have applied different estimation methods to equation (1), although we concentrate our discussion in one specification. Panel data model with fixed or random effects have been used to estimate the parameters when an unobservable is added to the model to contemplate the effect of the firm specific variables and the time constants. The choice the fixed effect model with the Student \( t \) distribution considering the multicollinearity has based on the recursive predictive criterion which provides a better prediction of the results. As the expected and more probable, the selected model is that the firm specific component can be correlated with the explanatory variables, the fixed effect.

The analyses of the results combine the estimated parameter with the financial characteristics of the firms which are classified by capital intensity. The classification of firms by capital intensity plays an important role in the results obtained in this study. We believe that this classification offers an excellent way of controlling and separating the effects of financial constraints from other factors.

Table 3 presents some descriptive statistics of the database used in this study. On average, the investment rate for firms with low capital-intensity is only 61.5% of the same rate for capital-intensive firms. For all the other 4 variables, capital-intensive firms show lower ratios; the differences are especially strong for sales and sales changes, with less capital-intensive firms presenting rations over 6 times higher than capital-intensive firms.

\(^{13}\) Other criteria are available but they present limitations for the task at hand, as shown in Kass (1992), Gelfand and Day (1994), Kass and Raftery (1995), and Weakliem (1999).
The use of external sources of funds is also higher for less capital-intensive firms, but the difference is not as large as for the other variables. As for cash flow, less capital-intensive firms present ratios 3 times higher than capital-intensive firms. It is clear that the cash flow, net income and working capital show higher values for less capital intensive firms. In this case the cash flow variable not represents a proxy for potential profit.

The higher mean value for sales for firms with lower capital intensity should be expected, since it is divided by capital stock. But it may represent a greater need of such firms to maintain a reserve of liquid assets in order to mitigate periods of high revenue volatility. The lower values for cash flow, net income, working capital and chance in working capital for more capital intensive firms could reflect greater liquidity requirements. Their liquidity coefficients could be larger, thus leading them to the considered as financially constrained.

On the other hand, if cash flow is signaling expectation of potential for future profitability, we might expect that less capital intensive firms would present a higher importance for cash flow, since are more profitable.

In general, it is clear from the data that less-capital-intensive firms, in general, rely more on sales and liquidity; while the differences in terms of external sources of funds is not as impressive. This is compatible with the findings of Scherer (1980) and Hsiao and Tahmiscioglu (1997).

The model in equation (1) was applied to the data, and the results are presented in Table 4.

As indicated by the ordinate predictive densities criterion, the fixed effects model was estimated with ridge regression with a normal and Student / distribution. For comparison purposes, the same model was estimated with a ridge regression and without it.

Cash flow is the key variable to make judgments about financial constraints. In the regression without to take into account the multicollinearity it is positive and significant for firms less capital-intensive, while for the more capital-intensive firms their coefficient is negative and significant. After correcting for multicollinearity in the ridge regression, the coefficient becomes non-significant for less capital-intensive firms, and positive and significant for capital-intensive firms. This result is in line with other studies in general and for Brazil in particular (Kalatizis, Azzoni and Achcar, 2007).

Considering the sales variable, the standard regression results indicate significant influence for both groups, and a higher importance for capital-intensive firms. The ridge regression results don’t change the sign and significance of the sales coefficient for less capital-intensive firms, but indicate a minor importance in investment decisions when their sales grow. On the other hand, for more capital-intensive firms the parameter of sale changes the signal.

Finally, the changes in the sizes of the coefficients for debt across groups also go in the same direction. For both regressions, the coefficients are significant for both groups, and larger for capital-intensive firms. In the case of the ridge regression, it is 10 times higher, but only 2 times for the standard regression.

Considering the estimated coefficients, it is clear that multicollinearity does produce a relevant effect in the estimation of parameters. Ignoring the problem would lead to conclude that capital-intensive firms are not financially constrained, and to underestimate the role of financing for these firms. It would also mislead the interpretation of the role of sales, with a very important change when multicollinearity is controlled for. This result indicates that the liquidity doesn’t actuate as a proxy for potential profit based on fact the lower capital intensive firms are more profitable. Hsiao and Tahmiscioglu (1997) report that if the liquidity affects the investment spending, then we could be expected higher coefficients for more capital-intensive firms due to high fixed costs and higher capital requirement for such companies, than less capital-intensive. According Devereux and Schiantarelli (1990), the higher cash flow coefficient for larger firms could reflect that such firms are more likely to have a relatively low cash flow and a more diversified ownership structure, which would tend increasing agency costs. For Schaller (1993), firms with a more concentrated structure of ownership is less dependent on cash flow than firms with a more diversified structure of ownership, due to the reduction in the conflict of interests between shareholders and managers of the firm, which reduced the cost of agency.

6. Conclusions

In this paper we have analyzed investment decisions of 373 large Brazilian firms in the period 1997-2004. We have investigated the role of multicollinearity among the explanatory variables in the assignment of the importance of different investment factors. We have
used a bayesian approach and have considered alternative specifications for the model, mainly, random versus fixed-effects, and normal versus Student $t$ distribution for the errors. The application of criterion for model selection has revealed that the best model is the fixed-effect with a Student $t$ distribution. The classification of firms by capital intensity plays an important role in the results obtained in this study. We believe that this classification offers a good way of controlling and separating the effects of financial constraint from other factors\textsuperscript{14}. When comparing the classification of firms in the sample used in this study by size and capital intensity, we observed that the rate of investment is more strongly correlated with the degree of capital intensity than the size of the firm. The correlation coefficient between investment rate and capital intensity is 0.0508, and between firm size and investment is 0.0069.

The main economic result of this study indicates that firms are subject to financial constraint, most notably capital-intensive firms. The ridge regression estimation revealed that capital intensive firms are financially constrained and rely more on external sources of funds for their investments. Their lower profitability rate and higher cash flow coefficients provide evidence that cash flow is not acting as a proxy for future profitability. Sales are negatively related to investment for these firms. This indicates that the variable sales withdraw importance of the cash flow variable in estimation without ridge regression. A comparison with these estimative with the ones obtained through a regression without consider multicollinearity indicates that the estimated coefficients vary importantly, which indicates that multicollinearity is indeed a problem in this kind of study.

\textsuperscript{14} Hirschman and Sirkin (1958) argue that: “The threat of obsolescence and the attractiveness of new better machines make the capitalists with expensive machinery more accumulation-minded than entrepreneurs with relatively little capital input. […] The emphasis on reinvestment can, therefore, yield an argument for investment in more capital-intensive industries than would indicate by a pure market calculation” (p. 470).
Table 3: Regression Results

- (I/K_{t-1}) is the dependent variable in the model.
- The second parameters' subscripts 1 and 2 are related to low and high capital intensity clusters, respectively. A student $t$ distribution with five degrees of freedom was used in the model.
- The symbols ***, ** and * indicate statistical significance at 10, 5 and 1%, respectively.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Without Ridge Regression</th>
<th>With Ridge Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixed Effects</td>
<td>Random Effects</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
<td>St. Dev.</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>0.189*</td>
<td>0.056</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>-0.332*</td>
<td>0.063</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>-0.042*</td>
<td>0.012</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>0.034*</td>
<td>0.009</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>0.032*</td>
<td>0.015</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>-0.14***</td>
<td>0.084</td>
</tr>
<tr>
<td>$\beta_{41}$</td>
<td>0.037*</td>
<td>0.003</td>
</tr>
<tr>
<td>$\beta_{42}$</td>
<td>0.436*</td>
<td>0.023</td>
</tr>
<tr>
<td>$\beta_{51}$</td>
<td>0.088*</td>
<td>0.012</td>
</tr>
<tr>
<td>$\beta_{52}$</td>
<td>0.179*</td>
<td>0.026</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.375</td>
<td>0.011</td>
</tr>
</tbody>
</table>
References


Appendix

Getting the joint a posterior densities for the parameter of the models is necessary to consider the expression below:
\[ \varepsilon_{it} = y_{it} - \alpha_i - \omega_l - \sum_{l=1}^{5} \beta_l y_{lit} \]

The joint a posterior densities for the parameters of the model considering normal and student t distributions are shown below:

### A.1. Model with a normal distribution

#### A.1.1. Fixed effect

\[
\pi(\alpha, \beta, \omega, \sigma^2 / y, x) \propto \prod_{i=1}^{N} \prod_{j=1}^{2} \exp \left\{ -\frac{1}{2\nu_y} \left( \beta_y - u_y \right)^2 \right\} \left( \sigma^2 \right)^{-\left(c_1+1\right)} \exp \left\{ -\frac{d_1}{\sigma^2} \right\} \left( \frac{NT}{2} \right)^{-\frac{k+1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{l=1}^{T} \left( \varepsilon_{it} \right)^2 \right\}
\]

#### A.1.2. Random-Effect

\[
\pi(\alpha, \beta, \omega, \sigma^2 / x, y) \propto \prod_{i=1}^{N} \prod_{j=1}^{2} \exp \left\{ -\frac{1}{2\nu_y} \left( \beta_y - u_y \right)^2 \right\} \left( \sigma^2 \right)^{-\left(c_1+1\right)} \exp \left\{ -\frac{d_1}{\sigma^2} \right\} \left( \frac{NT}{2} \right)^{-\frac{k+1}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{l=1}^{T} \left( \varepsilon_{it} \right)^2 \right\}
\]

### A.2. Model with a student t distribution

#### A.2.1. Fixed-Effect

\[
\pi(\alpha, \beta, \omega, \sigma^2 / y, x) \propto \prod_{i=1}^{N} \prod_{j=1}^{2} \exp \left\{ -\frac{1}{2\nu_y} \left( \beta_y - u_y \right)^2 \right\} \left( \sigma^2 \right)^{-\left(c_1+1\right)} \exp \left\{ -\frac{d_1}{\sigma^2} \right\} \prod_{i=1}^{T} \prod_{k=1}^{1} \frac{1}{\sqrt{\pi k}} \left[ \frac{1 + \left( \varepsilon_{it} \right)^2}{k\sigma^2} \right]^{\left(k+1\right)}
\]

#### A.2.2. Random-Effect

\[
\pi(\alpha, \beta, \omega, \sigma^2 / x, y) \propto \prod_{i=1}^{N} \prod_{j=1}^{2} \exp \left\{ -\frac{1}{2\nu_y} \left( \beta_y - u_y \right)^2 \right\} \left( \sigma^2 \right)^{-\left(c_1+1\right)} \exp \left\{ -\frac{d_1}{\sigma^2} \right\} \prod_{i=1}^{T} \prod_{k=1}^{1} \frac{1}{\sqrt{\pi k}} \left[ \frac{1 + \left( \varepsilon_{it} \right)^2}{k\sigma^2} \right]^{\left(k+1\right)}
\]