ERASMUS UNIVERSITEIT ROTTERDAM

Statistics Netherlands

Erasmus School of Economics

Master Thesis Econometrics & Management Science: Econometrics

Insights in the Underlying Complex Mechanisms of Dutch Companies using Agent Based Modelling

Author:
Michelle de Ruiter
Student number:
414429

Supervisor:
prof. dr. R. Paap
Second assessor:
dr. W. Wang

Abstract

This paper investigates the underlying complex mechanisms of the Dutch business sector using agent based modelling. Based on three stylized facts that have been observed in a data set of all firms in the Netherlands, the Complex Adaptive Trivial System (CATS) model by Gatti, Gallegati, Giulioni, and Palestrini (2003) is employed for the research. Several extensions are implemented in order to investigate whether the model can be improved in terms of replicating the empirical stylized facts. A calibration analysis using the Generalized Method of Simulated Moments has been performed to estimate the parameters of the original CATS model and its extended variants. Finally, an ex-post validation analysis compares the statistical characteristics of the simulated output with the empirical stylized facts. The results indicate that the CATS model is capable to reproduce all three stylized facts to a certain extent. Moreover, it is shown that some of the implemented extensions are an improvement of the original CATS model.

Keywords: Agent based modelling, Dutch firms, Stylized facts, Generalized Method of Simulated Moments, Validation

September 19, 2018

Contents

1 Introduction

In this paper the underlying complex mechanisms of Dutch firms are investigated using agent based modelling. Especially since the latest global economic crisis economists are investigating new types of modelling strategies. Mainstream economics is build upon the reductionist principle, which assumes that the economy behaves like an individual and that the behaviour of the system as a whole can be understood by simply magnifying a single individual's behaviour. However, in practice this does not always lead to the correct understanding of the aspects that cause the behaviour of the whole system. Agent based models simulate the actions of individual agents, such that the emerging patterns of behaviour that are not defined at the level of any individual agent can be assessed. In agent based modelling one starts with a simple model based on a few assumptions on the individual agents and through simulated interactions of these agents statistical emerging patterns may arise. This so-called holistic approach of modelling a complex system argues that the system as a whole is different from the sum of its components because of the interactions between the agents. Because agent based models embrace complexity by relying on simple assumptions, provide better understanding at the individual level, are easily extendable, are able to answer multiple questions across a complete system and are well suited to explain the behaviour of the whole system, they are very useful for investigation of the underlying complex mechanisms in the Dutch business sector.

Especially for policy makers it is important to understand these complex mechanisms of Dutch firms in order to provide reliable recommendations. Reactive policy deals with the major issue that it can only be determined whether the policy has been successful after it has been implemented. When it appears that the policy has different consequences than one would expect, these changes often cannot be recovered (Pijpers, 2018). Agent based models provide a solution in that we can change the assumptions beforehand, such that it can be investigated what will happen to the system as a whole before a policy change is implemented in the real world. Moreover, early warning indicators can be developed, such that it can be indicated beforehand when a sudden large change arises in the system (Buiten, de Jonge, & Pijpers, 2018). Especially for statistical institutes as Statistics Netherlands (SN), the Organization for Economic Cooperation and Development (OECD) and the Ministry of Economic Affairs and Climate Policy it is important to understand the complexity in Dutch firms to gain better insights in the range of policy options, the corresponding possibilities and the consequences of their policy recommendations. For the reasons discussed above, these institutes are highly interested in developing an agent based model to get insights in the behaviour of Dutch firms. This paper is the first that provides initial insights in the complex behaviour of the Dutch business sector using agent based modelling.

In order to determine a suitable agent based model for modelling the complex behaviour in the Dutch business sector it must first be determined which requirements have to be fulfilled. Therefore, the emerging patterns or stylized facts that are present in the behaviour of Dutch firms are investigated. A stylized fact is a generalized presentation of an economic phenomenon, which has been regularly observed in the empirical literature. Based on the stylized facts found in the Dutch business sector, the Complex Adaptive Trivial System (CATS) model first proposed by Gatti et al. (2003) is chosen as the starting point of this research. This model is chosen because it has proven to be capable of replicating several stylized facts that are also encountered in the behaviour of Dutch firms. The CATS model is widely used for modelling interacting heterogeneous agents, financial fragility and aggregate dynamics of firms. With the aim of improving the original CATS model such that it is a more realistic representation of the actual data of Dutch firms, this paper relaxes several assumptions of the original model. The relaxed assumptions involve the price, the productivity and the interest rate of firms. These variables will be modelled in a way such that it is expected that the stylized facts of the Dutch firms are replicated more accurately.

The next step in this research is to calibrate the parameters of the agent based models, such that the initial assumptions of the models match with those of the real data of Dutch firms. However, calibration of the parameters of these models is not straightforward in practice. Due to the complexities in agent based models it is very difficult to obtain an analytic solution for criterion functions, such that parameter estimation is forced to rely on simulations. In this paper, the Generalized Method of Simulated Moments is used to calibrate particular parameters of the agent based models. Followed by the recommendations of Winker, Gilli, and Jeleskovic (2007), an expression for the objective function using the covariance matrix of the empirical moments is employed. Based on the resulting values of the objective function it is also possible to gain initial insights in whether the implemented extensions result in a more realistic agent based model compared to the actual data. The final step in this research is to perform a validation analysis, such that it can be truly investigated how accurate the models represent the Dutch business sector. The stylized facts that are present in the Dutch business sector are compared to the statistical characteristics of the simulation output of the agent based models. Different statistical techniques are used for the comparison of the stylized facts, such that the performance of the CATS model and its extensions can be analyzed.

From the empirical literature it is known that firms follow several stylized facts. For example, Sutton (1997) and Dosi, Marsili, Orsenigo, and Salvatore (1995) show that the distribution of firms' size is right skewed, Stanley et al. (1996) and Bottazzi and Secchi (2003) show that Italian firms' growth rates follow a Laplace distribution, and Axtell (2001) has shown that the size of firms in the United States follow a Zipf distribution. A recent empirical study of Pijpers (2018) at Statistic Netherlands has shown that the latter stylized fact also holds true for Dutch firms. In order to model these stylized facts agent based models are becoming increasingly popular in the literature. For example, multiple researchers such as Dosi, Fagiolo, and Roventini (2010) and Assenza, Gatti, and Grazzini (2015) agree that agent based models are a worthy complement to the Dynamic Stochastic General Equilibrium model, which is used to explain economic growth, business cycles, and other economic phenomena. Fagiolo et al. (2017) and many other researchers agree that agent based modelling is becoming more attractive because, in contrast to many traditional economic modelling strategies, they provide "descriptive richness".

Multiple different agent based models have been developed to model the behaviour of the economy. For example, Catullo and Gallegati (2015) developed a multi-country agent based model based on technological change, and Caiani (2017) have build an agent based model that is based on technological innovation. There also exist very extensive agent based models, such as the one developed by Caiani, Catullo, and Gallegati (2017), who model an agent based stock flow consistent multi-country model including households, firms, banks, central banks and a government. This paper builds upon the original CATS model by Gatti et al. (2003), which in contrast to other agent based models, explicitly models the entry-exit process of firms. There already have been some investigations on this agent based model. For example, Gallegati, Giulioni, and Kichiji (2003) include a monopolistic bank in the simulated economy, and Bianchi, Cirillo, Gallegati, and Vagliasindi (2008) try to validate this model for Italian firms by setting up the model using empirical data. It is common in agent based modelling that one starts with an existing model and tries to improve it based on for example, economic theories or on the observed data set. The flexibility that these models entail is therefore one of the reasons that agent based modelling is being increasingly used in the literature.

However, as already mentioned briefly before, the flexibility of agent based models comes with a side-turn. That is, the calibration of the parameters of these models is still a hurdle in agent based modelling, even though many contributions have been made in the last decades. Gourieroux, Monfort, and Renault (1993) have introduced indirect inference, which has become the standard technique that is used for calibrating simple agent based models. Indirect inference relies on a few parameters and does not need much computation time. This technique makes use of simulation methods in order to estimate the parameters of an agent based model. Many researchers have employed this calibration technique, among whom Winker and Gilli (2004), Goldbaum and Mizrach (2008), De Jong, Verschoor, and Zwinkels (2010) and Chiarella, He, and Zwinkels (2014). Examples of simulation methods employed for indirect inference are the Simulated Minimum Distance (SMD) by Fabretti (2013), the Simulated Maximum Likelihood (SML) by Kukacka and Barunik (2017) and the Method of Simulated Moments (MSM).

The most commonly used simulation method is the MSM, where for each parameter combination the agent based model is simulated, such that the aggregate moments of the simulation output can be obtained. The distance between the aggregate simulated and empirical moments is then used to construct the objective function. The MSM finds the optimal combination of parameter values by minimizing the value of the objective function. However, there are still a few major drawbacks to this method, especially when the complexity of the model increases. The first problem is the formulation of the distance function, which rarely contains a closed-form expression due to the complexity of the agent based model. Therefore, this objective function is often subject to Monte Carlo simulations leading to increasing computation times. A second problem is that the estimation results depend on which and how many moments to include in the analysis, choices that are often arbitrary. Winker et al. (2007) recommend two general requirements for the selection of moments, such that they are robust and are able to discriminate between different agent based models. A possible solution to the above two problems is to use Bayesian methods for calibration of the parameters. Grazzini, Richiardi, and Tsionas (2017) have been the first that employed a Bayesian approach to estimate an agent based model. Although this solves the above two problems of the formulation of the objective function and the selection of moments, Lamperti, Roventini, and Sani (2018) argue that still a large number of Monte Carlo simulations have to be performed due to estimation of the likelihood function. All in all, the most problematic issue of calibrating an agent based model is the computation time, which increases substantially when the complexity of the model or the amount of parameters increases. Nowadays, efficient calibration of the initial parameters of agent based models is still an open challenge.

In the literature more attention has been paid to the validation techniques of an agent based model. Validation techniques have to be employed in order to investigate to what extent the agent based model is an accurate representation of the real data. The basis of ex post validation in agent based modelling is to investigating whether the model is able to reproduce the emerging patterns, or stylized facts of the observed data (Fagiolo et al., 2017). However, a lot of researchers are skeptic about the validation performance of agent based models because often the model does not rely on methodological standards (Richiardi, Leombruni, Saam, & Sonnessa, 2006). Besides investigating whether the model reproduces certain stylized facts, which is regarded as one of the most important model validation techniques by Ormerod and Rosewell (2009), several more advanced statistical validation techniques have been proposed. For instance, Marks (2013), Barde (2016), Barde (2017), Lamperti (2016) and Lamperti (2018) developed several similarity measures to validate and compare agent based models. Another technique of ex post validation is that of Bianchi et al. (2008), who use actual data as an initial set up of the model. Moreover, Guerini and Moneta (2017) provide several statistical measures which test the accuracy of the parameters of the simulated and the real data estimated by a Structural Vector Autoregression model. For the interested reader, Fagiolo et al. (2017) give a comprehensive overview of the current literature regarding the validation of agent based models.

This paper contributes to the current scientific literature in three different ways. The first is that agent based modelling has never been applied to model the behaviour of Dutch firms. Research has been done on companies in other countries using agent based modelling. Especially Italian firms have been investigated thoroughly with the CATS model. Therefore, it is interesting to investigate whether the CATS model is also capable of replicating the stylized facts that are observed in the Dutch business sector. If so, this strengthens the hypothesis that the CATS model is a good approximation of the behaviour of firms in general. In turn, modelling the behaviour of Dutch firms using agent based modelling gives insights in the Dutch business sector from a different perspective than with traditional economic modelling strategies.

The second way in which this paper contributes to the literature is by implementing the different extensions of the CATS model. Instead of fixed parameters, the price, productivity and interest rate variables are modelled in such a way that they are more realistic for the Dutch economy. In this way, it is expected that the CATS model is improved by being able to replicate the stylized facts observed for the Dutch firms more accurately. Moreover, because the extensions are implemented separately it is also possible to distinguish the performance of each extension. That is, it can be investigated whether each extension is a realistic assumption for the Dutch business sector. This allows for better insights in the mechanisms behind the Dutch economy.

Finally, many scientific papers that develop or apply an agent based model to explain phenomenon in their data set do not provide an explanation of the calibration procedure. In many cases the parameter values used are not provided, or they are given without any explanation. Therefore, I believe that it is of great relevance to include a detailed calibration analysis explaining the steps that are needed in order to estimate the optimal parameter values. Moreover, in this paper the parameters of an original agent based model and those of its implemented extensions are calibrated separately. In this way, it can be stated with more certainty that the additional parameters corresponding to the extended variables match accurately with the empirical data. Moreover, the resulting values of the objective function provide initial insights in whether relaxing these assumptions lead to an improvement compared to the original CATS model. All together, the agent based model, its extensions, the calibration procedure and the validation of the stylized facts are useful as a basis for further development of an agent based model that is suitable to represent the Dutch economy.

The remainder of this paper is organized as follows. Section 2 gives a brief description of the data that is used for this research and presents the stylized facts that can be observed in the Dutch business sector. In Section 3 the original CATS model representation of Gatti et al. (2003) is described, followed by the extensions that are implemented for this model. Moreover, the calibration and validation procedures are discussed. Section 4 provides a brief investigation on the calibration of the parameters and the validation of the stylized facts of the original and the adapted agent based model. Finally, in Section 5 the conclusion that can be drawn from the research including recommendations for further research are discussed. At the end of this report a list containing all symbols used in this paper is provided.

2 Data of Dutch firms

In this section the data of Dutch firms that is available for the research is briefly described and analyzed. Section 2.1 shortly describes how Statistics Netherlands has retrieved the data and gives a description of the necessary variables. Moreover, it is explained how is dealt with missing values. In Section 2.2 three stylized facts that are observed in the data set of Dutch firms are analyzed.

2.1 The data set

For this research, a panel data set provided by Statistics Netherlands (SN) is available that contains annual observations from 2010 till 2013 on 11 variables for all the firms in the Netherlands. The firms are represented with an identification number of the business unit containing 11 digits. At SN a business unit is defined as the actual actor in the production process that is characterized by autonomy, writability and external orientation. The business units are included in the General Business Register (ABR), defined as the system with a registration of identifying data in businesses, and can consist of multiple KvK-numbers (Dutch Chamber of Commerce). These KvK-numbers could be a natural person (that is, a human being or a legal persona, also called sole proprietorship) as well as a legal entity, defined as each non-natural person that is recognized as a legal subject under the law, for example a Public Limited Company, Limited Liability Company, religious denominations, ministries and municipalities. Moreover, these KvK-numbers can be linked to administrative units as, among others, the Tax Department. Sources used by the ABR are, for example, the Basic Business Register (BB), which is a partnership between the Chamber of Commerce, the Tax Department and Statistics Netherlands, or the Trade Register (NHR) of the Chamber of Commerce.

Based on these conditions, the data set from Statistics Netherlands provides information of $1,116,939$ firms in 2010, $1,197,494$ firms in 2011, $1,337,940$ firms in 2012 and $1,378,489$ firms in 2013. The different amount of firms in each year indicate that firms might go bankrupt, that new firms can arise or that companies might merge. The main focus in this research however is based on the year 2010. For the calculation of the empirical growth rates the year 2011 is also necessary. These cross-section data are chosen because a panel data set with a time horizon of four years is too short to perform accurate statistical analyses. Especially in agent based modelling this is a problem, because these models often need to simulate a certain amount of time periods before the variables start to represent the real world. Moreover, it is hard to determine what the time periods in agent based modelling exactly represent. Finally, validation analyses using, for example, econometric time series analysis cannot be performed accurately when the time horizon is too short. Therefore, this research is limited to the cross-section data set of Dutch firms in 2010.

The variables[1] in the data set that are of main importance for the research are

- **Employed persons**: Someone working for a company located in the Netherlands, or an institution or private household in the Netherlands. Persons employed are all persons having paid jobs, even for only a few hours a week and even if they: work legally as such, but without registration for income tax and social security ("undeclared work"), are temporarily not at work, but have continued receipt of wages or salary (for instance owing to illness or hold-ups due to frost), or are on a temporary unpaid-leave.
- **Business capital**: The fiscal business assets of the taxpayer at the end of the financial year.
- **Net turnover**: Business returns, excluding VAT (value added taxes) from the selling of goods and services to customers. Turnover is calculated after deduction of discounts, bonuses, returnable deposits and uncharged freight costs.

The data set provided by SN also includes information about the industrial classification code of firms (SBI), their size, legal form, fiscal profit, wages and salaries, whether firms are exporting or importing, have a foreign parent or subsidiary and which are mergers or acquisitions. For the interested reader, the description of these remaining variables can be found in Appendix A. These variables are used in this research to investigate whether they have any effect on omitting the firms that contain missing values.

About 20% of the same firms do not contain information about their business capital, tax profit, net turnover and wages and salaries. It is investigated whether there will be any effect on the distributions of the "remaining" variables when removing these firms from the data set. In doing so, the distribution of the remaining variables for all companies are plotted versus the distribution of the remaining variables without the companies that contain missing values. Substantial difference between these two distributions would indicate that there could be a particular cause for the missing values. Further investigation would suggest whether the firms with missing values can be removed from the data set. The distribution plots for the variables SBI-code, firm size, employed persons, legal form, export and import are shown in Appendix B. The distribution plots of the dummy variables indicating whether the firm has a foreign parent of subsidiary or has merged or been taken over are not shown because of their negligible contribution in the total amount of firms. In fact, only 0.270% of the firms have a foreign parent, 0.002% have a foreign subsidiary, 0.336% are mergers and 0.018% of the firms are acquisitions. The distribution plots in Appendix B show that for all considered variables the distribution of all the firms versus the distribution of the firms without the missing values are almost identical. This indicates that there are no particular missing data patterns and that the same distributional results apply when these firms are removed from the data set. After removal of the firms that contain missing values, there are 909,036 Dutch firms in 2010.

[1]The variables and their descriptions are based on the definitions provided by Statistics Netherlands.

2.2 Stylized facts

Based on several stylized facts described in the current literature, it is investigated whether these patterns also hold true for firms in the Netherlands. Not all known economic stylized facts can be considered due to limitations in the data set, i.e. not all variables are available. Moreover, stylized facts that involve dynamics are particularly hard to investigate for our data because only annual observation over a time span of four years are available. Therefore, this analysis is focused on the cross-sectional data in the year 2010. The following three stylized facts observed in the Dutch business sector are confirmed by the empirical literature.

1. The distribution of firm size follows a power law / Zipf distribution. (Gallegati et al., 2003; Axtell, 2001; Gabaix, Gopikrishnan, Plerou, & Stanley, 2003; Stanley et al., 1995; Fujiwara, Aoyama, Di Guilmi, Souma, & Gallegati, 2004).
This stylized fact has been investigated very thoroughly for business sectors in different countries, among which Italy (Gallegati et al., 2003) and the United States (Axtell, 2001). A power law behaviour in the distribution of firm size indicates that firms behave independently in a complex manner. This behaviour might be due to different phenomena present in business sectors. For example, mergers and acquisitions may lead to large jumps in the size of firms. Also, economies of scale can lead to a situation in which "the winner takes it all". Pijpers (2018) has shown that this stylized fact also holds true for the Dutch business sector by showing that the total net turnover of firms follows a power law probability distribution, which is stable over the years 2010 till 2013. In order to show that the net turnover of Dutch firms in 2010 indeed follows a power law distribution, the complementary cumulative distribution function (CCDF) of a real-valued random variable x, defined as

$$\bar{F}(x) = P(X > x) = 1 - F(x),$$

is presented in log-log form in the left plot of Figure 1. The graph clearly shows a power law behaviour, which is of the form $P(X > x) = cx^{\beta}$. The special case of $\beta = -1$ is known as the Zipf distribution (Axtell, 2001). In order to investigate whether the net turnover follows a Zipf distribution, the upper 5% of the observations are presented in the right plot of Figure 1, including the estimated regression line. By taking the logs the parameters can be estimated using a simple linear regression. The regression results are shown in Table 1. The estimated significant slope parameter equals $\hat{\beta} = -0.966$, which is very close to the value of the slope parameter of -1. Therefore, it indicates that the net turnover distribution of Dutch firms indeed follows a Zipf distribution and it can be concluded that this stylized fact holds true for the Dutch business sector.

2. Volatility of growth rates decreases when firm size increases. (Stanley et al., 1996; Gabaix et al., 2003; Gatti et al., 2003).
Multiple researchers have found that there is a particular relationship between the growth rates of firms and their size. It turns out that the growth rates of large firms have a lower variance than the growth rates of small firms. This also seem to make

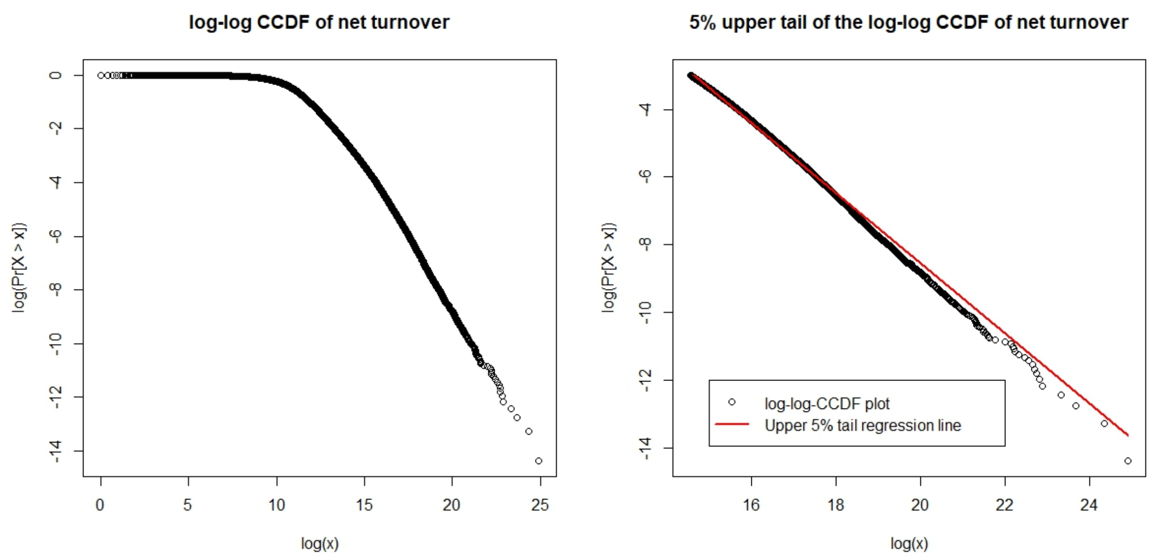

Figure 1: The left plot shows the Zipf or log-log CCDF plot of net turnover in 2010. The right plot shows the 5% upper tail of the log-log CCDF plot of net turnover including the fitted regression line.

Table 1: Regression results of the 5% upper tail distribution of the log-log CCDF of net turnover. The results are based on $909,036$ observations. (*** indicates significant at the 0.1% level.)

	Estimate	Std. Error	t-value	Pr(> |t|)
Intercept	11.745	0.001	10059	0.000***
x	-0.966	0.000	-3408	0.000***

sense because small firms are more likely to have large fluctuations in their growth rates than large firms (Coad, 2007). For Dutch firms this also appears to hold true. The growth rates of the firms are calculated as the percentage change in the business capital of all firms for the two subsequent years 2010 and 2011. The left plot in Figure 2 shows the scatter plot of the log of the growth rates of business capital of each firm of the year 2010-2011 versus the log of the number of employees. The relation between growth rates and employed persons in the years 2011-2012 and 2012-2013 are very similar and can be found in Appendix C. This indicates that this stylized fact is also stable over the years 2010-2013. It is clear that in all three subsequent years the growth rates become less volatile when the size of the firm increases.

In order to measure the exact relationship between firm size and the volatility of growth rates, the data is divided into bins of equal length. There are a total of 617769 observations after merging the firms in 2010 and 2011 and omitting the missing values. With a total amount of 1000 observations in each bin, there are 618 bins. In each bin the average of the log of employed persons (the average size) and the mean absolute deviation (MAD) of the log of the growth rates of business capital are calculated. Because there are observed a relatively small amount of outlying values in the left plot

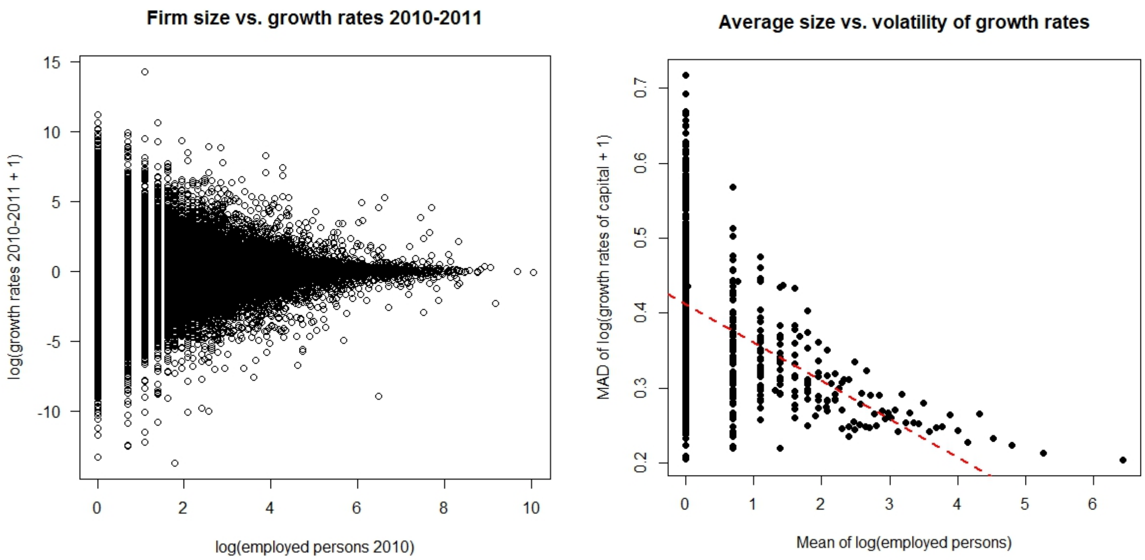

Figure 2: The left plot shows the log of number of employees of each firm in 2010 plotted against the log of the growth rates of business capital of each firm of the years 2010-2011. The right plot shows the average size of the firms in each bin plotted against the MAD of the log of the growth rates in each bin, including the estimated regression line.

of Figure 2, the MAD is chosen as a measure of volatility. This measure is a more robust one than the standard deviation, which squares the distances from the mean. When using the MAD, these relatively small number of outlying values are irrelevant. The right plot in Figure 2 plots the average size against the MAD of the growth rates. Based on the findings of Stanley et al. (1996), the regression

$$\mathbf{V} = g + h\mathbf{S} + \boldsymbol{\epsilon} \tag{1}$$

is performed, where $\mathbf{V}$ is the volatility of the log(growth rates) measured with the MAD in each bin, $\mathbf{S}$ is the average of the log(employed persons) in each bin, g and h are the intercept and slope parameters respectively, and $\boldsymbol{\epsilon}$ is the error term. The fitted regression line is shown in the right plot of Figure 2. The regression results of fitting the linear relation in (1) are presented in Table 2. The significant negative slope estimate $h = -0.051$ confirms what is observed in the left plot of Figure 2. That is, the volatility of growth rates decreases when firm size increases. Therefore, it can be concluded that this stylized fact holds for the Dutch business sector.

Table 2: Results of regressing the MAD of growth rates on the average size of firms in each bin. The results are based on 618 observations. (*** indicates significant at the 0.1% level.)

	Estimate	Std. Error	t value	Pr(> |t—)
Intercept	0.412	0.004	100.170	0.000 ***
S	-0.051	0.004	-13.660	0.000 ***

3. Growth rates follow a Laplace distribution. (Bottazzi & Secchi, 2003; Stanley et al., 1996; Bianchi et al., 2008).
Finally, a widely confirmed empirical result in the literature is that the growth rates of firms follow a "tent-shaped", or Laplace distribution. The probability density function of the Laplace distribution is

$$f(x|\mu, b) = \frac{1}{2b}\exp(-\frac{|x-\mu|}{b}), \tag{2}$$

where x is a random variable, μ is the location parameter and b is the scale parameter. The parameters of the Laplace distribution are estimated with maximum likelihood by taking the sample median as an estimate of the location parameter and the median absolute deviation as an estimate of the scale parameter.

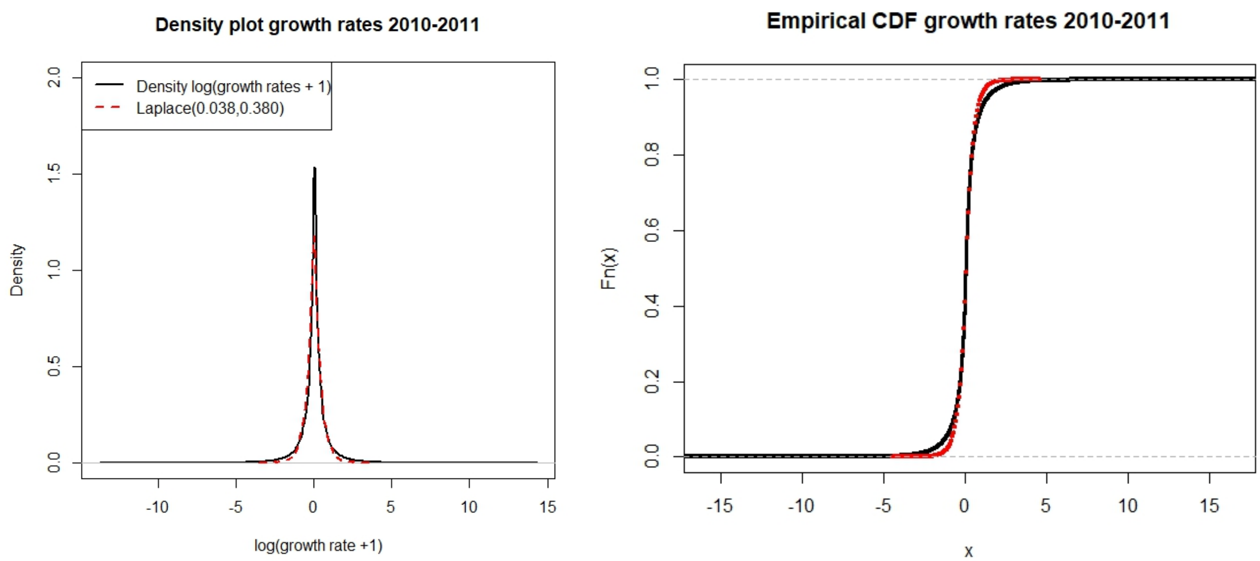

Figure 3: The left plot shows the density of the growth rates of business capital in the years 2010-2011 together with its estimated Laplace distribution. The right plot shows the empirical CDF plots of the growth rates of business capital in the years 2010-2011 (solid black line) versus its estimated Laplace distribution (dotted red line).

The left plot in Figure 3 shows the density function of the growth rates of business capital in the years 2010-2011 together with its estimated Laplace distribution. The estimated scale and location parameter are $\hat{\mu} = 0.038$ and $\hat{b} = 0.380$ respectively. The density plot shows that the observations in the sample are very concentrated around zero, leading to a distribution that is tent-shaped. This indicates that the growth rates might follow a Laplace distribution. However, it can be observed that the empirical density function shows a higher peak and slightly smaller tails than the estimated Laplace distribution. The same is observed for the growth rates in the years 2011-2012 and 2012-2013, of which the density plots are presented in Appendix D. The right plot in Figure 3 presents the cumulative distribution function (CDF) of the log growth rates in 2010-2011 and the CDF of its estimated Laplace distribution. The plots show a clear overlap, except for the tails. That is, the tails of the empirical distribution function are much wider than those of the Laplace distribution.

A statistical test is performed to test whether the empirical growth rates and its estimated Laplace distribution come from the same distribution. The two-sample Kolmogorov-Smirnof (KS) test is a nonparametric test that can be used for comparing two distributions. The KS test statistic is

$$D_{n,m} = \sup|F_{1,n}(x) - F_{2,m}(x)|, \tag{3}$$

where sup is the supremum function, $F_{1,n}$ and $F_{2,m}$ are the empirical distribution functions of the two samples and n and m are the sample sizes of the two distributions. The null hypothesis states that two samples are exactly from the same distribution. The null hypothesis is rejected if

$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{nm}},$$

where in general

$$c(\alpha) = \sqrt{-\frac{1}{2}ln(\frac{\alpha}{2})}$$

with α the significance level.

The first sample $F_{1,n}$ represents the empirical distribution function of the growth rates of business capital, with $n = 617,769$ observations. The second sample $F_{2,m}$ is the fitted Laplace distribution with the estimated location $\hat{\mu} = 0.038$ and scale parameter $\hat{b} = 0.380$ of the empirical growth rates, with $m = 10,000$ observations. The estimated Laplace parameters and the KS test statistics for the growth rates of the years 2010-2011 can be found in Table 3. The statistics for the years 2011-2012 and 2012-2013 are very similar and therefore presented in Appendix D for the interested reader.

Table 3: The Kolmogorov-Smirnof test statistic of comparing the empirical growth rate distribution in the years 2010-2011 with its estimated Laplace distribution. (*** indicates significant at the 0.1% level.)

n	m	$D_{n,m}$	p-value
617,769	10,000	0.037	0.000***

It can be observed from Table 3 that the KS test statistic $D_{n,m}$ is significant at the 0.1% level. This means that the test rejects the null hypotheses of equal distributions. This is not in line with the impressions in Figure 3, which showed similar density and CDF plots. These rejections are in fact due to the definition of the null hypothesis, stating that two samples are drawn from exactly the same distribution. Having such a large amount of sample observations n will therefore almost always lead to a rejection of the KS test. However, according to the distribution plots the empirical data of growth rates seem to match very well with the estimated distributions. Therefore, I assume that this final stylized fact also holds for the Dutch business sector. However, it has to be kept in mind that the tails of the empirical distribution deviate from its estimated Laplace distribution.

3 Methodology

In this section the methods and techniques that are necessary for the research are described. In Section 3.1 the main features of the CATS model of Gatti et al. (2003) are described. Section 3.2 discusses some extensions of this model in order to make it more realistic regarding the empirical data set. Section 3.3 describes the procedure of the Generalized Method of Simulated Moments to calibrate the initial parameters of the agent based model. Finally, in Section 3.4 the simulation outcomes of the model are validated by comparing them with the stylized facts observed in the Dutch business sector.

3.1 The CATS model

The starting point of modelling the underlying complex mechanisms in the Dutch business sector is the CATS model developed by Gatti et al. (2003). This model has been used in multiple studies, such as in Gallegati et al. (2003), Delli Gatti et al. (2004), Gatti et al. (2005), to accurately replicate various stylized facts present in multiple countries' business sectors. Bianchi et al. (2008) have shown that to a certain extent the CATS model is able to replicate the same three stylized facts for Italian firms that also occur in Dutch business sector. The stylized facts that are observed in the Italian economy are that the distribution of Italian firms' size can be described by a power law distribution, the growth rates follow a Laplace distribution and that there exists a negative relationship between the variance of the growth rates and the size of firms. Except for the exact values of the statistical characteristics, the stylized facts that have been observed in the behaviour of Dutch firms, which is discussed in Section 2.2, are the same as those encountered in the Italian economy. Therefore, the CATS model is chosen to model the emerging patterns of the Dutch business sector. The model is built from a financial point of view, that is, financial robustness of firms plays a crucial role. The model differentiates itself from others by modelling bankruptcies and the creation of firms as two different processes. This way of modelling heterogeneous agents is much more realistic than keeping the number of companies constant over time. These industrial dynamics are modelled such that it influences the financial robustness of companies, which in turn affects aggregate macroeconomic variables.

Gatti et al. (2003) state several basic assumptions in order to limit the complexity of the model. These assumptions include that firms use a constant return to scale technology, which means that production increases at the same proportional rate as capital, where capital is the only input variable of this model. This assumption includes that firms produce homogeneous goods with a constant productivity of capital. This means that the products of firms compete only by means of their price, and the amount of products available. The price is therefore chosen as a random variable with expected value equal to the market price. Moreover, firms can obtain credit in order to produce their output. The model assumes that firms have no limitations in obtaining external finance, which can only be supplied from banks at a uniform real interest rate. The expected profit that firms make is therefore equal to the expected revenue (the output multiplied by the price of a product) minus their dept commitments (the interest rate multiplied by the amount of capital stock). Other costs that firms have to deal with are capital adjustment costs, the costs that go along with altering the level of capital stock, and bankruptcy costs. The equity ratio, defined as the ratio of equity base to capital stock, is the variable that is used to proxy the financial robustness of firms. Firms may go bankrupt, which happens if the net worth of the company becomes negative. Finally, new firms can enter in the economy. It is assumed that the amount of entrant firms depends partly on a stochastic process, that is, a normal distribution, and partly on the performance of the market, which is approximated by the number of surviving firms.

In the remainder of this subsection the theoretical properties of the CATS model are described. The model specification in this subsection, including all assumptions and formulas, are developed by Gatti et al. (2003). They start with an environment that consists of N_t firms at any time period $t = 1, ..., T$. Because the only input that firms have is capital, the production function can be described as

$$Y_{it} = vK_{it}, \quad i = 1, ..., N_t, \quad t = 1, ..., T, \tag{4}$$

where Y_{it} is the output (production) of firm i at time period t, K_{it} is the capital stock of firm i at time period t and v is the constant productivity of capital. The price P_{it} at which firms sell their products is a random variable with expected value equal to the market price P_t, such that the relative price of each firm at each time period is defined as $u_{it} = P_{it}/P_t$. Therefore, the relative price is a random variable and, in the simulations, drawn from a uniform distribution. The revenue of a firm can be calculated as the relative price multiplied by the output of the firm, that is, $R_{it} = u_{it}Y_{it}$.

The goal of each firm is to maximize their expected profit, that is,

$$\max \mathrm{E}(\pi_t) = \mathrm{E}(R_{it} - rK_{it}), \tag{5}$$

where r is the constant real interest rate. However, not only dept commitments rK_{it} have to be paid, the firms also incorporate capital adjustment costs, defined as

$$CA_{it} = \frac{\gamma}{2}\frac{(K_{it} - K_{i,t-1})^2}{K_t}, \tag{6}$$

where K_t is the aggregate capital stock in time period t and $\gamma > 0$ is a constant parameter. Moreover, firms incur bankruptcy costs

$$CB_{it} = (\alpha_1 - \alpha_2 a_{i,t-1})Y_{it}, \tag{7}$$

where $a_{i,t-1}$ is the equity ratio one period ago and α_1 and α_2 are positive parameters. Firms differentiate from each other by their equity ratio, which measures the proportion of the amount of equity available to finance the firms' capital stock. The equity ratio can be defined as

$$a_{it} = A_{it}/K_{it}, \tag{8}$$

where A_{it} is the equity base of firm i at time period t. At time period 0, the amount of equity base is set equal for all firms. The model assumes that firms go bankrupt when their relative price u_{it} is lower than a threshold $\bar{u}_{it}$, such that the probability of bankruptcy increases with the interest rate r and the capital stock K_{it} and decreases with the amount of equity base $A_{i,t-1}$ one time period ago. This means that firms go bankrupt if

$$u_{it} < \frac{r}{v} - \frac{A_{i,t-1}}{vK_{it}} \equiv \bar{u}_{it}, \tag{9}$$

where v is a constant productivity factor. This allows us to define the objective function of firms as

$$\max\ E(\pi_{it}) - CA_{it} - CB_{it}Pr[u_{it} < \bar{u}_{it}], \tag{10}$$

where $Pr[u_{it} < \bar{u}_{it}]$ is the probability of going bankrupt. Gatti et al. (2003) solve this maximization problem and find expressions for the optimal rate of capital accumulation, that is,

$$\tau_{it} := \frac{K_{it} - K_{i,t-1}}{K_{it}} = \frac{1}{\gamma}[v - \rho(r, a_{i,t-1})], \tag{11}$$

where

$$\rho(r, a_{i,t-1}) := r\left(1 + \frac{\alpha_1}{2} - \frac{\alpha_2}{2}a_{i,t-1}\right) \tag{12}$$

is the bankruptcy cost augmented interest rate, which is different across firms through their equity ratio leading to a positive relationship between the optimal rate of capital accumulation in (11) and the financial fragility of a firm. Moreover, the maximization problem in (10) leads to the so-called law of motion of the equity base, that is,

$$A_{it} = A_{i,t-1} + \pi_{it} - CA_{it} \tag{13}$$

$$= A_{i,t-1} + u_{it}Y_{it} - rK_{it} - \frac{\gamma}{2}\frac{(K_{it} - K_{i,t-1})^2}{K_t}. \tag{14}$$

This means that the equity base is updated according to the equity base in the previous period plus the revenues and the dept commitments less the capital adjustment costs. Finally, the law of motion of the equity ratio can be obtained as

$$a_{it} = a_{i,t-1}(1 - \tau_{it}) + u_{it}v - r - \frac{\gamma}{2}\tau_{it}^2. \tag{15}$$

Averaging the equity ratio in (15) over all firms gives the aggregate equity ratio a_t, which depends on the variance of the aggregate equity ratio one period ago. These expressions are derived in Gatti et al. (2003) and are used for description of the moments of the distribution of the equity ratio. Changing the parameters cause a change in the distribution and therefore, this allows for calibration of the model described in Section 3.3.

Agent based modelling is relatively new in the scientific literature and the way in which these models are implemented it is not always clearly defined. This makes it hard to exactly replicate existing agent based models. In order to give a more clear overview of the way in which the agent based model is simulated, Figure 4 presents a flow chart of the CATS model. First, the initial variables and parameters must be initialized. Next, for all firms the necessary variables are updated and it is determined whether a firm goes bankrupt or not. If the firm survives, its capital stock and equity base are updated according to (11) and (13). Based on the surviving firms the number of entrants can be calculated, of which the equity base and capital stock are initialized. This process is repeated until the maximum number of time periods is reached.

3.2 Extending the CATS model

The ultimate goal whereof this paper is trying to contribute to is understanding how the mechanisms in the Dutch business sector exactly work. Although (agent based) models are only a simple representation of the real world and are most likely not able to capture all of the complexities in an economy, they serve as a useful benchmark for explaining various economic mechanisms. The CATS model has proven to be able to reproduce several stylized facts that are also encountered in real economies (Gallegati et al., 2003; Delli Gatti et al., 2004; Gatti et al., 2005). However, it would be of interest to investigate whether relaxing certain assumptions of the standard CATS model by Gatti et al. (2003) would lead to a more realistic model that is even better able to explain the behaviour of Dutch firms. If so, this will indicate whether these extended mechanisms are present in the Dutch business sector.

The assumptions that are relaxed involve the variables price, productivity and the interest rate, which are assumed to be constant values for all firms and all time periods in the standard CATS model. In the extended model, these variables are made dependent on the size of a firm in a certain time period. The exact implementation of these extensions are based on either scientific literature or experts' opinions. One aspect that must be noticed is that firm size is defined differently in the real data set as in the agent based model. The size of Dutch firms is measured by the number of employed persons, while in the CATS model firm size is based on the equity ratio. In the extended model firm size is divided into three categories; small, medium and large, where small firms are assigned to a lower equity ratio than large firms. However, these different definitions of firm size will not necessarily cause problems when validating the model. That is, if the agent based model is capable of reproducing the stylized facts present in the Dutch business sector, this indicates that the different definitions of firm size probably do not have impact on the validation outcomes. Moreover, this might indicate that the size definition of the agent based model corresponds to the empirical definition in a sense that the number of employees in Dutch firms is, among others, dependent on the equity ratio. However, this relationship can not be investigated because there is no information available about the equity ratio of Dutch firms. Anyhow, although firm size is defined differently in the extensions than in the real world, this will most likely not have impact on the ultimate goal of these extensions, which is to make the model more realistic. In turn, it is expected that the agent based model is able to reproduce

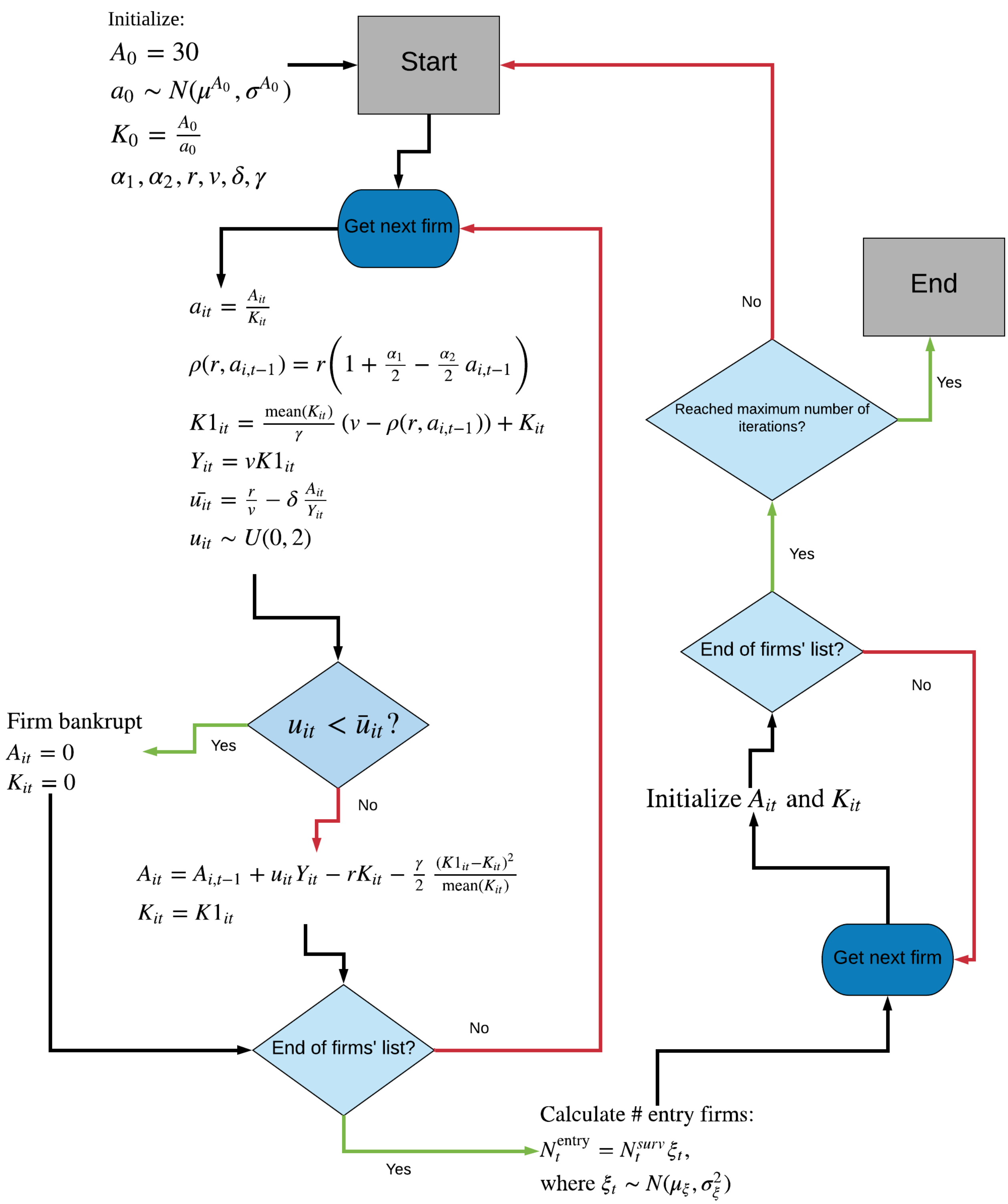

Figure 4: Flow chart of the CATS model.

the stylized facts in Section 2.2 more accurately.

The first extension is based on that of Bianchi et al. (2008), who adapt the original CATS model to be more realistic for their data set of Italian firms. Instead of equal levels of risk as in Gatti et al. (2003), they assume different risk levels in terms of firms' price, depending on the size of the firm. While Bianchi et al. (2008) discriminate only between small and large firms, this paper discriminates between three categories of firms' size. Based on the implementation of Bianchi et al. (2008) the price level u_{it} is generated according to three different random processes, that is,

$$u_{it} \sim \begin{cases} U(\mu^s, \sigma^s) & \text{if the firm is small,} \\ U(\mu^m, \sigma^m) & \text{if the firm is medium,} \\ U(\mu^l, \sigma^l) & \text{if the firm is large,} \end{cases} \tag{16}$$

where μ^S and σ^S, $S \in s, m, l$ are the mean and the variance of the uniform distributions. Following the recommendation of Bianchi et al. (2008) the average price levels are larger for small firms than for large firms, that is, $\mu^s > \mu^m > \mu^l$. Moreover, the volatility of the price level is smaller for large firms than for small firms, that is, $\sigma^s > \sigma^m > \sigma^l$. The latter assumption is similar to the stylized fact that is observed in the Dutch business sector explained in Section 2.2, where the volatility of growth rates decreases with firms' size.

Another extension that is implemented to the original CATS model by Bianchi et al. (2008) is that firms' productivity is different across firms and time periods. They achieve this by making the productivity a nonlinear function of its past productivity. This is different from the model described in Section 3.1, which assumes a constant value for all firms in all time periods. Based on the implementation of Bianchi et al. (2008), in this paper the productivity ϕ_{it} of firm i at time period t is calculated as

$$\phi_{it} = \begin{cases} \phi_{i,t-1} + \frac{M}{x}\phi^z_{i,t-1} & \text{if the firm is small,} \\ \phi_{i,t-1} + \frac{M}{x+y}\phi^z_{i,t-1} & \text{if the firm is medium,} \\ \phi_{i1} & \text{if the firm is large,} \end{cases} \tag{17}$$

where x, y and z are positive parameters and $M \sim U(\mu^{\text{prod}}, \sigma^{\text{prod}})$. From (17) it is clear that the productivity of small firms is in general more volatile than that of medium firms, which is in turn more volatile than the constant productivity of large firms. This implementation is chosen because it is in line with the second stylized fact observed in the Dutch business sector and therefore, extending the model in this way promises more realistic outcomes. The dynamic process of the productivity can be explained by stylized facts observed in other countries. Unfortunately, our data set does not contain a reasonable amount of time periods to investigate whether this dynamic relationship holds true. However, based on the current empirical literature (Bianchi et al., 2008) and economic sense, it is a realistic assumption to implement. To strengthen this statement the productivity is allowed to decrease over time, which is in contrast to Bianchi et al. (2008), who generate the productivity as an increasing function for small firms.

Finally, Gatti et al. (2003) assume that firms can finance all the production they want by obtaining as much financial credit they need at a constant real interest rate. A more realistic assumption for modelling the Dutch business sector would be to make the amount of credit available a function of firms' size, and thus of the equity ratio. Because the equity ratio can be seen as a measure of financial robustness, or the health of a firm, it would make sense to make the interest rate dependent of firms' equity ratio. This means that banks provide a lower interest rate to firms with a higher equity ratio because these firms are more likely to fulfill their dept commitments. In contrast, higher interest rates are assigned to firms with lower equity ratios. Summarizing, instead of a constant real interest rate r for all firms, it is calculated as

$$r_{it} = \begin{cases} r + \zeta_1 & \text{if the firm is small,} \\ r & \text{if the firm is medium,} \\ r - \zeta_2 & \text{if the firm is large,} \end{cases} \tag{18}$$

where ζ_1 and ζ_2 are positive parameters.

It is expected that implementing these extensions in the original model leads to a model that is more realistic and therefore, that these extended models are able to replicate the emerging patterns more accurately. However, this will also increase the complexity of the model and more importantly, it will increase the amount of additional parameters that need to be estimated. (Implementing the extensions leads to an additional six parameters for the price extension, six for the productivity extension, and two for the interest rate extension.) This causes problems when one wants to calibrate the model, discussed in the next subsection.

3.3 Calibration

Ex ante validation or calibration is a necessary step to find the optimal parameters of an agent based model. The initial parameter values should be chosen such that they represent the individual behaviour of the agents. However, finding the optimal combination of parameter values is not that straightforward in practice. A standard way of finding optimal parameters in a model would be to perform maximum likelihood estimation and do a grid search over the parameter values until the likelihood function is maximized. However, finding a closed-form expression of the likelihood function is a problem in agent based models because of the complexities, such as nonlinearities, that are incorporated in the model. Therefore, calibration of the parameters in an agent based model is often based on simulation techniques. The standard calibration technique used for agent based modelling is called indirect inference. This paper will employ the indirect inference technique called the Method of Simulated Moments (MSM) introduced by McFadden (1989). In this method the initial parameters are calibrated by matching the aggregate moments resulting from the agent based model with the empirical moments of the data set of Dutch firms. The optimal combination of parameter values is the combination that results in the minimal distance between the aggregate simulated moments and the empirical moments.

However, there are several drawbacks to this calibration procedure. The first is that the method is only practically applicable if the complexity of the model and the amount of parameters is not too large. Otherwise, the computation time will increase substantially because in order to find the optimal parameter vector, one needs to perform multiple Monte Carlo simulations of the agent based model, each time for different combinations of the parameter values. Therefore, only a part of the vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is estimated, that is, $\boldsymbol{\theta}_1$, and the other parameters $\boldsymbol{\theta}_2$ are set fixed a priori. Another drawback of the Method of Simulated Moments is that there is in general no analytic solution for the objective function, and is therefore dependent on numerical approximations. In this paper, the methodology of Winker et al. (2007) is followed, who set up the objective function using the Generalized Method of Moments. Moreover, the ad hoc selection of the sample moments is a problem. Different choices for the sample moments lead to different optimal parameter vectors. Following Winker et al. (2007), the sample moments can be chosen depending on which general statistical characteristics one may want to replicate with the agent based model. For example, if one wants to reproduce a distribution containing heavy tails, the kurtosis is a straightforward choice for a sample moment statistic. Other choices that need to be made when estimating the parameters in an agent based model using the Method of Simulated Moments are the amount of moment statistics to include and the number of simulations one wants to perform. Both choices involve a trade-off between accuracy and computing time.

The remainder of this subsection describes how the agent based models are calibrated using the Generalized Method of Simulated Moments, introduced by Winker et al. (2007). Suppose the true vector of moments is $\mathbf{m} = (m_1, ..., m_k)$. However, the true moments are unknown and can be estimated from the observed data sample $\mathbf{X}$ by using a bootstrap technique. The estimated vector of bootstrapped moments is denoted by $\mathbf{m}^e = (m_1^e, ..., m_k^e)$. Because the data set provided by Statistics Netherlands provides a very large amount of observations, one would expect that there is barely uncertainty in the estimated moments of the empirical data, and therefore, that using a bootstrap technique is unnecessary. This is true for first and second order moments, such as the mean and the variance. For example, the relative error in estimating the mean is roughly $\frac{1}{\sqrt{N}}$, where N is the sample size. For higher-order moments the estimation uncertainty also decreases when the sample size N increases, but to a much slower degree. This is due to the fact that higher-order moments require more degrees of freedom. Therefore, for higher-order moments a larger sample size is required in order to obtain an estimate that is of similar quality. For the reasons that will be described in the following Section 4.1, the calibration procedure is also based on higher-order moments. If this uncertainty of the higher-order moments is ignored, it is impossible to make realistic comparisons between the empirical moments and the simulated moments. Therefore, it is necessary to use bootstrapping for estimation of the empirical moments.

The vector of true moments can also be estimated from the simulated data generated by the agent-based model, which gives the simulated vector of moments $\mathbf{m}_i^s(\boldsymbol{\theta})$ for each replication $i = 1, ..., I$ for a given vector of parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_l)$. Following the introduction of the method of moments by Stern (2000), the standard method of moments condition is then equal to

$$\mathrm{E}[\mathbf{m}^s|\boldsymbol{\theta}] = \mathbf{m}. \tag{19}$$

The idea of the Method of Simulated Moments is to approximate the solution to the expression in (19) by replacing the expected value by the average over the simulated moments. That is, the method of moment aims to find the optimal parameter vector $\hat{\boldsymbol{\theta}}$, such that

$$\frac{1}{I}\sum_{i=1}^{I}[(\mathbf{m}_i^s)|\boldsymbol{\theta} - \mathbf{m}] = \mathbf{0}. \tag{20}$$

As mentioned before a part of the vector of parameters is set fixed a priori and the other part is calibrated. Therefore, we need a sufficient number of moments (Winker et al., 2007). When the number of moment conditions k is larger than the amount of parameters l, the Generalized Method of Simulated Moments is a natural choice for optimizing the parameter vector. This method finds the optimal parameter vector $\hat{\boldsymbol{\theta}}$ by minimizing the weighted sum of squares

$$\frac{1}{I}\mathbf{G}_I^{'}\mathbf{W}\mathbf{G}_I, \tag{21}$$

where

$$\mathbf{G}_I(\boldsymbol{\theta}) = \sum_{i=1}^{I}[(\mathbf{m}_i^s)|\boldsymbol{\theta} - \mathbf{m}] \tag{22}$$

and $\mathbf{W}$ is a $k \times k$ positive definite weighting matrix.

The uncertainty due to the estimation of the simulated moments $\mathbf{m}^s$ resulting from the agent based model is captured by the moment condition $\mathbf{G}_I(\boldsymbol{\theta})$ in (22) and can be reduced by increasing the number of simulations I of the agent based model. The uncertainty due to the estimation of the bootstrapped moments $\mathbf{m}^e$ also needs to be taken into account, and this is captured in the weighting matrix $\mathbf{W}$. According to Heij et al. (2004) the weighting matrix $\mathbf{W}$ would be estimated most efficiently when it is chosen equal to the inverse covariance matrix of the moment condition, that is, $\mathrm{cov}^{-1}(\mathbf{G}_I(\boldsymbol{\theta}))$. However, $\boldsymbol{\theta}$ is the quantity that is unknown and has to be estimated, and because $\mathbf{G}_I(\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$, the optimal choice for $\mathbf{W}$ is not possible. Therefore, this paper follows the expression of the objective function introduced by Winker et al. (2007), where the weighting matrix $\mathbf{W}$ in (21) is estimated by the inverse of the covariance matrix of the estimated bootstrap moments, that is, $\mathbf{W} = \mathrm{Var}^{-1}(\mathbf{m}^e) \equiv \boldsymbol{\Sigma}^{-1}$, where $\boldsymbol{\Sigma}$ can be estimated from the bootstrap distribution of $\mathbf{m}$. This expression for the weighting matrix allows us to take into account the uncertainty in the estimated moments $\mathbf{m}^e$. Moreover, taking into account the correlations between the bootstrapped moments is a realistic choice, because it is highly unlikely that the bootstrapped moments are independent as they are statistics of the same distribution. Finally, the simulation results in Winker et al. (2007) have shown that this expression of the objective function is able to find a good approximation of the optimal parameter combination in a large parameter subspace. Winker et al. (2007) do recommend to improve this objective function in further research, however this has not yet been investigated so far.

Using the definitions stated above, the objective function can be defined as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \frac{1}{I}\hat{\mathbf{G}}_I^{'}\hat{\mathbf{W}}\hat{\mathbf{G}}_I \\ &= \Big[\frac{1}{I}\sum_{i=1}^{I}[(\mathbf{m}_i^s)|\boldsymbol{\theta} - \bar{\mathbf{m}}^e]\Big]^{'}\hat{\mathbf{W}}(\mathbf{m}^e)\Big[\frac{1}{I}\sum_{i=1}^{I}[(\mathbf{m}_i^s)|\boldsymbol{\theta} - \bar{\mathbf{m}}^e]\Big].\end{aligned} \tag{23}$$

The objective function in (23) is calculated for different parameter settings and the parameter values that correspond to the lowest value of the objective function are optimal. Algorithm 1 summarizes the calibration procedure using the Generalized Method of Simulated Moments for two parameters α_1 and α_2, while keeping the other parameters $\boldsymbol{\theta}_2$ fixed. One important requirement for this calibration procedure to make sense is that the observed data sample $\mathbf{X}$ and the simulated data sample $\mathbf{X}^s$ have the same economic interpretation.

Algorithm 1: Calibration of the ABM

Input : The observed data sample $\mathbf{X}$
Output: The optimal parameter estimates α_1^* and α_2^*.

```
for b in 1 : B do
    Take a sample with replacement from X;
    Calculate the moments m_b^e of this sample;
end
Calculate the average of the moments m̄^e = 1/B Σ_{b=1}^B m_b^e;
Calculate weigthing matrix
Ŵ(m^e) ≡ Σ̂^{-1} = 1/B Σ_{b=1}^B (m_{s,b}^e − m̄^e)(m_{j,b}^e − m̄^e), s, j = 1, ..., k;

Determine the subspaces Θ_1 and Θ_2 for the parameters α_1 and α_2;
for α_1 ∈ Θ_1 do
    for α_2 ∈ Θ_2 do
        Set θ = (α_1, α_2, θ_2);
        for i = 1 : I do
            Run the model and obtain the simulated data set X^s;
            Calculate the simulated moments m_i^s(θ);
        end
        Calculate estimate of the objective function L(θ);
    end
end
Determine the optimal parameter estimates: (α_1^*, α_2^*, θ_2) = min L(θ).
```

3.4 Validation

A necessary step to investigate whether the simulated output of the agent based models represents the emerging patterns of the Dutch business sector is to perform ex-post validation techniques. The most straightforward output validation technique is to see whether the agent based model has correctly replicated the stylized facts that are observed in the real world data. Although this is often already accounted for in the design of the model, because most agent based models are partly based on replicating these stylized facts, replicating various stylized facts as observed in the Dutch business sector is not that straightforward. The CATS model is especially chosen for analyzing the complex mechanisms in Dutch firms because it is shown in different researches, such as those of Bianchi et al. (2008) and Bianchi, Cirillo, Gallegati, and Vagliasindi (2007), that the CATS model is able to replicate several stylized facts that are present in the Italian business sector. Using minor adaptions of the model they have shown that the CATS model is capable of reproducing some of the emerging patterns present in the empirical data, with a few exceptions. However, Bianchi et al. (2008) recommends to further investigate whether the CATS model is able to replicate the stylized facts by among other things improving the specification of the model and better calibrating one or more of the initial parameter values. Therefore, performing the validation procedure by comparing the stylized facts that are present in the Dutch business sector with the simulated output of the CATS model is interesting for an additional three reasons. The main reason is that this research is the first in investigating this agent based model for Dutch firms. It has been shown that the CATS model is able to capture the regularities observed for Italian firms, but it is unknown whether it is also capable of reproducing that of the Dutch business sector. The second reason is that several adaptions have been implemented regarding the way the price, productivity and the interest rate are modelled. Using the validation techniques it can be explored whether these specifications of the model are an improvement. Moreover, by comparing the stylized facts resulting from the agent based models with the different extensions it is possible to compare the implemented extensions with each other. In this way, conclusions can be drawn about which model performs best in terms of replicating the stylized facts. This also allows for better insights in the mechanisms behind the Dutch business sector. For example, if the CATS model with the price extensions replicates the stylized facts of Dutch firms better than the original CATS model does, this indicates that the way in which the price is generated as in Section 3.2 is a more realistic choice for the Dutch economy. The final reason is that in this paper a calibration study for several key parameters is performed. According to the recommendations of Bianchi et al. (2008) it is therefore expected that the results of comparing the stylized facts will be promising. However, this of course needs to be investigated in further detail. Therefore, the three stylized facts discussed in Section 2.2 are compared to the simulated output of the CATS model and its extended price, productivity and interest rate variants after the calibration analysis.

4 Results

In this section the results of the analyses are briefly discussed. Section 4.1 presents the results of the calibration procedure to estimate certain initial parameters of the CATS model and its extended variants. In Section 4.2 a validation procedure is performed by comparing the stylized facts of the Dutch business sector to the simulation outcomes of the agent based models, which are initialized with the optimal calibrated parameters.

4.1 Calibration

The original and the extended CATS models are calibrated based on the distribution of the growth rates of business capital. This means that the parameters of the models are calibrated such that the statistical characteristics of the distribution of simulated growth rates match best with the characteristics of the Laplace distribution of the observed data of Dutch firms. This stylized fact of Laplace distributed growth rates is chosen for the calibration analysis because the growth rates of capital can be calculated for the CATS model as well as the data of Dutch firms. The other stylized facts explained in Section 2.2 are left out of the analysis in order to limit the complexity and the computation time of the calibration algorithm described in Section 3.3. More importantly, the other stylized facts are not taken into account because they involve the size of firms, which is not defined in the same way for the empirical data as for the simulated data. As explained in Section 3.2 the CATS model defines size in terms of the equity ratio of a firm, while in the observed data set the size of firms is defined as the total number of employees. Although these two measures of firm size could be compared to get an idea of whether the agent based model is able to replicate the stylized facts, for the sake of accuracy (and simplicity) the calibration analysis is restricted to the stylized fact that can be compared by definition. Moreover, due to the fact that the data of Dutch firms is limited to a time horizon of four years, which is too short to use for performing time series analysis, this paper restricts the calibration procedure to the growth rates of capital observed for each firm in the year 2010. In order to compare this cross-section with the output of the agent based model, only one time period is used from the simulation output, after running the complete model. The time periods 100 and 101 are chosen for calculating the simulated growth rates of business capital for each firm. This relatively early time period is chosen because with particular combinations of the parameters all firms in the model went bankrupt after a considerable short amount of time periods, making it impossible to calculate the objective function. Other "early" time periods have been checked and this did not result in major differences with regard to the calibration outcomes.

The selection of moments used to calibrate the agent based model can be regarded as a study on its own. In this paper, the criteria for selecting the moments by Winker et al. (2007) is followed. The most important aspect when selecting the moments is the eventual goal of the analysis. The theoretical criteria of Winker et al. (2007) include that the moments should be robust and that the moments should be able to discriminate between different models and different parameter settings. The robustness criteria is checked by taking $10,000$ bootstrap samples from the actual data and comparing the statistics of the moments of the first 5000 samples with those of the second 5000 samples. When the relative difference between these sample moments is small enough, the moments can be considered as robust. According to Winker et al. (2007), the second criteria depends on three factors. The first factor is the variance of the estimated moments of the observed data, which can be taken into account by considering the bootstrap variance of the estimated moments of the real data. This bootstrap procedure is implemented in the calibration algorithm described in Section 3.3. The second factor is the uncertainty of the expected estimated moments of the simulated data. This uncertainty can be accounted for by increasing the number of simulations of the agent based model. However, when the complexity of the agent based model increases, the computation time of one simulation will increase too. Because this procedure has to be replicated for each parameter combination, the number of simulations cannot be increased to a very large extent. Therefore, one has to make a trade-off between computing time and accuracy. The third factor is the identification of the model parameters given the moments, which is not straightforward in agent based models because of their complexity. Therefore, Winker et al. (2007) advise to include a large amount of moments such that the probability of identification of the model increases.

For this calibration study, the goal is to find optimal parameters such that the characteristics of the distribution of growth rates are most accurately replicated by the agent based model. Taking this goal and the above theoretical criteria into account, the moments that are chosen for this calibration analysis are the mean, the standard deviation, the median, the mean absolute deviation (MAD), the skewness and the kurtosis of the growth rates of capital. The mean and the standard deviation are standard statistical measures for describing the shape of a distribution. The median and the MAD are the maximum likelihood parameters of the Laplace distribution. The skewness and kurtosis are chosen because they describe the shape of the distribution in terms of symmetry and "tailedness" respectively. The robustness criterion stated above is met for all moments. Other statistics have been considered, such as the Kolmogorov-Smirnof statistic. However, implementing this moment in the algorithm described in Section 3.3 caused the computing time to increase substantially[2] and is therefore not selected. All the simulations in this section are performed with $B = 10,000$ bootstraps with a sample size equal to 1000, which is equal to the number of initial firms in the agent based models. The number of repetitions I varies among the different calibrations, depending on the size of the parameter space that is searched over. The number of moments is equal to six, which is always larger than the number of parameters that will be estimated in the following models. Therefore, for all calibration procedures the Generalized Method of Simulated Moments discussed in Section 3.3 can be applied.

[2] Including the KS-statistic in the bootstrap procedure takes more than two hours of computing time. Running this procedure for different parameter combinations and each simulation of the model would increase the computation time dramatically. Therefore, I believe that including this moment is not worth the additional accuracy that would be obtained.

4.1.1 The standard CATS model

The optimal parameter estimates for the CATS model have been obtained using the calibration algorithm described in Section 3.3. A necessary condition for identification of the parameters is that the number of moment conditions is larger or equal to the number of parameters to be estimated. Because there are six moments chosen to calibrate the parameters on, a maximum of six parameters can be subject to calibration in order to be identified. In principle, all of the initial parameters of the CATS model can be chosen to be subject to estimation, that is, the choice of which parameters to estimate is rather arbitrary. This paper chooses to restrict the calibration analysis to the estimation of two parameters, α_1 and α_2, which are both used for calculating the bankruptcy costs in the CATS model. In this paper the calibration analysis of the standard CATS model is restricted to two parameters due to the computation time. If the maximum number of six parameters would have to be estimated, the agent based model would have to be simulated for I repetitions for each parameter combination. This would extremely increase the computation time. Moreover, a full calibration procedure is not the aim of this research and could be considered for further research.

Table 4: Parameter settings for the original CATS model.

Notation	Interpretation	Value
N_0	Initial number of agents	1000
C	Maximum number of agents in economy	2000
T	Number of time periods	500
μ^{e}	Mean of the entry process distribution	0.122
σ^{e}	Variance of the entry process distribution	0.03
A_0	Initial equity base	30
μ^{A_0}	Mean of the initial equity ratio distribution	0.5
σ^{A_0}	Variance of the initial equity ratio distribution	0.1
μ^{P}	Mean of the price generating distribution	0
σ^{P}	Variance of the price generating distribution	2
v	Productivity of capital	0.1
γ	Parameter for the capital adjustment costs	0.1
α_1	Parameter for the bankruptcy costs	$\mathbf{\Theta}_1 = [0.98, ..., 1.03]$
α_2	Parameter for the bankruptcy costs	$\mathbf{\Theta}_2 = [1.97, ..., 2.02]$

The remaining parameters included in the model can be found in Table 4. The values of these parameters are chosen equal to that in the research of Gatti et al. (2003) for Italian firms. Choosing these values of the initial parameters can be justified by the fact that Bianchi et al. (2008) showed that the CATS model is able to reproduce the stylized facts observed in the Italian economy that are also encountered for Dutch firms. It must be noted that the maximum number of agents in the economy is set equal to 2000 for computational reasons. When increasing the number of agents the computation time of simulating the model once increases substantially. Therefore, setting the maximum number of agents equal to the empirical sample size leads to a calibration procedure that is unfeasible. Moreover, although in the real world there exist of course much more companies, the simulation results of the CATS model in Bianchi et al. (2008) has proven that this agent based model is capable of reproducing several stylized facts using a limited number of interacting heterogeneous agents.

The left plot in Figure 5 shows the values of the objective function for the different parameter combinations of α_1 and α_2, performed with $I = 50$ simulations of the model[3]. The parameter range that is searched over is around the parameter settings of α_1 and α_2 of Gatti et al. (2003), equalling $\alpha_1 = 1$ and $\alpha_2 = 2$. The exact range is shown in Table 4. From the left plot it can be seen that the objective function is minimal for values of α_1 between 0.98 and 0.99 and for values of α_2 between 2.00 and 2.02. The lowest value of the objective function is equal to 359.69, which indicates that the hypothesis of the model producing the same statistical moments of the simulated growth rates as the bootstrap distribution of $\mathcal{L}$ is rejected. This conclusion can be drawn because the objective function in (23) is approximately χ^2 distributed normalized with I degrees of freedom, where I is the number of simulations. This χ^2 distribution holds approximately because the objective function in (23) is an expression with the sum of squares of multiple moments, divided by the uncertainty measure $\hat{\mathbf{W}}$. The expected value of a standardized χ^2 distribution is equal to 1 and even though the exact critical values (weakly) depends on the degrees of freedom, values of the objective function with a magnitude of 300 will certainly lead to a rejection of the hypothesis that the model is an accurate representation of the real data.

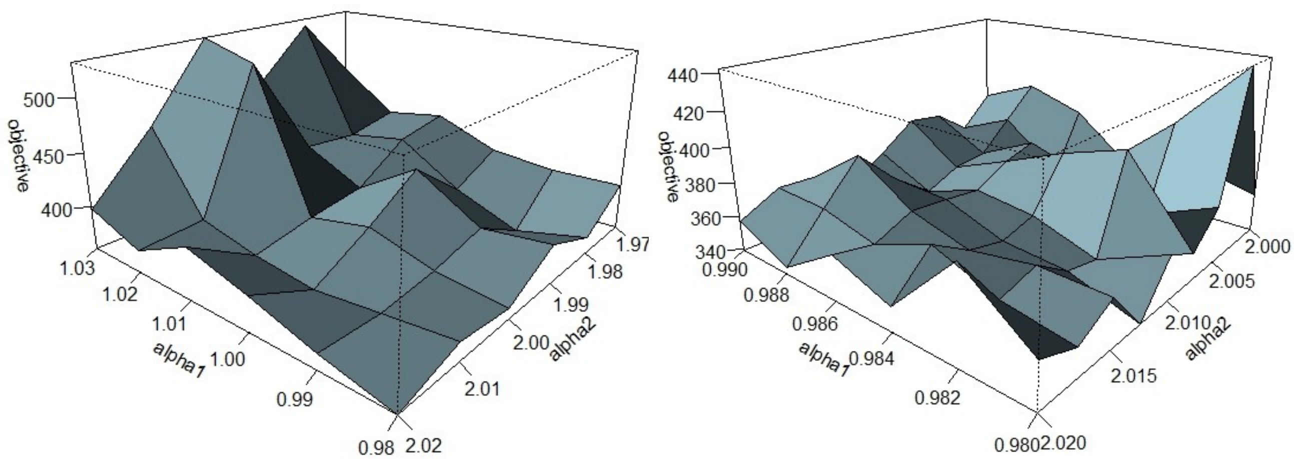

Figure 5: The values of the objective function for different parameter combinations of α_1 and α_2.

The right plot in Figure 5 zooms in to the optimal subspaces of α_1 and α_2 to obtain more accurate optimal parameter values. The calibration analysis has been performed on the parameter subspaces $\mathbf{\Theta}_1 \in [0.980, 0.982, ..., 0.988, , 0.99]$ and $\mathbf{\Theta}_2 \in$

[3]Running the algorithm took approximately 3,7 hours of computing time with an Intel(R) Core(TM) i7-2760QM CPU @ 2.40GHz x64-processor.

$[2.0000, 2.0025, ..., 2.0175, 2.02]$. For these parameter combinations there does not seem to be a clear pattern indicating which range of values minimize the objective function. This could indicate that with this accuracy of the parameter values, the objective function has reached a limit where it can not decrease any further. The optimal parameters, with an objective value of 339.70, are equal to $\alpha_1 = 0.9800$ and $\alpha_2 = 2.0125$. Although this value of the objective function is still rejected, it has improved compared to the optimal value obtained with the previous range of parameter values. However, other parameter settings or choice of moments might lead to different calibration outcomes.

4.1.2 The extended CATS model

Using the optimal parameter values of α_1 and α_2 obtained from the calibration results above, the extensions described in Section 3.2 are implemented in the CATS model and calibrated on the growth rates of business capital of Dutch firms. The extensions are calibrated separately because that allows us to investigate whether an individual extension improves the CATS model and therefore, whether it is beneficial to implement it in the model. Moreover, it allows us to compare the extensions with each other such that insights can be obtained about which extension performs best. Besides, performing a calibration procedure of all the extensions implemented at the same time increases the parameter space one has to search over substantially. Although calibrating all extension in one might lead to different results than calibrating them one by one, this large parameter space results in an analysis that is regarded as too time consuming for this research, and therefore might be subject to further research.

The price extension Calibration of the additional parameters in the price extension in (16) discussed in Section 3.2 would lead to a 6-dimensional parameter search due to the mean and variance required by the uniform distribution of small, medium and large firms. Although the restriction that the average price and volatility of the price are decreasing in firm size has been imposed, which leads to a substantial decrease of the parameter space, the computation time of searching through all of these parameter values is still regarded as too time-consuming. Therefore, in this analysis the parameters of the price distribution of large firms are set fixed to the parameter values used in Gatti et al. (2003) for Italian firms, that is, $\mu^l = 0$ and $\sigma^l = 2$, and the calibration analysis is limited to the parameters of the distribution for small and medium firms.

Because the CATS model is very sensitive to changes in the parameter values as in Table 4, first a rough sensitivity analysis has been performed in order to obtain an indication of the boundaries of the parameter values of μ^s, μ^m, σ^s and σ^m, such that the model does not break down to a situation in which all the firms go bankrupt. This analysis is performed by running the CATS model for different values of the parameters and observe the behavior of the aggregate output in each time period. When it was found that at a particular value of one of the parameters the aggregate output decreased in a relatively short time period to zero and continued to be zero, indicating that there are no surviving firms nor entrants in the market, these parameters are not included in

Table 5: Search space of the parameters of the price distributions.

Parameter	Search space
μ^s	[0.04,0.06,0.08,0.1,0.12]
μ^m	[0.02,0.04,0.06,0.08,0.1]
μ^l	0
σ^s	[2.05,2.075,2.10,2.125,2.150]
σ^m	[2.025,2.05,2.075,2.10,2.125]
σ^l	2

the parameter search of the calibration procedure. From this analysis it turned out that the mean of the uniform distribution is bounded between 0 and 0.14 and the variance is bounded between 1.96 and 2.18. Therefore, the parameter search of μ^s, μ^m, σ^s and σ^m is restricted to lie in these intervals. Table 5 shows the parameter values that have been searched through in the calibration analysis, where the parameters are restricted to be decreasing in firm size, that is, $\mu^s > \mu^m > \mu^l$ and $\sigma^s > \sigma^m > \sigma^l$, leading to a total of 225 parameter combinations.

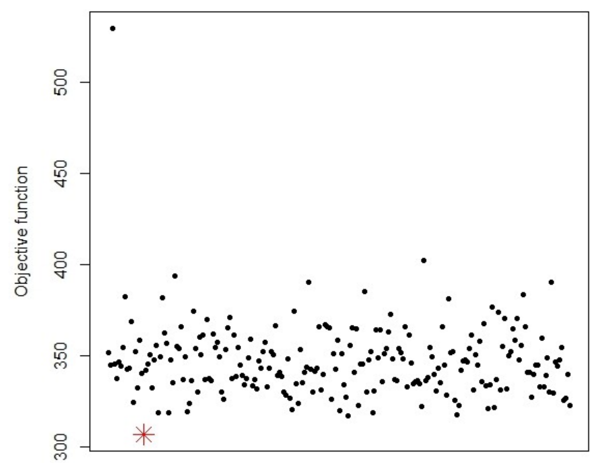

Figure 6: Values of the objective function for different parameter combinations of μ^s, μ^m, σ^s and σ^m. The red star indicates the minimum value of the objective function.

The calibration analysis has been performed in a similar way as described in Algorithm 1 in Section 3.3, only in this case for a 4-dimensional parameter space. Running the algorithm for $I = 30$ simulations results in 225 values of the objective function, each for a different combination of the parameters of the price distribution[4]. The result of the calibration analysis is visualized in Figure 6, where the points represent the values of the objective function, each for a different combination of the parameter values μ^s, μ^m, σ^s and σ^m. The x-axis therefore does not have any meaning, instead the graph is a tool for visualizing the values of the objective function. All values of the objective

[4]Running this analysis took approximately 23 hours.

function are between 300 and 400, except for one outlying value of $529, 9$, which belongs to the parameter combination $\mu^s = 0.040$, $\mu^m = 0.020$, $\sigma^s = 2.100$ and $\sigma^m = 2.025$. This outlier might be due to different causes. It could be the case that this particular combination of the parameter values leads indeed to relatively worse simulated moments compared to the bootstrapped moments, indicating that these parameter values are less suitable for replicating the stylized facts of Laplace distributed growth rates of business capital than the other parameter combinations. However, it is more likely that this outlying value of the objective function is due to remaining Monte-Carlo sampling variance (Winker et al., 2007). This variance can be reduced by increasing the number of simulations I. Due to the time consuming procedure this is left for further research.

The minimum value of the objective function is equal to 306.7, indicated with a red star in Figure 6. The optimal parameter combination that results in this minimum value is $\mu^s = 0.060$, $\mu^m = 0.020$, $\sigma^s = 2.075$ and $\sigma^m = 2.050$. This minimum value of the objective function is lower than the minimum value obtained by the calibration procedure of the standard CATS model with the parameters α_1 and α_2. Although this value still rejects the hypothesis of equal simulated and bootstrapped moments, it indicates that the price extension has lead to an improvement of the CATS model by better reproducing the statistical characteristics of the distribution of growth rates compared to the growth rates distribution of Dutch firms.

The productivity extension The nonlinear dynamic relationship of productivity discussed in Section 3.2 is based on the implementation by Bianchi et al. (2008) of productivity in the CATS models. They aim at validating the CATS model for Italian firms by discriminating between small and large firms, where for small firms the productivity is modelled as $\phi_{it} = \phi_{i,t-1} + \frac{M}{2}\phi^2_{i,t-1}$, where $M \sim U(0, 1)$ and for large firms productivity is equal to the productivity at the beginning of the time horizon, that is, $\phi_{it} = \phi_{i1}$. In this paper, this implementation is extended for small, medium and large firms, where small and medium firms are distinguished by their volatility of productivity by the division factors x and $x + y$ in (17) respectively. Therefore, these parameters are subject to calibration. Moreover, the value of the power z has to be estimated. Based on a rough sensitivity analysis performed in a similar way as explained above for the price extension, the parameter space of z is limited to the values 2, 3 and 4. For values of the power z greater than 4 the change in productivity becomes so small that it is regarded as negligible. For the sake of computation time the other parameters of the productivity extension in (17) are set fixed a priori. That is, productivity at time period $t = 1$ is set equal to the original parameter settings of the CATS model, shown in Table 4, such that $\phi_{i1} = v := 0.1 \ \forall \ i$, $i = 1, ..., N$. The parameters μ^{prod} and σ^{prod} of the uniform distribution determining the stochastic component in the productivity process, are set equal to -1 and 1 respectively. Recommended by expert opinions at SN[5] these values are chosen instead of those of Bianchi et al. (2008), such that the

[5]Thanks to Gert Buiten, Frank Pijpers, Edwin de Jonge, Ron Vellekoop, Sjoerd Hooijmaaijers and Rico Konen.

productivity is able to decrease over time, which is more realistic than a continuously increasing production function.

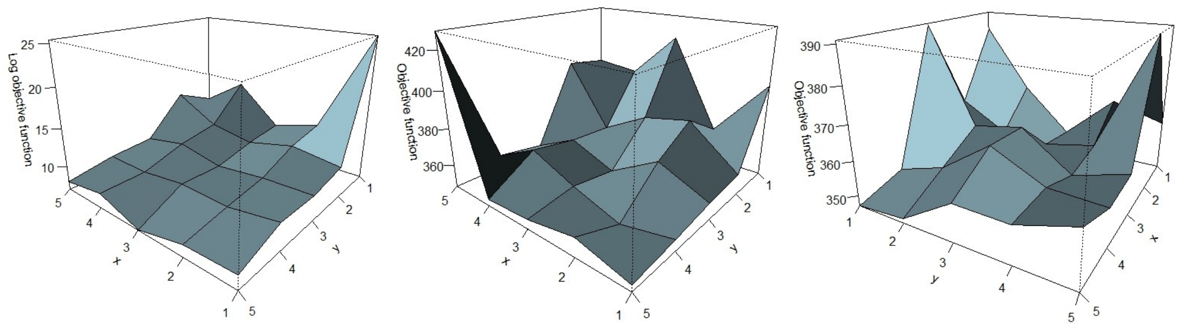

Figure 7: The objective function for different parameter combinations of x and y in the productivity equation. The left plot is the result of the calibration procedure with $z = 2$, the middle plot with $z = 3$ and the right plot with $z = 4$.

The parameter subspaces of x and y are chosen equal to $x, y \in [1, 2, 3, 4, 5]$, such that there are 25 parameter combinations for each value of the power z, investigated using $I = 30$ simulations[6]. The values for the parameter search of x and y are chosen such that the average change in volatility between small and medium firms can vary. For example, small firms can be twice as volatile as medium firms, but also five times as volatile. The left plot in Figure 7 shows the objective values for the different parameter combinations of x and y for $z = 2$. This quadratic form of the production function leads to extremely large values, including outlying values of the objective function, and therefore the log of the objective values is plotted. The minimum value of the objective for the case that $z = 2$ is equal to 1053.31, which is three times as large as that of the original CATS model with optimal parameters α_1 and α_2. This proves that the model is very sensitive to changes in the original parameter settings of Table 4 and indicates that the simulation outcomes are not in line with the empirical data when the productivity deviates too much across firms and over time. The middle and the right plot in Figure 7 show the objective function resulted from the parameter combinations of x and y with $z = 3$ and $z = 4$ respectively. In this case, the deviations from the original parameter setting are smaller, resulting in values of the objective function with a magnitude of 300. When analyzing the middle plot in Figure 7 where $z = 3$ it can be observed that the objective function is especially smallest for larger values of x, whereas in the right plot where $z = 4$ the objective seems to be minimal for smaller values of x and larger values of y. However, the objective values are very similar which makes it hard to distinguish a clear pattern of which combination of parameters performs best. In the case where $z = 3$ the minimal value of the objective function is equal to 347.92 at the parameter combination $x = 4$ and $y = 2$, whereas in the case where $z = 4$ the combination $x = 1$ and $y = 5$ results in the optimal value of the objective equal to 347.08. This indicates that the latter combination leads to the most optimal result,

[6] The total computation time was approximately 3.4 hours.

although the difference between the optimal objectives is very small. Because the two values of the minimal objective functions lie so close together, also the average over all the objective values is calculated for the case where $z = 3$ and where $z = 4$. This results in an average of 370.03 for the case where $z = 3$ and an average of 361.57 where $z = 4$. Therefore, it can be concluded that a power of 4 in the productivity equation leads to more similar simulated moments of the empirical data than a power of 3, although they both reject the agent based model.

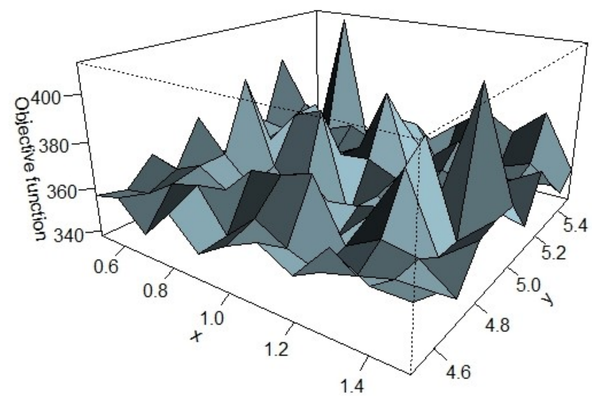

Figure 8: The objective function for the parameter combinations $x \in [0.5, 0.6, ..., 1.5]$ and $y \in [4.5, 4.6, ..., 5.5]$ for the case where $z = 4$ in the productivity equation.

Based on the previous findings that the optimal objective function is found at the parameter combinations $x = 1$, $y = 5$ and $z = 4$, the calibration procedure has been repeated for parameter subspaces around these values, such that x ranges from 0.5 to 1.5 and y ranges from 4.5 to 5.5 in steps of 0.1. This results in 121 parameter combinations, each investigated using $I = 30$ simulations of the model[7]. The resulting values of the objective function are presented in Figure 8. It can be noticed from the plot that all values of the objective function are within a range of approximately 340 to 400, indicating that there are no extreme values resulting from these particular parameter combinations. Moreover, there can not be observed a clear pattern which indicates whether large or small values of the parameter values specifically result in low values of the objective function. The optimal objective is equal to 337.93 corresponding to the optimal calibrated parameters $x = 1.0$ and $y = 5.0$. These optimal parameter values were also found in the previous calibration analysis shown in the right plot of Figure 7. Apparently, the optimal parameter values had already been found and increasing the accuracy of the parameter values did not result in any other conclusions. The optimal value of the objective function that has been found is now equal to 337.93 instead of 347.08 found in the previous analysis. This difference in optimal objective values can be explained by the Monte Carlo variance, which can be avoided by increasing the number of simulations I of the agent based model, explained at the beginning of this Section.

[7] The computation time was approximately 4.5 hours.

The interest rate extension Finally, the interest rate extension is implemented. The analysis of this extension includes two cases, one where the change in interest rate is equal for small firms and large firms, that is, $\zeta \equiv \zeta_1 = \zeta_2$ and the other allows for different deviations of the interest rate for small and large firms. The standard interest rate r charged to medium sized firms is chosen equal to the original parameter setting of 0.1 as in Table 4. The first case where the deviations from the standard interest rate are restricted to be equal is rather used as a sensitive analysis to indicate what range of parameter values minimizes the objective function than that it is used as a true calibration analysis with the aim of finding the optimal parameters. From this result the second case where the parameter values are not restricted can be investigated in further detail.

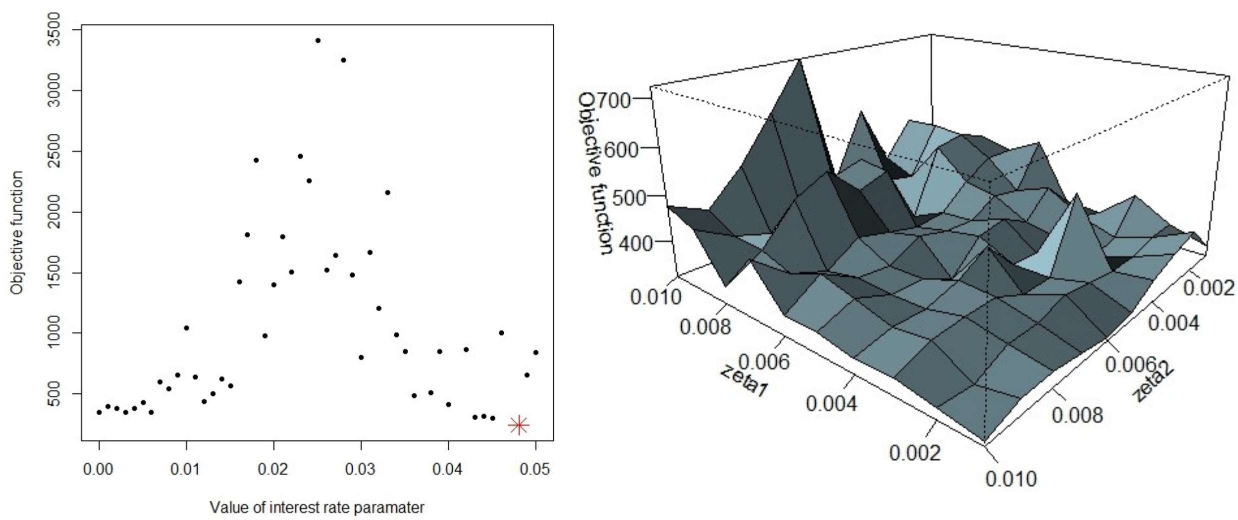

Figure 9: The values of the objective function for different values of the change in the interest rate. The red star indicates the minimum value of the objective function.

The left plot in Figure 9 shows the values of the objective function for different parameter values of ζ, ranging from 0.00 to 0.05 in steps of 0.001. Running the calibration algorithm for 51 parameters $I = 30$ simulations resulted in the objective values shown in Figure 9. There were observed three outliers that are removed from the plot in order to better visualize the remaining observations. The removed outliers were values of the objective function ranging from approximately $15,000$ to $70,000$ and occurred at values of ζ equal to 0.037, 0.041 and 0.0047. The outlying values are most likely due to the simulation uncertainty and not to the particular parameter combinations that cause the simulated moments to be relatively very different from the bootstrapped moments. This can be concluded because the other parameter values result in an objective value that stays within a particular range. In order to investigate if this is true one should increase the number of simulations, which is left for further research due to the increasing computation time.

From the plot it can be observed that the objective function with values of the interest rate between 0.01 and 0.04 is extremely large. For values of the interest rate between 0.00 and 0.01 the objective function is relatively small, ranging from 0 to 500. The minimum value of the objective function is found at $\zeta = 0.048$ with a value of 245.34 and it is indicated with a red star in Figure 9. However, when considering all the values of the objective function it is not straightforward that the parameters values around this optimal value of $\zeta = 0.048$ are indeed leading to small objective values. In contrast, it seems more likely that the parameters values between 0.00 and 0.01 lead in general to the small values of the objective function. Further investigation shows that the average value of the objective function for $0.00 < \zeta < 0.01$ is equal to approximately 500, whereas the average for $0.040 < \zeta < 0.05$ equals approximately 580, which is indeed larger.

For the second case where ζ_1 is allowed to be different from ζ_2 a similar calibration procedure has been performed. Using the same range of parameter values as in the left plot of Figure 9 in steps of 0.01, resulting in 25 parameter combinations of ζ_1 and ζ_2, $I = 30$ simulations have been performed. However, the values of the objective function resulting from these Monte-Carlo simulations resulted in extremely large values. This further strengthens the prognosis stated above that this range of parameter values is too wide for obtaining simulated moments that correspond relatively well with the bootstrapped moments. Therefore, the calibration procedure has been performed again with the parameters ζ_1 and ζ_2, now ranging from 0.001 to 0.01 in steps of 0.001. This results in 100 parameter combinations, each leading to a value of the objective function using $I = 30$ simulations of the model[8]. The values of the objective function are visualized in the right plot of Figure 9. For these range of values of the adjusted interest rate the objective functions stays within the "acceptable" bounds, similar to those obtained with the original CATS model. Further investigation of this plot shows that the objective is smallest for rather small values of ζ_1. The minimum value of the objective function results from the parameter combination $\zeta_1 = 0.005$ and $\zeta_2 = 0.006$ equalling 318.61. This value is smaller than that of the original CATS model, indicating that the interest extension is indeed useful to implement in order to obtain more accurate simulation results in comparison with the real data of Dutch firms.

Summarizing, the original CATS model is calibrated for the parameters α_1 and α_2 on the growth rates of business capital. Using the calibrated parameters of the original CATS model, the extensions discussed in Section 3.2 have been implemented and their associated parameters have been calibrated. An overview of the results of the calibration procedures is presented in Table 6, which includes the model that has been used in the simulations, the associated parameters that have been calibrated, the optimal parameter values resulting from the calibration procedure and the corresponding minimized objective function. The optimal objective values are all of magnitude 300, which clearly leads to a rejection of the hypothesis that the model is a good representation

[8]The computing time for each calibration procedure was approximately 2.5 hours.

of the real data. It is not surprising that a simple agent based model like the CATS model is not able to exactly model the entire Dutch business sector with all complexities that come with it. Although the proposed extensions do not change this result, their value of the objective function indicates that they are an improvement of the original CATS model. Modelling the price as three different stochastic distributions instead of one, modelling the interest rate dependent of firm size instead of keeping it constant across firms results and modelling the productivity as a nonlinear dynamic relationship instead of a constant value all result in lower optimal objective functions than in the original CATS model. This means that these extensions are beneficial to implement in the CATS model. Comparing the objective values of the extensions separately indicates that the model with the price extension outperforms the other models, whereas the productivity extension results in the highest value of the objective function.

Table 6: Overview of the calibration results.

Model	Calibrated parameter	Optimal parameter	Optimal objective
Original CATS	α_1	0.980	339.70
	α_2	2.0125	
Price extension	μ^s	0.060	306.70
	μ^m	0.020	
	σ^s	2.075	
	σ^m	2.050	
Productivity extension	x	1.0	337.93
	y	5.0	
	z	4.0	
Interest rate extension	ζ_1	0.005	318.61
	ζ_2	0.006	

However, the exact value of the objective function is subject to remaining Monte Carlo sampling variance, the choice of moments and the number of moments and therefore it should be mentioned that it is hard to state that the above conclusions apply in general. This is especially true because the optimal objective values do not differ much across the models, in particular the comparison of the original CATS model and the productivity extension. Therefore, in order to further investigate whether these conclusions hold in general one should perform multiple calibration studies for different number and choice of moments, and increase the number of simulations. However, this highly increases the computation time, which is exactly the problem of calibration in agent based modelling.

4.2 Validation

In this paragraph the simulation outcomes of the original CATS model and its extensions are analyzed by means of the stylized facts observed in the Dutch business sector in Section 2.2. This allows for investigating whether the CATS model is able to replicate the stylized facts, but also to test whether the extended variants of the model are an improvement compared to the original CATS model. In this way, it is possible to draw conclusions about which model performs best in replicating the stylized facts observed in the Dutch business sector. The simulation outcomes are obtained with the initial optimal calibrated parameters as in Table 6 in Section 3.3. The time periods 100 and 101 are used for comparing the simulation outcomes with the empirical data. These time periods are chosen because the calibration analysis is also based on the simulation outcomes at these time periods. Moreover, the simulation outcomes are all obtained using the same seed. Different time periods and different seeds did not result in major differences concerned the simulation outcomes.

Distribution of firm size follows a power law / Zipf distribution. In order to verify whether the simulated distributions of firm size, measured in net turnover, by the CATS model and its extended variants also follow a power law distribution, their log-log CCDFs are presented in Figure 10. The shape of the distribution plots clearly indicate that there is a power law behaviour present in the simulated net turnovers. However, there are some irregularities observed in the tails of these density functions. Whereas the observations of the log-log CCDF of the simulated net turnover by the model with the productivity extension seem to show a relatively linear pattern in the tails, the observations in the tails of the log-log CCDF of the original CATS model and the models with the price and the productivity extension deviate from this linearity. In order to investigate this more accurately, a linear regression is performed on the upper 5% of the observations, where the regression line is shown with a red line in the figures. The slope of the regression line represents the parameter β as explained in Section 2.2, that is, $P(X > x) = cx^{\beta}$. The regression results are presented in Table 7.

Table 7: Regression results and the Z-statistic of the 5% upper tail distribution of the log-log CCDF of the empirical and the simulated net turnovers by the CATS model and its extended variants. (*** indicates significant at the 0.1% level.)

		Estimate	Std. Error	t-value	Pr(> |t|)
Original CATS	Intercept	2.061	0.041	50.270	0.000***
	x	-0.276	0.010	-27.130	0.000***
Price extension	Intercept	2.149	0.008	260.600	0.000***
	x	-0.270	0.002	-134.500	0.000***
Productivity extension	Intercept	2.527	0.022	112.340	0.000***
	x	-0.133	0.006	-23.930	0.000***
Interest rate extension	Intercept	2.328	0.035	65.730	0.000***
	x	-0.179	0.009	-20.440	0.000***

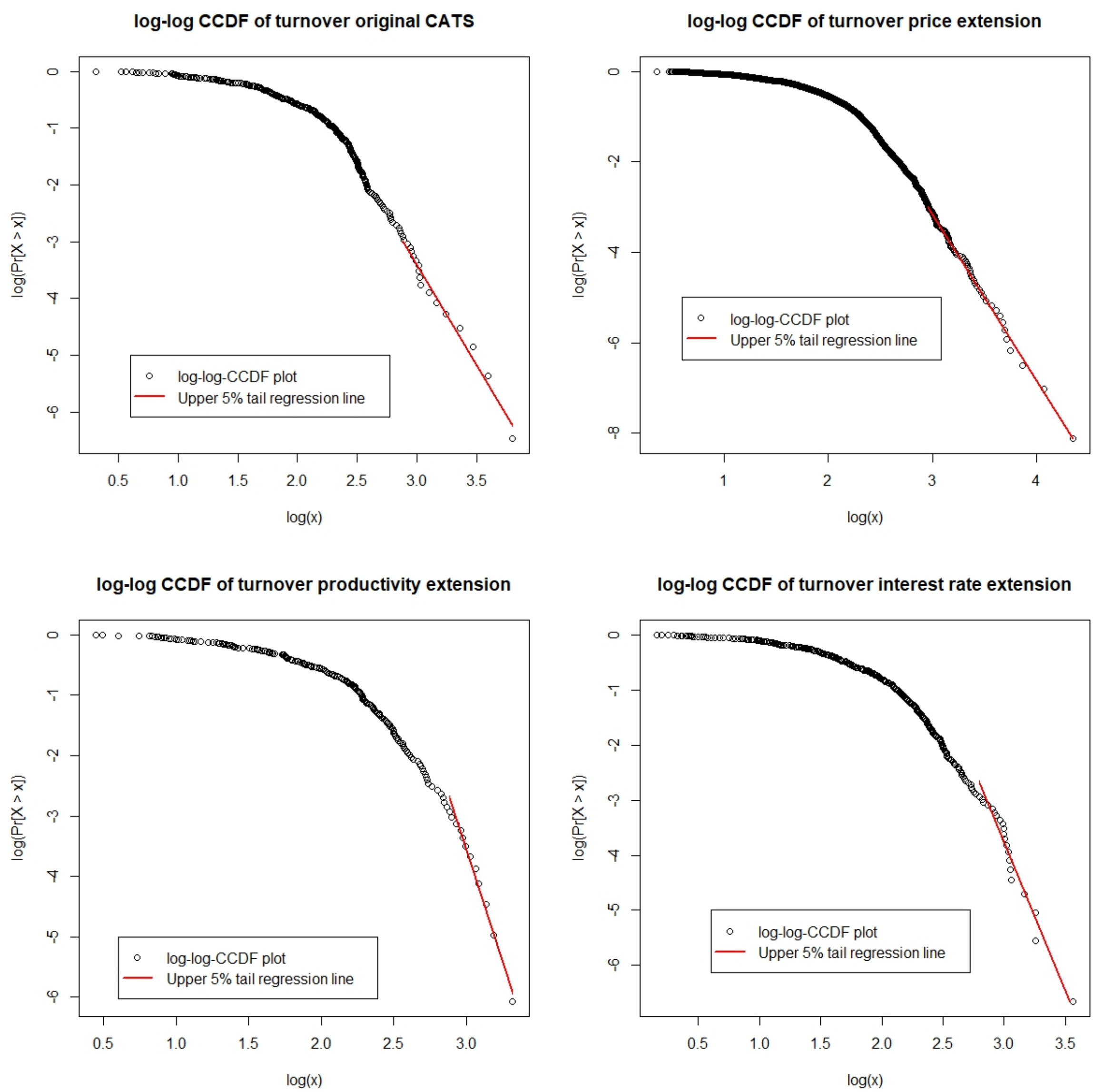

Figure 10: Log-log CCDF plots of the simulated net turnovers resulted from the four different models. The plots include the 5% upper tail estimated regression line.

First, all estimated regression parameters are significant which allows us to draw conclusions about the estimated parameter values. As shown in Table 7 the estimated slope parameters of the simulated turnover distributions are between -0.133 and -0.276. It has been investigated whether different time periods and seeds result in major differences regarding the characteristics of the distribution of turnover. The means of the estimated slope parameters over all time periods for different seeds has been investigated and all resulted in estimated intercept parameters of around 2 and estimated slope parameters of around -0.2. Unfortunately, these estimated parameter values do not correspond to the empirical CCDF of net turnover, equalling 11.745 and -0.966 for the intercept and slope parameter respectively. Whereas the distribution of the turnover of Dutch firms follows a Zipf distribution, the simulated distributions of the agent based models do not because their estimated slope parameter $\hat{\beta}$ is not equal to -1. According to Clogg, Petkova, and Haritou (1995) the correct test statistic for comparing the slope parameters of two linear regressions is

$$Z = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{\mathrm{SE}(\hat{\beta}_1)^2 + \mathrm{SE}(\hat{\beta}_2)^2}}, \tag{24}$$

where $\mathrm{SE}(\beta_1)$ and $\mathrm{SE}(\beta_2)$ represent the standard errors of the estimated slope parameters $\hat{\beta}_1$ and $\hat{\beta}_2$. To test whether the estimated slope parameter of the upper 5% tail distribution of the log-log CCDF of the empirical β_{emp} and the simulated β_{sim} net turnovers are equal, the null hypothesis

$$H_0: \quad \beta_{\text{emp}} = \beta_{\text{sim}}$$

is stated. The values of Z for the different simulation models are presented in Table 8. It is clear that the null hypothesis is rejected for all agent based models. This proves that the slope estimates from the empirical and simulated upper tail distributions are not equal.

Table 8: Z-statistic of comparing the slope estimate of the 5% upper tail distribution of the log-log CCDF of the empirical and the simulated net turnovers by the CATS model and its extended variants. (*** indicates significant at the 0.1% level.)

	Z	p-value
Original CATS	-69.000	0.000***
Price extension	-348.000	0.000***
Productivity extension	-138.833	0.000***
Interest rate extension	-87.444	0.000***

However, when comparing the results of the agent based models with each other, it can be concluded that the original CATS model and the model with the price extension result in estimated slope parameters of -0.276 and -0.270, which are closest to the estimated slope parameter of the empirical turnover distribution. Even though this result is not similar to the estimated parameter value of the empirical distribution, it indicates that these models outperform the other agent based models, which result in even smaller parameter values of -0.133 and -0.179 for the model with the productivity and the interest rate extension respectively. These slope estimates are further away from the empirical estimated slope parameter of -0.966. This is also in line with the results observed in the calibration study in Section 3.3, where the model with the price extension resulted in the lowest value of the objective function. Moreover, it can be concluded that, in terms of the slope parameter of the tail of the log-log CCDF of the simulated turnovers, the extended models do not outperform the original CATS model as the models with the price, the productivity and the interest rate extension result in a larger value of the estimate slope parameter.

Summarizing, it can be concluded that the CATS model and its extensions are able to capture the power law behaviour present in the distribution of firm size. Therefore, this agent based model is capable of reproducing the first stylized fact that is present in the Dutch business sector. However, the original CATS model nor its extensions are able to reproduce the special case of the power law distribution where the slope parameter equals -1, that is, the Zipf distribution, which does apply to the empirical data. As explained in Section 2.2 the power law behaviour in business sectors might be due to the presence of mergers and acquisitions and although the CATS model accounts for bankruptcies and the emergence of companies, mergers and acquisitions are not explicitly modelled. This might be an explanation for the different values of the estimated slope parameters. Further research may prove whether accounting for this phenomenon will indeed lead to a Zipf distribution of firm size in the CATS model.

Volatility of growth rates decreases when firm size increases. In Section 2.2 it is shown that for Dutch firms the volatility of growth rates of business capital stabilizes when firm size, measures in number of employees, increases. The CATS model nor its extended variants do not model the number of employees and therefore, another measure of firm size is necessary to validate this stylized fact. Following Bianchi et al. (2008) who also analyze this stylized fact using the CATS model, the turnover is used as a measure for the size of firms. Figure 11 shows the relationship between the growth rates of capital in time period $t = 100$, $t = 101$ and the turnover in time period $t = 100$ simulated with the original CATS model and its price, productivity and interest rate extension. From all models similar patterns can be observed in the relationship between turnover and growth rates. That is, the growth rates are concentrated in a slightly increasing trend, starting negative between 0 and approximately 10 and then become positive between approximately 10 and around 20. Besides that, there are some values that lie around this concentrated 'cloud' of observations. Therefore, it seems that there are two regimes present in the relationship between simulated growth rates and simulated turnover. The first contains a large population with a relatively small dispersion, and the second consist of a few companies that deviate substantially from the large population.

When measuring the volatility of the growth rates by for example the standard deviation, these largely deviating firms cause the results to be unstable. Even for more robust measures of volatility such as the median absolute deviation (MAD) sharp peaks are observed that do not represent the true relationship between the growth rates and the turnover. A solution to this problem is to remove the second regime with the largely deviating firms from the observations. In order to do this, the turnover of firms is divided into 14 bins. For each bin, Z-scores for each firm are calculated by subtracting the mean of the growth rates in that bin from the growth rate of each firm and divide by the standard deviation of the growth rates in that bin. Subsequently, the 20% of the companies of which the Z-score in absolute value is largest are removed from the

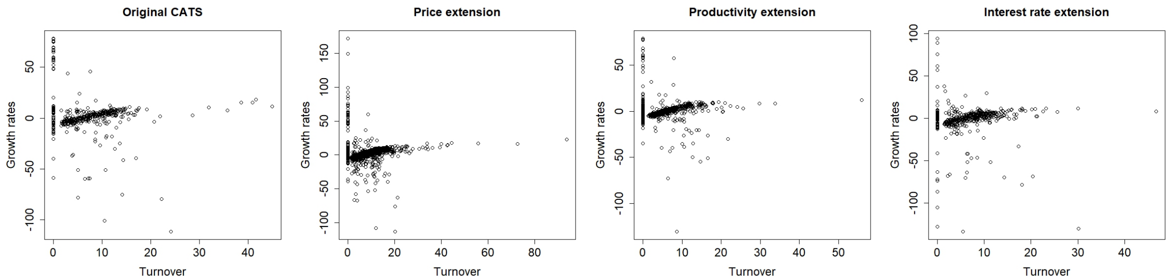

Figure 11: Firm size measured in turnover in period 100 versus the growth rates of business capital of period 100 and 101 simulated by the original CATS model and its extended variants.

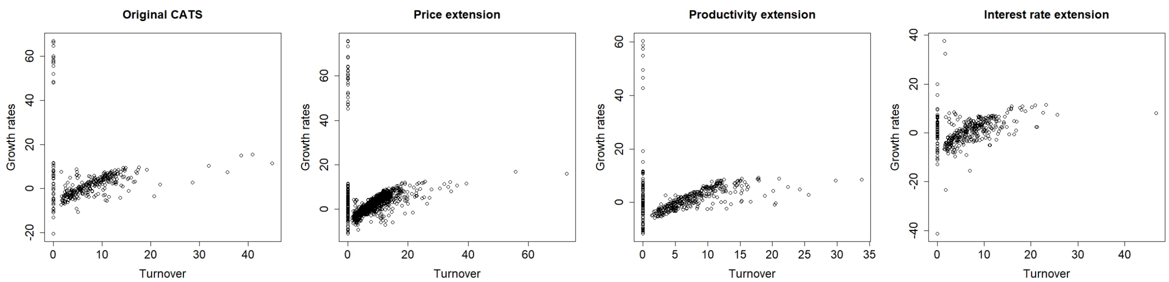

Figure 12: Firm size measured in turnover in period 100 versus the growth rates of business capital of period 100 and 101 simulated by the original CATS model and its extended variants without the 20% largest deviating firms, measured by Z-scores of growth rates.

data set. Figure 12 shows the scatter plots of turnover versus growth rates simulated with the models without the 20% largest deviating firms. These figures allow for better visualization of the upward sloping pattern discussed earlier. It can be noticed that the dispersion of growth rates at a turnover of zero is extremely large for all models. One of the differences between the models is the range of the growth rates at zero turnover. For example, the growth rates resulting from the interest rate extension show the most negative values at zero turnover, and the growth rates obtained by simulating the CATS model with the price extension result in the largest positive values at zero turnover. Moreover, the plurality of firms simulated by all models show growth rates that start negatively at approximately -10 and increase with turnover to approximately 10. Every model produces a few firms with a turnover larger than approximately 20, where the CATS model with the price and productivity extension generate more firms with this relatively large turnover than the original CATS model and its interest rate extension.

In order to investigate more accurately whether the volatility of growth rates of business capital decreases with firm size, the mean of turnover and the MAD of the growth rates in each bin are calculated without the 20% largest deviating firms. The MAD is chosen as a measure of volatility because it is a more robust statistic than the standard deviation, such that deviations due to a small amount of outlying values are irrelevant. The results are presented in Figure 13. The number of bins is different across the models because due to the entry-exit process of the CATS model the number of

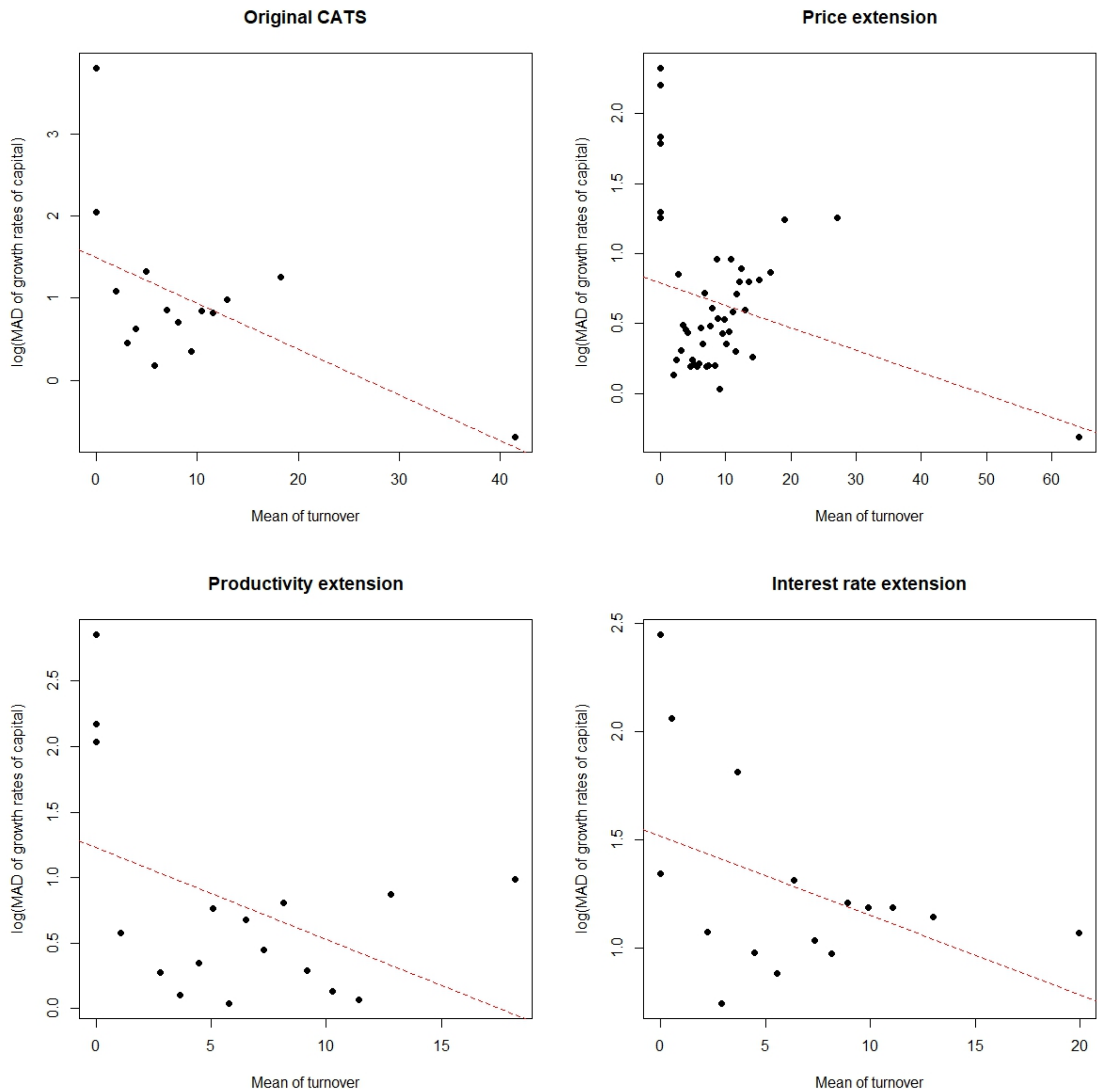

Figure 13: Firm size measured in turnover in period 100 versus the growth rates of business capital of period 100 and 101 simulated by the original CATS model and its extended variants without the 20% largest deviating firms, measured by Z-scores of growth rates.

firms is different in each period, and therefore also change when variations to the model are applied. The number of bins are calculated such that there are $(1 - 0.2) \times 30 = 24$ firms in each bin, where 0.2 represents the 20% most deviating firms that are removed from the data set, and the sub sample size of 30 is chosen randomly. When the total number of firms at time period $t = 100$ is not divisible by 24, the remaining firms are placed into the bin with the highest turnover, because the data is sorted on turnover before the division into bins takes place. Therefore, in Figure 13 the MAD of the growth rate belonging to the highest turnover mostly consists of less firms than the other MAD values.

Table 9: Regression results of the relationship between firm size and the volatility of growth rates of the empirical data and the simulated data generated by the CATS models and its extended variants. (*** indicates significant at the 0.1% level, ** at the 1% level and * at the 10% level.)

		Estimate	Std. Error	t-value	Pr(> |t|)
Original CATS	Intercept	1.492	0.296	5.042	0.000 ***
	x	-0.056	0.022	-2.548	0.024 **
Price extension	Intercept	0.792	0.105	7.570	0.000 ***
	x	-0.016	0.008	-2.030	0.048 **
Productivity extension	Intercept	1.232	0.298	4.133	0.001 ***
	x	-0.070	0.037	-1.875	0.080 *
Interest rate extension	Intercept	1.517	0.170	8.944	0.000 ***
	x	-0.037	0.020	-1.797	0.094 *

From the scatter plots in Figure 13 it is observed that the original CATS model produces the smallest amount of firms, whereas the CATS model with the price extension produces the largest amount. In the price extension model a clear 'cloud' of observations can be observed between an average turnover of approximately 0 and 20. In contrast, the other models show much more dispersion between the observations. In order to measure the relationship between the simulated turnover and volatility of growth rates, a linear regression is fitted based on the recommendations Stanley et al. (1996), who state that the volatility of growth rates linearly decreases with firm size. The regression is perform in a similar way as for the Dutch data in (1), except for the definition of size, which is expressed in number of employees in the empirical data set and defined as turnover in the simulated data sets. The estimated regression lines are visualized with red dotted lines in Figure 13 and the regression results are presented in Table 9. All agent based models show a significant negative slope estimate, proving that the models are able to simulate the linear decreasing relationship between firm size and the volatility of their growth rates. It can therefore be concluded that the models are able to capture this stylized fact, which is also present in the Dutch business sector.

However, it is also of interest to know how well the relationship is reproduced by the agent based models in comparison with the Dutch business sector. Moreover, it is interesting whether the different variants of the CATS model are an improvement to the original CATS model in terms of replicating this stylized fact of Dutch firms. In Section 2.2 the mean of the log of the number of employed persons was plotted against the MAD of the log of the growth rates of Dutch firms. Fitting the linear regression on these empirical observations resulted in a significant estimated slope parameter of -0.051. Although the definition of firm size is different for the empirical and the simulated data, the estimates of the slope parameters produced by the agent based models (especially by the original CATS model) seem to be quite similar to that of the Dutch business sector. In order to investigate whether these estimated slope parameters are indeed very similar, the Z-statistic is computed as in (24). The statistics and their p-values are presented in Table 10.

From the results it can be observed that only the Z-statistic of the CATS model with the price extension is significant, indicating that the null hypothesis of equal slope parameters of the relationship between firm size and volatility of growth rates of the empirical data compared to that of the simulated data generated with the model with the price extension must be rejected. The p-values corresponding to the Z-statistics of the other agent based model are in contrast much higher, and do not reject the null hypothesis of equal slope parameters. This strengthens what was observed in the scatter plots of Figure 13, where the model with the price extension is the only one that generates little dispersion between the observations, and the downward sloping regression line seems to be particularly due to the single observation with a low MAD value in the bin with the highest average turnover. Moreover, the estimated slope parameter of the price extension equals -0.016, which is compared to the other models furthest away from that of the empirical data. In contrast, the original CATS model shows the highest p-value of 0.411, which could have been expected because the corresponding estimated slope parameter equals -0.056, almost identical to that of the empirical data equalling -0.051.

Table 10: Z-statistics and p-values of comparing the estimated slope parameters of the relationship between firm size and the volatility of growth rates of the empirical data and the simulated data generated by the CATS models and its extended variants. (*** indicates significant at the 0.1% level.)

	Z-statistic	p-value
Original CATS	0.224	0.411
Price extension	-3.913	0.000***
Productivity extension	0.511	0.305
Interest rate extension	-0.686	0.246

From the analysis above it can be concluded that the CATS model and all of its extended variants are able to replicate the linearly decreasing relationship between firm size and the volatility of their growth rates of business capital. However, compared to the slope estimate of the Dutch data set not all models are able to replicate the exact value of the slope parameter. The CATS model with the price extension clearly generates an estimated slope parameter that deviates too much from that of the Dutch data. In contrast, the original CATS model is able to almost perfectly replicate the value of the slope estimate. The models with the productivity and interest rate extension are not able to outperform the original CATS model, but they are able to reproduce a relatively similar slope parameter of the relationship between firm size and growth rates compared to that of the Dutch business sector.

Growth rates are Laplace distributed In Section 2 it is shown that the growth rates of business capital of Dutch firms follow a Laplace distribution, which is one of the commonly known stylized facts in the empirical literature. To investigate whether the CATS model and its extensions are also able to produce a Laplace distribution of growth rates of business capital, the densities of the simulated growth rates produced by the original CATS model and their extensions are compared to the densities of the empirical growth rates of business capital. The distribution plots are shown in Figure 14, where in reality the tails of the empirical growth rate distributions are much wider. This can also be observed in Figure 3 in Section 2.2. However, in order to visualize the differences between the simulated and the empirical growth rate densities more accurately, the axis in the figures are limited to those of the simulated growth rates of business capital. From Figure 14 it can be observed that for all the agent based models the growth rates follow indeed a Laplace distribution. In fact, the densities of the simulated growth rates all seem to follow a similar location parameter as that of the density of the empirical growth rates. However, the scale of the densities of all simulated growth rate densities deviates from that of the density of the growth rates of Dutch firms. It seems that the CATS model with the interest rate extension deviates the least from the empirical growth rate distribution in terms of the scale parameter.

Table 11: Location and scale parameter and the KS test statistic of the simulated growth rate distributions.

	$\hat{\mu}$	$\hat{b}$	D
Dutch firms	0.038	0.380	-
Original CATS	0.022	0.075	0.327
Price extension	0.023	0.081	0.320
Productivity extension	0.017	0.076	0.333
Interest rate extension	-0.002	0.090	0.291

In order to investigate this more accurately, the estimated location parameters $\hat{\mu}$ and scale parameters $\hat{b}$ of the empirical and simulated growth rates are presented in Table 11. The table also contains the value of D, used for the Kolmogorov-Smirnof test statistic, which tests whether the simulated growth rate distributions from the models are equal to the empirical growth rate distribution. For this validation analysis it is chosen not to perform a simulation study. As can be seen from the densities and the estimated Laplace parameters, the parameter values of the simulated and the empirical growth rates distributions are very different. Moreover, obtaining the simulated distributions for different seeds did not lead to large differences in the results. Therefore, taking into account the additional uncertainty from the different seeds is regarded as unnecessary for this analysis. The values of the Laplace parameters shown in Table 11 confirm what was observed in the density plots in Figure 14. Indeed, the location parameters $\hat{\mu}$ of the simulated growth rates are close to that of the growth rates of Dutch firms. The values all lie around zero, slightly skewed to the right, except for the growth rates resulting from the CATS model with the interest rate extension, which is only slightly

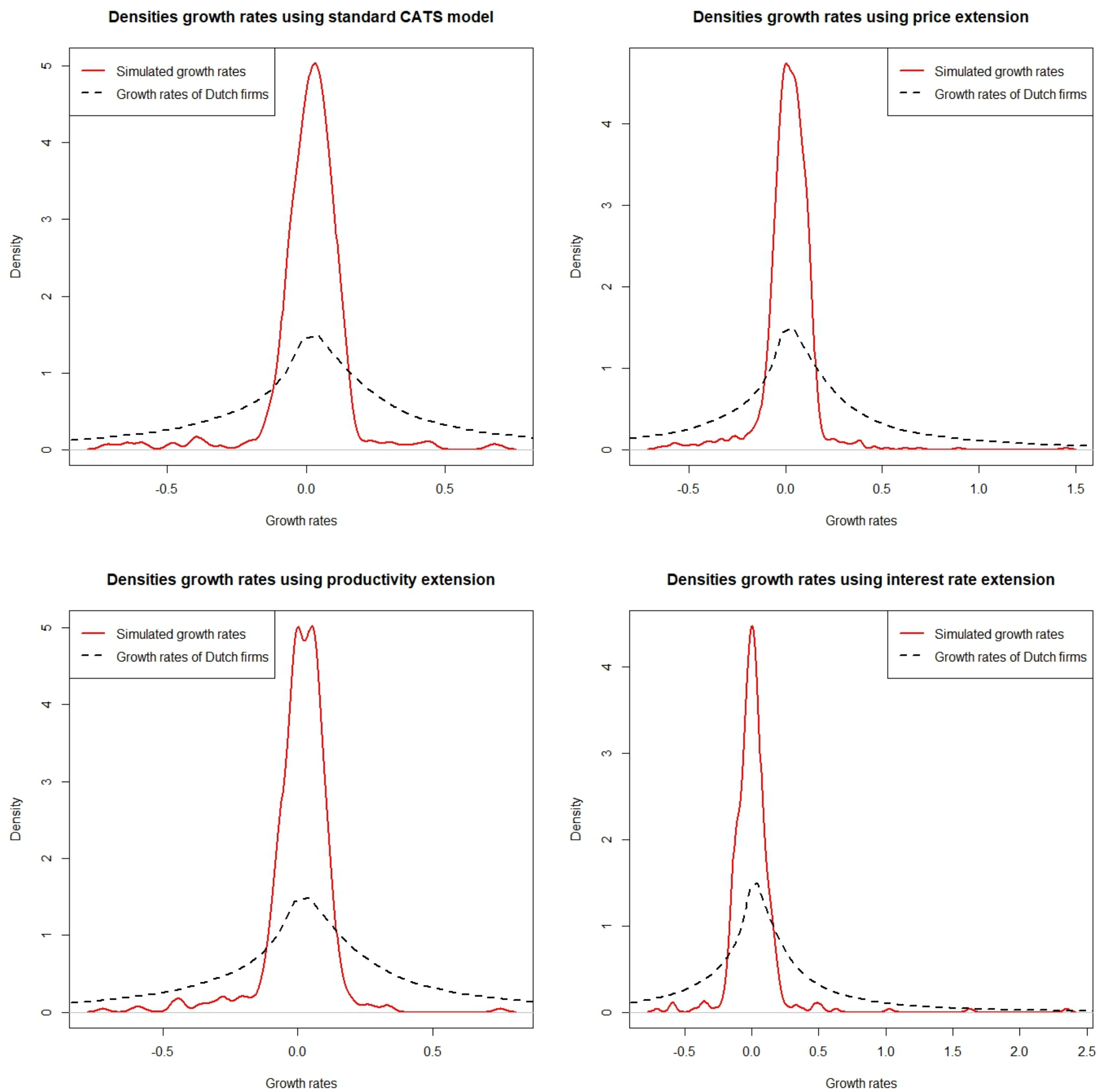

Figure 14: Density plots of the empirical growth rates and the simulated growth rates for the original CATS model and the extended CATS models using the calibrated parameters.

skewed to the left. The growth rates obtained from simulating the CATS model with the price extension results in the closest value of the location parameter compared to the growth rates of Dutch firms. This is in line with what is observed in Section 3.3, where the price extension resulted in the lowest value of the objective function. However, again the results are very similar and due to the uncertainty resulting from the drawbacks explained in the final paragraph of Section 3.3, it is hard to draw conclusions about which model generates the most accurate variables compared to empirical data.

The values of the estimated scale parameters $\hat{b}$ of the simulated growth rates are, in contrast to the estimated location parameters, not similar to that of the estimated scale parameters. The estimated scale parameter of Dutch firms is equal to 0.380, whereas those of the simulated growth rates are between 0.075 and 0.090, around five times as small. Therefore, the Kolmogorov-Smirnof test rejects the null hypothesis of equal distributions of the simulated and the empirical growth rates. In fact, this large difference of the scale parameters also explains the large values of the objective functions resulted from the calibration procedure in Section 3.3. Apparently, the CATS model nor their extensions are able to produce similar scale parameter values of the growth rates as the empirical growth rates distribution. This might be due to particular mechanisms that are present in the Dutch business sector but which are not captured by the agent based models. Many other explanations can be given for this difference, such as the assumptions stated by the models or the initial fixed parameter values of the models.

5 Conclusion

In this research the underlying complex mechanisms present in the Dutch business sector have been investigated using agent based modelling. Provided with a data set that contains information about all the firms in the Netherlands, three stylized facts have been observed and investigated in further detail. Based on these stylized facts a suitable agent based model is examined, which is the CATS model introduced by Gatti et al. (2003). This model is chosen as the starting point of investigating the complexity in Dutch firms because it is shown by different studies that it is capable of reproducing several stylized facts, among which the ones observed in the Dutch business sector. Three different extensions are implemented in the CATS model of which it is believed that they provide a more realistic representation of the Dutch business world. Comparing the simulation output of the CATS model and its extensions with the real world data would allow for investigation of whether these agent based models are indeed capable of replicating the stylized facts observed in the Dutch economy. This would in turn provide us with better insights in the complex mechanisms that are present in the behaviour of Dutch firms.

Although agent based modelling provides great flexibility in modelling complex systems, there is also a major drawback to this simulation method. Even though quite a few contributions have been made in the current literature, the estimation of the initial parameters is still regarded as a problematic issue in agent based models. This is due to the complexity that is present in agent based models, leading to the fact that there is no analytic solution for the objective function. However, in order to match the initial parameter values of the agent based model with the Dutch business sector as accurate as possible a calibration study is necessary. Therefore, this paper employs a simulation method called the Generalized Method of Moments with an expression of the objective function introduced by Winker et al. (2007) to investigate the optimal values of the initial parameter values. First, two parameters of the original CATS have been calibrated on the Dutch business sector. Using these optimal parameter values, the parameters necessary for the extensions have been calibrated. In this way, comparison of the extended variants of the CATS model with the original model can be investigated at its best. After calibration of the initial parameters the simulation outcomes of the agent based models are compared to the data of Dutch firms by investigating the stylized facts and their statistical characteristics.

Investigation of the data set of Dutch firms resulted in the presence of three stylized facts that are confirmed by the empirical literature. The first is that the distribution of firm size follows a power law distribution. In fact, the requirements for the Zipf distribution, which is a special case of a power law, are being fulfilled. The second stylized fact is that the volatility of the growth rates of firms decreases with firm size. Finally, the growth rates of firms follow a Laplace distribution. The calibration analysis is based on the final stylized fact, because the growth rates of business capital can be best compared for the simulation outcomes and the empirical data. Calibration of the parameters of the original CATS model and its extended variants all result in an optimal objective function that is too high to accept the hypothesis of the model being a good representation of the real world. However, this is not surprising because we are only in the infancy of understanding the complex mechanisms that are present in the behaviour of Dutch firms using agent based modelling. The calibration study does provide us with an indication that the implemented extensions are an improvement of the original CATS model because of their lower optimal objective value. The CATS model with the price extension results in the lowest value of the objective function, proving that this extension outperforms the other models in terms of replicating the moments of the growth rates of business capital observed in the Dutch business sector.

The optimal initial parameters resulting from the calibration study have been used to simulate the agent based models and obtain the output variables that can be used to compare with the stylized facts in the Dutch business sector. From the validation analysis it can be concluded that all the models are capable of replicating the stylized facts of Dutch firms to a certain extent. The power law behaviour in firm size is also found in the simulation outcomes of all agent based models. However, the models are not capable of reproducing the special Zipf distribution that is present in the Dutch business sector. An explanation for this might be the absence of modelling mergers and acquisitions in the CATS model. The original and the price extended CATS models perform best compared to the parameter value of the empirical data set. The second stylized fact where the volatility of the growth rates decreases with firm size is also being fulfilled by the agent based models. Although the exact values of the linearly decreasing relationship are not the same as that of the Dutch business sector, they are relatively close. Especially the original CATS model performs well in replicating this stylized fact. Finally, the simulated growth rates of business capital all follow a Laplace distribution. The estimated location parameters are relatively close to that of the empirical growth rates distribution, especially for the model with the price extension. However, the estimated slope estimates of the simulated growth rates deviate substantially from that of the Dutch growth rates of capital. This might be due to one or more mechanisms that are present in the Dutch business sector but are not captured by the agent based models. This might also explain the large values of the objective functions in the calibration study.

It can be concluded that a relatively simple agent based model as the CATS model is capable of replicating all of the three stylized facts that are observed for the Dutch business sector. This proves that this agent based model is a good starting point of explaining several complex mechanisms that are present in the behaviour of Dutch firms. Moreover, some of the implemented extensions have shown to be an improvement of the original CATS model for some characteristics of the stylized facts. However, there are also several drawbacks to modelling the Dutch business sector with agent based modelling. First, the CATS model is very sensitive to changes in the initial parameter values and to different specifications of the model. For the original CATS model but especially for the implemented extensions, this makes it particularly hard to investigate the optimal combinations of initial parameters that lead to simulation outcomes that are a good representation of the real world data. One could solve this problem by evaluating the objective function for all possible parameter combinations. However, this is not feasible in agent based modelling due to the computation time that increases dramatically when increasing the number of parameters that are subject to calibration. Moreover, calibration of the parameters in an agent based model is subject to estimation uncertainty due to the selection of moments and the number of simulations of the agent based model. However, increasing the number of moments or the number of simulations also lead to a substantial increase in computation time. Therefore, it is necessary to impose restrictions when calibrating the parameters of an agent based model. In turn, this estimation uncertainty leads to the fact that it is hard to draw general conclusions about whether any of the implemented extensions are an improvement of the original CATS model.

In order to investigate whether the implemented extensions are in general an improvement of the original CATS model, several solutions are possible. The first is to perform a similar calibration analysis to a different data set of another countries' firms that contain the same stylized facts. If the same results apply, it can be concluded with more certainty that the extensions are an improvement. Another solution is to increase the number of moments in the calibration study, such that also other stylized facts are taken into account when calibrating the parameters. Moreover, increasing the number of simulations of the agent based model would lead to a decrease in estimation uncertainty. The latter two solutions would also lead to more accurate initial parameter values, which in turn lead to more representative simulation outcomes compared to the stylized facts observed in the empirical data set. Because the CATS model has already shown that it is capable of capturing most aspects of the stylized facts present in the Dutch business sector, further investigation of the parameter space and the specification of this model would definitely be useful in further research. It would also be interesting to extend the specification of the model by including for example the presence of importing and exporting firms, or mergers and acquisitions. If so, this model could be very interesting for policy analyses, such as the development of early warning indicators. For such purposes it would be of great interest to investigate the dynamic properties of the behaviour of Dutch firms. Therefore, applying this agent based model to a data set of Dutch firms that contains information on a longer time horizon would definitely broaden the research to another level. By all means, this agent based model provides us with an endless amount of possibilities to further investigate the complex underlying mechanisms in the Dutch business sector.

List of Symbols

Symbol	Description
β	Slope parameter of the power law distribution
$\mathbf{V}$	Volatility of the log(growth rates) measured with the MAD in each bin
$\mathbf{S}$	Average of the log(firm size) in each bin
g	Intercept parameter of the regressing $\mathbf{V}$ on $\mathbf{S}$
h	Slope parameter of the regressing $\mathbf{V}$ on $\mathbf{S}$
$\boldsymbol{\epsilon}$	Error term of regressing $\mathbf{V}$ on $\mathbf{S}$
μ	Location parameter of the Laplace distribution
b	Scale parameter of the Laplace distribution
$D_{n,m}$	The Kolmogorov-Smirnof test statistic
$F_{1,n}$	The empirical distribution function of the growth rates of business capital
$F_{2,m}$	The fitted Laplace distribution of the growth rates of business capital
Z	The Z-statistic for comparing two slope estimates
N_t	Number of firms at time period t
T	Number of time periods
Y_{it}	Output/Production of firm i at time period t
v	Constant productivity of capital
K_{it}	Capital stock of firm i at time period t
P_{it}	Price of firm i at time period t
P_t	Market price at time period t
u_{it}	Relative price of firm i at time period t
R_{it}	Revenue of firm i at time period t
ϕ_t	Profit at time period t
r	Constant real interest rate
CA_{it}	Capital adjustment costs of firm i at time period t
γ	Constant positive parameter for the capital adjustment costs
CB_{it}	Bankruptcy costs of firm i at time period t
α_1, α_2	Constant positive parameters for the bankruptcy costs
a_{it}	Equity ratio of firm i at time period t
$\bar{u}_{it}$	Threshold price for bankruptcy of firm i at time period t
A_{it}	Equity base of firm i at time period t
τ_{it}	Optimal rate of capital accumulation of firm i at time period t
ρ_{it}	Bankruptcy cost augmented interest rate of firm i at time period t
μ^{s}, σ^{s}	Mean and variance of the uniform distribution of the price extension for small firms
μ^{m}, σ^{m}	Mean and variance of the uniform distribution of the price extension for medium firms
μ^{l}, σ^{l}	Mean and variance of the uniform distribution of the price extension for large firms
ϕ_{it}	Productivity of firm i at time period t
x, y, z	Positive parameters used in the productivity extension
μ^{prod}, σ^{prod}	Mean and variance of the uniform distribution used in the productivity extension
r_{it}	Interest rate of firm i at time period t as a function of firm size
ξ, ξ_1, ξ_2	Positive parameters used for the interest extension
$\mathbf{X}$	Observed data sample
$\mathbf{X}^s$	Simulated data sample
$\boldsymbol{\theta}$	Vector of parameters of the agent based model

$\boldsymbol{\theta}_1$	Vector of parameters subject to calibration
$\boldsymbol{\theta}_2$	Vector of parameters fixed a priori
$\mathbf{m}$	True vector of moments
$\mathbf{m}^e$	Estimated vector of bootstrapped moments
$\mathbf{m}^s$	Simulated vector of moments
I	Number of replications of the agent based model
$\mathbf{W}$	$k \times k$ positive definite weighting matrix
$\mathbf{G}_I$	The matrix $\sum_{i=1}^{I}[(\mathbf{m}_i^s)\vert\boldsymbol{\theta} - \mathbf{m}]$
$\boldsymbol{\Sigma}$	Covariance matrix of the estimated bootstrapped moments
k	Number of moment conditions
l	Number of parameters subject to calibration
$\boldsymbol{\Theta}_i$	Search space of the parameter subject to calibration
B	Number of bootstraps
$\mathcal{L}(\boldsymbol{\theta})$	The objective function of the Generalized Method of Simulated Moments

Appendices

A Remaining available variables

- **SBI-code**: This represents the Dutch Standard Industrial Classification, the Dutch classification of economic activities used by SN since 1993 to list companies by their main activity. Companies in an industry or branch can also carry out other activities, called side activities, in addition to this activity. The SBI-code contains multiple levels which are indicated with a maximum of five digits. The level of the first four digits closely corresponds to the classification used in the EU for economic activities, that is, the Nomenclature General des Activités économiques dans la Communauté Européenne (NACE). The level of the first two digits correspond to those of the United Nations classification (ISIC).
- **Firm size**: Firms' size is derived from the total number of employees in a company. This variable is determined based on employees on the payroll including cooperating firms, owners and family members. This total amount is also referred to as the number of Employed Persons.
- **Legal form**: The legal form of the business unit is indicated with a code, which can refer to, for example, a private or limited liability company, sole proprietor or to a general partnership.
- **Balance tax profit calculation**: The balance of the fiscal profit calculation is the positive or negative outcome resulting from the calculation of the income or corporation tax, in the current or previous tax year.
- **Wages and salaries according to National account**: The compensation for the employee who has worked in a given period and which is payable by the employer, including the wage tax and social premiums paid by the employer on behalf of the employee.
- **Coverage ratio WIA (Work and Income According to Labour Capacity Act)**: The degree to which units of activity related to the business unit cover the business unit for the profit return.
- **Export**: The total amount of export of a business unit according to the tax declaration.
- **Import**: The total amount of import of a business unit according to the tax declaration.
- **Foreign parent/subsidiary**: A dummy variable indicating whether the firm has a foreign parent and/or a foreign subsidiary. A foreign parent means that a foreign company has the predominant control. A foreign subsidiary means that the firm has predominant control over a foreign company.
- **Mergers and acquisitions**: A dummy variable indicating whether the firm has merged with another firm of has been taken over by another firm.

B Missing data patterns

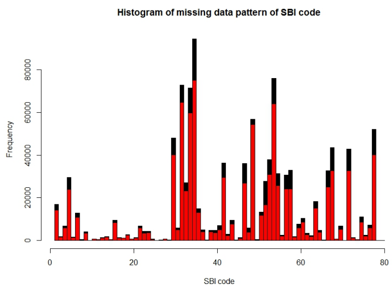

Figure 15: Distribution plot of the SBI-code of all firms (black) versus the distribution plot of the SBI-code of the firms without missing data values (red).

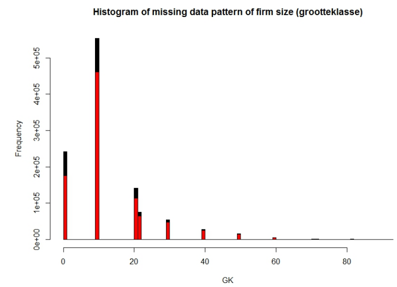

Figure 16: Distribution plot of the size of all firms versus the distribution plot of the size of the firms without missing data values.

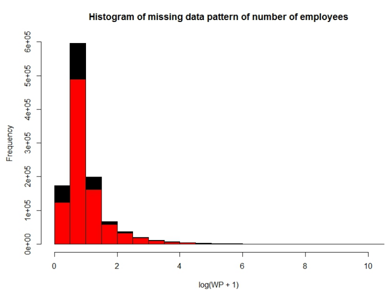

Figure 17: Distribution plot of the employed persons of all firms (black) versus the distribution plot of the employed persons of the firms without missing data values (red).

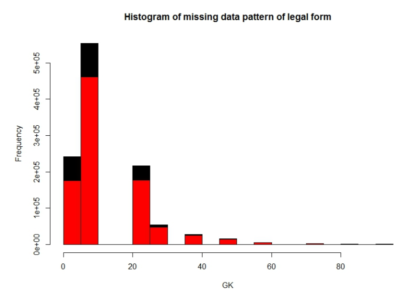

Figure 18: Distribution plot of the legal forms of all firms (black) versus the distribution plot of the legal forms of the firms without missing data values (red).

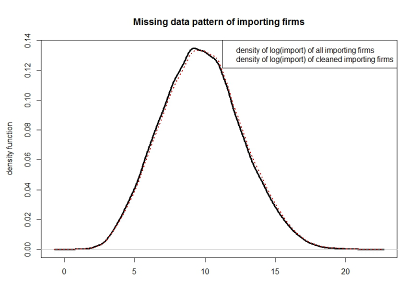

Figure 19: Distribution plot of the amount of import of all firms versus the distribution plot of the import of the firms without missing data values.

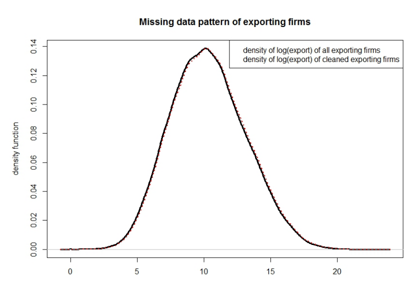

Figure 20: Distribution plot of the amount of export of all firms versus the distribution plot of the export of the firms without missing data values.

C Firm size vs. growth rates in 2011-2012 and 2012-2013

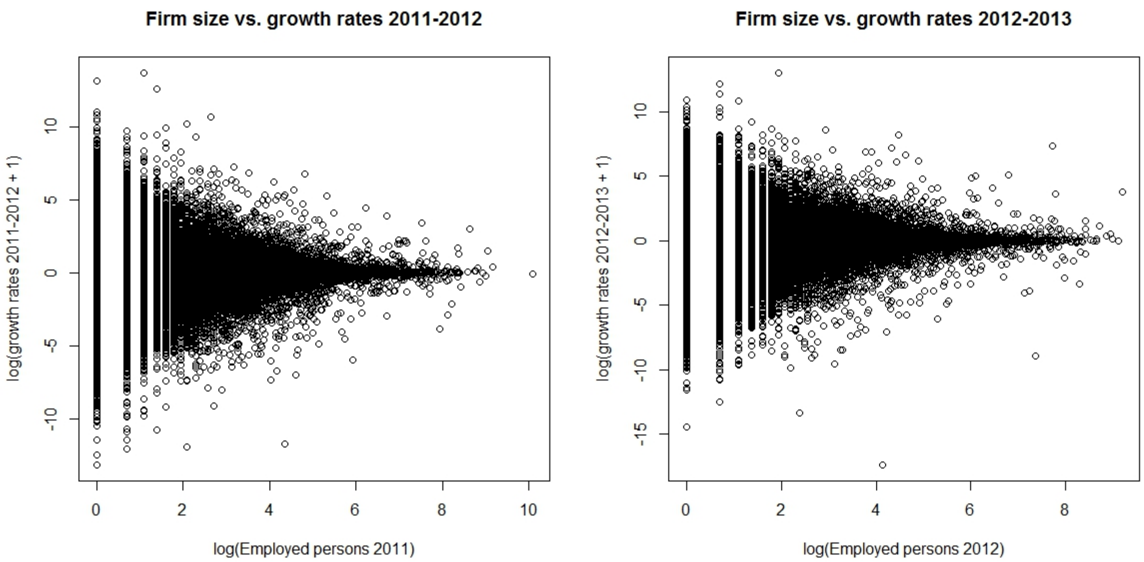

Figure 21: The log of the growth rates of business capital of each firm of the years 2011-2012 and 2012-2013 plotted against the log of number of employees of each firm in the base year, i.e. the size of the firm.

D Distribution of growth rates in 2011-2012 and 2012-2013

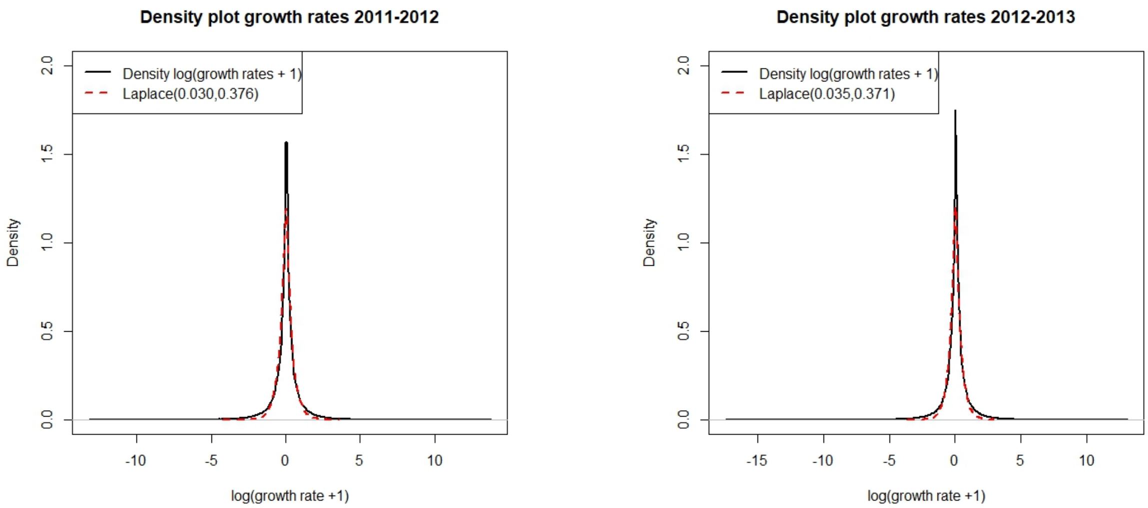

Figure 22: Density plot of the growth rates of business capital of the years 2011-2012 and 2012-2013 versus their estimated Laplace distributions.

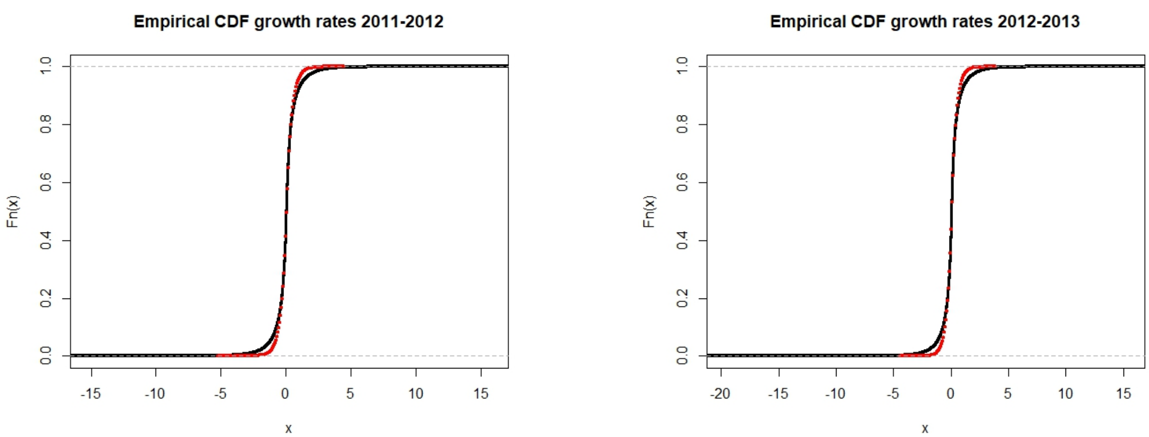

Figure 23: Empirical CDF plots of the growth rates of business capital of the years 2011-2012 and 2012-2013 (solid black lines) versus their estimated Laplace distributions (dotted red lines).

Table 12: Estimated Laplace parameters and the Kolmogorov-Smirnof test statistic of the log(growth rates + 1) for the years 2011-2012 and 2012-2013. (*** indicates significant at the 0.1% level.)

year	$\hat{\mu}$	$\hat{b}$	n	m	$D_{n,m}$	p-value
2011-2012	0.030	0.376	743.700	10,000	0.040	0.000***
2012-2013	0.035	0.371	797.568	10,000	0.040	0.000***

References

Assenza, T., Gatti, D. D., & Grazzini, J. (2015). Emergent dynamics of a macroeconomic agent based model with capital and credit. *Journal of Economic Dynamics and Control*, *50*, 5–28.

Axtell, R. L. (2001). Zipf distribution of us firm sizes. *science*, *293*(5536), 1818–1820.

Barde, S. (2016). Direct comparison of agent-based models of herding in financial markets. *Journal of Economic Dynamics and Control*, *73*, 329–353.

Barde, S. (2017). A practical, accurate, information criterion for nth order markov processes. *Computational Economics*, *50*(2), 281–324.

Bianchi, C., Cirillo, P., Gallegati, M., & Vagliasindi, P. A. (2007). Validating and calibrating agent-based models: a case study. *Computational Economics*, *30*(3), 245–264.

Bianchi, C., Cirillo, P., Gallegati, M., & Vagliasindi, P. A. (2008). Validation in agent-based models: An investigation on the cats model. *Journal of economic behavior & organization*, *67*(3-4), 947–964.

Bottazzi, G., & Secchi, A. (2003). Why are distributions of firm growth rates tent-shaped? *Economics Letters*, *80*(3), 415–420.

Buiten, G., de Jonge, E., & Pijpers, F. P. (2018, May). *Toepassingsmogelijkheden van complexiteitstheorie bij het cbs.* (Discussion paper)

Caiani, A. (2017). Innovation dynamics and industry structure under different technological spaces. *Italian Economic Journal*, *3*(3), 307–341.

Caiani, A., Catullo, E., & Gallegati, M. (2017). The effects of fiscal targets in a currency union: A multi-country agent based-stock flow consistent model.

Catullo, E., & Gallegati, M. (2015). Emerging economic product complexity: a multi-country agent based model.

Chiarella, C., He, X.-Z., & Zwinkels, R. C. (2014). Heterogeneous expectations in asset pricing: Empirical evidence from the s&p500. *Journal of Economic Behavior & Organization*, *105*, 1–16.

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, *100*(5), 1261–1293.

Coad, A. (2007). Testing the principle of 'growth of the fitter': the relationship between profits and firm growth. *Structural Change and economic dynamics*, *18*(3), 370–386.

De Jong, E., Verschoor, W. F., & Zwinkels, R. C. (2010). Heterogeneity of agents and exchange rate dynamics: Evidence from the ems. *Journal of International Money and Finance*, *29*(8), 1652–1669.

Delli Gatti, D., Di Guilmi, C., Gaffeo, E., Giulioni, G., Gallegati, M., & Palestrini, A. (2004). Business cycle fluctuations and firms' size distribution dynamics. *Advances in Complex Systems*, *7*(02), 223–240.

Dosi, G., Fagiolo, G., & Roventini, A. (2010). Schumpeter meeting keynes: A policy-friendly model of endogenous growth and business cycles. *Journal of Economic*

Dynamics and Control, *34* (9), 1748–1767.

Dosi, G., Marsili, O., Orsenigo, L., & Salvatore, R. (1995). Learning, market selection and the evolution of industrial structures. *Small Business Economics*, *7* (6), 411–436.

Fabretti, A. (2013). On the problem of calibrating an agent based model for financial markets. *Journal of Economic Interaction and Coordination*, *8* (2), 277–293.

Fagiolo, G., Guerini, M., Lamperti, F., Moneta, A., Roventini, A., et al. (2017). Validation of agent-based models in economics and finance. *LEM Papers Series*, *23*.

Fujiwara, Y., Aoyama, H., Di Guilmi, C., Souma, W., & Gallegati, M. (2004). Gibrat and pareto–zipf revisited with european firms. *Physica A: Statistical Mechanics and its Applications*, *344* (1-2), 112–116.

Gabaix, X., Gopikrishnan, P., Plerou, V., & Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature*, *423* (6937), 267.

Gallegati, M., Giulioni, G., & Kichiji, N. (2003). Complex dynamics and financial fragility in an agent-based model. *Advances in Complex Systems*, *6* (03), 267–282.

Gatti, D. D., Di Guilmi, C., Gaffeo, E., Giulioni, G., Gallegati, M., & Palestrini, A. (2005). A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility. *Journal of Economic behavior & organization*, *56* (4), 489–512.

Gatti, D. D., Gallegati, M., Giulioni, G., & Palestrini, A. (2003). Financial fragility, patterns of firms' entry and exit and aggregate dynamics. *Journal of economic behavior & organization*, *51* (1), 79–97.

Goldbaum, D., & Mizrach, B. (2008). Estimating the intensity of choice in a dynamic mutual fund allocation decision. *Journal of Economic Dynamics and Control*, *32* (12), 3866–3876.

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of applied econometrics*, *8* (S1), S85–S118.

Grazzini, J., Richiardi, M. G., & Tsionas, M. (2017). Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control*, *77*, 26–47.

Guerini, M., & Moneta, A. (2017). A method for agent-based models validation. *Journal of Economic Dynamics and Control*, *82*, 125–141.

Heij, C., De Boer, P., Franses, P. H., Kloek, T., Van Dijk, H. K., et al. (2004). *Econometric methods with applications in business and economics*. OUP Oxford.

Kukacka, J., & Barunik, J. (2017). Estimation of financial agent-based models with simulated maximum likelihood. *Journal of Economic Dynamics and Control*, *85*, 21–45.

Lamperti, F. (2016). Empirical validation of simulated models through the gsl-div: an illustrative application. *Journal of Economic Interaction and Coordination*, 1–29.

Lamperti, F. (2018). An information theoretic criterion for empirical validation of simulation models. *Econometrics and Statistics*, *5*, 83–106.

Lamperti, F., Roventini, A., & Sani, A. (2018). Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control*, *90*, 366–389.

Marks, R. E. (2013). Validation and model selection: Three similarity measures compared. *Complexity Economics*, *2*(1), 41–61.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, 995–1026.

Ormerod, P., & Rosewell, B. (2009). Validation and verification of agent-based models in the social sciences. In *Epistemological aspects of computer simulation in the social sciences* (pp. 130–140). Springer.

Pijpers, F. P. (2018, January). *Applications of complexity theory in official statistics.* (Discussion paper)

Richiardi, M., Leombruni, R., Saam, N., & Sonnessa, M. (2006). A common protocol for agent-based social simulation.

Stanley, M. H., Amaral, L. A., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., ... Stanley, H. E. (1996). Scaling behaviour in the growth of companies. *Nature*, *379*(6568), 804–806.

Stanley, M. H., Buldyrev, S. V., Havlin, S., Mantegna, R. N., Salinger, M. A., & Stanley, H. E. (1995). Zipf plots and the size distribution of firms. *Economics letters*, *49*(4), 453–457.

Stern, S. (2000). Simulation-based inference in econometrics: motivation and methods. *Simulation-Based Inference in Econometrics: Methods and Applications', Cambridge*, 9–37.

Sutton, J. (1997). Gibrat's legacy. *Journal of economic literature*, *35*(1), 40–59.

Winker, P., & Gilli, M. (2004). Applications of optimization heuristics to estimation and modelling problems. *Computational Statistics & Data Analysis*, *47*(2), 211–223.

Winker, P., Gilli, M., & Jeleskovic, V. (2007). An objective function for simulation based inference on exchange rate data. *Journal of Economic Interaction and Coordination*, *2*(2), 125–145.