Macroeconomic implications of labor market frictions
Sedláek, P.

Link to publication

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Fluctuations in economic activity, business cycles, are a fact of life. An important factor shaping the character of economic fluctuations is the labor market. However, the labor market does not operate smoothly. Neither workers, nor firms are all the same. Therefore, it often requires a substantial amount of time and resources to look for a job or a suitable worker and to agree on the terms of employment. These search and matching frictions are the reason why unemployed workers and vacant jobs coexist.

This thesis focuses on business cycles, the welfare costs related to them, and how developments in the labor market impact the overall economy. In particular, it provides a theoretical model showing that fluctuations in economic activity can be very costly for society, because they can decrease the level of output. Moreover, empirical results suggest that increased severity of search and matching frictions can explain up to one fifth of the unemployment run-ups during the most severe recessions. Finally, recent studies have shown that young firms are important for aggregate job creation. Based on a theoretical model, in which firm age is the main driver of a firm’s growth, the thesis shows that government support of business startups increases output. On the contrary, subsidizing existing firms reduces aggregate output as the selection process of successful firms is disrupted and average firm productivity declines.

Petr Sedláček (1982) obtained his MPhil degree in econometrics and macroeconomics at the Tinbergen Institute in Amsterdam. He also holds an MA degree in Economics of international trade and European integration from the Joint European Studies Program and an MSc degree in Economic policy and macroeconomic analysis from the University of Economics in Prague. In 2008 Petr began work on his PhD thesis at the University of Amsterdam, where he currently holds a postdoctoral researcher position.

MACROECONOMIC IMPLICATIONS OF LABOR MARKET FRICTIONS
Macroeconomic Implications of Labor Market Frictions

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 9 december 2011, te 11:00 uur

door

Petr Sedláček

geboren te Praha, Tsjechische Republiek
Promotiecommissie

Promotor: Prof. dr. W.J. den Haan

Overige Leden: Prof. dr. P.A. Gautier
dr. B. Hobijn
dr. F.J.G.M. Klaassen
Prof. dr. S.J.G. van Wijnbergen

Faculteit Economie en Bedrijfskunde
Věnováno mámě a tátovi
To mum and dad
Acknowledgements

In August 2006 I arrived in The Netherlands to begin my graduate studies at the Tinbergen Institute. After the first month, when I was still regularly wearing shorts, I came to two conclusions: The Netherlands is a country with pretty good weather and studying at the Tinbergen Institute will be hard work. Only one of these turned out to be true.

After two years of course work I joined the University of Amsterdam and started working on my thesis. Clear requirements on what and when to study were gone and in their place came the scary phenomenon of independent research. I would like to thank my advisor, Wouter den Haan, for guiding me through the endless options of this new reality and for sparking my fascination with macroeconomics. I am grateful for his advice and encouragement, the unprecedented attention he paid to my texts, his availability for discussions (related to research) and for pointing out that sometimes I just ”wasn’t listening”. My interaction with Wouter, however, did not end at the university doors. Not only did he often come to parties, eventually growing cautious when I was in charge of the refreshments, but he also generously offered his house to me and my girlfriend Emily, when he was away. Wouter, thanks for all that you have done for me!

Over the course of the last five years in Amsterdam I have met many people, mostly in the dual role of colleagues and fellow beer-drinking buddies, but also outside of the world of economics. From the first group I am especially glad that Zoltan and Matija were my flat and office mates, respectively. I enjoyed being the occasional wingman for Markus, conducting banker-talk with Hulusi and Emrah, analyzing basketball with Robert, owning a boat with Pia, having laughs with Marloes and Fujin, and losing money to all the poker friends. At the university I am happy to have met, learned from and had drinks with Alex, Arturo, Boe, Christian, Franc, Joris, Massimo, Naomi, Pieter, Siert, Sweder, Tim, Vincent, Ward and many others.

From the second group, the main thanks goes to my current flat mate (and future
wife), Emily. The English in my papers is much better owing to the proofreading skills of Emily and her father Bob. More importantly though, I have benefited unmeasurably from Emily’s ability to put my stress regarding iterations that refused to converge into perspective of real life. Moreover, Emily considerably increased my bargaining power (from zero) against Wouter in discussions over the use of my time on Saturdays. Thanks to Emily, my social circle was also enlarged by friends who are less enthusiastic about economic models. I am especially happy to have met my paranymph Maurice, his wife Asifa and their newborn daughter Amber Zahra, who made it into the acknowledgements just before this thesis went to print.

Finally, I would like to thank my family. I am extremely grateful to my parents, who have always encouraged and supported me and have never doubted my decision, even though it meant that I moved to a different country. I am glad that my brother, a lawyer, managed to work longer hours than me throughout the years, making me feel better about my personal life. Finally, I am happy for the support I got from all my grandparents, and for their continuous worry about whether or not I am studying hard enough.

October 2011, Amsterdam
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5.4 Model performance</td>
<td>109</td>
</tr>
<tr>
<td>3.6 Endogenous separations with firing costs</td>
<td>111</td>
</tr>
<tr>
<td>3.6.1 Matching model with firing costs</td>
<td>112</td>
</tr>
<tr>
<td>3.6.2 Wage bargaining</td>
<td>114</td>
</tr>
<tr>
<td>3.6.3 Closing the model</td>
<td>115</td>
</tr>
<tr>
<td>3.6.4 Match efficiency variation</td>
<td>115</td>
</tr>
<tr>
<td>3.6.5 Calibration</td>
<td>116</td>
</tr>
<tr>
<td>3.6.6 Model performance</td>
<td>117</td>
</tr>
<tr>
<td>3.6.7 Model-based match efficiency</td>
<td>118</td>
</tr>
<tr>
<td>3.7 Concluding remarks</td>
<td>119</td>
</tr>
<tr>
<td>Appendix to Chapter 3</td>
<td>121</td>
</tr>
<tr>
<td>3.A The Kalman filter</td>
<td>121</td>
</tr>
<tr>
<td>3.B The extended Kalman filter</td>
<td>121</td>
</tr>
<tr>
<td>3.C Diagnostic tests</td>
<td>122</td>
</tr>
<tr>
<td>3.C.1 Serial correlation</td>
<td>123</td>
</tr>
<tr>
<td>3.C.2 Homoscedasticity</td>
<td>123</td>
</tr>
<tr>
<td>3.C.3 Normality</td>
<td>123</td>
</tr>
<tr>
<td>3.D More robustness checks and further details</td>
<td>124</td>
</tr>
<tr>
<td>3.D.1 Different state processes</td>
<td>124</td>
</tr>
<tr>
<td>3.D.2 Estimating on subsamples and with different frequencies</td>
<td>126</td>
</tr>
<tr>
<td>3.D.3 Alternative functional forms</td>
<td>126</td>
</tr>
<tr>
<td>3.D.4 Alternative estimation procedure</td>
<td>128</td>
</tr>
<tr>
<td>3.D.5 Using information from the vacancy yield</td>
<td>129</td>
</tr>
<tr>
<td>3.D.6 Extending the sample period</td>
<td>129</td>
</tr>
<tr>
<td>3.D.7 Monte Carlo experiment</td>
<td>130</td>
</tr>
<tr>
<td>3.E Endogeneity</td>
<td>131</td>
</tr>
<tr>
<td>3.F Effects of firing costs</td>
<td>131</td>
</tr>
<tr>
<td>3.F.1 The case of a uniform distribution</td>
<td>132</td>
</tr>
<tr>
<td>4 Firm Age, Business Cycles, and Aggregate Labor Market Dynamics</td>
<td>135</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>135</td>
</tr>
<tr>
<td>4.2 Related research</td>
<td>140</td>
</tr>
<tr>
<td>4.3 Firm age and job creation in the U.S. economy</td>
<td>142</td>
</tr>
<tr>
<td>4.3.1 Definitions</td>
<td>143</td>
</tr>
<tr>
<td>4.3.2 Firm age characteristics across business cycles</td>
<td>143</td>
</tr>
<tr>
<td>4.3.3 Business cycle characteristics</td>
<td>148</td>
</tr>
<tr>
<td>4.3.4 Summary of empirical facts</td>
<td>152</td>
</tr>
</tbody>
</table>
List of Figures

2.1 Projects operating in a world without business cycles .............. 23
2.2 Effect of arbitrarily small business cycles on boundary projects .... 24
2.3 Projects affected by business cycles .................................. 35
2.4 Projects affected by business cycles in the presence of financial frictions 59
2.5 Fraction of total wages earned by recently non-employed workers ...... 74

3.1 Kalman smoothed states: benchmark ................................. 86
3.2 Job finding probability .................................................. 89
3.3 Unemployment rate ..................................................... 89
3.4 Decomposition of the cumulative drop in the job finding rate .......... 91
3.5 Indication of the degree of non-linearity ................................ 94
3.6 Match efficiency, robustness: alternative functional forms ........... 96
3.7 Vacancies, robustness: alternative estimation .......................... 97
3.8 Match efficiency, robustness: alternative estimation .................... 97
3.9 Match efficiency, robustness: using vacancy yield data ................ 99
3.10 Match efficiency, robustness: extended sample ....................... 99
3.11 Match efficiency, robustness: Monte Carlo experiment ................. 101
3.12 IRFs to a positive one-standard-deviation technology shock, 1st calibration 110
3.13 Idiosyncratic productivity cut-offs, cases with and without firing costs . 112
3.14 IRFs to a positive one-standard-deviation technology shock, model with firing costs ................................................. 117
3.15 Estimated match efficiency and its model prediction .................... 119
3.16 Kalman smoothed states: benchmark and other specifications ........ 125
3.17 Kalman smoothed states: benchmark and subsamples .................. 127

4.1 Firm shares according to size and age ................................... 144
4.2 Differential employment growth rate and the unemployment rate ...... 149
4.3 Actual-counterfactual unemployment decomposition; old firms ........ 150
4.4 Model time-line ......................................................... 154
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5</td>
<td>True value function and its perturbation approximation; illustration</td>
<td>165</td>
</tr>
<tr>
<td>4.6</td>
<td>Comparing the true and approximate solutions</td>
<td>166</td>
</tr>
<tr>
<td>4.7</td>
<td>Firm shares according to age for small firms, model and data</td>
<td>173</td>
</tr>
<tr>
<td>4.8</td>
<td>Firm shares according to size for young firms</td>
<td>173</td>
</tr>
<tr>
<td>4.9</td>
<td>Firm exit rates according to age, model and data</td>
<td>175</td>
</tr>
<tr>
<td>4.10</td>
<td>Job creation rates according to age, model and data</td>
<td>175</td>
</tr>
<tr>
<td>4.11</td>
<td>Job destruction rates according to age, model and data</td>
<td>176</td>
</tr>
<tr>
<td>4.12</td>
<td>Decomposition of IRF of the unemployment rate</td>
<td>178</td>
</tr>
<tr>
<td>4.13</td>
<td>IRFs to an aggregate productivity shock</td>
<td>180</td>
</tr>
<tr>
<td>4.14</td>
<td>KS algorithm, accuracy plot for $q$</td>
<td>191</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Impact of arbitrarily small business cycles for projects at the boundary ........ 32
2.2 Impact of business cycles for projects at the boundary ....................... 33
2.3 Welfare costs of business cycles .................................................. 47

3.1 Parameter estimates ................................................................. 85
3.2 Contributions to job finding probability volatility ............................... 88
3.3 Parameter values: 1st calibration .................................................. 108
3.4 Parameter values: 2nd calibration ............................................... 109
3.5 Standard deviations of variables: US data and different calibrations .......... 111
3.6 Parameter values: 1st calibration with firing costs ............................. 116
3.7 Standard deviations of variables: US data and model with firing costs ...... 118
3.8 Parameter estimates: different state space representations .................... 124
3.9 Parameter estimates: different subsamples ..................................... 126
3.10 Parameter estimates in vacancy yield equation ................................. 129

4.1 Firm exit/entry rates and shares according to firm age, 1992-2009 ........... 146
4.2 Job creation and destruction rates by firm age, 1992-2009 .................... 147
4.3 Job creation and destruction shares by firm age, 1992-2009 .................... 147
4.4 Unemployment rate decomposition; 2008-2009 .................................. 152
4.5 Model parameters ................................................................. 168
4.6 Firm growth regression on size and age ........................................ 171
4.7 Standard deviations and cross-correlations of selected variables; U.S. data 182
4.8 Standard deviations and cross-correlations of selected variables; HF model 183
4.9 Standard deviations and cross-correlations of selected variables; RF model 184
4.10 Results of subsidizing young firms .............................................. 186
Chapter 1

Introduction and Overview

And it came to pass at the end of two full years, that Pharaoh dreamed: and, behold, he stood by the river. And, behold, there came up out of the river seven well-favored kine and fat-fleshed; and they fed in a meadow. And, behold, seven other kine came up after them out of the river, ill-favored and lean-fleshed; and stood by the other kine upon the brink of the river. And the ill-favored and lean-fleshed kine did eat up the seven well-favored and fat kine. So Pharaoh awoke.

Genesis 41:1-4

This excerpt from the Bible is from the story when Joseph was imprisoned in Egypt. During that time Pharaoh had the dream, but did not understand what it meant and also none of his wise men was able to interpret it. Pharaoh learned of a prisoner, Joseph, and his skills in interpreting dreams and therefore had him summoned. Joseph interpreted the dream as foretelling that seven years of abundance would be followed by seven years of famine and advised Pharaoh to store surplus grain during the years of abundance. Pharaoh did as Joseph advised and Egypt managed to survive the seven years of famine that followed.

While this story was not the reason why I became interested in economics, it probably is the first documented business cycle forecast. More importantly, it shows that people have been concerned with fluctuating economic conditions, business cycles, for a long time. This interest has not died out over the centuries. On the contrary, after two decades of relatively smooth economic development, the most recent crisis has sparked
new intensive research in this topic.

An important factor in shaping the character of fluctuations in economic activity is the labor market. People work in order to be able to buy food, a house or to purchase any other good or service. For some people, even work alone brings pleasure. PhD students are a great example of such individuals. On the other hand, firms need workers to operate their machines in factories, to drive trucks that deliver the goods to supermarkets, to sell those goods to customers etc. However, the labor market, as many other markets, does not operate smoothly. Neither workers, nor firms are all the same. Therefore, it often requires a substantial amount of time and effort to look for a job or a suitable worker and to agree on the terms of employment. These search and matching frictions are the reason why unemployed workers and vacant jobs coexist.¹

My thesis focuses on business cycles, the welfare costs related to them, and how developments in the labor market impact the overall economy. I pay special attention to the different characteristics of both workers and firms in determining aggregate labor market outcomes. The ease (or difficulty) with which workers are able to find jobs or the way different firms react to overall business conditions both drive movements in aggregate employment and output. Understanding these links can then help evaluate alternative policy measures aimed at boosting employment or reducing the adverse effects of economic downturns.

Overview

While business cycles seem to interest the general public and there is an underlying feeling that they are detrimental for the economy, economists have struggled to show that fluctuations in economic activity actually decrease the wellbeing of society. Intuitively, is it not the case that times of recessions are "averaged out" by high economic activity during booms? Why should then governments and central banks be so concerned about "smoothing out" the cycle? In an influential contribution, Robert Lucas (1987) calculated how much better off an economy would be if its fluctuating consumption stream were replaced by its average consumption level. Put differently, Lucas asked

¹They are also the reason why Peter Diamond, Dale Mortensen and Christopher Pissarides obtained their Nobel prize in 2010.
what fraction of consumption would a person living in the postwar U.S. economy be willing to give up in return for a life in a world that has the same average consumption level, but no economic fluctuations. The answer is *one tenth of a percentage point*. This means that if our hypothetical person has an average consumption level of say $100, he is willing to give up 10 *cents* of his consumption each period in return for a life without business cycles. Lucas’s conclusion turned out to be frustratingly difficult to disprove in realistic models and these tiny costs of business cycles have intrigued the profession ever since.

The second chapter of my thesis, which is joint work with Wouter den Haan, focuses on the question of why business cycles could be more costly for the economy than Lucas predicted. The novelty of our approach is that we consider the interaction between costs to entry and a friction resulting in inefficient project shutdowns.

To make our argument, we compare a typical economy we live in with a hypothetical world without fluctuations in economic activity. In both economies, there is a range of projects (either jobs, or entire firms) that create output. Each project is different in terms of its productivity and its startup costs.

Another important ingredient in our framework is the belief that the world we live in does not always operate efficiently. Specifically, decisions to startup or continue a project are subject to ”inefficiencies” that force some profitable projects to shut down. For example, entrepreneurs typically seek external funds for their business. Often it is quite difficult for an entrepreneur to convince the bank of the project’s profitability, especially during bad times. Not being able to obtain the funds required to run a profitable project, either because the entrepreneur has no credit history, or simply because the bank does not believe his claims, is an example of an inefficiency.

In a world without business cycles, only projects with productivity above a certain threshold are able to overcome the inefficiencies and operate. Projects below that threshold do not even startup as they would be immediately forced out of business by the inefficiency. In a world with business fluctuations, this threshold is lower in booms and higher in recessions. In other words, it is relatively easier to overcome the inefficiency in good times and relatively more difficult to overcome it in bad times (it

\footnote{For Douglas Adams fans, this is as close as one can get to the question, to which the answer is 42.}
is harder for the entrepreneur to convince the bank of a project’s profitability during a downturn and vice versa). However, this means that some projects that were able to happily operate in a world without business cycles are forced by the inefficiency to shut down during recessions, which substantially lowers their expected duration and thus their expected benefits. It makes a huge difference whether you think you can operate your business for the rest of your life, or whether you think you will only survive until the next recession.

So what? On the one hand, some projects which can operate in a world without business cycles are forced out of business during recessions. On the other hand, however, some projects which could not operate in a world without fluctuations are able to startup during booms as they have an easier time overcoming the inefficiency. It seems that things average out again and the costs of business cycles will be negligible also in this model.

However, so far I have ignored the last crucial ingredient in our story - the startup costs. Remember that each project differs not only in terms of productivity, but also in terms of startup costs. Given that any reasonable entrepreneur compares the expected benefits of starting up a project with the associated costs, a reduction in the expected duration of a project can be fatal. Indeed for some projects which can operate without the presence of business cycles, but which are forced to shut down during recessions, the expected benefits become too low to cover the associated costs. Therefore, in the presence of business cycles it no longer makes sense to startup such projects.

Put differently, the projects which can operate in a world without business fluctuations, but do not even startup in the presence of economic cycles, permanently reduce the level of output. Our estimates suggest that such a permanent decrease in output can easily exceed several percent and in some cases can be even much larger. Hence, the interaction between inefficiencies and startup costs in our simple model generates costs of business cycles that are several orders of magnitude higher than the estimate of Lucas.

Having established that fluctuations in economic activity may come with large costs to society, Chapter 3 looks at how the severity of search and matching frictions, \textit{match efficiency}, affects the unemployment rate and in turn the magnitude of business
cycles. For example, the most recent recession was accompanied with a peculiar fact which raised many eyebrows within the economic profession. As in any other recession, unemployment shot up and the number of vacant jobs decreased. However, typically as the economy starts recovering the number of vacant jobs starts increasing pulling unemployment down as it is easier for people to find work again. This time, while vacancies increased during the recovery, the unemployment rate remained stubbornly high. Thus, while there seemed to have been jobs in the economy ready to be filled and many unemployed ready to work, the two somehow did not click.

One of the culprits in these developments is a reduction in match efficiency. For instance, a lot of construction workers lost their jobs during the recession. During the recovery, many jobs in health care and education sectors were opened up, but obviously these were not suitable for the unemployed construction workers. Another reason why mismatch might have increased has to do with problems in the housing market. During the housing slump, it became more difficult for people to move for jobs as they had a harder time refinancing their mortgages (as analyzed for instance by Sterk, 2010). These are examples of why search and matching frictions, match efficiency, can fluctuate over the business cycle.

Chapter 3 focuses on the ease (or difficulty) with which unemployed people find jobs and how it is affected by match efficiency and in turn how this affects the unemployment rate. Specifically, using data from the U.S., I employ econometric techniques to estimate how the probability with which an unemployed worker finds a job is affected by match efficiency. Although match efficiency is not directly observed in the data, the econometric technique is able to extract its estimate. It turns out that match efficiency is estimated to move together with the business cycle. This means that recessions are times when unemployed workers have a harder time finding jobs not only because there are less vacant jobs and more unemployed competing for them, but also because search and matching frictions become more severe. Quantitatively, match efficiency accounts for about 1/4 of the movements in the probability with which an unemployed person finds a job. Put into a different perspective, during the most severe recessions the unemployment rate increased on average by 3.5 percentage points. 1/5 of this run-up is due to a drop in match efficiency alone.
To further analyze the role of match efficiency in determining aggregate labor market dynamics, Chapter 3 also examines a simple theoretical model. I show that measured match efficiency moves together with the business cycle in a model incorporating search and matching frictions, workers with different skills and firms which are free to flexibly hire and dismiss employees.

Finally, Chapter 4 examines the role of firm heterogeneity for developments on the labor market. There is a long list of studies focusing on the link between a firm’s size and its growth. The general conclusion of these studies is that small firms grow faster than larger ones and that small firms are important for job creation in the economy. However, recent studies suggest that it is not firm size, but rather the age of a businesses that matters for firm growth. This questions the current way of thinking and poses new challenges as the link between a firm’s age and its growth has received very little attention. In Chapter 4, I explicitly investigate (both empirically and theoretically) the role of firm age for business growth and the importance of young firms for aggregate labor market dynamics.

Recent studies focusing on firm age show that young firms have higher exit rates and conditional on survival they grow faster than older ones, young firms create relatively more jobs, job creation and destruction rates fall with firm age and while young firms are mainly small, small firms are not all young. I extend these findings by showing that, compared to old firms, employment growth in young businesses is more volatile and that business start-ups are important for determining unemployment rate developments.

The main contribution of Chapter 4 is a novel general equilibrium model incorporating labor market frictions and heterogeneous firms which aims to capture the above empirical facts. Firms in this model differ in their productivity which evolves persistently over time. They are free to enter the economy and if their individual business conditions are so bad that it no longer pays to operate, they choose to shut down. Furthermore, depending on the economy wide and firm specific conditions businesses also choose whether to expand or shrink their workforce.

This model is shown to be consistent with the above-mentioned empirical facts relating to firm age. The key to understanding the model’s performance is the inherent selection process of successful firms. Only the relatively more efficient firms are able
to survive, expand and grow old. Therefore, younger firms exhibit a higher risk of shutting down and thus also a higher rate of job destruction. Moreover, younger firms are mainly small and as such they can take advantage of the lower worker turnover and the associated costs (in absolute terms) resulting in them having relatively more resources for expansion compared to older businesses. The model is also consistent with the dynamics of aggregate labor market variables, such as the unemployment rate, the vacancy rate and the probability of finding a job. The model correctly replicates the co-movement of these variables over the business cycle as well as predicting realistic sizes of their fluctuations.

The model is then used to analyze the impact of a government policy aimed at supporting young firms as drivers of job creation. Such measures were recently proposed under the "Startup America" initiative of the White House. Within the model, government intervention can also be justified, since firms receive a relatively small fraction of output and thus entry could be inefficiently low and firm exit inefficiently high.

The results suggest that subsidizing firm entry increases welfare as a higher number of new firms reduces unemployment and increases the level of output. However, if the government focuses its resources only on subsidizing existing young firms welfare decreases. The reason is that such a subsidy enables relatively less productive firms to survive and crowd out entrants. Lower firm entry together with higher survival rates of relatively less efficient firms results in lower average firm productivity. Moreover, the reallocation process of workers from relatively less productive firms to more efficient businesses is hampered. The overall impact of these effects is that the output level in the economy falls and thus welfare is reduced. Hence, the model suggests that government policies should focus on reducing barriers to firm entry, quickly withdrawing their support thereafter so that the economy can pick its own winners.
Chapter 2

Inefficient Continuation Decisions, Entry Costs, and the Cost of Fluctuations

Abstract

Fluctuations in firms’ revenues reduce firms’ survival chances and are costly from a social welfare point of view even when agents are risk neutral if (i) the decision to continue operating a firm is not efficient so that fluctuations lead to inefficient reductions in firms’ life expectancy and (ii) the shortening of firms’ life expectancy reduces firm entry due to (for example) the presence of entry costs. Conservative estimates for the per-period costs of moderate fluctuations, like business cycles, are between 0.18% and 2.12% of GDP, but some calculations result in estimates exceeding 30% of GDP.\

2.1 Introduction

Fluctuations are a fact of life. They come in many varieties such as idiosyncratic, sectoral, regional, and aggregate fluctuations. This paper documents that even modest fluctuations, like business cycles, are quite costly in a very simple framework with risk neutral agents and the following quite standard features. First, creating a project requires a fixed startup cost. Second, projects are not all the same. Here projects have different productivity levels and different startup costs, but our mechanism would

*This chapter is joint work with Wouter J. den Haan.
also operate for alternative forms of heterogeneity. Third, the decision to start or continue operating an existing project is subject to inefficiencies, that is, "frictions" prevent some profitable projects from operating. Fourth, the fluctuations affect the severity of the inefficiency, either positively or negatively. Using this framework, we show that fluctuations are costly because they deter entry and lower the average level of output produced. Whereas it has been difficult to develop models in which moderate fluctuations have non-negligible per capita costs, the costs of business cycles in our framework correspond to a permanent drop in output that can easily exceed several percentage points and possibly could be substantially higher. In contrast, the classic Lucas (1987) paper reports an estimate for the costs of business cycles that is less than one tenth of a percentage point of consumption when the coefficient of risk aversion is equal to 10.

Before providing intuition for the mechanism, we motivate the key underlying features of our framework. Starting a "project", whether it is a company, a plant, or a job, is almost never costless and entry/startup costs are part of many economic models. Regarding the inefficient decision to start or continue operating a project, one can think of the inability to obtain financing, the inability to motivate workers or prevent them from shirking, the inability to write contracts that prevent the employer from exploiting the employee, or the inability to adjust real wages sufficiently downward. Finally, it is a natural feature of models with inefficiencies that the impact of the inefficiency fluctuates over time. This feature can be illustrated using the net worth channel. In a recession, firms' net worth levels drop. At these lower equity levels, firms are more likely to exploit the convexity (due to limited liability) in the payoffs and increase the amount of risk undertaken. Consequently, it will be more difficult for firms to obtain credit. Similarly, the ability for financial intermediaries to channel funds from savers to firms may very well be weakened during a recession due to an increase in (the impact of) frictions.

The reason why fluctuations are costly in our framework is quite intuitive. There

\footnote{As for example in Townsend (1979), Bernanke and Gertler (1989) Kiyotaki and Moore (1997), Carlstrom and Fuerst (1998), and Bernanke, Gertler, and Gilchrist (1999).}

\footnote{See Shapiro and Stiglitz (1984).}

\footnote{See Ramey and Watson (1997).}
is a stochastic aggregate variable $\Phi_p$ that affects firms’ productivity levels, and firms’ revenues determine to what extent projects are affected by the inefficiency. Consider projects whose characteristics are such that they are at risk of inefficiently being discontinued. Positive (negative) movements in $\Phi_p$ decrease (increase) the number of projects that make inefficient continuation decisions. If there are no entry costs, then there is no robust reason why the positive effects of an increase in $\Phi_p$ would not offset the negative effects of a decrease in $\Phi_p$.\(^4\) With entry costs, however, this is no longer true. The reason is that fluctuations in $\Phi_p$ reduce projects’ expected lifetime; low surplus projects are now terminated when the economy enters a recession. This means that startup costs have to be paid more often. More importantly, fluctuations in $\Phi_p$ bring about a reduction in the number of created projects. These firms are marginal in terms of being able to overcome the inefficiency, but have positive value from a social welfare point of view. Consequently, their disappearance is welfare reducing.

Our model does not rely on high risk aversion to explain why moderate fluctuations like business cycles are costly. In fact, we assume that agents are risk neutral. As pointed out in Lucas (2003), if agents are highly risk averse, then the question arises why high risk aversion does not show up in, for example, the diversification of individual portfolios, the level of insurance deductibles, or the wage premiums of jobs with high earnings risk.

The framework used is simple and contains only a small set of structural parameters and for most of them it is not difficult to consider a set of plausible values. One important ingredient in our quantitative assessment is the mass of projects that are not created in a world with business cycles, but are created in a world without business cycles. Since these projects are not observed in the actual world with business cycles, we have to find a way to estimate this mass. Our identification procedure consists of two elements. First, economic theory pins down exactly the productivity levels and startup costs at which entry would occur, also in the world without business cycles that we do not observe. Second, in a way that will be made more precise below, we determine the mass of these projects basically by assuming that there are no sudden

\(^4\)As in Caballero and Hammour (2005), the inefficiencies would make recessions more contractionary and booms more expansionary. Our paper, however, is about the costs of business cycles, not about the costs of recessions.
changes in how different types of projects are distributed in the relevant area.

Our paper fits into a line of research that investigates the effect of uncertainty on the level or growth rate of output, which Lucas (2003) refers to as "... a promising frontier on which there is much to be done".\textsuperscript{5} Besides the assumption of linear utility, our framework differs from related papers in that we focus on different characteristics to generate the relationship between volatility and the level of real activity, namely entry costs and an inefficiency in the decision to operate a project; two simple features often found in the literature.

The rest of this paper is organized as follows. In section 2.2, we develop our framework. In section 2.3, we highlight a key discontinuity in our framework, namely that business cycles of arbitrarily small magnitude have a large impact on some projects. In section 2.4, we discuss the impact of business cycles of ordinary magnitude. In section 2.5, we discuss which elements are necessary for our mechanism to work. In section 2.6, we discuss our approach to obtain quantitative estimates for the costs of business cycles. The results are presented in section 2.7. In section 2.8, we discuss modifications of the model that lead to larger costs of business cycles. The literature is discussed in section 2.9 and the last section concludes.

2.2 Model

In this section, we present a generic model that is characterized by startup costs and an inefficiency in the decision to operate and continue the project. The model developed here is very simple and abstract, which makes it easy to explain why fluctuations are costly if there are both startup costs and inefficient continuation decisions. Although the model is abstract, we think that the highlighted interaction between startup costs and inefficient continuation decisions is so fundamental that it will play a role in many fully specified models with these two features. Support for this assertion is provided in Appendix 2.A in which we discuss two different models that incorporate an agency problem into a business cycle model.

\textsuperscript{5}In section 2.9, we discuss the related theoretical literature and provide references for papers that establish empirical support for the view that business cycles do not leave the long-run growth path unchanged.
2.2.1 Agents

The economy is inhabited by utility-maximizing risk-neutral agents. We assume that agents are risk neutral to accentuate that business cycles in our model are costly even if agents are risk neutral.

2.2.2 Productivity of active and inactive projects

There is a continuum of projects indexed by $i$. Projects are owned by a single agent. We make this last assumption to make clear that the presence of the highlighted mechanism does not rely on inefficiencies in how the costs and/or revenues of the projects are shared between different agents. Although not important for the qualitative results, such additional inefficiencies can make the mechanism quantitatively more important. This is documented in section 2.8.

Project $i$ is characterized by a startup cost, $\phi_c(i)$, and a productivity parameter, $\phi_p(i)$; both are assumed to be constant through time. A project can be active or inactive. To produce market output the project has to be activated, which requires paying the startup costs. Each period, individual active projects could be afflicted by an exogenous shock that will inactivate the project. This shock occurs with probability $1 - \rho$. The project can be reactivated by paying the startup costs again. Production of active project $i$, $y_t(i)$, is given by

$$y_t(i) = \phi_p(i)\Phi_{p,t},$$

(2.1)

where $\Phi_{p,t}$ is aggregate productivity. From now on, we suppress the $i$ index, but the reader should keep in mind that $\phi_p$ and $\phi_c$ vary across projects and all the other variables, including $\Phi_{p,t}$, do not.

When a project is not active, then it does not generate any market production. The project’s revenues are in this case equal to $\mu$. We assume that there are no transfers, such as subsidies, to agents with an inactive project. This means that $\mu$ only consists of benefits such as home production or the value of leisure. We do not include transfers in the analysis, because this additional inefficiency would only distract attention from the main mechanism.
2.2.3 Aggregate fluctuations

Two different assumptions about $\Phi_{p,t}$ are considered. Under the first assumption, $\Phi_{p,t}$ is constant through time and equal to 1. In this case, the projects are heterogeneous, but face an unchanging macroeconomic environment. Under the second assumption, $\Phi_{p,t}$ is a stochastic variable that varies across time with an unconditional mean equal to 1. The most common interpretation of $\Phi_{p,t}$ is that it is an aggregate shock that is common to all projects. In this case, fluctuations in $\Phi_{p,t}$ correspond to business cycle fluctuations. For simplicity, we assume that $\Phi_{p,t}$ can take on only two values, $\Phi_+$ in a boom and $\Phi_-$ in a recession. The probability of transitioning out of a boom, $1 - \pi$, is equal to the probability of transitioning out of a recession. This implies that the expected durations of staying in a boom and a recession are equal to each other.\(^6\) Since $E[\Phi_{p,t}] = 1$, it also implies that $\Phi_+ - 1 = 1 - \Phi_- = \Delta\phi_p$.

2.2.4 Inefficiencies and the operating/continuation decision

Two decisions have to be made before production can take place. First, the decision has to be made whether to activate the project by paying the startup cost $\phi_c$. Second, the decision has to be made whether to operate the project and continue to the next period.\(^7\) We start with the second decision.

The key aspect of the model is that operating the project is hampered by an inefficiency. The inefficiency is that a project can only continue operating if the period $t$

\(^6\)NBER recessions are shorter than NBER expansions. But the classification used by the NBER is not symmetric. The reason is that the NBER does not classify a recession as a period when observed growth is below a long-run average, but as a period when observed economic activity is sufficiently bad. For our purpose, it makes more sense to consider HP-filtered residuals. Within our sample from 1947Q1 to 2010Q4, we find that from 1949Q1 to 2008Q3 there are 16 complete recessions (periods with negative HP-filtered residuals surrounded by positive residuals) and 16 complete booms (periods with positive HP-filtered residuals surrounded by negative values). The average durations are equal to 7.1 and 7.9 quarters for recessions and booms, respectively. This corresponds roughly to the assumption adopted here that the expected duration of a boom is roughly equal to the expected duration of a recession.

\(^7\)Our model shares these two decisions with labor market matching models. In fact, our setup is similar to the one in Robin (2010). We differ from this literature in that we simplify the analysis by setting the matching probability equal to one and we introduce an inefficiency in the continuation decision.
revenues are sufficiently high. In particular, we require that

\[ \phi_p \Phi_{p,t} \geq \chi_t. \]  

(2.2)

We will refer to the requirement given in equation (2.2) as the efficiency requirement. Projects that do not satisfy this condition cannot operate and generate \( \mu \) instead of \( \phi_p \Phi_{p,t} \). The value of \( \chi_t \) could be a constant, but we allow for the possibility that \( \chi_t \) varies with \( \Phi_{p,t} \). That is, \( \chi_t = \chi(\Phi_{p,t}) \). When \( \Delta \Phi_p = 0 \), then the value of \( \chi_t \) is equal to \( \chi(1) \). With some abuse of notation we will refer to \( \chi(1) \) as \( \chi \).

**Assumption 2.1.**

\[ \mu < \min \{ \chi(\Phi_-), \chi(\Phi_+) \}. \]  

(2.3)

This assumption ensures that the efficiency requirement plays a role in both a boom and a recession.

It is important to realize that the efficiency requirement is not part of technology, but is capturing a private inefficiency. An inefficiency is a private inefficiency if the participant(s) in the project is (are) worse off if the project does not remain active. The following examples clarify what is and what is not a private inefficiency. Suppose that the owner of an inactive project not only receives \( \mu \), the value generated by an inactive project, but also receives a transfer from the government equal to \( \chi - \mu > 0 \). Then it would be socially inefficient for a project with \( \phi_p = \chi > \mu \) to become inactive, since the government has to finance the transfer to this inactive project, while the inactive project is only producing \( \mu \). But it is not privately inefficient to stop operating, since the owner of the project is not worse off. Now suppose that the government without good reason prohibits projects to operate if \( \chi \) exceeds \( \phi_p \) and \( \chi > \mu \). Then projects with \( \mu < \phi_p < \chi \) would face a private inefficiency. The reason is that the inefficiency makes the beneficiaries of these projects worse off, since it causes these projects to generate \( \mu \) instead of the higher \( \phi_p \).

\(^8\)In Section 2.5, we explain why the inefficiency has to be a private inefficiency.
2.2.5 Interpretation of the inefficiency

To understand the model, the reader may want to hold on to this last simple interpretation of the efficiency requirement. But the mechanism we highlight in this paper does not depend on this particular interpretation. In fact, the mechanism manifests itself in the presence of several different types of inefficiencies. The reason is that models with inefficiencies typically imply a condition like our efficiency requirement. That is, if the firm’s own internal resources are not high enough, then the firm cannot overcome inefficiencies due to, for example, moral hazard problems.

In Appendix 2.A, we describe two explicit dynamic models with agency problems in which agents carefully consider both the current and future consequences of their actions. In both models, the inefficiency leads to a condition equal to or similar to our efficiency requirement. It is easy to think of other types of inefficiencies. For example, another obvious reason for private inefficiencies is the presence of sticky wages, either because of social norms, inefficient bargaining, or efficiency wages.

2.2.6 Inefficiencies and the cut-off level for $\phi_p$

Let $\tilde{\phi}_{p, bc}$ be the value of $\phi_p$ such that equation (2.2) holds with equality when there are business cycles, that is, when $\Delta \phi_p > 0$. Thus,

$$\tilde{\phi}_{p, bc} \Phi_{p,t} = \chi_t \text{ or } \tilde{\phi}_{p, bc} = \frac{\chi_t}{\Phi_{p,t}}.$$

In most business cycle models with inefficiencies, the value of $\chi_t$ would be less cyclical than $\Phi_{p,t}$. The reason for this property is that the value of $\chi_t$ is typically affected by the value of the alternatives to producing market output. For example, this could be the value of the project when resources are diverted from the regular production process. The value of these alternatives is typically assumed to be less sensitive to $\Phi_{p,t}$ than market production, i.e., $\phi_p \Phi_{p,t}$. The value of $\tilde{\phi}_{p,t}$ would then be countercyclical. In this case, the fraction of agents that is affected by the inefficiency decreases in a boom and increases in a recession. Our mechanism does not depend on this property. It also operates if $\chi_t$ is more procyclical than $\Phi_{p,t}$, i.e., when the fraction of agents that is affected by the inefficiency is procyclical. All that is needed for our channel to be
operative is that \( \tilde{\phi}_{p,t} \) is cyclical, either procyclical or countercyclical. The only case that has to be ruled out is that \( \chi_t \) is proportional to \( \Phi_{p,t} \) as stated in the following assumption.

**Assumption 2.2.** One of the following two conditions holds:

\[
\begin{align*}
(i) & \quad \frac{\tilde{\phi}_{p,bc}(\Phi_{p,t})}{\chi_t} = \frac{\chi_t}{\Phi_{p,t}} = \frac{\chi_t(\Phi_{p,t})}{\Phi_{p,t}} \text{ decreases with } \Phi_{p,t} \quad \text{or} \\
(ii) & \quad \frac{\tilde{\phi}_{p,bc}(\Phi_{p,t})}{\chi_t} = \frac{\chi_t}{\Phi_{p,t}} = \frac{\chi_t(\Phi_{p,t})}{\Phi_{p,t}} \text{ increases with } \Phi_{p,t}
\end{align*}
\]

This assumption is trivially satisfied if \( \chi_t \) is constant. To focus the discussion, we describe the framework and the results under the assumption that \( \tilde{\phi}_{p,bc} \) is decreasing with \( \Phi_{p,t} \). The impact of frictions then increases during economic downturns, which we feel is the more natural possibility. Formal results are stated, however, using the more general condition given above.

### 2.2.7 Activation decision

Projects can be activated instantaneously by paying a start-up cost, \( \phi_c \). When the project has been idle for some time, then \( \phi_c \) would have to be paid once more to restart it. That is, one cannot simply mothball the project and restart it as if there had been no interruption. One possible reason for this is that it may take some time and effort before the project is operating at its potential productivity of \( \phi_p \) again.\(^9\)

For a project with productivity level \( \phi_p \) and startup costs \( \phi_c \), activation will occur if

\[
N_{bc}(\phi_c, \phi_p, 1, \Phi_{p,t}) - \phi_c \geq \mu + \beta E_t [N_{bc}(\phi_c, \phi_p, 0, \Phi_{p,t+1})],
\]

where \( E_t \) is the expectation conditional on period-\( t \) information, \( N_{bc}(\phi_c, \phi_p, 1, \Phi_{p,t}) \) is the discounted value of the project’s current and future earnings when the startup costs have been paid, and \( N_{bc}(\phi_c, \phi_p, 0, \Phi_{p,t}) \) is the discounted value of the project’s earnings.

---

\(^9\)For some projects, the value of \( \phi_c \) may be low. For example, in the US car industry it is not uncommon to leave capital idle or underutilized for some time and recall former workers when economic conditions improve. In our benchmark calibration, we set the lower bound of the distribution of \( \phi_c \) equal to zero, which clearly would accommodate the possibility of low (re)starting costs.
when the startup costs have not been paid. When the startup costs are not paid, then it remains possible to start the project at a future date.

Assumption 2.3.

\[(i) \quad 0 < \beta < 1, \quad 0 < \rho < 1, \quad 0 < \pi < 1, \quad \phi_c \geq 0\]

\[(ii) \quad \frac{(1+\Delta \Phi_p) \chi - \mu}{1-\beta \rho \pi} < \frac{\chi - \mu}{1-\beta \rho} \quad \text{.} \tag{2.7} \]

\[(iii) \quad \Delta \Phi_p < \frac{\chi - \mu}{\chi} \]

The first part of the assumption simply ensures that parameters do not take on nonsensical values. The second part is less trivial, but is also a weak assumption. It ensures that the NPV of the sequence of surplus values a project earns during one single stretch of high \(\Phi_+\) values does not exceed the NPV of the sequence of surplus values this project earns over its complete natural life if there are no business cycles to possibly end it prematurely. To see that this is a weak assumption, suppose that \(\beta = 0.99, \rho = 0.9,\) and \(\pi = 0.5.\) Then this condition is satisfied as long as the surplus in a boom is not more than 409% above the surplus value in a world without business cycles. If the second part of assumption 2.3 does not hold, then one particular negative consequence of business cycles is no longer present. It is not necessary, however, for business cycles to be costly. The third part of this assumption also restricts the magnitude of aggregate business cycles. It is made for convenience and means that we have to consider less types of projects in proposition 2.1.

2.2.8 Interpretation of a project

The two key aspects of a project are that (i) a startup cost has to be paid to activate the project and (ii) inefficiencies make it impossible to continue to operate the project if the project does not generate enough revenues.

There are many undertakings with these two criteria. For example, a job is characterized by these two criteria. First, to create a new job requires some investment,
both to create the job itself and to find a suitable worker. Second, inefficiencies could occur both in terms of ensuring cooperation between the employer and the employee and in terms of securing financing. When the project is a job, then $\mu$ would capture the value of leisure and/or home production.

But there are other ways to interpret projects, both on a smaller and on a larger scale. An example of a possible interpretation on a smaller scale would be an additional task for an existing worker. If creation of the task requires an up-front investment and introduces agency problems, then it would fit the description of a project in our model. Alternatively, one can think of much bigger projects such as the opening of a mine or starting a company.

### 2.2.9 Calibration

The value of $\beta$ is set equal to the standard value of 0.99. We follow Krusell and Smith (1998b) and set $\pi$ equal to 0.875, which means that the expected duration of a boom and a recession is equal to 8 quarters.\(^{11}\) The parameters $\mu$ indicates the value generated by an inactive project. Shimer (2005a) uses a value of leisure that is equal to 40% of market production. But his measure refers to all benefits that an unemployed worker receives, whereas here $\mu$ indicates the actual value generated by an inactive project. In case of a job it should exclude any type of transfer such as unemployment benefits; it should only include the value of home production and possibly the utility gain if a worker does not work. As our benchmark, we assume that half of the number used by Shimer (2005a) consists of actual net benefits generated by an unemployed worker. The value used by Shimer (2005a) is considered to be too low by some.\(^ {12}\) Hall (2006) estimates the flow value of leisure forgone to be equal to 43%, and we consider this as an alternative estimate.\(^ {13}\)

The parameter $\rho$ controls the rate of exogenous destruction and $1/(1 - \rho)$ is the expected duration of the project if the duration is not affected by business cycle considerations. Obviously, there are many types of projects and different expected lifetimes.

\(^{11}\)As pointed out in footnote 6, this corresponds roughly to observed durations.
\(^{12}\)See Mortensen and Nagypáll (2007) for a discussion.
\(^{13}\)Personally we value our leisure a lot less, but maybe we are overestimating our market production levels.
If the project is a company, a plant, a mine, or a ship, then the expected duration could easily exceed 10 years or 40 quarters, which would correspond to a value of $\rho$ equal to 0.975. But there are also projects with much shorter durations. The values of $\rho$ considered are equal to 0.875, 0.9167, and 0.975, which correspond with expected durations of respectively 2, 3, and 10 years.

These are the only parameter values we need for the results reported in the next section that focuses on the impact of arbitrarily small fluctuations in $\Phi_{p,t}$. In particular, we do not have to take a stand on the numerical value of $\chi$, we only have to assume that the value is such that assumptions 2.1 and 2.3 are satisfied.

### 2.3 The impact of arbitrarily small aggregate fluctuations

In this section, we first describe a slightly more general version of the efficiency requirement and introduce some terminology. Next, we describe the economy when there are no fluctuations in $\Phi_{p,t}$, i.e., when $\Delta\Phi_p = 0$. Finally, we discuss the impact on the economy of arbitrarily small fluctuations in $\Phi_{p,t}$.

#### 2.3.1 Preliminaries

The ultimate purpose of this section is to analyze the impact on the economy when $\Delta\Phi_p$ increases from 0 to an arbitrarily small number. By considering small changes in $\Delta\Phi_p$, we highlight the property of the model that small business cycles have large effects on some borderline projects. The inequality in equation (2.2) is a weak inequality, but there is no reason why it should not be a strict inequality. Whether it is a weak or a strict inequality does not matter when $\Delta\Phi_p$ is not arbitrarily small.\(^\text{14}\) But it does matter when we consider the case of arbitrarily small values for $\Delta\Phi_p$. The following assumption removes the ambiguity and ensures that the analysis of arbitrarily small fluctuations captures all aspects of regular fluctuations.\(^\text{15}\)

\(^{14}\)Unless there happens to be point mass exactly at the cut-off points.

\(^{15}\)We only need this assumption for the analysis of arbitrarily small fluctuations and then only for projects with a value of $\phi_p$ equal to $\tilde{\phi}_{p,\text{bc}}$. 

Assumption 2.4. For each combination of $(\phi_c, \phi_p)$, there is one project that faces the requirement that
\[
\phi_p \Phi_{p,t} \geq \chi_t
\] (2.8)
and one project that faces the requirement that
\[
\phi_p \Phi_{p,t} > \chi_t.
\] (2.9)

Definition 2.1. Projects that face the efficiency requirement given in equation (2.8) are referred to as "weak-inequality" projects and projects that face the efficiency requirement given in equation (2.9) are referred to as "strict-inequality" projects.

Since agents are risk neutral, business cycles only affect agents’ utility if aggregate fluctuations lead to different decisions. If decisions are not affected, then business cycles just make revenues more volatile, which is not important for risk-neutral agents. There are two types of projects that are affected by business cycles as made precise in the following two definitions.

Definition 2.2. Cyclical projects have the property that they can overcome inefficiencies in a boom, but not in a recession. Thus, cyclical projects are projects for which\(^\text{16}\)
\[
\frac{\chi_t}{1 + \Delta \phi_p} \leq \phi_p < \frac{\chi_t}{1 - \Delta \phi_p}
\]
if the project is a weak-inequality project
and
\[
\frac{\chi_t}{1 + \Delta \phi_p} < \phi_p \leq \frac{\chi_t}{1 - \Delta \phi_p}
\]
if the project is a strict-inequality project.

Definition 2.3. Timed-entry projects are projects such that (i) the NPV of activating the project is higher than the NPV of not activating in a boom and (ii) the NPV of activating the project is lower than the NPV of not activating in a recession. Thus,

\(^{16}\)The definition is given for the case that part (i) of assumption 2.2 holds. If part (ii) holds instead, then one would simply have to change the position of $1 + \Delta \phi_p$ and $1 - \Delta \phi_p$ in this definition.
timed-entry projects are projects for which the following holds:

\[ N_{bc}(\phi_c, \phi_p, 1, 1 + \Delta \phi_p) - \phi_c \geq \mu + \beta E_t [N_{bc}(\phi_c, \phi_p, 0, \Phi_{p,t+1})] \]

and

\[ N_{bc}(\phi_c, \phi_p, 1, 1 - \Delta \phi_p) - \phi_c < \mu + \beta E_t [N_{bc}(\phi_c, \phi_p, 0, \Phi_{p,t+1})]. \]

### 2.3.2 Economy with no aggregate fluctuations

In this subsection, we assume that \( \Delta \phi_p = 0 \), which implies that \( \Phi_{p,t} \) is constant through time, which in turn implies that \( \chi_t \) is constant as well (and equal to \( \chi \)). If \( \Phi_{p,t} \) is constant, then a firm either always overcomes the existing inefficiencies or never overcomes them. That is, the cut-off level for \( \phi_p \) is constant and given by

\[ \tilde{\phi}_{p, \text{no-bc}} = \chi. \] (2.10)

Projects with \( \phi_p < \tilde{\phi}_{p, \text{no-bc}} \) do not enter, since they can never overcome the inefficiency. Projects with \( \phi_p \geq \tilde{\phi}_{p, \text{no-bc}} \) will enter as long as the startup costs are low enough, where projects with a higher value for \( \phi_p \) allow for activation at higher values of \( \phi_c \).

This version of the model is graphically described in Figure 2.1. Projects in the shaded area are activated and produce market output, since their value of \( \phi_p \) exceeds \( \chi \) and their startup costs are low enough. We refer to the cut-off level for \( \phi_c \) in the world without business cycles as \( \tilde{\phi}_{c, \text{no-bc}} \). If productivity is high enough to satisfy the efficiency requirement, then \( \tilde{\phi}_{c, \text{no-bc}} \) is equal to the difference between the NPV of the revenues of the activated project and the NPV of the revenues of the not activated project. Thus, if \( \phi_p > \chi \) (\( \phi_p \geq \chi \)) and the firm is a strict(weak)-inequality project, then

\[ \tilde{\phi}_{c, \text{no-bc}}(\phi_p) = \frac{(\phi_p - \mu)}{1 - \beta \rho}. \] (2.11)

If the efficiency requirement is not satisfied, then entry will not occur, no matter how low the entry costs are. To economize on notation, we will typically drop the argument and simply write \( \tilde{\phi}_{c, \text{no-bc}} \).

The top panel of Figure 2.2 plots the NPV (before the startup cost has been paid)
as a function of $\phi_c$ for projects with $\phi_p = \chi$ and when $\Delta\phi_p = 0$. Strict-inequality projects with $\phi_p = \chi$ just do not meet the efficiency requirement. Consequently, their NPV is equal to $\mu/(1 - \beta)$ independent of the project’s value for $\phi_c$. Weak-inequality projects with $\phi_p = \chi$ just do meet the efficiency requirement. Their value decreases with $\phi_c$ until $\phi_c$ is so high that the value of activating is equal to the value of not activating, $\mu/(1 - \beta)$.

Now consider a weak-inequality project with $\phi_p = \chi$ and $\phi_c = \tilde{\phi}_{c,\text{no-bc}}$. The location of this project is identified with the letter A in the figure. At $\phi_c = \tilde{\phi}_{c,\text{no-bc}}$, the owner of this project would be (just) indifferent between activating and not activating the project in a world without business cycles. Thus, the NPV of this weak-inequality project is equal to the NPV of a strict-inequality project with the same values of $\phi_p$ and $\phi_c$, which does not enter in a world without business cycles. As the value of $\phi_c$ drops, it begins to matter whether a project with $\phi_p = \chi$ is a weak or a strict-inequality project and the NPV of the weak-inequality project increases whereas the NPV of the strict-inequality project remains the same. This is an important observation. The crux
Figure 2.2: Effect of arbitrarily small business cycles on boundary projects

NPV \textsubscript{no–bc} and NPV \textsubscript{bc}

Notes: The top panel plots the NPV when there are no business cycles and when there are arbitrarily small business cycles for projects that are just affected by or are just not affected by inefficiencies. The bottom panel plots the difference between the two NPV values. Point A corresponds to point A in the other figures.
of the matter is that two projects that are identical except that one just can and one just cannot overcome the efficiencies can have very different NPV values.

2.3.3 Quantitative impact of arbitrarily small fluctuations

When considering arbitrarily small fluctuations, we only have to consider projects with values of \( \phi_p \) and/or \( \phi_c \) that are equal to the cut-off points, i.e., projects with either \( \phi_p = \phi_{p,\text{no-bc}} = \chi \) or \( \phi_c = \phi_{c,\text{no-bc}} \). Projects with \( \phi_p = \phi_{p,\text{no-bc}} = \chi \) are cyclical projects and projects with \( \phi_c = \phi_{c,\text{no-bc}} \) are timed-entry projects.\(^{17} \) The impact on cyclical projects is discussed in Section 2.3.3 and the impact on timed-entry projects is discussed in Section 2.3.3.

Impact of arbitrarily small business cycles on cyclical projects

We have to distinguish between three different types of cyclical projects. First, some projects are permanently driven out of existence by business cycles. This happens even for arbitrarily small values of \( \Delta \Phi_p \). Second, some projects temporarily stop producing because of business cycles. Third, some projects that never produce in a world without business cycles, will at times produce in a world with business cycles. We now discuss these three types of cyclical projects.

**Permanent stop of market production when \( \phi_c \) is high.** We start the analysis by considering the weak-inequality project with \( \phi_p = \chi \) and \( \phi_c = \phi_{c,\text{no-bc}} \). A weak-inequality project with \( \phi_p = \chi \) is productive enough to satisfy the efficiency requirement. Since \( \phi_c = \phi_{c,\text{no-bc}} \), the NPV of this project is by definition equal to the NPV of a project that is not activated. The owner would be indifferent between activating and not activating. Now suppose that there are business cycles, but that the size of the fluctuations, i.e., the value of \( \Delta \Phi_p \), is arbitrarily small. If the economy is in a recession, i.e., \( \Phi_{p,t} = 1 - \Delta \Phi_p \), then the project cannot overcome the inefficiency and it clearly would not make sense to pay the startup costs and activate the project. Now suppose that the economy is in a boom, i.e., \( \Phi_{p,t} = 1 + \Delta \Phi_p \). As long as the economy is in a boom, then the project can overcome the inefficiencies and produce \( \chi \) which

\(^{17} \)Both the weak and the strict-inequality projects with \( \phi_p = \chi \) are cyclical projects.
strictly exceeds $\mu$. However, the project cannot remain active when the economy gets into a recession, since the project’s internal resources are not sufficient to overcome the inefficiencies during a recession. This means that there is a *discrete* reduction in the expected life time of the project *even* when the value of $\Delta \phi_p$ is arbitrarily small. In particular, it drops from $1/(1-\rho)$ to $1/(1-\rho + \rho(1-\pi))$. Consequently, the owner of the project with $\phi_p = \chi$ and $\phi_c = \tilde{\phi}_{c,\text{no-bc}}$ would strictly prefer not to activate the project when $\Delta \phi_p > 0$. The NPV of activating also drops for projects with a lower value of $\phi_c$ and the cut-off point value of $\phi_c$ for projects with $\phi_p$ equal to $\chi$ in the presence of business cycles, $\tilde{\phi}_{c,\text{bc}}$, is given by

$$
\tilde{\phi}_{c,\text{bc}} (\phi_p, \Phi_+) = \frac{(1 + \Delta \Phi_p) \phi_p - \mu}{1 - \beta \rho \pi} < \frac{\phi_p - \mu}{1 - \beta \rho} = \tilde{\phi}_{c,\text{no-bc}}.
$$

The value of $\tilde{\phi}_{c,\text{bc}} (\phi_p, \Phi_-$) would be negative for cyclical projects. Since startup costs are assumed to be positive, we do not have to consider this possibility. Unless stated otherwise, the variable $\tilde{\phi}_{c,\text{bc}}$ will refer to the formula given in equation (2.12), that is, to the cut-off level of cyclical projects for which the expected duration is shortened by business cycles.

For any positive value of $\Delta \phi_p$, no matter how small, $\tilde{\phi}_{c,\text{bc}}$ is below $\tilde{\phi}_{c,\text{no-bc}}$ with a discrete amount. A project with $\phi_p = \chi$ and $\phi_c = \tilde{\phi}_{c,\text{no-bc}}$ is faced with a discrete drop in the value of *activating* the project when $\Delta \phi_p$ increases from 0 to a positive value, even an arbitrarily small number. The NPV of this project with $\phi_p = \chi$ and $\phi_c = \tilde{\phi}_{c,\text{no-bc}}$ does not drop, however, since the option of never activating the project remains available and the NPV of a project with $\phi_c = \tilde{\phi}_{c,\text{no-bc}}$ is by definition equal to the NPV of not activating. In contrast, the NPV of projects with $\phi_p = \chi$ and $\phi_c < \tilde{\phi}_{c,\text{no-bc}}$ will display a discrete reduction in value when $\Delta \phi_p$ takes on a positive value.

The top panel of Figure 2.2 plots the NPV of projects with $\phi_p = \chi$ as a function of $\phi_c$, both when $\Delta \phi_p$ is positive (and arbitrarily small) and when $\Delta \phi_p$ is equal to zero. The bottom panel of this figure plots the change in the NPV value when business cycles of arbitrarily small value are introduced. Consider values for $\phi_c$ such that $\phi_c > \tilde{\phi}_{c,\text{bc}}$, i.e., $\phi_c$ is too high to make entry worthwhile given the reduction in expected duration.
brought about by the increase in $\Delta \Phi_p$. The negative impact of business cycles is larger for smaller values of $\phi_c$, since the projects with lower values for $\phi_c$ are more profitable, which means that the loss of not activating the project is larger.

**Cyclical projects with low values for $\phi_c$.** Next, we consider cyclical projects with values of $\phi_c$ that are less than $\tilde{\phi}_c,bc$. These projects cannot overcome the inefficiencies in a recession, *but* the values of their startup costs are so low that it is still worth activating the project even though the expected duration of the project is now shortened. When $\Delta \Phi_p > 0$, then there is no difference between the strict-inequality and the weak-inequality projects with $\phi_p = \chi$. But there is a difference when $\Delta \Phi_p = 0$. We first consider the weak-inequality projects and then the strict-inequality projects.

**Temporarily driven out of producing market output.** Consider weak-inequality projects, with $\phi_p = \chi$ and $\tilde{\phi}_c \leq \tilde{\phi}_c,bc$. These projects can always overcome the inefficiency when $\Delta \Phi_p = 0$. This means that the expected duration is equal to $1/(1 - \rho)$. An arbitrarily small increase in $\Delta \Phi_p$ affects these projects in two ways. First, during recessions they only generate $\mu$ instead of the larger $\chi$. Second, the project has to stop producing during a recession and the startup costs have to be paid again at the beginning of the boom. The NPV of these projects is higher in a world without business cycles than in a world with business cycles, as is documented in the top panel of Figure 2.2. The magnitude of the drop in value is shown graphically in the bottom panel of the same figure.\(^\text{18}\)

**Temporarily driven into producing market output.** Now consider strict-inequality projects with $\phi_p = \chi$ and $\tilde{\phi}_c \leq \tilde{\phi}_c,bc$. For these projects, the presence of business cycles is beneficial as long as their value of $\phi_c$ is low enough. The reason is that in a world without business cycles they always generate $\mu$, whereas in a world with business cycles

---

\(^{18}\)In the graph, the project with $\phi_c = 0$ has the smallest loss among the weak-inequality projects with $0 \leq \phi_c \leq \tilde{\phi}_c,bc$. But this depends on parameter values, and the project with $\phi_c = 0$ could also have the largest loss. What matters is how the NPV decreases with $\phi_c$ when there are and when there are no business cycles. On the one hand you only pay entry costs when you start in a boom which means that the NPV is less sensitive to $\phi_c$ in a world with business cycles. On the other hand you are expected to pay them more often at future dates in a world with business cycles. This would make the NPV more sensitive to $\phi_c$ in a world with business cycles. If $\rho = 0.5/(\beta(1 - 0.5\pi))$, then the effects exactly offset each other and the loss is the same for all weak inequality projects with $0 \leq \phi_c \leq \phi_c,bc$.  

they generate $\chi > \mu$ in a boom and their startup costs are low enough to take advantage of this. The benefits are reduced by the fact that startup costs have to be paid. Consequently, the welfare gains are largest for the project with zero startup costs, as is shown in the bottom panel of Figure 2.2.

It is important to realize that the welfare losses of the weak-inequality projects exceed the welfare gains of the strict-inequality projects. In terms of output, the losses and the gains exactly offset each other. But the shortening in the expected duration of a project means that the total amount of startup costs paid is higher when $\Delta \phi_p > 0$ than when $\Delta \phi_p = 0$. Thus, when $\phi_c \leq \phi_{c,bc}$, the loss of a weak-inequality project together with the gain of a strict-inequality project is a loss except when $\phi_c = 0$. This is documented in the lower panel of Figure 2.2. When $\phi_c = 0$, then the gain of the weak-inequality project does exactly off-set the loss of the strict-inequality project, because it does not matter that the entry costs have to be paid more often.

**Magnitude of the impact of tiny business cycles on cyclical projects.** We measure the impact of business cycles on individual projects as the permanent increase (or decrease) in per-period income that would make the owner of a project in a world with business cycles as well off as the owner of the same project in a world without business cycles. To standardize the measure we scale by $\phi_p$. The formula is given in the following definition.

**Definition 2.4.** The impact of business cycles on an individual project is given by

$$L(\phi_c, \phi_p, \Delta \phi_p) = (1 - \beta) \left( N_{no-bc}(\phi_c, \phi_p, 0) - E[N_{bc}(\phi_c, \phi_p, 0, \Phi_{p,t})] \right),$$

where

$$E[N_{bc}(\phi_c, \phi_p, 0, \Phi_{p,t})] = \frac{NPV_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) + NPV_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p)}{2}.$$ 

The following proposition gives the results for projects at the boundary, i.e., when $\phi_p = \tilde{\phi}_{p,no-bc} = \chi$.

**Proposition 2.1.** Suppose that (i) $\Delta \phi_p > 0$, (ii) assumptions 2.1 and 2.3 are satisfied,
and (iii) the first part of 2.2 holds. Then the following properties hold:

1. \[
\begin{align*}
\tilde{\phi}_{c, bc} &= \frac{\phi_p (1 + \Delta \phi_p) - \mu}{1 - \beta \rho \pi} \text{ if } \tilde{\phi}_{p, bc} (\Phi_+) \leq \phi_p < \tilde{\phi}_{p, bc} (\Phi_+), \\
\tilde{\phi}_{c, no-bc} &= \frac{\phi_p - \mu}{1 - \beta \rho} \text{ if } \phi_p \geq \tilde{\phi}_{p, no-bc}, \\
\tilde{\phi}_{c, bc} &< \tilde{\phi}_{c, no-bc} \text{ if } \tilde{\phi}_{p, no-bc} \leq \phi_p < \tilde{\phi}_{p, bc} (\Phi_+).
\end{align*}
\]

2. For any cyclical projects with \(\phi_p > \tilde{\phi}_{p, no-bc}\) and for weak-inequality projects with \(\phi_p = \tilde{\phi}_{p, no-bc}\), the change in welfare is given by

\[
L(\phi_c, \phi_p, \Delta \phi_p) = \begin{cases} 
\frac{1-\Delta \phi_p - \mu/\phi_p}{2} + \frac{\phi_c}{\phi_p} \left(1-\beta \rho \pi - 2(1-\beta \rho)\right) > 0 & \text{if } 0 \leq \phi_c \leq \tilde{\phi}_{c, bc} \\
1 - (1 - \beta \rho) \frac{\phi_c}{\phi_p} - \frac{\mu}{\phi_p} > 0 & \text{if } \tilde{\phi}_{c, bc} < \phi_c < \tilde{\phi}_{c, no-bc} \\
0 & \text{if } \phi_c \geq \tilde{\phi}_{c, no-bc}
\end{cases}
\]

For cyclical projects with \(\phi_p < \tilde{\phi}_{p, no-bc}\) and for strict-inequality projects with \(\phi_p = \tilde{\phi}_{p, no-bc}\), the change in welfare is given by

\[
L(\phi_c, \phi_p, \Delta \phi_p) = \begin{cases} 
\frac{\mu/\phi_p - (1+\Delta \phi_p)}{2} + \frac{\phi_c}{\phi_p} \left(1-\beta \rho \pi\right) < 0 & \text{if } 0 \leq \phi_c \leq \tilde{\phi}_{c, bc} \\
0 & \text{if } \tilde{\phi}_{c, bc} < \phi_c < \tilde{\phi}_{c, no-bc} \\
0 & \text{if } \phi_c \geq \tilde{\phi}_{c, no-bc}
\end{cases}
\]

The average of an affected weak-inequality and an affected strict-inequality project with

\[\text{If the second part of condition 2.2 holds, then a similar set of formulas are valid, but the role of the boom and the recession are switched.}\]
\( \phi_p = \tilde{\phi}_{p, \text{no-bc}} \) is given by

\[
L(\phi_c, \phi_p, \Delta \phi_p) = \begin{cases} 
\frac{1}{2} \left( -\Delta \phi_p + \frac{\phi_c}{\phi_p} \beta \rho (1 - \pi) \right) & < 0 \quad \text{if } 0 < \phi_c < \phi_p \Delta \phi_p \\
\frac{1}{2} \left( -\Delta \phi_p + \frac{\phi_c}{\phi_p} \beta \rho (1 - \pi) \right) & = 0 \quad \text{if } \phi_c = \phi_p \Delta \phi_p \\
\frac{1}{2} \left( -\Delta \phi_p + \frac{\phi_c}{\phi_p} \beta \rho (1 - \pi) \right) & > 0 \quad \text{if } 0 < \phi_c \leq \phi_{c, \text{bc}} \\
1 - (1 - \beta \rho) \frac{\phi_c}{\phi_p} - \frac{\mu}{\phi_p} & > 0 \quad \text{if } \phi_{c, \text{bc}} < \phi_c < \phi_{c, \text{no-bc}} \ \\
0 & \quad \text{if } \phi_c = \phi_{c, \text{no-bc}} \ \\
\end{cases}
\]

3. \[ \lim_{\Delta \phi_p \to 0} \tilde{\phi}_{c, \text{bc}} = \phi_p - \mu \frac{1 - \beta \rho}{1 - \beta \rho \pi} < \phi_{c, \text{no-bc}} = \phi_p - \mu \frac{1 - \beta}{1 - \beta \rho}. \]

4. For weak-inequality projects with \( \phi_p = \tilde{\phi}_{p, \text{no-bc}} \), the change in welfare of introducing arbitrarily small business cycles is given by

\[
\lim_{\Delta \phi_p \to 0} L(\phi_c, \phi_p, \Delta \phi_p) = \begin{cases} 
1 - \mu \phi_p + \phi_c \left( \frac{1 - \beta \rho \pi - 2(1 - \beta \rho)}{2} \right) & > 0 \quad \text{if } 0 \leq \phi_c \leq \phi_{c, \text{bc}} \\
1 - (1 - \beta \rho) \frac{\phi_c}{\phi_p} - \frac{\mu}{\phi_p} & > 0 \quad \text{if } \phi_{c, \text{bc}} < \phi_c < \phi_{c, \text{no-bc}} \\
0 & \quad \text{if } \phi_c \geq \phi_{c, \text{no-bc}} \ \\
\end{cases}
\]

and for strict-inequality projects with \( \phi_p = \tilde{\phi}_{p, \text{no-bc}} \) by

\[
\lim_{\Delta \phi_p \to 0} L(\phi_c, \phi_p, \Delta \phi_p) = \begin{cases} 
\frac{\mu}{\phi_p - 1} + \phi_c \left( \frac{1 - \beta \rho \pi}{2} \right) & < 0 \quad \text{if } 0 \leq \phi_c \leq \phi_{c, \text{bc}} \\
0 & \quad \text{if } \phi_{c, \text{bc}} < \phi_c < \phi_{c, \text{no-bc}} \\
0 & \quad \text{if } \phi_c \geq \phi_{c, \text{no-bc}} \ \\
\end{cases}
\]

For projects affected by business cycles, the average impact for arbitrarily small changes

---

\(^{20}\) If \( \phi_c \leq \phi_{c, \text{bc}} \), this is the average of the weak-inequality and the strict-inequality project. If \( \phi_{c, \text{bc}} < \phi_c \leq \phi_{c, \text{no-bc}} \), this is just equal to the effect of the weak-inequality project.
is given by

$$\lim_{\Delta \Phi_p \to 0} L(\phi_c, \phi_p, \Delta \Phi_p) = \begin{cases} 
\frac{1}{2} \frac{\phi_c}{\phi_p} \beta \rho (1 - \pi) = 0 & \text{if } \phi_c = 0 \\
\frac{1}{2} \frac{\phi_c}{\phi_p} \beta \rho (1 - \pi) > 0 & \text{if } 0 < \phi_c \leq \tilde{\phi}_{c, bc} \\
1 - (1 - \beta \rho) \frac{\phi_c}{\phi_p} - \frac{\mu}{\phi_p} > 0 & \text{if } \tilde{\phi}_{c, bc} < \phi_c < \tilde{\phi}_{c, no-bc} \\
0 & \text{if } \phi_c = \tilde{\phi}_{c, no-bc} 
\end{cases}$$

The following observations can be made. First, parts 3 and 4 of this proposition make clear that the impact of business cycles of arbitrarily small magnitude is not arbitrarily small for the affected projects, i.e., the projects at the boundary. Second, business cycles are beneficial for some projects, because they allow them to overcome inefficiencies during some periods. Third, the combined loss for a strict-inequality and a weak-inequality project is equal to zero when $\phi_c = 0$, but the combined impact is a strictly positive loss when $\phi_c > 0$ as long as $\phi_c < \tilde{\phi}_{c, no-bc}$. The maximum combined loss, $L_{max}$, is attained at $\phi_c = \tilde{\phi}_{c, bc}$ and is given by the following expression:

$$L_{max} = \frac{1}{2} \frac{\phi_p - \mu}{\phi_p} \beta \rho (1 - \pi) - \frac{1}{1 - \beta \rho \pi}.$$

Table 2.1 reports the gains and losses for cyclical projects using the parameter values discussed at the end of Section 2.2. There are non-trivial numbers. For some projects the impact is as high as 62.1%.

Table 2.2 is the analogue of Table 2.1 for a typical value of $\Delta \Phi_p$. We set $\Phi_+ - 1$ (and thus $1 - \Phi_-$) equal to 0.007, which means that the standard deviation of $\Phi_{p_{it}}$ is equal to 0.007, a standard value used in the literature. The numbers are similar to those of Table 2.1. That is, what matters for individual projects is in the first place the presence of business cycles and whether it affects their ability to satisfy the efficiency requirement. The magnitude of business cycles is of secondary importance.
Table 2.1: Impact of arbitrarily small business cycles for projects at the boundary

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\phi_p = \tilde{\phi}_{\text{p, no-bc}}$</th>
<th>$\phi_c = \tilde{\phi}_{\text{c, bc}}$</th>
</tr>
</thead>
</table>
| $\rho = 0.875$ | $\tilde{\mu} = 0.2$  
$[35.8\%, 40.0\%]$  
$[40.0\%, 44.1\%]$  
$[40.0\%, 62.1\%]$ | $\mu = 0.43$  
$[25.5\%, 28.5\%]$  
$[28.5\%, 31.4\%]$  
$[28.5\%, 44.3\%]$ |
| $\rho = 0.9167$ | $\tilde{\mu} = 0.2$  
$[-40.0\%, 0\%]$  
$[-40.0\%, 0\%]$  
$[-40.0\%, 0\%]$ | $\mu = 0.43$  
$[0\%, 25.5\%]$  
$[0\%, 31.4\%]$  
$[0\%, 44.3\%]$ |
| $\rho = 0.975$ | $\tilde{\mu} = 0.2$  
$[0\%, 0\%]$  
$[0\%, 0\%]$  
$[0\%, 0\%]$ | $\mu = 0.43$  
$[0\%, 0\%]$  
$[0\%, 0\%]$  
$[0\%, 0\%]$ |

Notes: This table gives the welfare losses corresponding to a marginal increase in $\Delta \Phi_p$ starting at $\Delta \Phi_p = 0$ for individuals with a project for which either $\phi_p = \tilde{\phi}_{\text{p, no-bc}}$ or $\phi_c = \tilde{\phi}_{\text{c, bc}}$. $\mu$ is the value of $\mu$ relative to the value of $\phi_p$, thus, $\mu = \tilde{\mu} \phi_p$. Each cell contains the smallest and the largest outcome found across the possible values for $\phi_c$.

Impact of arbitrarily small business cycles on timed-entry projects

Now consider projects at the other boundary, i.e., when $\phi_c = \tilde{\phi}_{\text{c, no-bc}}$. These projects have a value of $\phi_p$ such that inefficiencies are not an issue, but the value of $\phi_c$ is such that the owner is indifferent between activating and not activating the project when there are no business cycles. We will see that there is a fundamental difference between these boundary projects and the cyclical projects discussed in the previous subsection.

If $\phi_c = \tilde{\phi}_{\text{c, no-bc}}$, then the owner of this project is (by definition) indifferent between activating and not activating in a world without business cycles, i.e., when $\Phi_{p,t} = 1 \quad \forall t$. The owner of this project would strictly prefer to activate the project when the economy is in a boom, i.e., when $\Phi_{p,t} > 1$ even when at some point in the future the economy gets into a recession, i.e., when at some point $\Phi_{p,t} < 1$. The reason is that the startup costs are the same in both cases, but the NPV of the revenues are higher if the projects starts in a boom than if it starts in a world without business cycles. Thus $N_{bc}(\tilde{\phi}_{\text{c, no-bc}}, \phi_p, 0, 1 + \Delta \Phi_p) > N_{\text{no-bc}}(\tilde{\phi}_{\text{c, no-bc}}, \phi_p, 0) = \mu / (1 - \beta)$ for $\Delta \Phi_p > 0$. If the economy is in a recession, i.e., when $\Phi_{p,t} < 1$, then the owner would strictly prefer to
Table 2.2: Impact of business cycles for projects at the boundary

<table>
<thead>
<tr>
<th>( \rho = 0.875 )</th>
<th>( \rho = 0.9167 )</th>
<th>( \rho = 0.975 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cyclical projects:</strong> ( \phi_p = \tilde{\phi}_{p,\text{no-bc}} )</td>
<td>( \phi = \tilde{\phi}_{c,\text{bc}} )</td>
<td>( \phi = \tilde{\phi}_{c,\text{bc}} )</td>
</tr>
<tr>
<td>( \hat{\mu} = 0.2 )</td>
<td>( \phi_c \leq \tilde{\phi}_{c,\text{bc}} )</td>
<td>( \phi_c \leq \tilde{\phi}_{c,\text{bc}} )</td>
</tr>
<tr>
<td>( \hat{\mu} = 0.43 )</td>
<td>( 35.4%, 39.7% )</td>
<td>( 39.7%, 43.8% )</td>
</tr>
<tr>
<td>( \hat{\mu} = 0.43 )</td>
<td>( 25.1%, 28.2% )</td>
<td>( 28.2%, 31.1% )</td>
</tr>
<tr>
<td><strong>strict-inequality project with ( \phi_c \leq \tilde{\phi}_{c,\text{bc}} )</strong></td>
<td>( \hat{\mu} = 0.2 )</td>
<td>( \phi = \tilde{\phi}_{c,\text{bc}} )</td>
</tr>
<tr>
<td>( \hat{\mu} = 0.43 )</td>
<td>( -40.4%, 0% )</td>
<td>( -40.4%, 0% )</td>
</tr>
<tr>
<td>( \hat{\mu} = 0.43 )</td>
<td>( -28.9%, 0% )</td>
<td>( -28.9%, 0% )</td>
</tr>
<tr>
<td><strong>timed-entry project:</strong> ( \phi_c = \tilde{\phi}_{c,\text{no-bc}} )</td>
<td>( \phi = \tilde{\phi}_{c,\text{no-bc}} )</td>
<td>( \phi = \tilde{\phi}_{c,\text{no-bc}} )</td>
</tr>
<tr>
<td>( \hat{\mu} = 0.2 )</td>
<td>( -0.19%, 0% )</td>
<td>( -0.16%, 0% )</td>
</tr>
<tr>
<td>( \hat{\mu} = 0.43 )</td>
<td>( -0.19%, 0% )</td>
<td>( -0.16%, 0% )</td>
</tr>
</tbody>
</table>
In contrast to the results for cyclical projects, the loss function for the timed-entry projects is not characterized by discontinuities and the impact of business cycles is equal to 0 for arbitrarily small values of $\Delta \Phi_p$. Table 2.2 reports that the individual welfare consequences of business cycles for these projects remain small when a realistic value for $\Delta \Phi_p$ is considered, especially compared with the numbers for the cyclical projects.

### 2.4 Impact of regular aggregate fluctuations

In this section, we analyze the impact of business cycles when fluctuations are not arbitrarily small. The analysis here is actually very similar to the one in the previous section that focused on arbitrarily small fluctuations except that as fluctuations get larger they affect more projects and it is no longer true that only projects with values of $\phi_c$ or $\phi_p$ at cut-off levels are affected. We start with a graphical representation. Next, we discuss the quantitative consequences of business cycles for the cyclical and timed-entry projects. We will show that the results from the last section on the impact of business cycles on projects with values of $\phi_c$ or $\phi_p$ at cut-off levels provide a lower (upper) bound of the negative (positive) consequences of business cycles for projects with other values for $\phi_c$ or $\phi_p$.

For the analysis in this section, the distinction between strict-inequality and weak-inequality projects is no longer necessary.\(^{23}\) From now on we assume that the efficiency requirement is given by equation (2.2), unless explicitly stated otherwise.

#### 2.4.1 Graphical representation of regular business cycles.

In Section 2.3.3, we identified four types of projects that are affected by business cycles of arbitrarily small magnitude. Those are three types of cyclical projects and timed-entry projects. When considering business cycles of non-trivial magnitude we can distinguish the exact same four groups, but there are some differences with the analysis of arbitrarily small business cycles. The main difference is that business cycles

---

\(^{23}\)Recall that this distinction was introduced to capture both the gains and the losses of business cycles in a framework with arbitrarily small fluctuations.
Figure 2.3: Projects affected by business cycles

Notes: The shaded areas in this graph indicate the projects that are affected by business cycles. Light grey: Cyclical projects that operate during a boom and do not operate during a recession. Projects in the "gain" ("loss") area never (always) operate in a world without business cycles. Grey: Cyclical Projects that can overcome inefficiencies during a boom, but their entry costs are too high to make entry worthwhile given that inefficiencies will force exit during a recession. Black: Timed-entry projects. Point A corresponds to point A in the other figures.

of non-trivial magnitude not only affect projects at the boundary, i.e., not only projects with $\phi_p = \tilde{\phi}_{p,\text{no-bc}}$ or $\phi_c = \tilde{\phi}_{c,\text{no-bc}}$.

This is documented in Figure 2.3, which illustrates where the four different types of projects are located. The timed-entry projects have a value of $\phi_c$ just above or just below $\tilde{\phi}_{c,\text{no-bc}}$. Their value for $\phi_p$ is above $\tilde{\phi}_{p,\text{bc}}(\Phi_+)$ so that they can overcome inefficiencies even during downturns. Cyclical projects can have a value of $\phi_c$ above or below $\tilde{\phi}_{c,\text{bc}}$. Affected cyclical projects with a value of $\phi_c$ below $\tilde{\phi}_{c,\text{bc}}$ can have a value of $\phi_p$ below or above $\tilde{\phi}_{p,\text{no-bc}}$. In contrast, all affected cyclical projects with a value of $\phi_c$ above $\tilde{\phi}_{c,\text{bc}}$ have a value of $\phi_p$ above $\tilde{\phi}_{p,\text{no-bc}}$. The corresponding projects with a value of $\phi_p$ below $\tilde{\phi}_{p,\text{no-bc}}$ can overcome inefficiencies during a boom, but there entry costs are too high to benefit from this given that inefficiencies would force them to cease operations during a recession. In the remainder of this section, we discuss the quantitative impact of non-trivial business cycles.
2.4.2 Impact of regular business cycles on timed-entry projects

For values of $\Delta \Phi_p$ that are not arbitrarily small, the set of timed-entry projects is not limited to projects with $\phi_c = \tilde{\phi}_{c,\text{no-bc}}$, but includes the set $I_{\text{timed entry}}$ defined by

$$I_{\text{timed entry}} = \left\{ (\phi_c, \phi_p) : \phi_p > \max \tilde{\phi}_{p,bc}(\Phi_p), \tilde{\phi}_{c,bc}(\phi_p, \Phi_-) < \phi_c \leq \tilde{\phi}_{c,bc}(\phi_p, \Phi_+) \right\}. \quad (2.13)$$

The duration of these projects is not shortened by business cycles. Consequently, the formula for $\tilde{\phi}_{c,bc}(\Phi_p)$ is not given by equation (2.12), but is given by a system of equations.\(^{24}\) We have to distinguish between the projects in $I_{\text{timed entry}}$ with values of $\phi_c$ above $\tilde{\phi}_{c,\text{no-bc}}$ and those with values of $\phi_c$ below $\tilde{\phi}_{c,\text{no-bc}}$. We will now discuss these two groups in turn, each time keeping the value of $\phi_p$ fixed.

Consider projects with values of $\phi_c$ above $\tilde{\phi}_{c,\text{no-bc}}$ and below $\tilde{\phi}_{c,bc}(\Phi_+)$. These projects would never produce market output in a world without business cycles. Although $\phi_p > \mu$, their startup costs are too high to make entry profitable. In a world with business cycles, they would enter in a boom. Moreover, they would continue producing when the economy gets into a recession. The smaller the value of $\phi_c$, the larger the gains from business cycles. In particular, projects with $\phi_c = \tilde{\phi}_{c,bc}(\Phi_+)$ do not benefit at all from business cycles and projects with $\phi_c = \tilde{\phi}_{c,\text{no-bc}}$ benefit the most. Thus, the welfare gains reported in Proposition 2.2 and Tables 2.1 and 2.2 are an upper bound for the gains for projects with values of $\phi_c$ in between $\tilde{\phi}_{c,\text{no-bc}}$ and below $\tilde{\phi}_{c,bc}(\Phi_+)$.\(^{24}\)

Now consider projects with values of $\phi_c$ below $\tilde{\phi}_{c,\text{no-bc}}$ and above $\tilde{\phi}_{c,bc}(\Phi_-)$. In a world without business cycles, these projects always immediately enter. In a world with business cycles, they delay entry during a recession. The larger the value for $\phi_c$, the more valuable the benefit to delay entry. So the largest gain is again achieved by projects with $\phi_c = \tilde{\phi}_{c,\text{no-bc}}$. Consequently, the welfare gains reported in Proposition 2.2 are an upper bound for the welfare gains achieved by all timed-entry projects.

**Proposition 2.3.** If (i) $(\phi_c, \phi_p) \in I_{\text{timed entry}}$, (ii) the values of $\beta$, $\rho$, and $\pi$ are in

\(^{24}\)See Appendix B.1.2 in Den Haan and Sedlacek (2009).
between 0 and 1, and (iii) \( \Delta \phi_p > 0 \), then

\[
L(\hat{\phi}_{c,na-bc}, \phi_p, \Delta \phi_p) < L(\phi_c, \phi_p, \Delta \phi_p) \leq 0 \text{ for } \phi_c \neq \hat{\phi}_{c,na-bc}.
\]

As discussed in Section 2.3.3, even these maximum gains are relatively small. Note that the gains are equal to zero at the boundaries of \( I_{\text{timed entry}} \), i.e., when \( \phi_c = \hat{\phi}_{c,bc}(\phi_p, \Phi_-) \) or \( \phi_c = \hat{\phi}_{c,bc}(\phi_p, \Phi_+) \).

**Three reasons why timed-entry projects are unlikely to be important**

Table 2.2 documents that the welfare gains of business cycles for projects with a value of \( \phi_c \) equal to \( \hat{\phi}_{c,na-bc} \) are small and these are an upper bound for the welfare gains of the other timed-entry projects, as documented in Proposition 2.3. This is not too surprising. These are marginal projects (in a true economic sense) and business cycles just make it possible to create a bit of value by alternative timing.

Another reason why timed-entry projects are unlikely to be important is that these projects are truly marginal from an economic point of view. The question arises how many of such projects exist. Why bother starting a business and producing market output when you only marginally improve upon your outside option? Cyclical projects are also marginal, but not from an economic point of view; if these projects overcome the inefficiency, then there are substantial rents to be earned.

The last reason why timed-entry projects are unlikely to be important is that the presence of such projects is not a robust outcome of models with inefficiencies. In particular, it seems reasonable that timed-entry projects are especially prone to inefficiencies given that these are truly marginal projects. For example, consider the following alternative formulation of the efficiency requirement:

\[
\phi_p \Phi_{p,t} - r \phi_c \geq \chi_t, \quad (2.14)
\]

where \( r = 1 - \beta \rho \) is the interest rate. In Appendix 2.A, this efficiency requirement is derived using a model with financial frictions. In this specification of the efficiency requirement, the net per-period revenues matter, i.e., the revenues after amortization of the entry costs. With this alternative efficiency requirement, there would be no
timed-entry projects at all. In contrast, the same three types of cyclical projects still exist. This case is discussed in more detail in Appendix.2.A

2.4.3 Impact of regular business cycles on cyclical projects

Let $I_{\text{cyclical}}$ be the set of cyclical projects. Thus,

\[
I_{\text{cyclical}} = \left\{ \phi_c, \phi_p : \tilde{\phi}_{p,\text{bc}}(\Phi_+) \leq \phi_p < \tilde{\phi}_{p,\text{bc}}(\Phi_-) \wedge \phi_c \leq \tilde{\phi}_{c,\text{no-bc}} \right\}.
\] (2.15)

The welfare consequences of business cycles for cyclical projects with $\phi_p = \chi = \tilde{\phi}_{p,\text{no-bc}}$ are given in Proposition 2.1. How do these welfare consequences compare with the welfare consequences for the cyclical projects with $\phi_p \neq \tilde{\phi}_{p,\text{no-bc}}$? Keeping the value of $\phi_c$ fixed, then the welfare losses of projects with $\phi_p > \tilde{\phi}_{p,\text{no-bc}}$ are bigger than the losses of the weak-inequality projects with $\phi_p = \tilde{\phi}_{p,\text{no-bc}}$ and the welfare gains of projects with $\phi_p < \tilde{\phi}_{p,\text{no-bc}}$ are smaller than the gains of the strict-inequality projects with $\phi_p = \tilde{\phi}_{p,\text{no-bc}}$.

First, consider projects with $\phi_p < \tilde{\phi}_{p,\text{no-bc}}$ and $\phi_c \leq \tilde{\phi}_{c,\text{bc}}$. The entry costs of these projects are low enough so that entry is profitable even though they only survive until the next recession. In a world with business cycles, these projects generate output equal to $(1 + \Delta \Phi_p)\phi_p > \mu$ during a boom and $\mu$ during a recession, whereas they always generate $\mu$ in a world without business cycles.26 The larger the value of $\phi_p$, the larger the gains (for equal values of $\phi_c$). Thus, the welfare gain attained by strict-inequality projects with $\phi_p = \tilde{\phi}_{p,\text{no-bc}}$ is an upper bound for the gains achieved by cyclical projects.

Next, consider projects with $\phi_p \geq \tilde{\phi}_{p,\text{no-bc}}$. In this case, the welfare loss attained by weak-inequality projects with $\phi_p = \tilde{\phi}_{p,\text{no-bc}}$ are a lower bound for the losses suffered by cyclical projects. Business cycles permanently reduce market output of cyclical

\[\text{25}\text{Here we still adopt the assumption that } \tilde{\phi}_{p,\text{bc}}(\Phi_p) \text{ is countercyclical. If } \tilde{\phi}_{p,\text{bc}}(\Phi_p) \text{ would be procyclical, then the set is given by}
\]

\[
I_c = \left\{ \phi_p : \tilde{\phi}_{p,\text{bc}}(\Phi_-) \leq \phi_p < \tilde{\phi}_{p,\text{bc}}(\Phi_+) \right\}.
\]

\[\text{26}\text{The result that } (1 + \Delta \Phi_p)\phi_p > \mu \text{ follows directly from Condition 2.1.}\]
projects when \( \phi_c \) is above \( \tilde{\phi}_{c,bc} \) and temporarily (namely during recessions) when \( \phi_c \) is below \( \tilde{\phi}_{c,bc} \). Consequently, given the value for \( \phi_c \), the loss is larger when the value of \( \phi_p \) is larger.

The following proposition summarizes these results.

**Proposition 2.4.** Suppose that (i) \( (\phi_c, \phi_p) \in I_{cyclical} \), (ii) assumptions 2.1, 2.2, and 2.3 are satisfied, and (iii) \( \Delta \phi_p > 0 \). Then

\[
L(\phi_c, \tilde{\phi}_{p,no-bc}, \Delta \phi_p) < L(\phi_c, \phi_p, \Delta \phi_p) \text{ for } \phi_p \neq \tilde{\phi}_{p,no-bc}.
\]

### 2.5 Necessary ingredients

In this subsection, we explain why both entry costs and inefficient operating decisions are needed for business cycles to be costly.

#### 2.5.1 Why are entry costs essential?

Suppose that entry costs are equal to zero for all projects. This would mean that the whole graph would collapse onto the \( x \)-axis in Figure 2.3. If there are no entry costs, then projects cannot be permanently driven out of business. In our model, business cycles shorten the expected duration of the project. For some projects this makes the entry costs too high relative to the reduced expected revenue stream. This cannot matter when \( \phi_c \) is equal to zero. If there are no entry costs, then business cycles allow some additional projects to produce during a boom and make production impossible for some projects during a recession. The gain for a project just below \( \tilde{\phi}_{p,no-bc} \) and the loss for a project just above \( \tilde{\phi}_{p,no-bc} \) roughly offset each other.\(^{27}\) Given that we are in the lower tail of the distribution, it is possible that there are more projects with values of \( \phi_p \) above \( \tilde{\phi}_{p,no-bc} \) than below \( \tilde{\phi}_{p,no-bc} \). Then business cycles could be costly even when entry costs are equal to zero. We prefer to build our argument, however,

\(^{27}\)As documented in proposition 2.1, the gain of the strict-inequality project and the loss of the weak-inequality project exactly off set each other for arbitrarily small fluctuations when \( \phi_p = \tilde{\phi}_{p,no-bc} \). For larger fluctuations, there actually would be a net gain for the two projects with \( \phi_p = \tilde{\phi}_{p,no-bc} \). But note that the output gained by a project with \( \phi_p < \tilde{\phi}_{p,no-bc} \) is less than the output lost by a project with \( \phi_p > \tilde{\phi}_{p,no-bc} \). So for all affected projects there still could be a net loss.
without relying on such distributional assumptions.

2.5.2 Why are inefficient operating decisions essential?

There are two aspects of the inefficient operating decision that are important. First, the inefficiency makes it impossible to offset worsened conditions during a recession with improved conditions during a boom. That is, the inefficiency specified in equation (2.2) has to hold at each point in time, not just on average. This means that the reason for the efficiency requirement must be a private inefficiency. If there are no private inefficiencies, then there is no reason why the higher revenues in a boom could not offset the lower revenues in a recession. Second, in the presence of inefficient operating decisions, marginal projects have a positive surplus when defined relative to the true outside option. This means that it is costly from a social welfare point of view that business cycles prevent such projects from being created.

In an earlier version of this paper, we describe in detail the model without the efficiency requirement. In that case, business cycles are beneficial for the following reason. Without the efficiency requirement, the entry and continuation decisions in a world with business cycles could be identical to those in a world without business cycles. If these decisions are the same, then the revenues would be more volatile in the world with business cycles, but—given our assumption of risk neutrality—the NPVs would the same. Consequently, business cycles cannot make agents worse off. In fact, when the outside option $\mu$ is acyclical (or at least less cyclical then $\phi_p\Phi_{p,t}$), then the possible revenues, $\min\{\mu, \phi_p\Phi_{p,t}\}$ is a convex function of $\Phi_{p,t}$ and increased volatility in $\Phi_{p,t}$ would increase the expected value.

2.6 Approach to measure overall cost of business cycles

In this section, we outline the procedure used to obtain a quantitative estimate for the welfare costs of business cycles for the economy as a whole. In the first subsection, we

---

discuss the key assumptions underlying our procedure. In the second subsection, we describe the basic idea. Details are given in Appendix 2.B.

### 2.6.1 Key elements underlying our approach

The reader may wonder whether a credible quantitative answer can be provided with the simple framework presented here. For example, different types of projects may face different types of inefficiencies, which would be associated with different values of $\chi_t$. Moreover, even if all projects would face the same type of inefficiency, then the value of $\chi_t$ would still not have to be the same for all individual projects. For example, projects with a higher value of $\phi_p$ could have a higher value of $\chi_t$.

Before explaining our strategy, we make explicit what we do not do. We do not focus on one particular type of inefficiency, calibrate the function $\chi(\Phi_{p,t})$, or possibly $\chi(\Phi_{p,t}, \phi_c, \phi_p)$, and finally calibrate the distribution of $\phi_c$ and $\phi_p$. Just the calibration of the distribution of $\phi_p$ and $\phi_c$ would be very difficult given that what matters is the mass in a very specific, relatively small, part of the distribution. Moreover, by focusing on only one type of inefficiency, the results are likely to provide an incomplete picture of the impact of business cycles.

**First element: Countercyclical efficiency requirement**

Business cycles shorten the duration of projects when $\tilde{\phi}_{p,bc}(\Phi_{p,t})$ is countercyclical (and projects are eliminated during recessions), but also when $\tilde{\phi}_{p,bc}(\Phi_{p,t})$ is procyclical (and projects are eliminated during expansions). Business cycles are costly under both assumptions, but having a countercyclical number of projects does not seem very plausible. Therefore, to calibrate the model we assume that $\tilde{\phi}_{p,bc}(\Phi_{p,t})$ is countercyclical.

**Second element: Assumption about importance of inefficient continuation**

If the actual economy gets into a recession, then there are two reasons why aggregate output decreases. First, the output of existing projects decreases. Second, the number of projects decreases. For example, a recession goes together with job destruction. But even projects of existing workers could be eliminated. Thus, there is an intensive as
well as an extensive margin to changes in output, just like there are two margins to changes in the total number of hours worked. We discuss below how to obtain estimates for the extensive margin of output changes. But this is not sufficient for our analysis. We have to identify which part of the adjustment along the extensive margin is due to inefficiencies.\(^{29}\) In terms of the terminology of the paper, we have to determine which of the observed changes in the number of projects are due to “cyclical projects”. The numbers presented are based on the assumption that \textit{all} observed changes along the extensive margin are due to inefficiencies.

There are, of course, other possible reasons why projects are terminated during downturns and restarted during expansions. One possible reason is the presence of timed-entry projects. Timed-entry projects continue operating during a downturn, but timed-entry projects that faced the exogenous separation shock during a recession are only restarted during the next expansion. Timed-entry projects could, thus, be responsible for some of the observed cyclical movements in the number of projects. In Section 2.4.2, we gave reasons why these projects are unlikely to matter much quantitatively. A second possible reason for adjustment along the extensive margin not related to inefficiencies is that projects are terminated during a recession because the value of \(\phi_p\Phi_-\) is too low relative to the value of the true outside option, \(\mu\). For example, during a recession, the value of market production could be too low compared to the value of leisure and home production. We doubt very much that most workers that become unemployed during an economic downturn prefer unemployment over being employed and we ignore this possibility. If the reader thinks otherwise and for example believes that say only 50% of all observed movements along the extensive margin is due to inefficiencies, then he/she should multiply the presented numbers by one half.

\(^{29}\)Here inefficiencies include a wide range of possibilities including sticky wages, the inability to overcome moral hazard problems related to the financing of positive NPV projects, and the motivation of the participants in the project.
Third element: Lower tail of the distribution

For a given value of $\phi_c$, we assume that

$$\int_{\tilde{\phi}_{p,\text{no-bc}}}^{\tilde{\phi}_{p,bc}} \phi_p f(\phi_p|\phi_c) d\phi_p \geq \int_{\tilde{\phi}_{p,\text{no-bc}}}^{\tilde{\phi}_{p,bc}} \phi_p f(\phi_p|\phi_c) d\phi_p,$$

where $f(\phi_p|\phi_c)$ is the density of $\phi_p$ conditional on $\phi_c$. That is, the output produced by projects above $\tilde{\phi}_{p,\text{no-bc}}$ is higher than the output produced by projects below $\tilde{\phi}_{p,\text{no-bc}}$. A sufficient condition for this to be true is that $\partial f(\phi_p|\phi_c)/\partial \phi_p \geq 0$, which is likely to be true given that we are in the left tail of the distribution.

Fourth element: Cyclical changes in output along the extensive margin

The assumption that changes in output along the extensive margin are due to inefficiencies simplifies the analysis considerably. But we still have to take a stand on how to calculate which part of the observed cyclical changes in output is due to changes along the intensive margin (changes in output due to changes in productivity of existing projects) and which part is due to the extensive margin (changes in output due to change in the number of projects). We follow two approaches to estimate the magnitude of the latter, i.e., $Y_{\text{cyclical}}/Y$. The first (and simple) approach is discussed here. In Appendix 2.D, we discuss a more elaborate method that is based on a panel data set for German wages. With this procedure we obtain a quantitatively similar estimate.

Using US total nonfarm employment from 1948Q1 to 2007Q4, we find measures for the volatility of detrended employment ranging from 0.0145 to 0.0388. In our model, in which there are two regimes, a boom and a recession, the difference in employment levels between a recession and a boom would then range between 2.90% and 7.76%. The corresponding numbers for the cyclical change in output are 3.36% to 8.30%. The observed cyclical change in the number of employed workers is used as a proxy for the cyclical change in the number of projects. To determine the importance of the extensive margin for output, we need to know how productive the workers are that are

---

30The standard deviation is equal to 0.0145 (0.0220) when data are detrended using the Hodrick-Prescott filter with a smoothing coefficient equal to 1,600 (100,000). It is equal to 0.0388 when a linear trend is used. If we extend the sample up to 2010Q4 then the ranking of these three numbers remain the same and their magnitudes are equal to 0.0146, 0.0231, and 0.0487.
associated with these changes in employment along the extensive margin.

Let $N_{\text{cyclical}}$ stand for the number of projects that stop operating during a recession, let $N$ stands for the time series average number of all projects, and let $\phi_{p,\text{ave}}$ stand for the average productivity of the $N_{\text{cyclical}}$ cyclical projects. This means that

$$\frac{Y_{\text{cyclical}}}{Y} = \frac{N_{\text{cyclical}}}{N} \phi_{p,\text{ave}}.$$

Suppose that $\phi_{p,\text{ave}}$ is equal to 40%. This seems conservative. Combined with the observed range for $N_{\text{cyclical}}/N$, this implies a range for $Y_{\text{cyclical}}/Y$ from 1.16% to 3.10%. This means that roughly 36% of the observed cyclical variation in GDP is assumed to be due to inefficiencies, i.e., to $Y_{\text{cyclical}}/Y$. This does not seem to be an excessive estimate for the role of inefficiencies. The higher this number, the higher the costs of business cycles.

**Fifth element: Using what is known to extrapolate**

With the assumptions made above, it is possible to get an estimate for the importance of cyclical projects with values of $\phi_c$ below $\tilde{\phi}_{c,\text{bc}}$, that is, values of $\phi_c$ that are low enough to make entry worthwhile in a boom even though the project has to be terminated during a recession. Here we discuss how to obtain an estimate for the importance of the cyclical projects with values of $\phi_c$ in between $\tilde{\phi}_{c,\text{bc}}$ and $\tilde{\phi}_{c,\text{no-bc}}$, that is values of $\phi_c$ that are such that entry is no longer attractive in a world with business cycles while entry is attractive in a world without business cycles. We will refer to the total output level of these projects as $Y_{\text{cyclical-PL}}$. Here PL stands for permanent loss, since in the world with business cycles their projects are never activated.

The difficulty with these projects is that they are never observed in the actual world, which does have business cycles. Still, several characteristics of this set of projects are known. First, the productivity levels of these cyclical projects are in the upper half of the range of values for $\phi_p$ associated with observed cyclical projects. This means

---

31 According to Goldin and Katz (2008), those without a college degree earn wages that are roughly half of those with a college degree. Thus, if all non-cyclical workers are as productive as college graduates and all cyclical workers are as productive as those without a college degree then we still underestimate the productivity of the cyclical workers.

32 This is documented in Figure 2.3. The observed cyclical projects are in two areas referred to as
that we only have to know the mass of projects in between \( \tilde{\phi}_{c, bc} \) and \( \tilde{\phi}_{c, no-bc} \) relative to the mass below \( \tilde{\phi}_{c, bc} \). Second, the values of \( \tilde{\phi}_{c, bc} \) and \( \tilde{\phi}_{c, no-bc} \) are known functions of parameters that can be calibrated. The key aspect of this part of our procedure is the assumption that conditional on \( \phi_p \), the mass in between \( \tilde{\phi}_{c, bc} (\phi_p) \) and \( \tilde{\phi}_{c, no-bc} (\phi_p) \) relative to the mass in between 0 and \( \tilde{\phi}_{c, bc} (\phi_p) \), is equal to the length of the interval in between \( \tilde{\phi}_{c, bc} (\phi_p) \) and \( \tilde{\phi}_{c, no-bc} (\phi_p) \) relative to the length of the interval in between 0 and \( \tilde{\phi}_{c, bc} (\phi_p) \).

2.6.2 Implementation: the bottom line

In Appendix 2.B, we give the exact formulas, but one can obtain a quite accurate estimate using the following simple procedure.

1. Using the formulas in the second part of Proposition 2.1, obtain a minimum and a maximum value of the impact of business cycles for projects with \( 0 \leq \phi_c \leq \tilde{\phi}_{c, bc} \). Take the average value.

2. These welfare consequences are expressed relative to the affected projects own productivity levels. Combining the estimate from the first step with the estimate for \( Y_{\text{cyclical}} / Y \) gives the total cost of business cycles due to cyclical projects with \( 0 \leq \phi_c \leq \tilde{\phi}_{c, bc} \).

3. Using the formulas in the second part of Proposition 2.1, obtain a minimum and a maximum value of the impact of business cycles for projects with \( \tilde{\phi}_{c, bc} < \phi_c \leq \tilde{\phi}_{c, no-bc} \). Take the average value.

4. The magnitude of \( (\tilde{\phi}_{c, no-bc} - \tilde{\phi}_{c, bc}) \) relative to the magnitude of \( (\tilde{\phi}_{c, bc} - 0) \) determines the magnitude of \( Y_{\text{cyclical-PL}} / Y \) relative to the magnitude of \( Y_{\text{cyclical}} / Y \). This makes it possible to calculate \( Y_{\text{cyclical-PL}} / Y \).

5. Combining the estimate of step 3 with \( Y_{\text{cyclical-PL}} / Y \) gives the total cost of business cycles due to cyclical projects with \( \tilde{\phi}_{c, bc} < \phi_c \leq \tilde{\phi}_{c, no-bc} \).

"gain" and "loss". The cyclical projects that are not observed are in the area referred to as "permanent loss", which is associated with the higher values of \( \phi_p \).
For example, suppose that the average individual welfare costs are equal to 15.3% and 31.0% for cyclical projects with \(0 \leq \phi_c \leq \tilde{\phi}_{c,\text{bc}}\) and \(\tilde{\phi}_{c,\text{bc}} \leq \phi_c \leq \tilde{\phi}_{c,\text{no-bc}}\), respectively. These numbers are averaged across values of \(\phi_c\) and when \(0 \leq \phi_c \leq \tilde{\phi}_{c,\text{bc}}\) across the gains and losses of the affected projects.\(^{33}\) Suppose that \(Y_{\text{cyclical}}/Y\) is equal to 3.1%, the upper end of our range of estimates. This means that the welfare costs for cyclical projects with \(\phi_c \leq \tilde{\phi}_{c,\text{bc}}\) relative to aggregate output is equal to \(15.3 \times 0.031 = 0.474\%\).

The magnitude of \(\left(\tilde{\phi}_{c,\text{no-bc}} - \tilde{\phi}_{c,\text{bc}}\right) / \left(\tilde{\phi}_{c,\text{bc}} - 0\right)\) together with the value of \(Y_{\text{cyclical}}/Y\) imply that the output that could be produced by projects with \(\phi_c > \tilde{\phi}_{c,\text{bc}}\) is equal to 5.3% of aggregate output. Combining this number with the individual welfare costs of 31.0% means that the welfare costs for cyclical projects with \(\phi_c > \tilde{\phi}_{c,\text{bc}}\) is equal to \(31.0 \times 0.053 = 1.643\%\). The total cost of business cycles would then be equal to \(0.474\% + 1.643\% = 2.12\%\).

### 2.7 Quantitative impact of business cycles

The top panel of Table 2.3 reports the welfare costs of business cycles. In Lucas (1987), the welfare costs of business cycles are estimated to be less than 0.1% when the coefficient of relative risk aversion is equal to 10.\(^{34}\) In the basic version of our model, we find welfare costs ranging from 0.13% to 2.12% with risk-neutral agents. Roughly one fourth is due to cyclical projects with values of \(\phi_c\) below \(\tilde{\phi}_{c,\text{bc}}\) (that have to pay startup costs more often) and the remainder is due to cyclical projects with values of \(\phi_c\) above \(\tilde{\phi}_{c,\text{bc}}\) (that are permanently driven out of business by business cycles). The results are quite sensitive to \(\rho\), the value of the survival rate of projects in a world without business cycles.\(^{35}\) In our model, business cycles are costly because they shorten the duration of projects. Consequently, the higher the expected duration of projects in

\(^{33}\)The example is based on the case with \(\rho = 0.975\) and \(\hat{\mu} = 0.2\) documented in Table 2.2. For example, at \(\phi_c = 0\) the average loss of a weak and strict-inequality project is equal to \((39.65 - 40.35)/2\). At \(\phi_c = \tilde{\phi}_{c,\text{bc}}\) this average is equal to \((61.9547 + 0)/2\). The average of these two numbers is equal to 15.3.

\(^{34}\)The costs of business cycles according to the Lucas formula are equal to \(0.5\sigma_c^2\gamma\), where \(\sigma_c\) is the standard deviation of the cyclical component of aggregate consumption and \(\gamma\) is the coefficient of relative risk aversion. Thus, if \(\sigma_c\) is equal to 0.013 and \(\gamma\) is equal to 10, then the costs of business cycles are equal to 0.084%.

\(^{35}\)Note that we cannot calibrate this parameter to observed separation rates, because in the real world economic downturns are important for separations.
Table 2.3: Welfare costs of business cycles

<table>
<thead>
<tr>
<th></th>
<th>( \rho = 0.875 )</th>
<th>( \rho = 0.9167 )</th>
<th>( \rho = 0.975 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no entrepreneurs; ( \phi_{c,\text{min}} = 0; \frac{Y_{\text{cyclical}}}{Y} = 1.16% )</td>
<td>( \hat{\mu} = \bar{\mu} = 0.2 )</td>
<td>0.18%</td>
<td>0.28%</td>
</tr>
<tr>
<td>no entrepreneurs; ( \phi_{c,\text{min}} = 0; \frac{Y_{\text{cyclical}}}{Y} = 3.1% )</td>
<td>( \hat{\mu} = \bar{\mu} = 0.2 )</td>
<td>0.49%</td>
<td>0.74%</td>
</tr>
<tr>
<td>Extended model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no entrepreneurs; ( \phi_{c,\text{min}} = \tilde{\phi}<em>{c,\text{no-bc}}/5; \frac{Y</em>{\text{cyclical}}}{Y} = 3.1% )</td>
<td>( \hat{\mu} = \bar{\mu} = 0.2 )</td>
<td>0.71%</td>
<td>1.22%</td>
</tr>
<tr>
<td>with entrepreneurs; ( \phi_{c,\text{min}} = 0; \frac{Y_{\text{cyclical}}}{Y} = 3.1% )</td>
<td>( \hat{\mu} = \bar{\mu} = 0.2 )</td>
<td>0.91%</td>
<td>1.38%</td>
</tr>
<tr>
<td>with entrepreneurs; ( \phi_{c,\text{min}} = \tilde{\phi}<em>{c,\text{bc}}/5; \frac{Y</em>{\text{cyclical}}}{Y} = 3.1% )</td>
<td>( \hat{\mu} = \bar{\mu} = 0.2 )</td>
<td>1.41%</td>
<td>2.47%</td>
</tr>
</tbody>
</table>

Notes: The table reports the permanent percentage increase in GDP needed to make the welfare of agents living in the economy with business cycles equal to the welfare of agents living in the economy without business cycles. \( \phi_{c,\text{min}} \) is the lower bound of the distribution of \( \phi_c \). \( Y_{\text{cyclical}}/Y \) is the amount of output that is earned by cyclical projects in a boom relative to total output. \( \hat{\mu} \) and \( \bar{\mu} \) are averages for the amount produced (in terms of leisure and/or home production) if the project does not operate as a fraction of market production for projects with values of \( \phi_c \) below and above \( \tilde{\phi}_{c,\text{bc}} \), respectively.

In a world without fluctuations, the stronger this effect will be. The highest value for \( \rho \) considered in Table 2.3 is 0.975, which corresponds to an expected duration of 10 years. If the value of \( \rho \) is increased from 0.975 to 0.99, then the welfare costs of business cycles increase from 2.12% to 3.77% (when \( \mu = 0.2 \)).

Although these numbers are a magnitude larger than the ones reported in Lucas (1987), they are still relatively small. In the next section, we discuss modifications of the model in which the costs of business cycles are substantially larger.

### 2.8 Extensions with larger effects

In this section, we discuss two modifications of the model. Both substantially increase the costs of business cycles.
2.8.1 Non-zero lower bound for $\phi_c$

The results presented above are based on the assumption that some projects can be created for free. This leads to a conservative estimate for the costs of business cycles.\textsuperscript{36} Quantitatively, this matters a lot. Instead, suppose that the lowest possible value for $\phi_c$ is equal to 20% of the highest value of $\phi_c$ at which entry is still profitable when inefficiencies and business cycles do not matter, thus, $\min\{\phi_c\} = \tilde{\phi}_{c,\text{no-bc}}/5$. As documented in Table 2.3, the modification increases the upper bound of our estimates for the welfare costs of business cycles from 2.12% to 15.5%.\textsuperscript{37}

This modification has a stronger effect on the outcome for higher values of $\rho$. The reason is that the lower bound of $\phi_c$ depends on the value of $\rho$. As the value of $\rho$ increases, the value of $\left(\tilde{\phi}_{c,\text{no-bc}} - \tilde{\phi}_{c,bc}\right) / \left(\tilde{\phi}_{c,bc} - \min\{\phi_c\}\right)$ increases, which means that the number of projects that would be resurrected if business cycles are eliminated increases relative to the level of observed cyclical projects.

2.8.2 Entrepreneurs and inefficient entry decision

Our current setup differs from the usual setup in which an entrepreneur pays the entry costs and makes the entry decision, but the revenues are shared with others, like workers. This model is in many aspects similar to the one described above and the equations are the same or almost the same. For example, the cut-off levels for the entry costs in the world without and with business cycles are now given by

\begin{align*}
\tilde{\phi}_{c,\text{no-bc}} & = \omega_e \frac{(\phi_p - \mu)}{1 - \beta \rho}, \\
\tilde{\phi}_{c,bc} & = \omega_e \frac{\phi_p (1 + \Delta \phi_p) - \mu}{1 - \beta \rho \pi}
\end{align*}

where $\omega_e$ is the entrepreneur’s share of the surplus. If $\omega_e$ is equal to 1, then there is no change in model.\textsuperscript{36}\textsuperscript{37}

\textsuperscript{36}The second part of proposition 2.1 shows that business cycles have a stronger negative impact on the combined well being of owners of strict-inequality and weak-inequality projects as $\phi_c$ increases.

\textsuperscript{37}In understanding these numbers it is important to keep in mind the following. When $\left(\tilde{\phi}_{c,\text{no-bc}} - \tilde{\phi}_{c,bc}\right) / \left(\tilde{\phi}_{c,bc} - \min\{\phi_c\}\right)$ increases and the output produced by cyclical projects with $\phi_c \leq \tilde{\phi}_{c,bc}$ remains equal to the calibrated value (either 1.16% or 3.1%), then the output that can be produced by projects with $\tilde{\phi}_{c,bc} \leq \phi_c \leq \tilde{\phi}_{c,\text{no-bc}}$ increases.
One might think that changing $\omega_e$ is not important since it would simply scale the curves for $\tilde{\phi}_{c, bc}$ and $\tilde{\phi}_{c, no-bc}$ and the relative values of these two cut-off levels play a key role in our calibration. But lowering the value of $\omega_e$ has large implications for the numerical results. The reason is the following. A lower value of $\omega_e$ implies that the entry costs relative to $\phi_p$ are lower. When calculating the welfare costs of business cycles, we take into account that cyclical projects that are permanently driven out of business because of business cycles no longer pay entry costs. This positive aspect of ceasing to produce is smaller when the entry costs paid are smaller, that is, it is smaller when $\omega_e$ is smaller. Thus, welfare costs of business cycles are larger for smaller values of $\omega_e$.

There is another reason why the welfare costs are larger when $\omega_e$ is smaller. When entry is efficient, i.e., when $\omega_e = 1$, then business cycles have a positive effect, because business cycles introduce the worthwhile option to postpone entry. As discussed above, this effect is quantitatively very small. As $\omega_e$ decreases, this small gain of business cycles can turn into a cost. The reason is the following. The option to wait still has positive value for the entrepreneur. The problem is that the entrepreneur does not take into account that by postponing entry he also postpones the worker getting his share of the revenues, which would exceed $\mu$ if entry is inefficient, i.e., when $\omega_e < 1$. The value of $\omega_e$, thus, determines whether business cycles are beneficial for timed-entry projects or not. Quantitatively, however, the impact of business cycles on timed-entry projects is always small.\(^{38}\)

What value for $\omega_e$ to use? This parameter captures the reward for the pure entrepreneurial activity of starting the project, thus, excludes revenues to other participants like workers and those that provide financing. Consequently, one would think that the value of $\omega_e$ is small. Here we set $\omega_e$ equal to 0.125. This is likely to be still high.\(^{39}\) If this value is indeed too high, then our estimates are conservative, since the lower $\omega_e$, the higher the costs of business cycles.

Table 2.3 reports the results for the model with entrepreneurs when $\min \{ \phi_c \}$ is

---

\(^{38}\)In this version of the paper, we ignore the timed-entry projects. In Den Haan and Sedlacek (2009), we explicitly take these projects into account, which makes the analysis much more cumbersome, but has virtually no effect on the results.

\(^{39}\)For example, Den Haan and Kaltenbrunner (2009) set $\omega_e$ equal to 0.0228 to match the observed employment volatility.
equal to 0 and when \( \min \{ \phi_c \} \) is equal to 20% of \( \tilde{\phi}_{c, \text{no-bc}} \). When the lower bound for \( \phi_c \) is equal to zero, then the upper bound of our estimates increase from 2.12% to 3.95%. If we set the lower bound for \( \phi_c \) equal to 20% of \( \tilde{\phi}_{c, \text{no-bc}} \) then the welfare costs of business cycles are equal to 34.6%.

### 2.9 Related literature

Following the classic Lucas (1987) paper, there have been numerous attempts to develop models in which business cycles are costly. One strand of the literature considers preferences in which fluctuations are more harmful to the agent. But if agents are truly highly risk averse, then—as pointed out in Lucas (2003)—the question arises why high risk aversion does not show up in, for example, the diversification of individual portfolios, the level of insurance deductibles, or the wage premiums of jobs with high earnings risk. A second strand of the literature considers the possibility that risk is not spread evenly across agents. When idiosyncratic risk is persistent, then this line of research generates estimates for the cost of business cycles that are an order of magnitude larger than those found by Lucas. The idea is that unemployment has very negative consequences for the individual and is a relatively rare event. Business cycles can be costly if risk sharing among agents is sufficiently limited, but the agents’ degree of risk aversion is again important.

Lucas’ calculations on the cost of business cycles are based on a comparison between two economies, one with and one without business cycles, that both have the same long-run growth path. But this is not necessarily the case. It could very well be that the presence of business cycles has long-term level or even growth effects. Empirical evidence for this view can be found in Ramey and Ramey (1995), Martin and Rogers (2000), Loayza, Ranciere, Serven, and Ventura (2007), Burnside and Tabova (2009), and Den Haan and Sedlacek (2009).

---

40See Den Haan and Sedlacek (2009), for details on how welfare costs of business cycles are calculated when \( \omega_e < 1 \).
41See Lucas (2003) for a summary.
42Examples of this line of research are Alvarez and Jermann (2005) and Tallarini (2000).
44These papers establish an unconditional negative correlation between business cycles and real
and middle income countries, the link also exists for OECD countries.

Our paper is related to a set of papers that build theoretical models in which fluctuations have level or growth effects. As shown in Ramey and Ramey (1991), Jones, Manuelli, and Stacchetti (2000), Epaualard and Pommeret (2003), and Barlevy (2004), there is a relationship between growth and volatility in endogenous growth models. Barlevy (2004) points out that for a reduction in volatility to have a quantitatively important effect on output, it is important that the increase in investment induced by a reduction in volatility not only increases the growth rate of consumption, but also does not lead to an initial reduction in the level of consumption.\footnote{Barlevy (2004) accomplishes this by introducing diminishing returns to investment into an endogenous growth model. The nonlinearity makes it possible for a reduction in fluctuations to have a positive effect on the growth rate, even if average investment levels, and thus the initial consumption level, are not affected by fluctuations.} Barlevy (2004) accomplishes this by introducing diminishing returns to investment into an endogenous growth model. The nonlinearity makes it possible for a reduction in fluctuations to have a positive effect on the growth rate, even if average investment levels, and thus the initial consumption level, are not affected by fluctuations.

Such a nonlinearity is also important in papers that show that volatility has a negative effect on the average level of real activity. Gali, Gertler, and Lopez-Salido (2007) consider a simple New-Keynesian model in which the efficiency losses due to mispricing in a recession are not offset by the efficiency gains in a boom. Business cycles are then welfare reducing, although the effects turn out to be small. Jung and Kuester (2008) and Hairault, Langot, and Osotimehin (2010) show that the matching model contains a nonlinearity that causes volatility in job finding rates to reduce average unemployment. In particular, increases in job finding rates during booms have less of an impact than decreases in job finding rates during recessions, because the unemployment rate is smaller during booms than recessions.

How does our explanation compare with the models of these papers? Our main mechanism is not that expansions and contractions have asymmetric effects on inefficiencies or asymmetric effects on entry decisions. That is, if we would add a third state to our model in which $\Phi_p$ is equal to $\Phi_0 \equiv (\Phi_+ + \Phi_-)/2$, then there is no robust reason for a positive correlation. For example, Levchenko, Ranciere, and Thoenig (2009) show that financial liberalization leads to an increase in both volatility and expected economic growth.

\footnote{Mertens (2008) shows how fluctuations can affect the average level of investment in a model without endogenous growth. In his model, agents have distorted beliefs and fluctuations worsen this distortion. Consequently, the risk premium is higher and the average capital stock is lower in a world with business cycles.}
for asymmetric effects if $\Phi_p$ increases from $\Phi_0$ to $\Phi_+$ or decreases from $\Phi_0$ to $\Phi_-$.\footnote{The model could generate such asymmetries, but the mass of projects would have to be distributed in a particular way.} The main reason why fluctuations are costly in our framework is that some projects are permanently driven out of business. The reason is that the inefficiency does not allow the worsening of the friction during a recession to be offset with a loosening during a boom. This loss does not introduce an asymmetry in generated business cycle fluctuations. Fluctuations are also costly for projects with an entry cost that is low enough to make entry worthwhile, even if the project is only viable during an expansion. For this type of project there is a negative and a positive effect in the sense that fluctuations end some projects during a recession but make additional projects viable during a boom. Even if the effects cancel out in terms of output, they do not in terms of welfare, since entry costs have to be paid more often in a world with fluctuations.

2.10 Concluding comments

According to our model, business cycles are costly for two types of reasons. First, since business cycles shorten the expected duration of projects and projects would pay startup costs less often in a world without business cycles. Second, business cycles make some projects completely impossible. In the simple version of our model, the first type of cost does not exceed 0.5% of aggregate output per period. In the experiments we considered, this type of cost is usually not much higher than this. But larger numbers are possible.\footnote{Suppose that (i) there is no mass below $\tilde{\Phi}_{p,\text{no-bc}}$, that is, there are no projects that gain from the presence of business cycles, (ii) cyclical fluctuations in output are due solely to changes in the extensive margin, and (iii) entry costs are equal to zero. An upper bound for these types of costs would then be the observed difference in the level of output during a boom and a recession, which is roughly 4%. If entry costs are not zero, then these costs could be higher than 4%.} How large the second type of cost is depends on how many projects would come into existence if there are no business cycles. In the real world, we never see these potential projects, so it is not easy to obtain an estimate. The estimates we report for the simple model are based on what we feel is a conservative ”extrapolation” of the environment we do observe.\footnote{In particular, we assume that the lower bound for startup costs is equal to zero and cyclical projects are only 40% as productive as other projects.} This leads to non-trivial costs, but—combined with the first type of costs—our maximum cost estimate is ”only” equal to 2.12% of
aggregate output. However, we also show that one can quite easily get much larger costs. This occurs when one agent is responsible for paying the startup costs, but the revenues of the projects are shared with others and when there are not many projects with very low entry costs.

We end the paper with the following somewhat disturbing observation. Our numerical work focuses on the costs of business cycles. But our time-varying shock, $\Phi_{p,t}$, does not have to be an aggregate shock. It also could be a sectoral, geographical, and even an idiosyncratic shock. Consequently, if even the costs of business cycles are already non-trivial, then the costs of all fluctuations could very well be a staggering number.
2.A Theories consistent with our efficiency requirement

In the main text, we simply imposed the friction that firms can only operate if

\[ \phi \Phi_{p,t} \geq \chi_t. \]  

(2.19)

In this section, we show that that two very different theories generate such a requirement. The first is the contractual fragility framework of Ramey and Watson (1997) and the second is a framework in which the entrepreneur has to borrow to finance the investment made and borrowing is subject to a standard agency problem. Section 2.A.1 discusses the contractual fragility framework and Section 2.A.2 discusses the framework with the financial friction. The contractual fragility specification results in a condition that is exactly equal to the one given in equation (2.19). The financial friction leads to a slightly different specification in which one can expect the costs of business cycles to be even higher.

2.A.1 Contractual fragility of Ramey and Watson (1997)

The physical environment of the model described here is identical to the one described in the main text. The difference is that the efficiency requirement is not simply imposed, but is an implication of the model. In Ramey and Watson (1997), both participants have the option to cheat. For our purpose, it is sufficient if just the entrepreneur has such an option. The cheating option may be privately attractive, but is inefficient from the relationship’s point of view. For example, the entrepreneur may deviate from the original business plan and choose one that is riskier, but gives him personally more prestige. It is also possible that he diverts funds to himself or acquaintances. The total current-period private benefits the entrepreneur can obtain by cheating are equal to \( \chi_e \) and these consist of the actual funds the entrepreneur receives, \( \chi_e \), plus any non-pecuniary benefits, \( \chi_e - \chi_{e,s} \). The entrepreneur obtains pecuniary benefits by extracting a larger share of the resources than agreed upon, for example, by not paying out overtime or by not promoting workers. Increases in prestige or human capital and
improvements of the entrepreneur’s network are examples of non-pecuniary benefits. If the entrepreneur chooses the alternative business plan, then output is equal to

$$\phi \chi \phi p \Phi p,t. \tag{2.20}$$

The alternative business plan causes the worker extra disutility equal to $\chi w \geq 0$.\(^{49}\) For the alternative choice to be inefficient, it must be the case that

$$\left(1 - \phi \chi\right) \phi p \Phi p,t > (\chi c - \chi c $$. \(\tag{2.21}\)

That is, the loss in the project’s revenues is larger than the net utility gain (when $\phi \chi < 1$) or the gain in revenues is smaller than the net utility loss (when $\phi \chi > 1$).

An entrepreneur is only willing to choose the original business plan if

$$\phi p \Phi p,t - w_t + \beta E_t \left[\rho N_e(\phi c, \phi p, 1, \Phi p,t+1) + (1 - \rho) N_e(\phi c, \phi p, 0, \Phi p,t+1)\right] \geq \chi c + \beta E_t \left[\rho N_e(\phi c, \phi p, 1, \Phi p,t+1) + (1 - \rho) N_e(\phi c, \phi p, 0, \Phi p,t+1)\right], \tag{2.22}$$

where $w_t$ is the wage rate of the worker under the original business plan.\(^{50}\) For simplicity, we assume that the entrepreneur’s choice to cheat does not affect his continuation value.\(^{51}\) Equation (2.22) can, then, be written as

$$\phi p \Phi p,t - w_t \geq \chi c. \tag{2.23}$$

The current-period benefits of the worker when he is not employed are equal to $\mu$.

\(^{49}\)The alternative business plan may not only require more effort from the worker, but may also force him to move to a different plant and may even result in dismissal.

\(^{50}\)Under the alternative business plan, the worker receives $\phi \chi \phi p \Phi p,t - \chi c $.

\(^{51}\)In the unfair world we live in, it is probably not an unrealistic assumption that the employer can impair his workers’ well being without this having much effect on the options available to him in the next period. The case when the entrepreneur’s continuation value is affected is discussed at the end of this section.
A worker is only willing to participate in a project if

\[ w_t + \beta E_t [\rho N_w(\phi_c, \phi_p, \Phi_{p,t+1}) + (1 - \rho)U_w(\phi_c, \phi_p, \Phi_{p,t+1})] \geq \mu + \beta E_t [U_w(\phi_c, \phi_p, \Phi_{p,t+1})], \]  

(2.24)

where \( w_t \) is the wage rate of the worker, \( N_w(\phi_c, \phi_p, \Phi_{p,t+1}) \) is the discounted value of current and future benefits that accrue to the worker when he starts next period in a relationship, \( U_w(\phi_c, \phi_p, \Phi_{p,t+1}) \) the discounted value of current and future benefits that accrue to the worker when he starts next period not being in a relationship. Since the matching probability is equal to 1, it does not matter whether you leave period \( t \) in a relationship or not; in period \( t + 1 \) the worker still has the freedom to choose what is best. Consequently, \( N_w(\phi_c, \phi_p, \Phi_{p,t+1}) = U_w(\phi_c, \phi_p, \Phi_{p,t+1}) \) and the condition given in equation (2.24) can simply be written as

\[ w_t \geq \mu. \]  

(2.25)

A necessary and sufficient condition to satisfy both the participation condition of the worker and the no-cheating condition of the entrepreneur is given by\(^{52}\)

\[ \phi_p \Phi_{p,t} \geq \chi_e + \mu. \]  

(2.26)

If we let \( \chi = \chi_e + \mu \), then we get exactly the condition in equation (2.2) used in the main text to model the friction.\(^{53}\)

In the remainder of this section, we provide some more intuition on why contractual fragility makes production impossible when \( \phi_p \Phi_{p,t} \) is less than \( \chi \). Consider the case when

\[ \mu < \phi_p \Phi_{p,t} < \chi_e + \mu. \]  

(2.27)

If an existing project does not operate when \( \phi_p \Phi_{p,t} > \mu \), then this is not efficient.

---

\(^{52}\)Since the entrepreneur can never earn any benefits outside of a relationship, there is no participation constraint for the entrepreneur.

\(^{53}\)For simplicity, we assumed that \( \chi_e \) and \( \mu \) are not time-varying, but the analysis here allows for these values to be time-varying.
However, for these values of $\phi_p \Phi_{p,t}$ it is not possible to both pay the entrepreneur enough so that he will not choose the alternative business plan and pay the worker enough so that his wage exceeds $\mu$. In this case, the project does not operate. The idea of the contractual fragility of Ramey and Watson (1997) is that no credible contracts can be written that will prevent the entrepreneur from cheating and implementing the inefficient alternative business plan. The entrepreneur may promise that he will pay the worker a wage above $\mu$ and that he will not implement the alternative business plan, but if $\phi_p \Phi_{p,t} < \chi_e + \mu$, then the entrepreneur cannot both pay the worker more than $\mu$ and satisfy his own incentive compatibility condition. The worker knows that the entrepreneur will face this dilemma and chooses not to work for this entrepreneur.

The beauty of the contractual fragility framework of Ramey and Watson (1997) is that it allows for the possibility that projects are not activated even though everybody would be better off if the project is activated. The reason is that it is not possible to write contracts that prevent cheating.

The framework described here results in a restriction that is identical to the one used in the main text. There are, of course, more general specifications. For example, the values of $\chi_e$ and $\chi_w$ could depend on $\phi_c$. A higher entry cost means larger investments in the firm and possibly more options for the entrepreneur to extract resources from the firm. The financial friction also predicts that $\chi$ depends positively on $\phi_c$ and we will argue that the costs of business cycles are higher when there is such positive dependence. But if the entrepreneur cannot continue operating his project after having exploited the worker and has to pay part of $\phi_c$ again, then this would dampen and possibly even overturn this effect.

2.A.2 Financial friction

There are many different models with financial frictions. In this section, we develop a model in which the friction is such that firms have to be sufficiently productive to make operating the project possible. The condition on firm productivity has a somewhat different form than the ad hoc efficiency requirement imposed in the main text. But this difference actually implies larger costs of business cycles.

We assume that the entrepreneur does not have any net worth and has to borrow
to finance the entry costs. Also, we assume that the entrepreneur simply rolls over this debt every period until he is hit by the exogenous destruction shock and he stops producing. Let the interest rate be equal to \( r \), which includes a premium for the fact that producers default on the debt obligation when they stop producing. For simplicity, we assume that all firms face the same interest rate, for example, because lenders cannot distinguish between different types of borrowers.\(^{54}\) A standard financial friction is a limit on the amount that can be collateralized. In particular, assume that the entrepreneur can extract \( \chi_e \) when he defaults on his loan. Examples of assets of the firm that cannot be collateralized are human capital, the value of the good will created, and the value of the networks built up.

The entrepreneur would not default on his loan if

\[
\phi_p \Phi_{p,t} - w_t - r \phi_c + \beta E_t [\rho N_e(\phi_c, \phi_p, 1, \Phi_{p,t+1}) + (1 - \rho) N_e(\phi_c, \phi_p, 0, \Phi_{p,t+1})] \geq \chi_e + \beta E_t [N_e(\phi_c, \phi_p, 0, \Phi_{p,t+1})] \tag{2.28}
\]

If the entrepreneur borrows the funds to finance the entry costs, then the value of the project’s revenues that accrue to the entrepreneur before or after entry costs have been paid are identical, that is, \( N_e(\phi_c, \phi_p, 1, \Phi_{p,t+1}) = N_e(\phi_c, \phi_p, 0, \Phi_{p,t+1}) \) and we can write the last equation as

\[
\phi_p \Phi_{p,t} - w_t - r \phi_c \geq \chi_e. \tag{2.29}
\]

Since \( w_t \geq \mu \), it must be true that

\[
\phi_p \Phi_{p,t} \geq \chi_e + \mu + r \phi_c \tag{2.30}
\]

and if we let \( \chi = \chi_e + \mu \), then we get

\[
\phi_p \Phi_{p,t} \geq \chi + r \phi_c. \tag{2.31}
\]

\(^{54}\)To ensure that even the amount borrowed does not reveal the type of the borrower, one could assume that one can scale each project. An entrepreneur with a high value for \( \phi_c \) can then invest in say half a project and, thus, hide that he has a project with a high value for \( \phi_c \).
Figure 2.4: Projects affected by business cycles in the presence of financial frictions

Notes: The shaded areas in this graph indicate the projects that are affected by business cycles. Light grey: Cyclical projects that operate during a boom and do not operate during a recession. Projects in the "gain" ("loss") area never (always) operate in a world without business cycles. Grey: Cyclical Projects that can overcome inefficiencies during a boom, but their entry costs are too high to make entry worthwhile given that inefficiencies will force exit during a recession. Point A corresponds to point A in the other figures.

This condition differs from the one we used in the main text, because it implies an upward sloping cut-off curve in the \((\phi_c, \phi_p)\) space, whereas the condition that \(\phi_p \Phi_{p,t} \geq \chi\) implies a vertical cut-off for the production decision.

**Disappearance of timed-entry projects**

Although the efficiency requirement looks somewhat different than the one used in the main text, the main implications are the same. There are two differences, both will make the costs of business cycles stronger. The first is that there are no longer timed-entry projects. The reason is the following. The entry condition when there are no business cycles and no frictions is given by

\[
N_{\text{no-bc}}(\phi_c, \phi_p, 1) - \phi_c \geq \mu + \beta N_{\text{no-bc}}(\phi_c, \phi_p, 0),
\]  

(2.32)
which implies that
\[ \phi_c \leq \tilde{\phi}_{c, \text{no-bc}} (\phi_p) = \frac{\phi_p - \mu}{1 - \beta \rho}. \]  
(2.33)

From equation (2.31) we get for the case of no business cycles that
\[ \phi_c \leq \hat{\phi}_{c, \text{no-bc}} (\phi_p) = \frac{\phi_p - \chi}{r}. \]  
(2.34)

The value of \( r \) is given by \( \beta \rho / (1 - \beta \rho) \).\(^{55}\) If we strengthen assumption 2.1 slightly by assuming that \( \mu < \beta \rho \chi \), then
\[ \phi_c \leq \beta \rho \frac{\phi_p - \chi}{1 - \beta \rho} = \hat{\phi}_{c, \text{no-bc}} (\phi_p) < \tilde{\phi}_{c, \text{no-bc}} (\phi_p) = \frac{\phi_p - \mu}{1 - \beta \rho}. \]  
(2.35)

This means that the entry condition given in equation (2.33) is not binding. As long as aggregate fluctuations are not too large this remains true for projects that continue until they are hit by the exogenous destruction shock. This means that timed-entry projects have disappeared out of the model. This case is illustrated in Figure 2.4. Whereas in Figure 2.3 the cut-off level of the inefficiency condition intersects with the entry condition, in Figure 2.4 the cut-off level condition of the inefficiency condition makes the entry condition redundant.

Quantitatively, we always found the results for timed-entry projects to be small. So the disappearance of the timed-entry projects cannot matter much for the results. What matters is that the cyclical projects are still there. To this question we turn next.

**Cyclical projects and larger welfare costs of business cycles**

Figure 2.4 documents that the same three groups of cyclical projects are present in this version of the model. In the previous subsection, we showed that the condition that the entry costs should be less than the NPV of the revenue stream is no longer binding if the efficiency condition is given by equation (2.31). That is, \( \hat{\phi}_{c, \text{no-bc}} (\phi_p) < \tilde{\phi}_{c, \text{no-bc}} (\phi_p) \). This inequality remains true in the presence of business cycles if business cycles do not shorten the project’s life expectancy. If projects can only satisfy the efficiency

\(^{55}\)The interest rate is higher than the discount rate, since the entrepreneur stops paying interest when he is hit by an exogenous destruction shock.
requirement in a boom, then the life expectancy is shorter. This results in a downward shift of $\tilde{\phi}_c$ and for some projects $\tilde{\phi}_{c, bc} (\phi_p, \Phi_+) < \hat{\phi}_{c, no-bc} (\phi_p, \Phi_+)$. This means that there are again projects that do not enter in a boom even though they satisfy the efficiency requirement.

Although the same three types of cyclical projects exist, the welfare costs of business cycles could very well be higher in this version of the model. Consider the cyclical projects that are permanently eliminated by business cycles. As in the model discussed in the main text, these are the projects with the higher values for $\phi_c$. But in this version of the model, they have higher levels of output, whereas in the model discussed in the main text they had the same level of output as one group of cyclical projects. If this is the correct model, then our calibration procedure may have underestimated the output potential of the projects that are permanently eliminated because of business cycles.

The analysis so far assumed that the interest rate was not affected by business cycles. But given that some firms only survive until the next recession one would expect the interest rate to be higher in a world with business cycles. For now, assume that all firms have the same interest rate. Then the increase in the interest rate would be relatively small. Nevertheless, it would mean that business cycles do not just put a mean-preserving spread on the efficiency condition given in equation (2.31), they actually make it more difficult to satisfy it on average.

The situation would be worse if lenders can identify firms that will only survive until the next recession. Then these firms will have to face a further downward shift in the efficiency condition, increasing the number of firms that are permanently eliminated.

2.B Implementation: detailed description

Our strategy to determine the net welfare losses for all cyclical projects is divided in two parts. In the first part, we determine the net loss for cyclical projects with values of $\phi_c$ below $\tilde{\phi}_{c, bc}$. In the second part, we determine the loss for cyclical projects with values of $\phi_c$ above $\tilde{\phi}_{c, bc}$. The joint density of $\phi_c$ and $\phi_p$ is denoted by $f(\phi_c, \phi_p)$.

1. To calculate the net loss for cyclical projects with $0 < \phi_c \leq \tilde{\phi}_{c, bc}$, we do the following:
(a) **Average of φ_c given φ_p.** The question arises what to use for the distribution of φ_c and φ_p conditional on the project being a cyclical project with φ_c ≤ ˜φ_{c,bc}. In this step, we deal with the relevant aspects of the distribution of φ_c. We assume that the lower bound for φ_c is given by 0. It may not be realistic to assume that there are projects with zero startup costs, but we obtain a conservative estimate by assuming that the lower bound is equal to 0. The reason is that the net welfare costs are lower if more projects have low values of φ_c. The upper bound for φ_c in this group of projects is given by ˜φ_{c,bc} which is given by

\[ ˜φ_{c,bc} = \frac{φ_p (1 + ∆φ_p) - µ}{1 - βρπ}. \]

Finally, we assume that the average value of φ_c, conditional on φ_p, for projects with a value of φ_c in between 0 and ˜φ_{c,bc} is equal to the average of the two endpoints, i.e., we assume it to be equal to \( (φ_{c,bc} + 0)/2 \). Besides this conditional average, we do not need any further information about the distribution of φ_c.
(b) The net loss for cyclical projects with $\phi_c \leq \widetilde{\phi}_{c,\text{bc}}$ is given by

$$
\int_{\phi_{p,\text{bc}}(\Phi_-)}^{\phi_{p,\text{bc}}(\Phi_+)} \int_{\phi_{c,\text{bc}}(\phi_p)}^{\phi_{c,\text{bc}}(\phi_p)} \frac{L(\phi_c, \phi_p, \Delta \phi_p) \phi_p}{Y} f(\phi_c, \phi_p) d\phi_c d\phi_p = (2.36)
$$

$$
\int_{\phi_{p,\text{no-bc}}}^{\phi_{p,\text{bc}}(\Phi_-)} \int_{\phi_{c,\text{bc}}(\phi_p)}^{\phi_{c,\text{bc}}(\phi_p)} (1 - \Delta \phi_p) \phi_p - \mu + \phi_c (1 - \beta \rho \pi - 2 (1 - \beta \rho)) f(\phi_c, \phi_p) d\phi_c d\phi_p \leq \frac{2Y}{2Y}
$$

$$
\int_{\phi_{p,\text{bc}}(\Phi_-)}^{\phi_{p,\text{bc}}(\Phi_+)} \int_{\phi_{c,\text{bc}}(\phi_p)}^{\phi_{c,\text{bc}}(\phi_p)} -\Delta \phi_p \phi_p + \phi_c \beta \rho (1 - \pi) f(\phi_c, \phi_p) d\phi_c d\phi_p = \frac{2Y}{4Y}
$$

$$
\int_{\phi_{p,\text{bc}}(\Phi_-)}^{\phi_{p,\text{bc}}(\Phi_+)} \int_{\phi_{c,\text{bc}}(\phi_p)}^{\phi_{c,\text{bc}}(\phi_p)} -2\Delta \phi_p + \phi_{c,\text{bc}} \beta \rho (1 - \pi) f(\phi_c, \phi_p) d\phi_c d\phi_p = \frac{2Y}{4Y}
$$

$$
-2\Delta \phi_p + \left(1 + \Delta \phi_p\right) \frac{-\hat{\mu}}{1 - \beta \rho \pi} \beta \rho (1 - \pi) \int_{\phi_{p,\text{bc}}(\Phi_-)}^{\phi_{p,\text{bc}}(\Phi_+)} \int_{\phi_{c,\text{bc}}(\phi_p)}^{\phi_{c,\text{bc}}(\phi_p)} \phi_p f(\phi_c, \phi_p) d\phi_c d\phi_p = \frac{2Y}{4Y}
$$

In the first row, the value of $L(\cdot)$ is multiplied by $\phi_p$ since $L(\cdot)$ is scaled by $\phi_p$. The expressions used for $L(\cdot)$ are given in proposition 2.1. The inequality in the next line is based on the assumption that the mass of projects below $\widetilde{\phi}_{p,\text{no-bc}}$ does not exceed the mass above $\widetilde{\phi}_{p,\text{no-bc}}$. Given that we are in the left tail of the distribution, this is unlikely to be the case. In the following step, we use that the average value of $\phi_c$, conditional on $\phi_p$, is given by the midpoint of the interval as discussed in part a. The remaining steps are simple algebra, where we define $\hat{\mu}$ is the mean value of the outside option of the project as a fraction of $\phi_p$ across all cyclical projects with $\phi_c$ below $\widetilde{\phi}_{c,\text{bc}}$.

The parameters $\Delta \phi_p$, $\beta$, $\rho$, $\hat{\mu}$, and $\pi$ can be relatively easily calibrated. See Section 2.6.1 and appendix 2.D for a discussion on how we calculate $Y_{\text{cyclical}}/Y$.

---

56See appendix 2.C for a derivation.
2. To calculate the net loss for cyclical projects with $\tilde{\phi}_{c,\text{bc}} < \phi_c \leq \tilde{\phi}_{c,\text{no-bc}}$ we do the following:

(a) *Average of $\phi_c$ given $\phi_p$.* As above, the average value of $\phi_c$ for projects with a value of $\phi_c$ in between $\tilde{\phi}_{c,\text{bc}}$ and $\tilde{\phi}_{c,\text{no-bc}}$ is assumed to be the average of the end points, that is, it is assumed to be equal to $(\tilde{\phi}_{c,\text{bc}} + \tilde{\phi}_{c,\text{no-bc}})/2$.

(b) The welfare loss of cyclical projects with a value of $\phi_c$ above $\phi_{c,\text{bc}}$ is given by

\[
\int_{\tilde{\phi}_{p,\text{no-bc}}}^{\tilde{\phi}_{p}(\Phi_\mu)} \int_{\tilde{\phi}_{c,\text{bc}}(\phi_p)}^{\tilde{\phi}_{c,\text{no-bc}}(\phi_p)} \frac{L(\phi_c, \phi_p, \Delta \phi_p) \phi_p}{Y} f(\phi_c, \phi_p) d\phi_c d\phi_p = (2.37)
\]

(c) *Output generated by cyclical projects with $\tilde{\phi}_{c,\text{bc}} < \phi_c \leq \tilde{\phi}_{c,\text{no-bc}}$.* If a project’s value of $\phi_c$ is above $\tilde{\phi}_{c,\text{bc}}$, then it is too high to make entry profitable in a world with business cycles. Consequently, we do not observe these projects in the real world. These projects, however, have one characteristic in common with projects we do observe and that is their productivity level, $\phi_p$. In particular, the productivity levels of these projects are in between $\tilde{\phi}_{p,\text{no-bc}}$ and $\tilde{\phi}_{p}(\Phi_\mu)$, which means that they are in the upper half of the productivity levels of the cyclical projects that produce $Y_{\text{cyclical}}$. Let $f(\phi_c | \phi_p)$ be the density of $\phi_c$ given $\phi_p$. To calculate the value of $Y_{\text{cyclical}-\text{PL}}$ given the value
of \( Y_{\text{cyclical}} \), we would need to know the value of

\[
\int_{\phi_{c,\ell}^{\text{bc}(\phi_p)}}^{\phi_{c,\ell}^{\text{no-bc}(\phi_p)}} f(\phi_c | \phi_p) d\phi_c
\]

that is, the mass of cyclical projects with a value of \( \phi_c \) above \( \tilde{\phi}_{c,\ell}^{\text{no-bc}} \) to the mass of cyclical projects with a value of \( \phi_c \) below \( \tilde{\phi}_{c,\ell}^{\text{bc}} \). We assume that a reasonable estimate for this ratio is to use the lengths of the intervals, thus

\[
\int_{\phi_{c,\ell}^{\text{bc}(\phi_p)}}^{\phi_{c,\ell}^{\text{no-bc}(\phi_p)}} f(\phi_c | \phi_p) d\phi_c \approx \frac{\tilde{\phi}_{c,\ell}^{\text{no-bc}}(\phi_p) - \tilde{\phi}_{c,\ell}^{\text{bc}}(\phi_p)}{\tilde{\phi}_{c,\ell}^{\text{bc}}(\phi_p) - 0}.
\]

This implies that

\[
\frac{Y_{\text{cyclical-PL}}}{Y} = \int_{\tilde{\phi}_{p,\ell}^{\text{bc}(\Phi)}}^{\tilde{\phi}_{p,\ell}^{\text{no-bc}(\Phi)}} \frac{\tilde{\phi}_p Y f(\phi_c, \phi_p) d\phi_c d\phi_p}{\tilde{\phi}_{c,\ell}^{\text{bc}(\phi_p)}} \int_{\phi_{c,\ell}^{\text{bc}(\phi_p)}}^{\phi_{c,\ell}^{\text{no-bc}(\phi_p)}} f(\phi_c | \phi_p) d\phi_c d\phi_p
\]

\[
= \int_{\tilde{\phi}_{p,\ell}^{\text{bc}(\Phi)}}^{\tilde{\phi}_{p,\ell}^{\text{no-bc}(\Phi)}} \frac{\tilde{\phi}_p Y f(\phi_c, \phi_p) d\phi_c d\phi_p}{\tilde{\phi}_{c,\ell}^{\text{bc}(\phi_p)}} \int_{\phi_{c,\ell}^{\text{bc}(\phi_p)}}^{\phi_{c,\ell}^{\text{no-bc}(\phi_p)}} f(\phi_c | \phi_p) d\phi_c d\phi_p
\]

\[
\approx X \int_{\tilde{\phi}_{p,\ell}^{\text{bc}(\Phi)}}^{\tilde{\phi}_{p,\ell}^{\text{no-bc}(\Phi)}} \frac{\tilde{\phi}_p Y f(\phi_c, \phi_p) d\phi_c d\phi_p}{\tilde{\phi}_{c,\ell}^{\text{bc}(\phi_p)}} \int_{\phi_{c,\ell}^{\text{bc}(\phi_p)}}^{\phi_{c,\ell}^{\text{no-bc}(\phi_p)}} f(\phi_c | \phi_p) d\phi_c d\phi_p
\]

\[
= X \frac{Y}{\tilde{\phi}_{p,\ell}^{\text{bc}(\Phi)}} \int_{\phi_{c,\ell}^{\text{bc}(\phi_p)}}^{\phi_{c,\ell}^{\text{no-bc}(\phi_p)}} \phi_p f(\phi_c, \phi_p) d\phi_c d\phi_p
\]

\[
> X \frac{Y_{\text{cyclical}}}{2 Y}.
\]

where

\[
X = \frac{1 - \tilde{\mu} - \beta \rho}{1 + \Delta - \tilde{\mu} - \beta \rho}.
\]

The inequality follows from the fact that the output levels in \( Y_{\text{cyclical-PL}} \) are all above \( \tilde{\phi}_{p,\ell}^{\text{no-bc}} \) whereas the output levels in \( Y_{\text{cyclical}} \) are both above and
Chapter 2

2.C Proofs

2.C.1 Part 1 of proposition 2.1

\( \tilde{\phi}_{c,\text{no-bc}} (\phi_p) \) is defined as the value of \( \phi_c \) at which the value of activating the project is equal to the value of never activating the project.\(^{58} \) Thus,

\[
N_{\text{no-bc}}(\tilde{\phi}_{c,\text{no-bc}}, \phi_p, 1) - \tilde{\phi}_{c,\text{no-bc}} = \frac{\mu}{1 - \beta}.
\]

The value of an activated project is given by

\[
N_{\text{no-bc}}(\phi_c, \phi_p, 1) = \phi_p + \beta \rho N_{\text{no-bc}}(\phi_c, \phi_p, 1) + \beta (1 - \rho) N_{\text{no-bc}}(\phi_c, \phi_p, 0).
\]

If assumption 2.1 is satisfied, then activating a project is profitable if the efficiency condition is satisfied and the startup costs are low enough. Thus,

\[
N_{\text{no-bc}}(\phi_c, \phi_p, 1) = N_{\text{no-bc}}(\phi_c, \phi_p, 0) + \phi_c \text{ when } \begin{cases} \phi_c \leq \tilde{\phi}_{c,\text{no-bc}} & \text{and} \\ \phi_p \geq \tilde{\phi}_{p,\text{no-bc}} \end{cases}.
\]

Combining the last three equations gives

\[
\tilde{\phi}_{c,\text{no-bc}} (\phi_p) = \frac{\phi_p - \mu}{1 - \beta \rho} \text{ if } \phi_p \geq \tilde{\phi}_{p,\text{no-bc}}.
\]

Next, we calculate the value of \( \tilde{\phi}_{c,\text{bc}} (\phi_p) \) for cyclical projects, that is, when \( \tilde{\phi}_{p,\text{bc}} (\Phi_+) \leq \phi_p < \tilde{\phi}_{p,\text{bc}} (\Phi_-) \).\(^{59} \) If the startup costs are low enough, then these projects are activated

---

\(^{57}\)See Section 2.6.1.

\(^{58}\)To economize on notation we typically write \( \tilde{\phi}_{c,\text{no-bc}} \) instead of \( \tilde{\phi}_{c,\text{no-bc}} (\phi_p) \).

\(^{59}\)The ordering is based on the first part of assumption 2.2. If the second part of this assumption would hold instead, then the formulas would be very similar, but the role of booms and expansions would be switched.
in a boom, but not in a recession.\footnote{Assumption 2.1 ensures that the value of $\mu$ is low enough to ensure that entry is profitable as long as the efficiency condition is satisfied and the entry costs are low enough.} Thus, if $\phi_c \leq \tilde{\phi}_{c,bc}(\phi_p)$, then

$$N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) = -\phi_c + \phi_p (1 + \Delta \phi_p) + \beta$$

\begin{align*}
N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) &= \left(\begin{array}{ccc}
\pi & \rho & (N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) + \phi_c) \\
+\pi & (1 - \rho) & N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) \\
(1 - \pi) & \rho & N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) \\
(1 - \pi) & (1 - \rho) & N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p)
\end{array}\right).
\end{align*}

By definition, $\tilde{\phi}_{c,bc}(\phi_p)$ is the value of $\phi_c$ such that

$$N_{bc}(\tilde{\phi}_{c,bc}, \phi_p, 0, 1 + \Delta \phi_p) = \frac{\mu}{1 - \beta}. \quad (2.45)$$

In a boom this type of project could either produce or not produce. The NPV would be equal to $\mu/(1 - \beta)$ for both choices. In a recession this project cannot produce, so the revenues are equal to $\mu$ until the economy gets out of a recession at which point the NPV by definition is equal to $\mu/(1 - \beta)$. Consequently,

$$N_{bc}(\tilde{\phi}_{c,bc}, \phi_p, 0, 1 - \Delta \phi_p) = \frac{\mu}{1 - \beta}. \quad (2.46)$$

Combining the last equations gives

$$\tilde{\phi}_{c,bc}(\phi_p) = \frac{\phi_p (1 + \Delta \phi_p) - \mu}{1 - \beta \rho \mu}. \quad (2.47)$$

Assumption 2.3 directly implies that

$$\tilde{\phi}_{c,bc}(\phi_p) < \tilde{\phi}_{c,no-bc}(\phi_p) \text{ if } \tilde{\phi}_{p,no-bc} \leq \phi_p < \tilde{\phi}_{p,bc}(\Phi_-). \quad (2.48)$$
2.C.2 Proof of part 2 of proposition 2.1

To calculate the formulas in this part of the proposition, we have to first calculate the relevant NPVs. First consider a world without business cycles. If the project cannot satisfy the efficiency condition or if the startup costs are too high, then the project is never activated and the revenues are \( \mu \) each period. Thus

\[
N_{no-bc}(\phi_c, \phi_p, 0) = \frac{\mu}{1-\beta} \text{ if } \phi_p < \tilde{\phi}_{p, no-bc} \text{ or } \phi_c > \tilde{\phi}_{c, no-bc}(\phi_p). \tag{2.49}
\]

If assumption 2.1 is satisfied, then activating is profitable if the efficiency condition is satisfied and the entry costs are low enough. From equations 2.41 and 2.42, it follows that

\[
N_{no-bc}(\phi_c, \phi_p, 0) = \frac{\phi_p - (1-\beta \rho) \phi_c}{1-\beta} \text{ if } \phi_p \geq \tilde{\phi}_{p, no-bc} = \chi \text{ and } \phi_c \leq \tilde{\phi}_{c, no-bc}(\phi_p). \tag{2.50}
\]

Now consider the case without business cycles. Here we focus on cyclical projects. If the project can never satisfy the efficiency conditions or if the startup costs are too high, then the project will not operate and revenues are equal to \( \mu \). Thus,

\[
\text{if } \phi_p < \tilde{\phi}_{p, bc}(\Phi_+) \text{ or } \phi_c > \tilde{\phi}_{c, bc}(\phi_p), \text{ then}
E[N_{bc}(\phi_c, \phi_p, 0, \Phi_{p,t})] = \text{NPV}_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) \tag{2.51}
= \text{NPV}_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) \tag{2.52}
= \frac{\mu}{1-\beta}.
\]

Now suppose that \( \phi_p \geq \tilde{\phi}_{p, bc}(\Phi_+), \phi_c \leq \tilde{\phi}_{c, bc}(\phi_p) \), and (since we focus on cyclical projects) \( \phi_p < \tilde{\phi}_{p, bc}(\Phi_-) \). The value for

\[
E[N_{bc}(\phi_c, \phi_p, 0, \Phi_{p,t})] = \frac{\left(N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) + N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p)\right)}{2}
\]
is equal to

\[ E \left[ N_{bc}(\phi_c, \phi_p, 0, \Phi_{p,t}) \right] = \frac{1}{2} \left( \frac{\phi_p \left( 1 + \Delta \phi_p \right) + \mu - (1 - \beta \rho \pi) \phi_c \left( 1 - \beta \right)}{1 - \beta} \right). \] (2.53)

This formula follows from equation (2.44) and

\[ N_{bc} \left( \phi_c, \phi_p, 0, 1 - \Delta \phi_p \right) = \mu + \beta \left( \pi \frac{N_{bc} \left( \phi_c, \phi_p, 0, 1 - \Delta \phi_p \right)}{N_{bc} \left( \phi_c, \phi_p, 0, 1 + \Delta \phi_p \right)} \right). \] (2.54)

The formulas in the proposition follow directly from combining the formulas for the appropriate NPVs.

### 2.C.3 Proof of proposition 2.2

This proposition focuses on jobs that can always overcome the efficiency condition and have a value of \( \phi_c \) that is exactly at the boundary. Thus,

\[ \phi_c = \tilde{\phi}_{c,\text{no-bc}} \text{ and } \phi_p > \max \tilde{\phi}_{p,\text{bc}} = \max \left\{ \frac{\chi (\Phi_+)}{\Phi_+}, \frac{\chi (\Phi_-)}{\Phi_-} \right\}. \] (2.55)

In a world without business cycles, the value of activating a project with \( \phi_c = \tilde{\phi}_{c,\text{no-bc}} \) would be equal to the value of not activating. Thus,

\[ N_{\text{no-bc}} \left( \phi_c, \phi_p, 0 \right) = \frac{\mu}{1 - \beta} \text{ if } \phi_c = \tilde{\phi}_{c,\text{no-bc}} \text{ and } \phi_p > \max \tilde{\phi}_{p,\text{bc}} = \max \left\{ \frac{\chi (\Phi_+)}{\Phi_+}, \frac{\chi (\Phi_-)}{\Phi_-} \right\}. \] (2.56)

In a world with business cycles, projects with \( \phi_c = \tilde{\phi}_{c,\text{no-bc}} \) and \( \phi_p > \max \tilde{\phi}_{p,\text{bc}} \) are timed-entry projects. Timed-entry projects are only activated during expansions, but an already activated project continues operating in a recession. The NPVs for timed-
Chapter 2

entry projects are given by the following equations:

\[ N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) = -\phi_c + \phi_p (1 + \Delta \phi_p) \]  \hspace{1cm} (2.57)

\[ + \beta \left( \begin{array}{c} \pi \rho N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) \\ + \lambda (1 - \rho) N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) \\ + (1 - \pi) \rho N_{bc}(\phi_c, \phi_p, 1, 1 - \Delta \phi_p) \\ + (1 - \pi) (1 - \rho) N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) \end{array} \right) , \]

\[ N_{bc}(\phi_c, \phi_p, 1, 1 + \Delta \phi_p) = N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) + \phi_c, \]  \hspace{1cm} (2.58)

\[ N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) = \mu + \beta \left( \begin{array}{c} \pi N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) \\ + (1 - \pi) N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) \end{array} \right) , \]  \hspace{1cm} (2.59)

\[ N_{bc}(\phi_c, \phi_p, 1, 1 - \Delta \phi_p) = \phi_p (1 - \Delta \phi_p) \]  \hspace{1cm} (2.60)

\[ N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) = N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) - \phi_c. \]  \hspace{1cm} (2.61)

Let

\[ D(\phi_c, \phi_p) = N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) - N_{bc}(\phi_c, \phi_p, 1, 1 - \Delta \phi_p) + \phi_c. \]

Then

\[ E[N_{bc}(\phi_c, \phi_p, 0, \Phi_{p,t})] = \frac{N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) + N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p)}{2}, \]  \hspace{1cm} (2.62)

\[ = \frac{- (1 - \beta \rho) \phi_c + \phi_p (1 + \Delta \phi_p) + \mu - \beta \rho (1 - \pi) D(\phi_c, \phi_p)}{2 (1 - \beta)}. \]

For these projects the cut-off level for \( \phi_c \) is given by

\[ \hat{\phi}_{c, \text{no-bc}} = \frac{\phi_p - \mu}{1 - \beta \rho}. \]  \hspace{1cm} (2.63)

Using this last expression we get that

\[ E[N_{bc}(\hat{\phi}_{c, \text{no-bc}}, \phi_p, 0, \Phi_{p,t})] = \frac{\phi_p \Delta \phi_p + 2 \mu - \beta \rho (1 - \pi) D(\phi_c, \phi_p)}{2 (1 - \beta)}. \]  \hspace{1cm} (2.64)
The expression for $D$ is calculated as follows. Working out the terms in the definition of $D$ we get the following

$$D(\phi_c, \phi_p) = \phi_c + N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) - N_{bc}(\phi_c, \phi_p, 1, 1 - \Delta \phi_p)$$
$$= \phi_c + \mu + \beta \left( \pi N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) + (1 - \pi) N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) \right)$$

$$- \phi_p (1 - \Delta \phi_p) + \beta \left( \pi \rho N_{bc}(\phi_c, \phi_p, 1, 1 - \Delta \phi_p) + \pi (1 - \rho) N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) + (1 - \pi) \rho N_{bc}(\phi_c, \phi_p, 1, 1 + \Delta \phi_p) + (1 - \pi) (1 - \rho) N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) \right)$$

$$= \phi_c + \mu + \beta \left( \pi N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) + (1 - \pi) N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) \right)$$

$$- \phi_p (1 - \Delta \phi_p) + \beta \left( \pi \rho N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) + \pi (1 - \rho) N_{bc}(\phi_c, \phi_p, 0, 1 - \Delta \phi_p) + (1 - \pi) \rho N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) + (1 - \pi) (1 - \rho) N_{bc}(\phi_c, \phi_p, 0, 1 + \Delta \phi_p) \right)$$

$$= \phi_c (1 - \beta \rho) + \mu - \phi_p (1 - \Delta \phi_p) + \beta \pi \rho D.$$  

From this, we get

$$D(\phi_c, \phi_p) = \frac{\phi_c (1 - \beta \rho) + \mu - \phi_p (1 - \Delta \phi_p)}{1 - \beta \pi \rho}.$$  

Moreover,

$$D(\phi_c, \phi_p) = \frac{\phi_p \Delta \phi_p}{1 - \beta \pi \rho} \text{ if } \phi_c = \tilde{\phi}_{c, \text{no-bc}} = \frac{\phi_p - \mu}{1 - \beta \rho}.$$  

If we combine this expression for $D(\phi_c, \phi_p)$ with the expression in equation (2.64) and the definition of the welfare loss, then we get that

$$L(\phi_c, \phi_p, \Delta \phi_p) = (1 - \beta) \left( \frac{\mu}{1 - \beta} - \frac{\phi_p \Delta \phi_p + 2 \mu - \beta \rho (1 - \pi) D(\phi_c, \phi_p)}{2 (1 - \beta)} \right).$$

$$= -\frac{\phi_p \Delta \phi_p (1 - \beta \rho)}{2 (1 - \beta \pi \rho)}.$$
2.D  Cyclical changes in output: Estimate of the extensive margin

In this appendix, we use a German panel data set on wage data to obtain a more direct estimate of that part of the cyclical change in output that is due to cyclical changes in the number of projects (the extensive margin). This data set does not exactly contain what we need. First, the data set gives information about employment positions, not about projects. That is, it is possible that new projects are created (eliminated) during a boom (recession) without additional workers being hired (fired). Second, the data set gives information about wages, whereas we would like to know the total value added created by the job, not just the wage component.

Nevertheless, we think that this data set provides direct evidence that cyclical changes in output along the extensive margin are nontrivial. We already know that the extensive margin is important for cyclical changes in total hours. But it is still possible that the workers being hired during a boom are not very productive, which would mean that these additional workers are not important in terms of explaining cyclical changes in output. The analysis here suggests that this is definitely not the case.

2.D.1  Data set used

We use the IAB monthly employment panel, a 2% representative subsample from the German social security and unemployment records. It is described in more detail in Jung and Kuhn (2009). The data set excludes self-employed and civil servants, but nevertheless covers 80% of the West German labor force.\footnote{In Den Haan and Sedlacek (2009), we document that aggregated wage data according to this panel data set follow true aggregate wages very closely.}

2.D.2  Constructing the estimate

As the economy gets out of a recession, the total wage sum earned increases. There are two reasons why this happens. First, there are workers that have found a job,
because the recession has ended. In our model, these are cyclical workers. Second, workers whose employment status is not affected by the recession earn higher wages as economic conditions improve. We will refer to these workers as "non-cyclical" workers. The objective is to determine which part of the increase in total wages, observed as the economy gets out of a recession is due to cyclical workers gaining employment and which part is due to non-cyclical workers earning higher wages.

To answer this question we do the following. For each month, the workers are divided into two groups, group A and group B.

A. Workers in group A did not have a non-employment spell in the last 24 months. To do this, we first determine whether the worker was employed 24 months ago. If the worker was employed \( T \) months ago and the total number of days the worker was "not employed" during the last \( T \) months was less than 30 days, then we include her/him. Thus, a worker that experiences a job-to-job transition, but takes a "vacation" in between the two jobs is not excluded from this group of workers.

B. Workers in group B are all other workers. These workers did experience a non-trivial period of non-employment during the last 24 months.

Clearly, all workers in group A are non-cyclical workers and all cyclical workers in our sample are in group B. The problem is that group B also contains non-cyclical workers. For example, new entrants to the labor force or workers who rejoin the labor force for personal reasons are also part of group B, but they are not cyclical workers. Below, we show how we deal with this aspect of the data set.

---

62 According to definition 2.2, cyclical projects are all projects for which the stance of the business cycle determines whether they can satisfy the efficiency requirement. This includes projects with values of \( \phi_e \) above \( \bar{\phi}_{c, bc} \) that will never operate in a world with business cycles. With cyclical workers we refer here to workers that belong to projects with values of \( \phi_e \) below \( \bar{\phi}_{c, bc} \). Those with of \( \phi_e > \bar{\phi}_{c, bc} \) would never show up in our sample.

63 They also could be timed-entry workers, but these are likely to be not very important as explained in Section 2.4.2.

64 To be precise, we check whether he was employed in the reference week 24 months ago. In Den Haan and Sedlacek (2009), we document that the results are robust to using 36 instead of 24 months.

65 The data set keeps track of an experience variable that counts the total number of days a worker has worked. By looking at the increase in this variable, it is easy to check which fraction of a particular period a worker was actually working.
Figure 2.5: Fraction of total wages earned by recently non-employed workers

Notes: This graph plots the German unemployment rate (left-side axis) together with the fraction of total wages earned by workers that recently (i.e., in the last 24 months) had a “non-employment spell” (right-side axis). Both series are quarterly averages of monthly series. The series are the HP-filtered series using a smoothing coefficient equal to $10^5$ plus the mean.

We find that on average the sum of wages earned by workers in group B is equal to 15% of the total wage sum. This is obviously a non-trivial number, but it is still possible that this 15% is mainly earned by new entrants and other non-cyclical workers and that the sum of wages earned by cyclical workers is small. To analyze the quantitative importance of cyclical workers we investigate how this fraction changes over the cycle. Such changes must by definition be due to cyclical workers. In particular, we look at the HP-filtered value of the ratio $f_t$, where

$$f_t = \frac{\text{sum of wages earned by workers in group B}}{\text{sum of all workers}}.$$  

Figure 2.5 plots this series, together with the filtered unemployment rate. For convenience, we have added the means of the two series to the filtered data. As the economy moves from the trough of the recession to the peak of a boom, the fraction of wages earned by group B workers goes from roughly 13% to 17%. This is true for both the ex-
pansion of the eighties as the expansion of the late nineties. It is this increase from 13% to 17%—a non-trivial quantity—that we use as our measure of what cyclical workers can produce. It would not be right to use this difference between the observed values at the troughs and peaks as our estimate of $Y_{\text{cyclical}}/Y$. In our model, we only have two states, so it would be more prudent to use the difference between the average value of (detrended) $f_t$ in a boom and the average value of (detrended) $f_t$ in a recession. This would give an estimate for $Y_{\text{cyclical}}/Y$ equal to roughly 2%.
Chapter 3

Match Efficiency Fluctuations and the Behavior of Job Finding Rates

Abstract

The assumption of a constant matching function, typically made in the literature, implies that a given number of searching workers and employers always leads to the same number of matches. This is unlikely to be true, for example, if the composition of searching workers changes over the cycle. This paper relaxes the assumption of constant matching function parameters and allows match efficiency to fluctuate. Using data on the job finding rate and unemployment, an unobserved components model is estimated where both match efficiency and vacancies are unobserved. The latter deals with the poor data availability of vacancies over most of the sample. Estimated match efficiency is procyclical and can explain about 25% of job finding rate fluctuations. Drops in match efficiency account for up to 20% of unemployment rate increases during the most severe recessions. Next, the paper shows that procyclical movements in measured match efficiency are present even in a simple matching model with endogenous separations due to a countercyclical rejection rate. A simple extension of introducing firing costs results in the model performing well quantitatively.

3.1 Introduction

A popular way to model flows from unemployment to employment is by using a simple (matching) function relating aggregate labor market variables, typically unemployment and vacancies. The advantage of simple matching functions is their ability to capture
the consequences of labor market heterogeneities in a parsimonious way. However, a constant matching function (commonly used in the literature) implies that a given number of searching workers and employers always leads to the same number of matches. This is unlikely to be true, for example, if the composition of searching workers changes with the business cycle.

This paper estimates a standard matching function while relaxing the assumption of constant parameters. Namely, the slope coefficient, interpreted as match efficiency, is allowed to vary. One can view time varying match efficiency as the Solow residual of the matching function. Hence, a parameter that captures fluctuations in hires that cannot be accounted for by observed unemployment and vacancies. Estimated match efficiency is found to be procyclical and it turns out to be an important driver of fluctuations in the rate at which unemployed workers find jobs. In the benchmark specification around 25% of the job finding rate variation can be explained by fluctuations in match efficiency. Furthermore, match efficiency declines account for up to 20% of the unemployment rate runups during the most severe recessions. Next, the paper shows that procyclical movements in measured match efficiency are present even in the simple matching model with endogenous separations because of a countercyclical rejection rate. A simple extension of introducing firing costs results in the model performing well also quantitatively.

Before providing intuition as to why match efficiency might be time varying, I briefly describe the estimation procedure and its caveats. Estimating the matching function on U.S. data and investigating the possibility of time variation in match efficiency is severely complicated by the lack of a good or sufficiently long data series on vacancies.\footnote{The typically used proxy for vacancies, dating back to 1951, is the Help Wanted Index. This index is constructed from help wanted ads (not number of job vacancies) in 51 newspapers across the U.S. and is therefore only a crude measure of vacancies.} To tackle this problem I specify and estimate an unobserved components model where both match efficiency and vacancies are treated as unobserved. Assumptions on the underlying processes together with additional information on vacancies at the very end of the sample from the Job Openings and Labor Turnover Survey (JOLTS) facilitate identification of the two unobserved states.\footnote{The JOLTS database provides high quality data on vacancies, but it dates back only to December of 2000, while the sample used in this paper starts in 1948.} Robustness checks suggest that the results
are not an artifact of a specific functional form, estimation procedure or sample period. Using additional information from the JOLTS database to further pin down the two unobserved states does little to the results and a Monte Carlo exercise documents that the benchmark specification can identify the unobserved states well.

One reason why match efficiency might fluctuate over the business cycle can be found in cyclical variations of labor market heterogeneity. A structural model can then shed light on the specific channel, form of heterogeneity, that drives match efficiency fluctuations. One such channel is variation in the endogenous rejection rate in the standard matching model with (a constant matching function and) endogenous separations. The mechanism is the following: in the standard endogenous separations model workers differ in their productivity levels. There exists a cut-off value for worker productivity below which employment relationships are no longer viable and thus they (endogenously) separate. Recessions are times when this cut-off increases, since a fall in aggregate productivity makes only the relatively more productive workers survive in their jobs. The same logic applies to unemployed workers who are matched with a vacancy. Those with productivity levels above the cut-off are accepted and form an employment relationship. On the other hand, those who are not productive enough are (endogenously) rejected and fall back into the unemployment pool. A positive rejection rate creates a wedge between the total unemployment pool and the part of the unemployment pool that is useful for forming employment relationships. Moreover, the countercyclical fluctuations of the rejection rate imply that in a recession the fraction of the unemployment pool useful for matching shrinks. Hence, in a downturn the aggregate job finding rate falls by more than would be implied by a constant matching function that takes into account the total number of unemployed and vacancies.

One can calibrate the above-mentioned model in such a way that it perfectly captures the match efficiency fluctuations observed in the data. However, such an attempt leads to the model grossly exaggerating the volatility of other endogenous variables, most significantly that of the separation rate. I document that incorporating firing costs the model explains match efficiency movements well, both qualitatively and quantitatively, while not exaggerating the volatility of other variables. The intuition is the following: firing costs drive a wedge between workers in existing employment relation-
ships and newly hired workers. The cut-off productivity level for workers in existing relationships is lower than in the case of no firing costs, since firms know that separations entail a cost. On the other hand, the cut-off probability for the newly hired workers is higher, since firms require a compensation for expected future firing costs. This, together with the distributional assumption of an upward sloping density in the neighborhood of the cutoff, makes the rejection rate more sensitive to aggregate fluctuations. The assumption on the distribution is not unreasonable considering that the cutoff values are in the lower tail. This simple extensions enables the model to explain about 60% of match efficiency fluctuations found in the empirical part, while not exaggerating the volatility of other variables.

This paper fits into a line of research studying and estimating the matching function. Petrongolo and Pissarides (2001) provide an excellent survey of this literature. Furthermore, it is related to a strand of literature trying to understand the influence of match efficiency on unemployment and/or explain the sources of match efficiency movements. Barnichon and Figura (2011b) undertake a steady state decomposition of the Beveridge curve using the vacancy series from Barnichon (2010) to estimate the contributions of firms’ actions (hiring and firing), demographics and match efficiency on unemployment movements.3 Their estimated match efficiency, however, displays a mainly counter-cyclical pattern, except for the most recent recession. Barnichon and Figura (2011a) further study the reasons behind match efficiency fluctuations and find that the composition of the unemployment pool is responsible for most of its movements in the years 1976-2006. Over 2007-2010, however, an increase in the dispersion of labor market conditions explains almost half of the match efficiency decline. The model mechanism behind match efficiency movements presented in this paper is related to Darby, Haltiwanger, and Plant (1985), who propose that declines in the job finding rate during recessions are due to an increase in the proportion of workers with low individual probabilities of exiting unemployment. Sterk (2010) presents an alternative mechanism. He builds a DSGE model combining a housing market, labor and financial frictions. His model predicts that a decline in house prices reduces geographical

3Barnichon (2010) constructs a vacancy proxy that takes into account internet posting after 1995. However, prior to 1995 the Help Wanted Index is taken as the ideal proxy for vacancies.
mobility which in turn leads to a drop in match efficiency.

This paper is organized as follows. Section 3.2 describes the estimation procedure and Section 3.3 shows the empirical results. Section 3.4 provides some robustness exercises. Then, Section 3.5 builds a matching model with endogenous separations and shows that it features procyclical match efficiency movements, but that calibrating the model to fit match efficiency fluctuations makes it exaggerate the volatility of other variables. Section 3.6 extends the model to include firing costs and shows that such a model does relatively well in explaining match efficiency fluctuations also quantitatively. Finally, Section 3.7 provides the conclusion.

### 3.2 Estimating match efficiency variation

The starting point of the estimation procedure is the definition of the job finding probability, $F_t = M_t/U_t$, where $M_t$ is the number of matches (unemployed workers who find a job) and $U_t$ is the number of unemployed in period $t$. As noted in the introduction, the typical way to model the number of matches is using a matching function $M_t = A_m(U_t, V_t)$, where $A$ is match efficiency and $V_t$ is the number of vacancies in period $t$. This paper allows $A$ to be time-varying and at the same time deals with the problematic nature of vacancies. One can view time varying match efficiency as the Solow residual of the matching function. Hence, a parameter that captures fluctuations in hires that cannot be accounted for by observed unemployment and vacancies. The main goal of the paper is to investigate if fluctuations in match efficiency are important for determining the aggregate job finding rate and in turn affecting the aggregate unemployment rate.

To this end, I specify a state-space model treating both match efficiency and vacancies as unobserved. I use quarterly data on the job finding rate taken from Shimer (2007) and the number of unemployed published by the BLS in the period 1948Q1-2007Q1. In addition, to help identify the two unobserved states, I use information from the JOLTS vacancies series (available from December 2000) assuming that they are a

---

4In the text I use the number of unemployed and unemployment interchangeably.
noisy observation of the underlying process.\footnote{Prior to this date the JOLTS vacancy data are treated as missing observations.}

### 3.2.1 State-space representation

In the general state space form a $m \times 1$ vector of observables, $y_t$, is related to a $q \times 1$ vector of unobserved states $s_t$ via the \textit{measurement equation}.

$$
y_t = \Theta_{0,t} + \Theta_{1,t}s_t + \epsilon_t, \tag{3.1}
$$

where $\Theta_{0,t}$ is an $m \times 1$ vector, $\Theta_{1,t}$ is an $m \times q$ matrix and $\epsilon_t$ is an $m \times 1$ vector of serially uncorrelated disturbances with mean zero and a covariance matrix $R$. The unobserved states are assumed to evolve according to a first-order Markov process (the \textit{transition equation})

$$
s_t = \Phi_{0,t} + \Phi_{1,t} s_{t-1} + \eta_t, \tag{3.2}
$$

where $\Phi_{0,t}$ is an $q \times 1$ vector, $\Phi_{1,t}$ is an $q \times q$ matrix and $\eta_t$ is an $q \times 1$ vector of serially uncorrelated disturbances with mean zero and covariance matrix $Q$.

In the model at hand there are two unobserved states ($q = 2$): match efficiency ($A_t$) and vacancies ($V_t$). In the benchmark model vacancies are assumed to be a random walk, while match efficiency is assumed to follow a stationary AR(1) process. The choice of the random walk on vacancies is motivated by its fundamentally close relationship to unemployment, for which one cannot reject a unit root in the given sample.\footnote{The ADF with 4 lags and an intercept (intercept and trend) can reject a unit root at the 11.9% (12%) level. For first differenced unemployment the unit root is rejected (in all specifications: with(out) intercept and intercept with trend) at the 0% level. Although one would not, a priori, expect unemployment to be nonstationary, in the (finite) sample at hand it is a good data description.} Similarly, for the typical vacancy proxy, the Help Wanted Index (HWI), one
also cannot reject a unit root.\footnote{The ADF test with 4 lags and an intercept (intercept with trend) rejects the unit root at the 11.4\% (40.5\%) level. For first differences it rejects at the 0\% level.} This is, arguably, a less compelling argument, because of the problematic nature of the HWI as a vacancy proxy. Although the above-mentioned evidence points to a process for vacancies that is I(1), not necessarily a random walk, allowing for a richer non-stationary structure does not change the results much as is shown in Appendix 3.D. Assuming match efficiency to be an AR(1) process then helps the identification by distinguishing it from the vacancy process. However, the appendix shows that an alternative specification where both states are random walks delivers similar results.

The two states are related to observed variables via two measurement equations $(m = 2)$: one for the job finding probability $(F_t)$ and one postulating that the JOLTS job openings series $(V^J_t)$ is a noisy observation of the vacancy state. The former is the main equation facilitating the identification of the two processes. The latter helps pin down further the properties of the vacancy state and especially their level. Remember, however, that the job openings data is available only from 2001Q1. The periods prior to that date can be conveniently handled by the Kalman filter as missing observations. Finally, I follow the literature and assume the matching function is Cobb-Douglas with constant returns to scale\footnote{Petrongolo and Pissarides (2001) survey the literature and conclude that the such a functional form has large empirical support.}

\begin{equation}
M_t = A_t U_t^{1-\mu} V_t^\mu.
\end{equation}

Denoting with small letters the natural logarithm of variables one can write the state space representation of the model as

\begin{equation}
\begin{bmatrix}
  f_t \\
  v_t^J
\end{bmatrix} = 
\begin{bmatrix}
  -\mu u_t \\
  0
\end{bmatrix} + 
\begin{bmatrix}
  1 & \mu \\
  0 & 1
\end{bmatrix} 
\begin{bmatrix}
  a_t \\
  v_t
\end{bmatrix} + \epsilon_t,
\end{equation}

\begin{equation}
\begin{bmatrix}
  a_t \\
  v_t
\end{bmatrix} = 
\begin{bmatrix}
  (1 - \rho_a)\overline{a} \\
  0
\end{bmatrix} + 
\begin{bmatrix}
  \rho_a & 0 \\
  0 & 1
\end{bmatrix} 
\begin{bmatrix}
  a_{t-1} \\
  v_{t-1}
\end{bmatrix} + \eta_t,
\end{equation}

where $\rho_a$ is the autoregressive coefficient of log match efficiency and $\overline{a}$ is its unconditional mean. Furthermore, the innovations of the state and measurement equations
are assumed to be jointly normally distributed with mean zero and variance covariance matrix
\[
E_t \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix} \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix} = \begin{pmatrix} Q & C' \\ C & R \end{pmatrix},
\]
where \( C \) is the \( 2 \times 2 \) cross-covariance matrix.

### 3.2.2 Estimation

Maximum likelihood (ML) is used to estimate the elasticity of vacancies in the matching function \( (\mu) \), the autoregressive coefficient and unconditional mean of log match efficiency \( (\rho_a \text{ and } \bar{a}) \) and all the elements of the variance covariance matrix of the innovations \( (R, C \text{ and } Q) \).

The Kalman filter is then employed to obtain smoothed states\(^{11}\) at the ML estimates. Furthermore, to overcome potential endogeneity problems, I use the first lag of the regressor as an instrument. Appendix 3.E provides explicit exogeneity tests supporting this procedure.

To start the minimization routine one must pick initial values. The starting values for \( \rho_a, \bar{a} \) and \( \mu \) are set to 0.9, −0.6 and 0.3, respectively. The initial values for the covariance matrices are based on error variances from an auxiliary regression of the job finding probability on observed labor market tightness (using the HWI as an indicator of vacancies) using data up until 1955Q4. Denote the error variance from the trial regression by \( W_f \). Furthermore, denote by \( W_v \) the variance of the (log) job openings series from the JOLTS database. The initial values for the covariance matrices are then
\[
R_{\text{init}} = \begin{pmatrix} \omega_{R,f}W_f & 0 \\ 0 & \omega_{Q,v}W_v \end{pmatrix}, \quad Q_{\text{init}} = \begin{pmatrix} \omega_{Q,f}W_f & 0 \\ 0 & \omega_{Q,v}W_v \end{pmatrix},
\]
The scaling parameters \( \omega_{i,j} \), where \( i = Q, R \) and \( j = f, v \), are found by a grid search that maximize the log-likelihood of the model. The initial value for the cross-covariance matrix \( C \) is a \( 2 \times 2 \) zero matrix. Robustness checks show that changing the initial values does little to the results.

Furthermore, to start the Kalman filter routine one must set the initial state vector \( s_0 \) and its covariance matrix \( P_0 \). Following Durbin and Koopman (2001) the former

\(^{10}\)The minimization itself is done using Chris Sims’ csminwel algorithm.

\(^{11}\)The term "smoothed" might be confusing later on when evaluating the volatility of the states. Note that it refers to running the Kalman filter "backwards". The estimates in period \( t \) are then based on not only past information, but also on information from observations \( t \) onwards.
Match Efficiency Fluctuations and the Behavior of Job Finding Rates

Table 3.1: Parameter estimates

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$-0.639$</td>
<td>$(0.025)$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$0.386$</td>
<td>$(0.032)$</td>
</tr>
<tr>
<td>$\rho_\alpha$</td>
<td>$0.696$</td>
<td>$(0.065)$</td>
</tr>
<tr>
<td>$10^3 R$</td>
<td>$1.15$</td>
<td>$-0.15$</td>
</tr>
<tr>
<td></td>
<td>$-0.15$</td>
<td>$0.95$</td>
</tr>
<tr>
<td>$10^3 Q$</td>
<td>$1.28$</td>
<td>$-0.38$</td>
</tr>
<tr>
<td></td>
<td>$-0.38$</td>
<td>$3.11$</td>
</tr>
</tbody>
</table>

Notes: The reported values of diagnostic tests are p-values, where the null hypothesis is a satisfaction of the assumption. Details on the tests used are in Appendix 3.C. The result of the serial correlation diagnostic test is based on 4 lags (1, 2 and 8 lags were also tested and could not reject the null). The homoscedasticity test is based on the first and last third of the sample (the first and last quarters were also used and could not reject the null). Standard errors are in brackets.

is set to the unconditional mean of the state vector, while the latter is set to a large number ($10^5$). This essentially means that there is large uncertainty about the initial state and the data is allowed to speak freely.

### 3.3 Estimation results

Table 3.1 provides the estimated parameter values as well as p-values of diagnostic tests related to the model residuals. The Cobb-Douglas elasticity on vacancies is estimated to be 0.39, which falls within the range reported in Petrongolo and Pissarides (2001) and it is close to the estimates in Shimer (2005b) and Barnichon (2009).

The diagnostic tests indicate that the model assumptions hold. It is necessary, however, to deal with 4 outliers$^{12}$ (using a single dummy variable) in order to satisfy the normality assumption. Note, however, that the other diagnostic tests still hold when the outliers are not treated. Moreover, the match efficiency and vacancy estimates are almost unchanged when the dummy variable is added. Figure 3.1 shows the estimated

$^{12}$The outliers are in quarters 1957Q4, 1958Q1, 1974Q4 and 1975Q1.
Figure 3.1: Kalman smoothed states: benchmark

(a) Match efficiency

(b) Vacancies

Notes: The Kalman smoothed match efficiency and vacancy estimates together with the JOLTS vacancy and HWI data. The vacancy estimate and the JOLTS data are appropriately scaled to ease comparison with the HWI. Shaded areas are NBER recessions.
smoothed vacancies and match efficiency which I discuss in detail next.

### 3.3.1 Match efficiency

Match efficiency varies substantially with a standard deviation of almost 5.0%. This value is the theoretical standard deviation based on the estimated parameters, hence

\[
\sigma_\alpha = \sqrt{\frac{Q(\lambda)}{1 - \rho_\alpha^2}}. \quad \text{13}
\]

Furthermore, match efficiency is procyclical with respect to the business cycle. The correlation coefficients of match efficiency with (log) unemployment and output are -0.44 and 0.59, respectively.\textsuperscript{14} This means that recessions are periods when unemployed workers on average have a harder time finding a job not only because the number of vacancies drops and there are more unemployed workers competing for a given vacancy, but also because the efficiency of the matching process declines.

Match efficiency drops, however, have different patterns across recessions. During the recessions in the late 50’s, mid 70’s and early 80’s match efficiency experienced the sharpest declines in the range of 5-6%. Smaller, but still sizeable falls in match efficiency happened during the recessions in the early 60’s and the new millennium with falls of around 4%. The 1990 recession is peculiar in that match efficiency kept on falling for a few quarters while the economy was already recovering. A similar pattern is apparent for the 2001 recession, where match efficiency picked up at the end of the recession, but showed a sharp (temporary) relapse. These developments reflect the jobless recoveries experienced after these two downturns. Even though output started to rise in the recovery phase, match efficiency remained low keeping down the job finding rate and thus dampening employment growth.

**How important is match efficiency on average?**

To answer this question, I decompose the variation of the job finding rate into contributions of match efficiency and labor market tightness. Such a decomposition is not trivial, since the two components are correlated. For a decomposition that appropriately disentangles the covariance term I follow Fujita and Ramey (2009). The starting

\textsuperscript{13}Computing the standard deviation would be misleading since the Kalman filter series are an expected value, rather than a realization.

\textsuperscript{14}The HP filter (with smoothing coefficient of 1,600) was used to extract the cyclical components of unemployment and output.
Table 3.2: Contributions to job finding probability volatility

<table>
<thead>
<tr>
<th></th>
<th>$\beta^A$</th>
<th>$\beta^\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st-differenced</td>
<td>0.326</td>
<td>0.674</td>
</tr>
<tr>
<td>HP-filtered (1600)</td>
<td>0.241</td>
<td>0.759</td>
</tr>
<tr>
<td>HP-filtered (10^5)</td>
<td>0.153</td>
<td>0.847</td>
</tr>
</tbody>
</table>

Notes: $\beta^A$ and $\beta^\theta$ indicate contributions of match efficiency and vacancies, respectively, in a variance decomposition of the job finding rate.

Point is a log deviation of the job finding rate from its trend value (denoted by bars)

$$\ln \frac{F_t}{\bar{F}_t} = \ln \frac{A_t}{\bar{A}_t} + \mu \ln \frac{\theta_t}{\bar{\theta}_t} + \omega_t,$$

where $\omega_t$ is an error coming from the detrending procedure. In general it will not be the case that the trend components of match efficiency and labor market tightness exactly add up to that of the job finding rate. The above can be expressed generically as

$$df_t = df_t^A + df_t^\theta + df_t^\omega.$$

One can then show that

$$\text{var}(df_t) = \text{cov}(df_t, df_t^A) + \text{cov}(df_t, df_t^\theta) + \text{cov}(df_t, df_t^\omega),$$

where the term $\text{cov}(df_t, df_t^A)$ gives the amount of variation in the job finding probability due to match efficiency appropriately taking into account its covariance with labor market tightness. Expressing this variation relative to total volatility in the job finding probability gives:

$$\beta^A = \frac{\text{cov}(df_t, df_t^A)}{\text{var}(df_t)}.$$

From (3.9) it is clear that $\beta^A + \beta^\theta + \beta^\omega = 1$. Table 3.2 shows the respective decompositions for HP-filtered (with smoothing coefficients of 1, 600 and $10^5$) and first-differenced data. The table documents how the influence of match efficiency differs depending on what frequency one focuses on. Match efficiency gains explanatory power as one focuses on higher frequencies (for first differenced data match efficiency explains up to 33%, while for HP-filtered data with smoothing coefficient $10^5$ the contribution
Figure 3.2: Job finding probability

Notes: "Varying match efficiency" refers to the actual job finding rate and "constant match efficiency" is a counterfactual job finding rate where match efficiency was fixed at its average value. Shaded areas are NBER recessions.

Figure 3.3: Unemployment rate

Notes: "Varying match efficiency" refers to the actual unemployment rate and "constant match efficiency" is a counterfactual unemployment rate where match efficiency was fixed at its average value. Shaded areas are NBER recessions.
of match efficiency is 15%). Zooming in on business cycle frequencies match efficiency accounts for 24% of the variation in the job finding probability, which is a nontrivial amount.

Figure 3.2 shows the job finding rate and its counterfactual generated under the assumption that match efficiency is fixed at its average value. Looking at troughs of the three recessions with the largest fall in match efficiency (in 1957, 1974 and 1981), the counterfactual job finding rate is 2-3 percentage points higher. In other words, during these recessions the fall in match efficiency pushed down further the probability of finding a job by up to 3 percentage points. Given the low cyclical level in the troughs, 3 percentage points amount to almost 10% of the job finding probability. A similar effect of comparable magnitude occurred also in the early 90’s, but this time a few quarters after the recession ended. Note, however, that even the counterfactual job finding rate picks up after the recession. Hence, although match efficiency contributed to a greater drop in the job finding probability after the recession, it was not the only reason for its delayed bounce-back and hence jobless recovery.

One can use this counterfactual job finding rate to construct a counterfactual unemployment rate. To do so I use the equilibrium expression $u_t = s_t / (s_t + f_t)$, where $u$ is the unemployment rate, $s$ the separation rate and $f$ the job finding rate. For the U.S. economy this “equilibrium” unemployment rate tracks to actual unemployment rate very closely. Figure 3.3 shows the actual and the counterfactual unemployment rate based on a job finding rate with a fixed match efficiency. On average the deterioration in match efficiency accounts for almost 10% of unemployment increases during recessions. However, in downturns with the largest match efficiency drops (1957, 1974 and 1982) the contribution of match efficiency to the unemployment rate runups was as high as 20.6%.

**How important is match efficiency during different recessions?**

To better understand the importance of match efficiency I decompose the cumulative fall of the job finding rate during each recession into contributions of match efficiency and labor market tightness. Log-differencing the definition of the job finding probability gives $dF_t \approx F_t(d \ln(A_t) + \mu d \ln(\theta_t))$. 
Figure 3.4: Decomposition of the cumulative drop in the job finding rate

Notes: The cumulative (log) drops in the job finding rate for each recession are decomposed into the contributions of match efficiency and labor market tightness. Adding the two contributions together gives the total cumulative (log) job finding rate drop.

Figure 3.4 shows these contributions to the cumulative drop of the (log) job finding rate during the recessions (the quarters prior to the starting dates of the recessions are indicated on the horizontal axis). It is apparent that the variance decomposition is hiding quite a bit of heterogeneity. The contribution of match efficiency to the job finding rate fall during the recessions in 1960 and 1981 is roughly half of the labor market tightness contribution. On the other hand, during the recessions in 1990 and 2001 match efficiency contributed only very slightly. This is related to the fact that match efficiency fell mostly after these recessions. Furthermore, the downturns with the highest contributions of match efficiency are also on average longer and deeper.\footnote{The recessions associated with the largest match efficiency drops are on average one quarter longer with real GDP growth falling by 1.5 percentage points more.}

It seems that match efficiency contributes to reductions in job finding rates more at the onset of recessions. Getting closer to the recovery phase match efficiency contributions slow down and in a few cases they even reverse before the end of the recession. This is related to the previous decomposition exercise where it is shown that match efficiency explains job finding rate fluctuations especially at higher frequencies. As
one focuses on longer fluctuations the effect of aggregate labor market tightness gains importance.

All the above points to the fact that match efficiency is an important determinant of job finding rate fluctuations. Therefore, specifications of the matching function should not be such that they rule out this channel by assumption.

### 3.3.2 Vacancies

Although vacancies are not the main focus of this paper, the estimated vacancies are of separate interest, because the estimate provides information about the behavior of vacancies over several business cycles. The typically used vacancy proxy, the HWI, dates back to 1951, but is increasingly inaccurate as internet vacancy posting took over in the later part of the sample. The HWI actually stopped being published in May 2008 and was replaced by the Online HWI.\(^{16}\) The more recent job openings data from the JOLTS database provide a much better indicator of vacancies, but they date back only to 2001 missing all the previous business cycles. On the contrary, the vacancy estimate in this paper enables a methodologically consistent comparison of labor market dynamics over several business cycles, including the more recent ones. For instance, studying movements of the Beveridge curve could shed some new light on the recent developments in the U.S. economy. However, the kind of analysis that this deserves is outside the scope of this paper. The following paragraphs are therefore only descriptive and a deeper investigation is left for future research.

The bottom panel of Figure 3.1 displays the estimated vacancy state and the HWI. At first sight, the dynamics of the two series are similar (correlation coefficient of 0.81). At the same time, the vacancy estimate is much smoother. Note, however, that the estimated vacancy series is the Kalman filter estimate (a conditional expectation) and not a realization. For a fair comparison one needs to compare the estimated theoretical standard deviation of the vacancy innovations to a suitable empirical counterpart. To this end one can assume that the HWI is also a random walk and use the standard

\(^{16}\)Barnichon (2009) attempts to link the two indices into a composite HWI. Apart from specific assumptions on the dispersion process of internet use that Barnichon needs to make, he also assumes that prior to 1995 the HWI was the ideal characterization of vacancies.
deviation of its first difference.\textsuperscript{17} Such a comparison shows that the benchmark vacancy series fluctuates less by approximately 10%. Alternatively, one can estimate the process for the HWI and use its innovation variance for the comparison. Based on inspecting the (partial) autocorrelation function and the Akaike and Schwarz information criteria, the HWI for the sample at hand is estimated to be an ARIMA(1,1,0). In this case the benchmark vacancy series has an innovation variance that is roughly 10% larger than that of the ARIMA(1,1,0) process on the HWI. Hence, the volatility of vacancies according to the estimated model is roughly equal to estimated volatility of the observed HWI.

Going back to the bottom panel of Figure 3.1, after 1990 there is a clear departure of the HWI and the estimated vacancies. This is arguably due to the spur in internet posting of vacancies as was also pointed out by Shimer (2005b) and Barnichon (2009).

Finally, the Beveridge curve is somewhat weaker for the estimated vacancy series. The correlation coefficient between the unemployment and vacancy rate (HP filtered with smoothing coefficient 1600) is $-0.9$ when using the HWI and $-0.72$ when using the estimated vacancies. Once again, one needs to keep in mind that the correlation can be affected by the fact that the vacancy series is a conditional expectation and not a realization.

### 3.4 Robustness checks

In this section I provide five robustness checks. First, I investigate whether alternative functional forms of the matching function result in similar match efficiency estimates. Second, an alternative estimation procedure is employed checking whether variation in match efficiency is not just a result of poor identification in the benchmark specification. Third, I estimate the model with additional information from the JOLTS database on the vacancy yield to further help pin down the unobserved states. Fourth, the sample period is extended to see whether the results hold also during the most recent severe downturn. Finally, a Monte Carlo exercise is conducted to document the ability

\textsuperscript{17}Although unit root tests do not imply that the HWI is a random walk, they show that the series is non-stationary in the given sample. ADF test with 4 lags and an intercept (intercept with trend) rejects the unit root at the 11.4% (40.5%) level. For first differences it rejects at the 0% level.
Notes: The figure scatter-plots the estimated (log) vacancy series and an artificial (log) job finding rate series created by using the average value of match efficiency and unemployment and the ML parameter estimates. In the benchmark specification the scatter-plot would be exactly linear with a slope of $\mu$.

of the benchmark procedure to identify unobserved match efficiency and vacancies. Other robustness checks, such as alternative assumptions on the unobserved states and estimation on different samples and using data at monthly frequencies are in Appendix 3.D.

### 3.4.1 Alternative functional forms

Here I repeat the benchmark estimation using two alternative matching functions also found in the literature. First, a standard CES specification $M_t = A_t (U_t^\xi + V_t^\xi)^{1/\xi}$ and second a specification proposed by den Haan, Ramey, and Watson (2000) $M_t = A_t \frac{U_t V_t}{(U_t^\gamma + V_t^\gamma)^{1/\gamma}}$. In both cases the state-space becomes non-linear in the first measurement equation. To deal with this caveat, I employ the extended Kalman filter (EKF), which essentially uses a first order approximation of the state-space system in the usual Kalman filter recursions (details in Appendix 3.B). Admittedly, there are more sophis-
ticated non-linear filters available. For the purpose at hand, where one is interested in a first glimpse whether or not the results change substantially with different matching functions, the extended Kalman filter is a natural choice. Moreover, the degree of non-linearity at the ML parameter estimates for these two functional forms is quite small (Figure 3.5) and thus a linear approximation can be expected to perform quite well. The state-space system together with further details can be found in Appendix 3.D.

Figure 3.6 shows the benchmark match efficiency estimate and those from the two alternative functional forms. The results based on both the CES specification and the matching function proposed in den Haan, Ramey, and Watson (2000) are very similar to the benchmark estimate. If anything, using the Cobb-Douglas matching function dampens match efficiency fluctuations.

\subsection{Alternative estimation procedure}

One could be worried that the benchmark specification attributes variation to match efficiency only due to poor identification. To check this, I propose an alternative two-step estimation procedure. First, assume that match efficiency is fixed and use data on the job finding rate and unemployment to obtain an estimate for implied vacancies. Second, use the JOLTS vacancy data to decompose the implied vacancies in the first step into the "true" underlying vacancies and match efficiency fluctuations.

In this way, one is not "forcing" match efficiency to vary. If match efficiency is truly constant, then the estimated vacancies in the first step should follow the JOLTS vacancy data closely. If, on the other hand, match efficiency varies over the business cycle, the estimated vacancy series in the first step will incorporate these fluctuations and deviate from the JOLTS vacancy data. The second step will then disentangle the two states. The details of the state-space system are in the Appendix 3.D.

Figure 3.7 shows how the first-step vacancy estimate differs from the JOLTS vacancy data (both series are demeaned to ease comparison). Figure 3.8 plots the benchmark match efficiency estimate and the one from the two-step procedure. The two-step procedure yields an estimate that is very similar to the benchmark estimate.
Figure 3.6: Match efficiency, robustness: alternative functional forms

Notes: "Benchmark" refers to the Cobb-Douglas specification. The other two functional forms are estimated using the Extended Kalman filter due to the non-linearity of the state-space. Shaded areas are NBER recessions.

\[ \tau = \frac{\lambda \Omega}{(\lambda + \Omega)^2} \]

\[ \tau = \frac{\lambda \Omega}{(\lambda + \Omega)^2} \]

Benchmark
Figure 3.7: Vacancies, robustness: alternative estimation

Notes: In the first step of the alternative estimation procedure match efficiency is forced to be constant and thus all of its (potential) variation is captured by the vacancy estimate.

Figure 3.8: Match efficiency, robustness: alternative estimation

Notes: In the second step of the alternative estimation procedure the vacancy estimate from the first step is decomposed into match efficiency and vacancies by using the JOLTS vacancy data. Shaded areas are NBER recessions.
3.4.3 Using information from the vacancy yield

It could be the case that the estimated match efficiency series is capturing mainly a cyclical component of the unobserved vacancies. Since both match efficiency and vacancies are procyclical, they have the same qualitative effect on the job finding rate. However, their impact on the job filling rate, $Q_t = M_t/V_t$, is of opposite signs. An increase in match efficiency increases the probability of filling a vacancy, since the overall process of matching is more efficient. However, an increase in aggregate vacancies decreases the aggregate vacancy filling probability, because there are more vacancies competing for a given number of unemployed.

To make sure that the estimated match efficiency and vacancy states are consistent with job filling rate data I augment the benchmark state-space system with a third measurement equation using data on the vacancy yield from the JOLTS database (details are provided in Appendix 3.D). Figure 3.9 compares the estimated match efficiency from the state-space system including the vacancy yield data and that from the benchmark specification. The two are very similar. Furthermore, the coefficient on match efficiency is 0.98, while that on vacancies is −0.28 (both are statistically significant). These not only have the expected signs, but also indicate that the estimation procedure does not simply ignore match efficiency fluctuations in the vacancy yield equation.

3.4.4 Extending the sample period

In the benchmark, the sample period ends in the first quarter of 2007. In this subsection, using a shortcut (building on Elsby, Michaels, and Solon (2009)) explained in Appendix 3.D, I extend the sample to the fourth quarter of 2010. The reader should, however, keep in mind that due to the way the job finding rate is calculated this subsection can serve only as a robustness check. A more careful analysis of the most recent data would require the job finding rate to be calculated by explicitly using detailed CPS data. Details on the estimation are provided in Appendix 3.D.

---

18 The vacancy yield is the flow of realized hires during the month per reported job opening at the end of the previous month; hence it is not exactly the job filling rate due to time aggregation issues. Nevertheless, Davis, Faberman, and Haltiwanger (2009) construct a measure of the job filling rate using the JOLTS vacancy yield data and conclude that "... the job-filling rate exhibits the same strong patterns as the vacancy yield."
Figure 3.9: Match efficiency, robustness: using vacancy yield data

Notes: Benchmark estimate and estimate based on a state-space system augmented with a third measurement equation using vacancy yield data. Shaded areas are NBER recessions.

Figure 3.10: Match efficiency, robustness: extended sample

Notes: The benchmark estimate and an estimate based on an extended sample. Shaded areas are NBER recessions.
Figure 3.10 shows the benchmark match efficiency estimate together with the one based on the extended sample. The most recent recession was characterized by a sharp fall in match efficiency, almost double that of the harshest drop in past recessions. Overall, however, the picture does not change and match efficiency remains procyclical (although it leads the most recent recession slightly).

3.4.5 Monte Carlo experiment

This subsection checks how well the benchmark procedure can identify the unobserved match efficiency and vacancies. To this end, I use the benchmark state-space structure, the maximum likelihood parameter estimates and the estimated unobserved states to construct 1,000 artificial data series for the job finding rate and observed vacancies at the end of the sample (representing the JOLTS vacancy data series). Then, for each of the 1,000 replications the benchmark state-space specification is used to estimate the parameters with maximum likelihood and recover the unobserved states. The details on the construction of the artificial data series and the estimation procedure are in Appendix 3.D.

Figure 3.11 shows the benchmark estimate of match efficiency together with the average across the 1,000 Monte Carlo replications. The shaded area indicates the 90% confidence bands. The Monte Carlo average is very close to the true underlying state (correlation coefficient of 0.95). The confidence area also clearly follows the procyclical pattern of the true state. However, it could still be that for some realizations the match efficiency estimate is much worse than the mean suggests. To check this, I count the number of Monte Carlo realizations for which the correlation of the estimate with the truth is above a certain level. 85% of the time the match efficiency estimate is correlated with the truth with a higher than 0.5 correlation coefficient and more than 99% of the time the correlation is positive.

In nine cases the correlation between the Monte Carlo estimate and the truth was negative, with a minimum of −0.17. All of these estimates were associated with one or more extreme random draws. Treating these draws with a dummy variable during

\footnote{A very similar picture is to be seen when one plots the median of the Monte Carlo replications, instead of the average.}
3.5 Matching model with endogenous separations

As noted in the introduction, match efficiency might be time-varying because of cyclical changes in labor market heterogeneity. This section documents how endogenous rejection in the standard matching model with a constant matching function implies procyclical fluctuations in measured match efficiency. One would think that the model can be calibrated such that it exactly fits observed match efficiency volatility. That is true. However, in doing so the model grossly exaggerates fluctuations in other endogenous variables, most significantly that of the separation rate. It is shown that in this setup there is a trade-off between realistic fluctuations for the separation rate and match efficiency. Next a simple extension is proposed that makes the model perform well in terms of capturing match efficiency variation while not exaggerating volatility.
of the separation rate.

Before describing the model, I provide intuition as to why match efficiency fluctuates procyclically in the endogenous separations model. Workers in this model are characterized by individual specific productivity levels. In each period there is a productivity threshold, below which a given job is not viable anymore and the employment relationship is terminated. In this environment, the individual probability of finding a job depends not only on aggregate variables (unemployment and vacancies), but also on the workers’ individual productivity. The aggregate job finding rate thus depends on the fraction of unemployed workers who are productive enough to form viable employment relationships. Recessions are times when the productivity threshold increases, since a fall in aggregate productivity makes some employment relationships with relatively less productive workers unsustainable. In other words, recessions are times when the part of the unemployment pool that can form employment relationships shrinks. In this environment a constant matching function taking into account only the total number of unemployed and vacancies underestimates the fall in the aggregate job finding rate during a downturn.

3.5.1 Model

Household behavior

The household consists of a continuum of workers of unit mass. Every period each worker draws a productivity level $p$ from a constant distribution $F$. The productivity draws are independently and identically distributed, hence there is no persistence in individual productivity levels. This feature of the model makes it tractable as the worker productivity distribution is constant and identical for both the unemployment and employment pools. The members of the household pool their incomes from employment and non-employment activities and spend it on consumption. The model abstracts from any investment or labor force participation decisions.

Formally the household maximizes expected life-time utility by choosing aggregate
consumption subject to its budget constraint

\[ E_t \left[ \sum_{j=0}^{\infty} \beta^j c_t^{1-\gamma} \right] \]

\[ s.t. \]

\[ c_t = \int_{\tilde{p}_t} f_t(w_t(p)n_t)dF(p_t) + bu_t + \Pi_t, \]

where total \( c_t \) is aggregate consumption, \( \int_{\tilde{p}_t} f_t(w_t(p)n_t)dF(p_t) \) is aggregate wage income, \( bu_t \) is non-employment income and \( \Pi_t \) are aggregate profits. Costs of posting vacancies are assumed to be paid to the household.

**Matching process**

Matching occurs at the end of the period and matched workers are available for production in the next period. Hence, workers that separate at the beginning of period \( t \) enter the unemployment pool and are ready to be re-matched in the same period.

Let \( u_t \) be the mass of unemployed workers available for matching and let \( v_t \) be the mass of vacancies being posted by firms at the end of period \( t \). The number of matches in period \( t \) is determined by a matching function

\[ m_t = A u_t^\mu v_t^{1-\mu}. \]

The choice of the Cobb-Douglas functional form with constant returns to scale is consistent with the empirical part of the paper. Notice, however, that \( m_t \) only gives the number of matched worker-vacancy pairs. It still depends on the workers productivity in the next period, whether or not the job is created. Hence, not only workers who are not matched with a vacancy remain in the unemployment pool, but so do those who are matched with a vacancy but are not productive enough.

The probability that a worker is matched with a vacancy in period \( t \) is defined as \( f_t = m_t/u_t \), while the probability that a firm with an open vacancy is matched with a worker in period \( t \) is \( q_t = m_t/v_t \). Remember, that these are not equal to the probabilities of finding a job and filling a vacancy, which are defined below in Section 3.5.2.
Employment relationships

An employment relationship consists of a worker and firm pair. Production is given by \( z_t p_{i,t} \), where \( z_t \) is the aggregate productivity shock and \( p_{i,t} \) is the worker specific productivity shock. The relationship can be severed exogenously before the shocks materialize and this happens with probability \( \rho_x \). After observing the aggregate and worker specific shocks the employment relationship decides whether to continue and produce or whether to separate. In the event of (exogenous or endogenous) separation there is no production and the worker joins the unemployment pool.

Endogenous separations

Next I provide the value functions describing the problem of firms and workers in the matching market. Denote with \( W_{i,t} \) the value at time \( t \) of being in a productive employment relationship for a worker with job specific productivity \( p_{i,t} \) (measured in current consumption units). This is given by

\[
W_{i,t} = w_{i,t} + E_t \left[ \beta_t (1 - \rho_x) \int_{\tilde{p}_{t+1}}^P (W_{t+1} - U_{t+1}) dF(p_{t+1}) + U_{t+1} \right],
\]

where \( w_{i,t} \) is the wage rate, \( \beta_t = \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} \) is the stochastic discount factor, \( \tilde{p}_{t+1} \) is the threshold value of the worker specific shock such that employment relationships with values of \( p_{i,t} \) below this threshold endogenously separate and \( P \) is the upper bound of the skill distribution. Hence, workers get a wage rate dependent on their idiosyncratic productivity levels plus the continuation value of exiting period \( t \) in an employment relationship.

The value of being in the matching pool for the worker \( U_t \) at time \( t \) is defined as

\[
U_t = b + E_t \left[ \beta_t f_t (1 - \rho_x) \int_{\tilde{p}_{t+1}}^P (W_{t+1} - U_{t+1}) dF(p_{t+1}) + U_{t+1} \right],
\]

\( ^{20} \)In the specification with iid idiosyncratic productivity shocks it does not matter whether the shock is worker or job specific. Both cases would be identical in terms of the functioning of the model. The interpretation of why match efficiency varies would, however, be different. In the present environment match efficiency varies because of procyclical fluctuations in the fraction of unemployed workers that are productive enough to find jobs. If idiosyncratic productivity shocks were job specific match efficiency would fluctuate because of procyclical variation in the fraction of productive enough jobs.
where the worker enjoys leisure and the outcome of home production worth b units of consumption, the value of being in an employment relationship tomorrow if successful in the matching process or otherwise the future value of remaining unemployed.

Denote with $J_{i,t}$ the value of a productive employment relationship for the firm employing a worker with idiosyncratic productivity $p_{i,t}$. This value is given by

$$J_{i,t} = z_t p_{i,t} - w_{i,t} + E_t \left[ \beta_t (1 - \rho_x) \int_{p_{t+1}}^\infty (J_{t+1} - V_{t+1}) dF(p_{t+1}) + V_{t+1} \right] ,$$  

(3.16)

where the firm gets profits from production plus the continuation value of leaving the period in an employment relationship.

The value of an unfilled vacancy $V_t$ is driven down to zero due to the assumption of free entry of firms. This gives then the vacancy posting condition

$$\frac{\kappa}{q_t} = E_t \left[ \beta_t (1 - \rho_x) \int_{\tilde{p}_{t+1}}^\infty J_{t+1} dF(p_{t+1}) \right] ,$$  

(3.17)

where vacancies are being posted until the expected future payoffs exactly equal the effective costs ($\kappa/q_t$).

When deciding whether or not to separate, the match weighs the payoffs of staying in the relationship against the outside option. Hence, the employment relationship continues when $W_{i,t} + J_{i,t} > U_t$. In other words, the threshold value $\tilde{p}_t$ is such that it makes the employment relationship exactly indifferent between continuing and separating

$$z_t \tilde{p}_t - b + E_t \left[ \beta_t (1 - \rho_x) (1 - f_t) \int_{\tilde{p}_{t+1}}^\infty (W_{t+1} - U_{t+1} + J_{t+1}) dF(p_{t+1}) \right] = 0 .$$  

(3.18)

Given $\tilde{p}_t$ the endogenous separation rate is $F(\tilde{p}_t)$ and total separations are defined as

$$\rho_t = \rho_x + (1 - \rho_x) F(\tilde{p}_t) .$$  

(3.19)

**Wage bargaining**

Wages are assumed to be set according to Nash bargaining and are thus such that

$$(1 - \eta) W_{i,t} - U_t = \eta J_{i,t},$$

where $\eta$ is the bargaining power of workers. Using (3.14) to
one can obtain the following expression for the wage

\[ w_{i,t} = \eta(z_t p_i + \kappa \theta_t) + (1 - \eta)b, \]  

(3.20)

where \( \theta_t = v_t / u_t \) is labor market tightness. The wage rate is a weighted average of firms’ revenues and savings on hiring costs and the foregone outside option, where the weights are determined by the relative bargaining strengths.

**Closing the model**

Let \( n_t \) be the mass of employed (producing) workers in period \( t \). Then, the law of motion for unemployment is given by

\[ u_{t+1} = (1 - f_t(1 - \rho_{t+1}))u_t + \rho_{t+1}n_t. \]  

(3.21)

Tomorrows unemployment pool thus consists of workers unsuccessful in finding a job (either because they did not match with a vacancy, or they did, but were not productive enough), plus newly separated workers employed in the previous period. Workers who were matched with a vacancy, but in the end did not start a production relationship (\( \rho_t f_{t-1} u_{t-1} \)) are denoted as rejected and hence \( \rho_t \) is the rejection rate. Note that in this model the rejection rate is identical to the separation rate. Since the labor force is set to \( u_t = 1 - n_t \).

Finally, aggregate output is determined by

\[ y_t = z_t n_t G(\bar{p}_t), \]  

(3.22)

where \( G(x) = E_t[p|p \geq x] = \int_x^\infty p \frac{dF(p)}{1 - F(x)} \) is the average productivity of workers with an idiosyncratic draw above \( x \).

**3.5.2 Match efficiency fluctuations**

The probability that an unemployed worker is employed in the next period is given by \( f^*_t = f_t(1 - \rho_{t+1}) \).\(^{21}\) Making the matching function explicit, one can write \( f^*_t = \)

\(^{21}\) Similarly, the probability that a firm fills a vacancy is \( q^*_t = q_t(1 - \rho_{t+1}) \).
A(1 − ρ_{t+1})u_t^{1−ρ}v_t^ρ. Measured match efficiency is then defined as

\[ A_t = A(1 − ρ_t). \] (3.23)

Therefore, unless the rejection rate is constant, measured match efficiency varies over time. In other words, in the model agents who are matched with a vacancy, but are not productive enough to start working contribute to a lower job finding rate. (3.23) provides a direct model counterpart to the match efficiency estimates from Section 3.3.

### 3.5.3 Calibration

The calibration procedure follows the principle proposed in Hagedorn and Manovskii (2008), where the bargaining power and outside option of workers are set such that the model can match the wage elasticity with respect to productivity and profit share observed in the data. First, I consider a calibration, where the volatility of the separation rate is targeted. Such a calibration leads to disappointing results in terms of explaining match efficiency volatility. To highlight the basic trade-offs at play, I consider a second calibration that makes the model exactly fit match efficiency fluctuations and show how such a calibration grossly exaggerates the volatility of other endogenous variables.

To facilitate the exposition of the calibration, I divide the parameters of the model into two groups - first, a group of parameters that are fixed across both calibrations, and second, parameters which are calibrated internally to match statistics in the data. The two calibrations differ in the statistics that are being matched and therefore the second group of parameters differs between calibrations. The parameter values are summarized in Table 3.3 and 3.4 for the 1st and 2nd calibration, respectively.

#### Fixed parameters

The model period is set to be one quarter. Standard choices are made for the discount factor, \( \beta = 0.99 \), the coefficient of relative risk aversion, \( \gamma = 1 \), the standard deviation of the aggregate productivity shock, \( \sigma_z = 0.007 \), and its autocorrelation coefficient, \( \rho_z = 0.95 \). The idiosyncratic productivity distribution is assumed to be log-normal, \( \log(p) \sim N(\mu_F, \sigma_F^2) \), with \( \mu_F \) normalized to 0. Finally, the elasticity of unemployment
Calibrated parameters

The second group of parameters contains match efficiency, $A$, the flow cost of vacancies, $\kappa$, the bargaining power of workers, $\eta$, the exogenous separation rate, $\rho_x$, the standard deviation of the worker specific productivity distribution, $\sigma_F$ and the value of leisure and home production $b$. These parameters are selected to match six statistics in the data.

In the case of the 1st calibration, the six targets consist of the following statistics. The mean job finding probability from Shimer (2007) that was also used in the empirical part (45.4%). An unemployment rate of 12% commonly used in the literature. Following den Haan, Ramey, and Watson (2000) and van Ours and Ridder (1992) the mean vacancy filling probability of 71%. A wage elasticity with respect to productivity of 0.45 and a profit ratio of 0.03 as in Hornstein, Krusell, and Violante (2005a). Finally, the 1st calibration targets the standard deviation of the separation rate equal to 0.061.

In the case of the 2nd calibration, the first five target statistics (job finding proba-
Table 3.4: Parameter values: 2nd calibration

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Target/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discount factor</td>
<td>$\beta$</td>
<td>0.99</td>
</tr>
<tr>
<td>Relative risk aversion</td>
<td>$\gamma$</td>
<td>1</td>
</tr>
<tr>
<td>Agg. shock persistence</td>
<td>$\rho_z$</td>
<td>0.95</td>
</tr>
<tr>
<td>Agg. shock st. dev.</td>
<td>$\sigma_z$</td>
<td>0.007</td>
</tr>
<tr>
<td>Idio. shock mean</td>
<td>$\mu_F$</td>
<td>0</td>
</tr>
<tr>
<td>Match elasticity</td>
<td>$\mu$</td>
<td>0.614</td>
</tr>
<tr>
<td>Match efficiency</td>
<td>$A$</td>
<td>0.574</td>
</tr>
<tr>
<td>Exogenous destruction</td>
<td>$\rho_x$</td>
<td>0.004</td>
</tr>
<tr>
<td>Vacancy posting costs</td>
<td>$\kappa$</td>
<td>0.282</td>
</tr>
<tr>
<td>Worker bargaining power</td>
<td>$\eta$</td>
<td>0.113</td>
</tr>
<tr>
<td>Worker outside option</td>
<td>$b$</td>
<td>1.020</td>
</tr>
<tr>
<td>Idio. shock st. dev.</td>
<td>$\sigma_F$</td>
<td>0.295</td>
</tr>
</tbody>
</table>

Notes: This calibration targets match efficiency volatility. $\hat{p}$ is the average productivity of the employment relationships, consisting of aggregate productivity $z$ and the average worker productivity $G(\tilde{p})$. $\epsilon_{W,\hat{p}}$ is the elasticity of wages with respect to productivity and $W/\hat{p}$ is the wage share.

bility, unemployment rate, vacancy filling probability, wage elasticity with respect to productivity and profit ratio) remain the same. However, instead of matching separation rate volatility, the model targets match efficiency volatility equal to 0.050.

3.5.4 Model performance

Under both calibrations the model is solved with first-order perturbation techniques. To understand the mechanics of the model Figure 3.12 shows the impulse responses to a positive one-standard-deviation shock to aggregate productivity for the 1st calibration, where separation rate volatility is targeted. All workers become more productive and therefore the idiosyncratic productivity threshold $\tilde{p}_t$ falls. This is directly reflected in a fall of the separation rate $\rho_t$, which leads to a fall in unemployment (also on impact), and a rise in employment and output. At the same time labor market tightness $\theta_t$ rises, which together with a fall in the rejection rate makes the job finding probability rise which reinforces the fall in unemployment.

Table 3.5 compares second order moments of labor market variables from the simulated model under both calibrations and the U.S. economy. The economy, under both calibrations, is simulated 1,000 times. Each time 1,237 quarters are simulated and the first 1,000 are dropped to obtain 237 quarters as in the empirical part. The simulated
Figure 3.12: IRFs to a positive one-standard-deviation technology shock. 1st calibration

Notes: The model is calibrated with zero firing costs and targets the volatility of the separation rate. Data are detrended with an HP filter with smoothing coefficient 1,600 and then the standard deviations are calculated for each of the 1,000 simulations. The reported statistics are averages over the 1,000 simulations.

The model under the 1st calibration replicates the second-order moments of unemployment, vacancies and the job finding rate quite well as documented in Hagedorn and Manovskii (2008). However, it is able to explain disappointingly little of the observed match efficiency variation (only about 8%). To highlight the basic trade-offs at hand, the 2nd calibration targets match efficiency volatility directly. In this case the volatility of other labor market variables is grossly exaggerated. The separation rate fluctuates 7 times, the unemployment rate 3 times and vacancies 2 times more than seen in the data.

The reason why the standard calibration fails so blatantly in explaining match efficiency fluctuations and why under the 2nd calibration other labor market variables become enormously volatile is that the rejection rate is identical to the separation rate. Hence, calibrating the separation rate (as is done in the standard calibration) completely pins down the rejection rate properties as well. Since the average level of separations is low and the volatility moderate, the volatility of $A(1 - \rho)$ is bound to be
Table 3.5: Standard deviations of variables: US data and different calibrations

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Model 1st</th>
<th>Model 2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.050</td>
<td>0.004</td>
<td>0.050</td>
</tr>
<tr>
<td>ρ</td>
<td>0.061</td>
<td>0.061</td>
<td>0.425</td>
</tr>
<tr>
<td>u</td>
<td>0.122</td>
<td>0.146</td>
<td>0.366</td>
</tr>
<tr>
<td>v</td>
<td>0.138</td>
<td>0.104</td>
<td>0.256</td>
</tr>
<tr>
<td>f</td>
<td>0.077</td>
<td>0.076</td>
<td>0.062</td>
</tr>
</tbody>
</table>

Notes: The empirical standard deviations are based on U.S. data in the period between 1951Q1 and 2007Q1. The data were seasonally adjusted, logged and detrended with an HP filter with smoothing coefficient 1,600. Unemployment is taken from the Current Population Survey published by the BLS, vacancies are the Help Wanted Index published by the Conference Board and the job finding rate and separation rates are taken from Shimer (2007). "1st" refers to the case with zero firing costs φ = 0 and when the volatility of the separation rate σ(ρ) is targeted. "2nd" refers to the case with zero firing costs φ = 0 and when the volatility of match efficiency σ(A(1 − ρ)) is targeted.

...small. The opposite reasoning holds under the 2nd calibration.22

3.6 Endogenous separations with firing costs

This section shows that a simple extension of the model improves its performance considerably. The introduction of firing costs drives a wedge between unemployed and employed workers. In this environment, the cut-off productivity level for workers in existing employment relationships is lower compared to the case without firing costs, since firms realize that separation entails a cost and are therefore willing to hold on to workers longer. At the same time, the cut-off for newly hired workers is higher compared to the case with no firing costs, since firms require compensation for future costs of firing. Given the distributional assumption of an upward sloping density in the neighborhood of the cut-offs, this means that a larger mass of unemployed is now affected by aggregate fluctuations. Figure 3.13 illustrates how the productivity thresholds change when firing costs are introduced. The volatility of the rejection and

---

22 Another reason why the model under the standard calibration fails to generate larger match efficiency fluctuations lies in the assumption of the idiosyncratic productivity shocks being iid. Introducing persistent idiosyncratic productivity shocks would arguably strengthen the effects on match efficiency. The reason is the following. In the standard case, in a recession newly separated workers can get a new job in the next period even if aggregate productivity remains low, since they can be "lucky" and draw a high value of productivity. With persistent idiosyncratic productivity shocks such high draws would be more unlikely and thus the match efficiency fall would be stronger and more persistent.
Notes: The figure shows part of the log-normal distribution and illustrates the effect of firing costs on the idiosyncratic productivity cut-offs.

Separation rate depends on the mass around the cut-off points; the larger the mass, the higher the volatility. Consequently, the slope of the density around the cut-off points is crucial. I assume that the density is upward-sloping in this area which is not unreasonable considering that the cut-off points are in the lower tail of the distribution. The assumption implies that the unemployment pool is populated by a relatively larger mass of marginal workers, compared to the employment pool.

In this setup it is the firing cost that drives the threshold for the unemployed into an area of greater mass. Different mechanisms, such as on the job training, would have a similar effect. Alternatively, one could try and model two productivity distributions for the (un)employed. Such an approach, however, is too complex given the illustrative aim of this section.

3.6.1 Matching model with firing costs

Household behavior, the matching process and the setup of employment relationships is the same as in Sections 3.5.1 to 3.5.1. The only exception is the budget constraint, (4.1), where total wage income is now more complex, which will be explained in Section 3.6.3. In what follows I describe the rest of the model.
Endogenous separations

The value of being in the unemployment pool $U_t$ at period $t$ is given by

$$U_t = b + E_t \left[ \beta_t f_t (1 - \rho_x) \int_{\tilde{p}_{t+1}}^p (W^N_{t+1} - U_{t+1}) dF(p_{t+1}) + U_{t+1} \right] ,$$  \hspace{1cm} (3.24)

where $\tilde{p}_{t+1}$ is the productivity threshold for newly matched workers. The threshold is the only difference with (3.15) in the model without firing costs.

The value of a job in period $t$ for newly matched and existing workers with idiosyncratic productivity level $p_{i,t}$ are

$$W^N_{i,t} = w^N_{i,t} + E_t \left[ \beta_t (1 - \rho_x) \int_{\tilde{p}_{t+1}}^p (W^E_{t+1} - U_{t+1}) dF(p_{t+1}) + U_{t+1} \right] ,$$  \hspace{1cm} (3.25)

$$W^E_{i,t} = w^E_{i,t} + E_t \left[ \beta_t (1 - \rho_x) \int_{\tilde{p}_{t+1}}^p (W^E_{t+1} - U_{t+1}) dF(p_{t+1}) + U_{t+1} \right] ,$$  \hspace{1cm} (3.26)

where $\tilde{p}_{t+1}$ is the productivity threshold for existing relationships. The only difference between (3.25) and (3.26) is in the wage rate, which is discussed in the next subsection.

Similarly, the value for the firm of being in a productive employment relationship with a newly hired and existing worker with individual productivity level $p_{i,t}$ is, respectively

$$J^N_{i,t} = z_t p_{i,t} - w^N_{i,t} + E_t \left[ \beta_t (1 - \rho_x) \int_{\tilde{p}_{t+1}}^p J^E_{t+1} dF(p_{t+1}) - F(\tilde{p}_{t+1}) \phi \right] ,$$  \hspace{1cm} (3.27)

$$J^E_{i,t} = z_t p_{i,t} - w^E_{i,t} + E_t \left[ \beta_t (1 - \rho_x) \int_{\tilde{p}_{t+1}}^p J^E_{t+1} dF(p_{t+1}) - F(\tilde{p}_{t+1}) \phi \right] ,$$  \hspace{1cm} (3.28)

where $\phi$ is the firing cost. The firing cost is assumed to be fully paid by the firm and wasteful. It is thus not a transfer payment to the worker, but rather a tax on the match in the event of separation. Such a specification is justified by the fact that firing costs are (at least partly) of administrative and legal nature, they include for instance loss of efficiency due to disruption of regular work flow etc.
Finally, the value of an open vacancy (imposing the free entry condition) is

$$\frac{\kappa}{q_t} = E_t \left[ \beta_t (1 - \rho_x) \int_{\tilde{p}_t} J_{t+1}^N dF(p_{t+1}) \right]. \quad (3.29)$$

The threshold for newly matched workers is such that the surplus of the new match is equal to zero

$$W^N(\tilde{p}_t^N) + J^N(\tilde{p}_t^N) - U_t = 0. \quad (3.30)$$

An analogous reasoning holds for existing employment relationships. However, in this case, one must take into account the firing cost in the outside option of the firm. Essentially, the surplus can be negative up to the value of the firing cost, since the firm saves this cost by holding onto the worker

$$W^E(\tilde{p}_t^E) + J^E(\tilde{p}_t^E) - U_t = -\phi \quad (3.31)$$

### 3.6.2 Wage bargaining

Workers coming from the unemployment pool do not possess any contract with the firm from the previous period. Therefore, if they do not come to an agreement with the firm over the wage, no firing costs need to be paid. Assuming Nash bargaining, the wage of newly matched workers is then a solution to $(1 - \eta)(W^N_{i,t} - U_t) = \eta J^N_{i,t}$. On the other hand, when the firm decides to fire a worker that has been in an employment relationship in the previous period it must pay firing costs. The wage of a worker in an existing employment relationship is then a solution to $(1 - \eta)(W^E_{i,t} - U_t) = \eta J^E_{i,t} + \phi$.

Using (3.24) to (3.29) one can show that the wages of newly hired workers and workers in existing relationships are, respectively

$$w^N_{i,t} = \eta(z_t p_{i,t} - \beta (1 - \rho_x) \rho_{t+1}^E \phi + \kappa \theta_t) + (1 - \eta) b, \quad (3.32)$$

$$w^E_{i,t} = \eta(z_t p_{i,t} + (1 - \beta (1 - \rho_x) \rho_{t+1}^E) \phi + \kappa \theta_t) + (1 - \eta) b, \quad (3.33)$$

where the structure is the same as in the model without firing costs. Newly hired workers, however, are penalized because of the threat of having to pay firing costs.
in the future. On the other hand, workers in existing employment relationships now have a higher wage compared to the case without firing costs, because their effective bargaining power increased, since firing them entails a cost for the firm.

3.6.3 Closing the model

Let the separation rate of existing employment relationships ($\rho^E_t$) and the rejection rate ($\rho^N_t$) be defined as, respectively

$$
\rho^E_t = \rho_x + (1 - \rho_x)F(\tilde{p}^E_t),
$$

$$
\rho^N_t = \rho_x + (1 - \rho_x)F(\tilde{p}^N_t).
$$

Then, the law of motion for unemployment is given by

$$
 u_{t+1} = (1 - f_t(1 - \rho^N_{t+1}))u_t + \rho^E_{t+1}n_t,
$$

where the only difference compared to (3.21) is that now one needs to distinguish between the separation and rejection rates. Finally, aggregate output is determined by

$$
y_t = z_n(t(\omega_tG(\tilde{p}^N_t) + (1 - \omega_t)G(\tilde{p}^E_t)),
$$

where $\omega_t = \frac{f_{t-1}u_{t-1}(1 - \rho^N_t)}{n_t}$ is the fraction of newly employed workers in total employment. Similarly, total wage income is now given by $n_t(\omega_tw(G(\tilde{p}^N_t), t)^N + (1 - \omega_t)w(G(\tilde{p}^E_t), t)^E)$.

3.6.4 Match efficiency variation

In the case with firing costs the probability that an unemployed worker finds himself employed in the next period is given by $f^*_t = f_t(1 - \rho^N_{t+1})$. Measured match efficiency is then defined as

$$
A_t = A(1 - \rho^N_{t+1}),
$$

where its fluctuations now depend on the rejection rate, which is not equal to the separation rate in this setup.
Table 3.6: Parameter values: 1st calibration with firing costs

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Target/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discount factor</td>
<td>$\beta$</td>
<td>0.99</td>
</tr>
<tr>
<td>Relative risk aversion</td>
<td>$\gamma$</td>
<td>1</td>
</tr>
<tr>
<td>Agg. shock persistence</td>
<td>$\rho_z$</td>
<td>0.95</td>
</tr>
<tr>
<td>Agg. shock st. dev.</td>
<td>$\sigma_z$</td>
<td>0.007</td>
</tr>
<tr>
<td>Idio. shock mean</td>
<td>$\mu_F$</td>
<td>0</td>
</tr>
<tr>
<td>Match elasticity</td>
<td>$\mu$</td>
<td>0.614</td>
</tr>
<tr>
<td>Firing costs</td>
<td>$\phi$</td>
<td>0.179 Bentotila and Bertola (1990), OECD</td>
</tr>
<tr>
<td>Match efficiency</td>
<td>$A$</td>
<td>0.605 f = 0.45, Shimer (2007)</td>
</tr>
<tr>
<td>Exogenous destruction</td>
<td>$\rho_x$</td>
<td>0.058 u = 0.12</td>
</tr>
<tr>
<td>Vacancy posting costs</td>
<td>$\kappa$</td>
<td>0.423 q = 0.71, den Haan et al. (2000)</td>
</tr>
<tr>
<td>Worker bargaining power</td>
<td>$\eta$</td>
<td>0.079 $\epsilon_{w,\bar{p}} = 0.45$</td>
</tr>
<tr>
<td>Worker outside option</td>
<td>$b$</td>
<td>1.004 $W/\bar{p} = 0.97$</td>
</tr>
<tr>
<td>Idio. shock st. dev.</td>
<td>$\sigma_F$</td>
<td>0.386 $\sigma(\rho) = 0.061$</td>
</tr>
</tbody>
</table>

Notes: This calibration targets separation rate volatility and introduces positive firing costs. $\bar{p}$ is the average productivity of the employment relationships, consisting of aggregate productivity $z$ and the average worker productivity $G(\tilde{p})$. $\epsilon_{w,\bar{p}}$ is the elasticity of wages with respect to productivity and $W/\bar{p}$ is the wage share.

3.6.5 Calibration

The Employment Protection Legislation index (EPL) published by the OECD is a comprehensive indicator and more precise than other alternatives. It is a weighted average of indicators capturing protection of regular workers against individual dismissals, requirements for collective dismissals and regulation of temporary employment. However, one needs to translate this index into a suitable model parameter. Bentotila and Bertola (1990) provide estimates of firing cost for France, Germany, Italy and the UK in the period between 1975 and 1986. Assuming that the EPL is proportional to the estimates provided by Bentotila and Bertola, one can get an estimate of firing costs for the U.S., since EPL data is readily available for the above countries and the U.S. economy. I take the UK estimate as a benchmark assuming that its institutional environment is closest to that of the U.S. economy. The implied firing costs are equal to 4.47% of annual wage. Hence, firing costs are set to $\phi = 4 \times 0.0447 \times \overline{w}^E = 0.179 \times \overline{w}^E$ for a quarterly model, where $\overline{w}^E$ is the steady state wage for workers in existing employment

---

\(^{23}\)For instance compared to the hiring and firing costs calculated by the World Bank in its “Doing Business studies”, the OECD indicator both covers a larger range of relevant aspects of LTC, and has more precise and differentiated sub-indicators.

\(^{24}\)Using the ”regular employment” EPL index.
Figure 3.14: IRFs to a positive one-standard-deviation technology shock, model with firing costs

Notes: The model is calibrated with firing costs \( \phi = 0.179 \) and targets the volatility of the separation rate. \( \hat{p}^E \) and \( \rho^E \) are the idiosyncratic productivity cut-off and the separation rate for workers in existing employment relationships, respectively. \( \hat{p}^N \) and \( \rho^N \) are the idiosyncratic productivity cut-off and the rejection rate for unemployed workers, respectively.

Using the above value for the firing costs, I recalibrate the model to fit the targets as in the 1st calibration which fits separation rate volatility. The resulting parameter values are summarized in Table 3.6.

### 3.6.6 Model performance

The mechanics of this model are the same as those with zero firing costs with one exception. With positive firing costs the separation and rejection rates are no longer identical. Figure 3.14 shows the impulse response functions of the productivity thresholds for existing and newly hired workers together with the associated separation and rejection rates. While both productivity thresholds drop by similar amounts, the separation rate fall is dampened while the rejection rate decrease is magnified because the density is upward-sloping in this part of the distribution. This difference between the magnitudes of the two responses of the separation and rejection rate is what makes the
Chapter 3

Table 3.7: Standard deviations of variables: US data and model with firing costs

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.050</td>
<td>0.028</td>
</tr>
<tr>
<td>ρ</td>
<td>0.061</td>
<td>0.061</td>
</tr>
<tr>
<td>u</td>
<td>0.122</td>
<td>0.149</td>
</tr>
<tr>
<td>v</td>
<td>0.138</td>
<td>0.102</td>
</tr>
<tr>
<td>f</td>
<td>0.077</td>
<td>0.077</td>
</tr>
</tbody>
</table>

Notes: The empirical standard deviations are based on U.S. data in the period between 1951Q1 and 2007Q1. The data were seasonally adjusted, logged and detrended with an HP filter with smoothing coefficient 1,600. Unemployment is taken from the Current Population Survey published by the BLS, vacancies are the Help Wanted Index published by the Conference Board and the job finding rate and separation rates are taken from Shimer (2007). The model is calibrated with firing costs $\phi = 0.179$ and it targets separation rate volatility $\sigma(\rho)$.

model able to better explain match efficiency fluctuations while still fitting separation rate volatility.

Table 3.7 compares model standard deviations with those in the US economy. The model calibrated in this way can explain 56% of the match efficiency volatility found in the data. At the same time, the volatility of other labor market variables are close to their empirical counterparts (only the volatility of unemployment is somewhat higher).

3.6.7 Model-based match efficiency

The previous section showed that the model can account for a sizable portion of match efficiency variation. Another way to view this is to compare the estimated match efficiency with its model-based counterpart. To this end, I use data for real GDP (logged and detrended with a HP filter with a smoothing coefficient of 1,600) and back out the implied aggregate productivity shock. This is done by inverting the policy function obtained when solving the model. I use this shock series to simulate the model. Note that the shock series is recovered without using labor market variables.

Figure 3.15 compares the estimated match efficiency and the one implied by the model using the backed-out technology shock. The model-based match efficiency series follows the estimated reasonably well (correlation coefficient 0.57), especially prior to 1990 (correlation coefficient of 0.67). In the case of the 1991 recession the economy experienced a jobless recovery and estimated match efficiency fell mostly after the end of the downturn. Since the only shock driving the model is backed out from real
Notes: The model based match efficiency is constructed in the following way. Using detrended real GDP data for the U.S. and the inverted policy rules of the model with firing costs one can back out the implied shock series consistent with the real GDP data. These shocks are then fed through the model which gives the “model-based” match efficiency. The “empirical estimate” series is the benchmark match efficiency estimate.

GDP, such different dynamics cannot be captured by the model. Similarly, during the recession in 2001 the fall in match efficiency as predicted by GDP growth was smaller than the estimated one, again pointing to a different character of the recession. It seems that after the onset of the Great Moderation output growth lost on importance in explaining match efficiency fluctuations.

Finally, I decompose the variance of the model-based job finding rate into contributions of match efficiency and labor market tightness as in Section 3.3.1. The model predicts that 19% of job finding rate fluctuations are driven by match efficiency. This is only slightly lower than the 24% found in the data.

3.7 Concluding remarks

A constant matching function, as is typically used in the literature, implicitly assumes that the labor market heterogeneity that it is aimed at capturing, is time invariant. This paper relaxes the assumption of constant parameters in the matching function. In
particular, the constant, which reflects match efficiency, is allowed to vary. Using data on the job finding rate and unemployment I specify an unobserved components model, where both match efficiency and vacancies (due to the poor data availability over a longer sample) are treated as unobserved states. The JOLTS vacancy series (available only at the end of the sample) provides additional information determining vacancies in the earlier part.

Estimated match efficiency varies procyclically over the business cycle and it can explain up to 25% of job finding rate fluctuations. Hence, recessions are periods when unemployed workers have a harder time finding jobs not only because there are less vacancies and more unemployed competing for them, but also because the process of matching workers to jobs is less efficient. The results are robust to several modifications and a Monte Carlo exercise documents that the empirical model is able to identify the unobserved states quite well.

One reason for varying match efficiency can be found in changes in labor market heterogeneity which the matching function is aimed at capturing. One such form of varying heterogeneity is endogenous rejection. A positive rejection rate (not all workers that get matched with a vacancy start producing in the next period) drives a wedge between the total unemployment pool and the part useful for forming employment relationships. Moreover, countercyclical fluctuations in the rejection rate imply that in recessions the part of the unemployment pool useful for matching shrinks. The aggregate job finding rate then falls by more than would be implied by a constant matching function that takes into account the total number of unemployed and vacancies. However, calibrating this model to fit match efficiency fluctuations leads to a gross exaggeration in the volatility of other variables, most significantly that of the separation rate. Introducing firing costs helps alleviate this issue and makes the model perform well also quantitatively.
3.A The Kalman filter

The state-space model is summarized by (3.1) and (3.2), which I rewrite here for convenience:

\[ y_t = \Theta_{0,t} + \Theta_{1,t} s_t + \Theta_{2,t} x_t + \epsilon_t, \quad (3.39) \]

\[ s_t = \Phi_{0,t} + \Phi_{1,t} s_{t-1} + \eta_t. \quad (3.40) \]

The Kalman filter recursions can then be written as

\[ s_{t|t-1} = \Phi_0 + \Phi_1 s_{t-1|t-1}, \quad (3.41) \]

\[ P_{t|t-1} = \Phi_1 P_{t-1|t-1} \Phi_1' + Q, \quad (3.42) \]

\[ Z_t = \Theta_1 P_{t|t-1} \Theta_1' + R + \Theta_1 C + C' \Theta_1', \quad (3.43) \]

\[ V_t = y_t - \Theta_0 - \Theta_1 s_{t|t-1} - \Theta_2 x_t, \quad (3.44) \]

\[ K_t = (P_{t|t-1} \Theta_1' + C) Z_t^{-1}, \quad (3.45) \]

\[ s_{t|t} = s_{t|t-1} + K_t V_t, \quad (3.46) \]

\[ P_{t|t} = P_{t|t-1} - K_t (\Theta_1 P_{t|t-1} + C'), \quad (3.47) \]

where the subscript \( t|t-1 \) indicates a prediction of the variable for period \( t \), using information available in period \( t-1 \). Similarly, \( t|t \) is the update of the period \( t \) forecast, when period \( t \) information is revealed.

3.B The extended Kalman filter

Let the non-linear state space be described by the following measurement and transition equation:

\[ y_t = h(s_t, x_t) + \epsilon_t, \quad (3.48) \]

\[ s_t = f(s_{t-1}) + \eta_t, \quad (3.49) \]

where \( f \) and \( h \) are non-linear functions. Note that in Section 3.4.1 \( f \) is in fact linear. For
the state-space system given in (3.48) and (3.49) the extended Kalman filter recursions are the following:

\[ s_{t|t-1} = f(s_{t-1|t-1}), \]  
\[ P_{t|t-1} = F_t P_{t-1|t-1} F_t' + Q, \]  
\[ Z_t = H_t P_{t|t-1} H_t' + R + H_t C + C' H_t', \]  
\[ V_t = y_t - h(s_{t|t-1}, x_t), \]  
\[ K_t = (P_{t|t-1} H_t' + C) Z_t^{-1}, \]  
\[ s_{t|t} = s_{t|t-1} + K_t V_t, \]  
\[ P_{t|t} = P_{t|t-1} - K_t (H_t P_{t|t-1} + C'), \]  

where \( F_t \) and \( H_t \) are the Jacobian matrices of the transition and measurement equations, respectively:

\[ F_t = \frac{\partial f}{\partial s} |_{s_{t-1|t-1}}, \]  
\[ H_t = \frac{\partial h}{\partial s} |_{s_{t|t-1},x_t}. \]

3.C  Diagnostic tests

The assumptions underlying the specified model is that the residuals are normally distributed with constant variance and no serial correlation. Following Durbin and Koopman (2001) one can apply diagnostic tests of these properties to the standardized prediction errors defined as:

\[ e_t = V_t Z_t^{-1}, \]

where it then follows that the standard deviation of \( e_t \) is 1.
3.C.1 Serial correlation

One can use the Ljung-Box test to investigate the presence of serial correlation in the residuals. Denote the residual autocorrelation of order $k$ as

$$r_k = \frac{\sum_{t=1}^{n-k} (e_t - \overline{e})(e_{t+k} - \overline{e})}{\sum_{t=1}^{n} (e_t - \overline{e})^2},$$  \hspace{1cm} (3.60)

where $\overline{e}$ is the mean of the residuals. The Ljung-Box statistic is then

$$Q(k) = n(n + 2) \sum_{l=1}^{k} \frac{r_l^2}{n - l},$$ \hspace{1cm} (3.61)

which is $\chi^2(k - w + 1)$ distributed, with $w$ being the number of estimated hyperparameters (elements in the disturbance variance matrix).

3.C.2 Homoscedasticity

The assumption of constant variance can be tested with the following test statistic:

$$H(h) = \frac{\sum_{t=n-h+1}^{n} e_t^2}{\sum_{t=1}^{h} e_t^2},$$ \hspace{1cm} (3.62)

where $h$ is typically set to the nearest integer to $n/3$. The statistic then tests whether the variance in the first third of the sample is equal to that in the last third of the sample. This statistic is then $F(h, h)$ distributed.

3.C.3 Normality

The assumption that the standardized prediction errors are normally distributed can be readily tested using the Jarque-Berra test. The test statistic is defined as

$$JB = n \left( \frac{S^2}{6} + \frac{(K - 3)^2}{24} \right),$$ \hspace{1cm} (3.63)

where $S$ denotes the skewness and $K$ the kurtosis of the standardized prediction errors. The test statistic is $\chi^2(2)$ distributed.
3.D More robustness checks and further details

3.D.1 Different state processes

In the benchmark specification match efficiency was assumed to be an AR(1) process, while the process for vacancies was postulated to be a random walk. In this section I check the robustness of the results against two alternative specifications for the underlying states. First, I estimate the model assuming match efficiency follows a random walk, while keeping the specification of vacancies as in the benchmark model. Second, I retain the AR(1) assumption on match efficiency, but I allow for a richer non-stationary structure for vacancies. Namely, I assume that the first difference of vacancies follows an AR(2) process. The level of vacancies can then be written as

$$v_t = (1 + \rho_1)v_{t-1} + (\rho_2 - \rho_1)v_{t-2} - \rho_2 v_{t-3} + \eta^v_t. \quad (3.64)$$

Table 3.8 shows the estimated parameters for the benchmark model and the two alternative specifications. All specifications deliver very similar results. Figure 3.16 shows the Kalman smoothed states for the three specifications. As with the model parameters, the smoothed states are also very close to each other.

---

25 This specification makes it harder to identify the two states, because both have the same process. To help with this issue I use information from the benchmark for the starting values of the Kalman filter.
Figure 3.16: Kalman smoothed states: benchmark and other specifications

(a) Match efficiency

(b) Vacancies

Notes: $AR(p)$ and $RW$ indicate that the given process is an autoregressive process with $p$ lags or a random walk, respectively. Shaded areas are NBER recessions.
Table 3.9: Parameter estimates: different subsamples

<table>
<thead>
<tr>
<th>parameters/sample</th>
<th>Full</th>
<th>after 1970</th>
<th>after 1985</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )  ( \beta ) ( \rho )</td>
<td>(-0.639 ) (0.386) (0.696)</td>
<td>(-0.577 ) (0.497) (0.897)</td>
<td>(-0.659 ) (0.327) (0.791)</td>
</tr>
<tr>
<td>( ) ( (0.025) ) (0.032) (0.065)</td>
<td>( (0.041) ) (0.053) (0.106)</td>
<td>( (0.127) ) (0.133) (0.125)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Standard errors in brackets.

3.D.2 Estimating on subsamples and with different frequencies

Here I use two different subsamples to check whether the results are not driven just by a certain part of the data. The first subsample uses data after 1970 and the second data after 1985. Figure 3.17 shows the Kalman smoothed states for the subsamples together with the benchmark. Table 3.9 then shows the estimated parameter values. There are slight differences in the parameters, but they are also estimated with less precision as one discards more data points. Overall, the dynamics of the states are quite robust over the different samples. Furthermore, virtually identical results are obtained using monthly frequencies.

3.D.3 Alternative functional forms

The two alternative functional forms considered in the main text change the measurement equation related to the job finding rate (the rest of the state-space system remains the same). In the case of the first alternative matching function, the measurement equation becomes

\[
f_t = \alpha_{1,t} + 1/\xi \log(\exp(u_t)^\xi + \exp(v_{1,t})^\xi) - u_t + \epsilon_{1,t}, \quad (3.65)
\]

and the case of the second specification it is

\[
f_t = \alpha_{2,t} + v_{2,t} - 1/\zeta \log(\exp(u_t)^\zeta + \exp(v_{2,t})^\zeta) + \epsilon_{2,t}. \quad (3.66)
\]
Figure 3.17: Kalman smoothed states: benchmark and subsamples

(a) Match efficiency

(b) Vacancies

Notes: The benchmark estimates and those based on shorter subsamples. Shaded areas are NBER recessions.
3.D.4 Alternative estimation procedure

The first step of the alternative estimation procedure assumes a constant matching function and extracts a vacancy series implied by only data on the job finding rate and unemployment. The state-space of this system is given by

\[ f_t = \alpha + \mu(v_t^* - u_t) + \epsilon_t, \]  
(3.67)

\[ v_t^* = v_{t-1} + \eta_t. \]  
(3.68)

The second step then takes \( v_t^* \) and decomposes it into match efficiency (with mean zero) and the underlying vacancy state. The state-space system (allowing still for some measurement error) is

\[ v_t^* = v_t + \frac{1}{\mu} \alpha_t + \epsilon_{v^*,t}, \]  
(3.69)

\[
\begin{bmatrix}
\alpha_t \\
v_t
\end{bmatrix} =
\begin{bmatrix}
\rho & 0 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
\alpha_{t-1} \\
v_{t-1}
\end{bmatrix} + \eta_{v^*,t}. \]  
(3.70)

Checking the degree of non-linearity

To check the nonlinearity I fix match efficiency and unemployment at their average values and construct an artificial job finding rate using the state-space specification and the respective ML estimates. Then a scatter plot between this counterfactual (log) job finding rate and (log) vacancies indicates the degree of non-linearity in the model (in the benchmark case the result would be a straight line with the slope of \( \mu \)). Figure 3.5 shows the scatter plots for the two matching functions. In both cases the curves are very close to being linear indicating that the degree of non-linearity is not large and a first order Taylor expansion can be expected to perform quite well. Repeating the exercise for different (fixed) values of unemployment and match efficiency delivers practically identical results.
Table 3.10: Parameter estimates in vacancy yield equation

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_0$</td>
<td>$-0.006$</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>$0.984$</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>$-0.284$</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>$0.365$</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
</tr>
</tbody>
</table>

Notes: Standard errors in brackets. In the regression where the dependent variable is the vacancy yield, $\gamma_0$, $\gamma_1$, $\gamma_2$ and $\gamma_3$ are the coefficients on the constant, match efficiency, vacancies and unemployment, respectively.

3.D.5 Using information from the vacancy yield

The benchmark state-space system is augmented by a third measurement equation that reads

$$i_t = \gamma_0 + \gamma_1 a_t + \gamma_2 v_t + \gamma_3 u_t + \epsilon_{i,t},$$

(3.71)

where $i_t$ is the log of the vacancy yield. Once again, data points prior to 2001Q1 when the JOLTS data were not available, are treated as missing observations.

Although the job filling rate based on a Cobb-Douglas matching would imply that $\gamma_0 = 0$, $\gamma_1 = 1$ and $\gamma_2 = \gamma_3$, here I do not impose such restrictions since the estimation uses vacancy yield data instead. Table 3.10 shows the parameter estimates and shows that although all four parameters are close to satisfying the above restriction, they do violate them from a statistical significance point of view.

3.D.6 Extending the sample period

In calculating the job finding rate according to Shimer (2007) one needs short term unemployment data. In 1994 there is a discontinuity in this series due to a methodological change. Shimer proposes to deal with this by multiplying the official count of unemployment by the short-term share in only the first and fifth rotation groups in the CPS panel (these groups are measured in the same way, even after the methodological change, as the full sample prior to 1994). Here, instead of using the CPS groups directly I follow Elsby, Michaels, and Solon (2009). The authors propose to multiply
the official count of unemployment with the era’s average of the ratio of the short-term share for the first and fifth rotation groups relative to the full sample’s short-term share. In their case this ratio is 1.155 with the sample ending in 2005Q1. I take the shortcut of assuming that this ratio has not changed dramatically in the last five years and use it to calculate the job finding rate up until 2010Q4. The reader should keep in mind, however, that this subsection is only illustrative and a more careful analysis would require the job finding rate to be calculated using the actual CPS data. Finally, the estimation over this longer period uses the JOLTS vacancies data after the revision in March 2011.\footnote{There are two more dummy variables included in the quarters 2009Q1 to 2010Q2. This is likely due to the inaccuracy of the job finding rate estimates in this period.}

### 3.D.7 Monte Carlo experiment

The artificial data series are constructed using the benchmark state-space system, the maximum likelihood parameter estimates and the estimated unobserved states. For convenience I repeat the measurement equations below

\[
\begin{bmatrix}
 f_{t}^{MC} \\
 v_{t}^{MC}
\end{bmatrix}
= \begin{bmatrix}
 -\mu_{t}^{ML} \\
 0
\end{bmatrix}
+ \begin{bmatrix}
 1 & \mu_{t}^{ML} \\
 0 & 1
\end{bmatrix}
\begin{bmatrix}
 a_{t} \\
 v_{t}
\end{bmatrix}
+ \epsilon_{t}^{MC},
\]  

(3.72)

where the superscript $MC$ indicates that the respective data series is one of the artificial series generated in the Monte Carlo exercise while the superscript $ML$ indicates the benchmark maximum likelihood parameter estimate. Finally, $u_{t}$ is period $t$ U.S. unemployment and $a_{t}$ and $v_{t}$ are the benchmark Kalman smoothed estimates of match efficiency and vacancies, respectively, and are fixed across Monte Carlo replications. The variance-covariance matrix of $\epsilon_{t}^{MC}$ is equal to benchmark ML estimate.

The estimation procedure is exactly as in the benchmark, where the initial mean of the Kalman states is set to the respective expected value and its variance is set to $10^5$. The initial conditions for the maximization are set to the true values, to ease the computational burden. The actual ML estimates in each Monte Carlo replication step will, however, be different from the true parameter values due to the small sample at hand.
3.E Endogeneity

A valid concern is that there are endogeneity problems in the first observation equation. Therefore, the model in the main text is estimated using lagged values of the regressor as an instrument, which is typically done in the literature. Such an instrument is valid only if there is no serial correlation in the residual. As was shown in the main text, the model satisfies the assumption of no serial correlation in the residuals. In addition, the Hausmann test on exogeneity of instruments cannot reject the null hypothesis of exogenous instruments at the 40% level when the instrument is the first lag of unemployment. In the case when the model is estimated on monthly data, the exogeneity tests suggest 4 lags as the appropriate instrument, which is consistent with the quarterly tests.

3.F Effects of firing costs

In the case of zero firing costs the separation rate exactly equals the rejection rate. However, introducing positive firing costs drives a wedge between the two, making the rejection rate larger than the separation rate. This section shows analytically how firing costs increase the idiosyncratic productivity threshold for newly matched workers, while reducing the threshold for workers in existing employment relationships.

The two equations defining the threshold values are (3.31) and (3.30). First note that one can write the following

\[ J^E(p_{t+1}) = J^E(\tilde{p}_{t+1}) - (J^E(\tilde{p}_{t+1}) + \phi) = z_{t+1}(1 - \eta)(p - \tilde{p}_{t+1}) - \phi, \]  

(3.73)

where the first equality follows from the threshold condition (3.31) and the fact that with Nash bargaining the job value \( J^E \) is proportional to total surplus. The second equality comes from observing that both \( J^E(p_{t+1}) \) and \( J^E(\tilde{p}_{t+1}) \) have all terms common apart from the value of idiosyncratic productivity. Substituting (3.73) into (3.30) and
(3.31) one can obtain analytical expressions for the thresholds.

\[
\tilde{p}_N^t = \frac{1}{z_t} \left[ b + \frac{n}{1-\eta} \kappa \theta_t - \beta_t (1 - \rho_x) (G(\tilde{p}^E_{t+1}) - \tilde{p}^E_{t+1}) + \phi \frac{\beta (1 - \rho_x) (1 + F(\tilde{p}^E_{t+1}) - \eta)}{1 - \eta} \right],
\]

\[
\tilde{p}_E^t = \frac{1}{z_t} \left[ b + \frac{n}{1-\eta} \kappa \theta_t - \beta_t (1 - \rho_x) (G(\tilde{p}^E_{t+1}) - \tilde{p}^E_{t+1}) - \phi \left( 1 - \beta_t (1 - \rho_x) \left( 1 + \frac{F(\tilde{p}^E_{t+1})}{1 - \eta} \right) \right) \right],
\]

where \(1 + F(\tilde{p}^E_{t+1}) - \eta > 0\) for any non-negative value of endogenous separations and \(1 - \beta(1 - \rho_x) \left( 1 + \frac{F(\tilde{p}^E_{t+1})}{1 - \eta} \right) > 0\) for low enough values of \(F(\tilde{p}^E_{t+1})\). The steady state effect of firing costs on the threshold for new matches is directly evident from (3.74). Firing costs make the firm demand higher productivity of new matches as a compensation for expected future separations. The opposite reasoning holds for existing matches, where the firm settles for lower productivity levels, because separations now entail a cost. Obtaining an analytical expression for the steady state threshold for existing employment relationships is, however, impeded by the assumption of the log-normal distribution. The following subsection shows this steady state effect analytically under the assumption of a uniform distribution for idiosyncratic productivity. Nevertheless, in all the analysis it was always the case that the threshold for existing employment relationships fell with higher values of firing costs.

### 3.F.1 The case of a uniform distribution

Assuming a uniform distribution over idiosyncratic productivity levels \(p\) and normalizing its lower bound to 0, the steady state threshold level for existing matches can be shown to be

\[
\tilde{p}^E = \frac{b + \frac{n}{1-\eta} \kappa \theta_t - \beta (1 - \rho_x) \overline{p} - \phi (1 - \beta (1 - \rho_x))}{1 - \beta (1 - \rho_x) (1/2 + \frac{\phi}{(1-\eta)\overline{p}})},
\]

where \(\overline{p}\) is the upper bound of the uniform distribution. Since \(1 - \beta (1 - \rho_x) > 0\) then for the threshold to fall with higher firing costs it must be that \(1 - \beta (1 - \rho_x) (1/2 + \frac{\phi}{(1-\eta)\overline{p}}) > 0\). This depends not only on the extent of the firing costs, but also on the width of
the uniform distribution. It holds true as long as \( \frac{\phi}{\beta} < \frac{2-1}{2\beta(1-\rho_x)}(1 - \eta) \). For example, assuming a tight distribution, where the upper bound is 1, then firing costs need to be smaller than 0.194. For comparison with the benchmark model, one needs to multiply this value by 2, since average idiosyncratic productivity is half of what it is in the main text.
Chapter 4

Firm Age, Business Cycles, and Aggregate Labor Market Dynamics

Abstract

Recent studies show that the well established negative relationship between a firm’s size and its growth rate vanishes once its age is taken into account. Furthermore, it has been documented that young businesses have higher exit rates and grow faster than older ones, job creation and destruction rates fall with a firm’s age and young firms create relatively more jobs. I extend these findings by showing that young firms are also more volatile than older ones and that business start-ups are important for unemployment rate developments. Next, I build a general equilibrium model with heterogeneous firms that is consistent with these cross-sectional facts and delivers realistic aggregate labor market dynamics. The model is then used to evaluate a government policy supporting young firms, a measure proposed under the recent ”Startup America” initiative of the White House. The results suggest that such a policy should focus mainly on reducing barriers to entry. Supporting existing firms disrupts the selection process of successful firms, reducing average firm productivity, and resulting in lower levels of output.

4.1 Introduction

The most recent economic downturn in the U.S. was accompanied by a severe deterioration of the labor market. From December 2007, the official start of the recession, until the end of the recession in June 2009, 7.5 million jobs were lost. Such a fall in the
absolute number of jobs is unprecedented in U.S. postwar history. The effects of the economic downturn, however, were not the same across different firms. Employment in both young and small firms fell much more than in older, larger businesses. Jobs declined by 10.4% in firms with fewer than 50 employees, while larger businesses reported a 7.5% drop (Sahin, Kitao, Cororaton, and Laiu, 2011). Similarly, firms 5 years of age and younger experienced an employment fall of 14.8%, while jobs in older firms were reduced by 2.8%.

There is a long list of studies focusing on the role of firm size for the growth of businesses and the importance of small firms for job creation. Recently, however, a few studies have pointed out that it is young firms that are important for aggregate job creation and that young firms tend to grow faster than older ones. Moreover, using U.S. data Haltiwanger, Jarmin, and Miranda (2010) suggest that there is no link between a firm’s size and its growth, once its age is taken into account.1 This questions the current way of thinking and suggests that implications based on the paradigm in which firm size is an important driver of business growth can be misleading. Examining this freshly highlighted link between firm growth and its age and the impact young firms have on aggregate labor market dynamics is exactly the topic of this paper. First, I provide empirical findings related to firm age, some of which have been established in other studies and some of which are to my knowledge new. Next, I build a model with labor market frictions and heterogeneous firms which is consistent with these cross-sectional facts and predicts realistic aggregate labor market dynamics. This model is then used for analyzing policy questions concerning firm age, such as those proposed under the recent ”Startup America” initiative of the White House.

Studies that have devoted their attention to analyzing the link between firm growth and its age have shown that young firms have higher exit rates and conditional on survival they grow faster than older businesses. Furthermore, job creation and destruction rates fall with firm age and young businesses create relatively more net jobs. Finally, young firms are found to be mainly small, but small firms are not necessarily young.

Using the most recent version of the Business Dynamics Statistics (BDS) database, which includes the latest recession, I confirm these findings and document two new

---

1Ibsen and Westergaard-Nielsen (2011) document the same for Danish data.
facts. First, young firms are more volatile than older ones contributing to unemployment increases during and right after recessions and boosting employment growth later on in expansions. Second, firm entry is important for developments of the unemployment rate. This is especially evident during the most recent recession when lower than average firm entry alone accounted for almost 1/5 of the observed unemployment rate increase.

The main contribution of this paper is a novel general equilibrium model with search and matching frictions and heterogeneous firms. The model allows for rich firm dynamics in which businesses of different age have endogenously different employment behavior. Based on the expected benefits of operating a firm, new firms endogenously enter the economy upon which they obtain a business idea (productivity). Firm specific productivity levels then persistently evolve over time. Based on aggregate and individual business conditions, active firms produce and decide on (costly) hiring of unemployed workers. If firm specific conditions are so bad that it no longer pays off to remain in operation, firms shut down. In addition, firms can shut down and workers can be dismissed for exogenous reasons.

Firms in this model operate a production technology that uses labor as its only input and features constant returns to scale. This assumption is standard in the literature when firms also use capital as a production factor. However, many labor market models abstracting from capital also use constant returns to scale production technologies. Moreover, assuming a decreasing returns to labor technology would make small firms automatically grow faster than larger ones (irrespective of their age), because they would be further away from their optimal size and could therefore take advantage of higher marginal products of labor. This would directly contradict the findings of Haltiwanger, Jarmin, and Miranda (2010) that within a given age group there is no systematic link between firm size and firm growth. The setup in this paper implies that a small and a large firm with the same level of efficiency will advertise the same number of jobs. A constant returns to labor technology, however, also means that the concept of optimal firm size vanishes. Nevertheless, the characteristics of the firm

\[^2\text{See for example the classic Mortensen and Pissarides (1994) paper and many others that build on it.}\]
specific efficiency levels together with the presence of exogenous worker dismissals result in a well-defined firm size distribution.

The costs related to hiring new workers are assumed to be convex. Assumptions of increasingly costly factor adjustment are common in many models. In the context of this paper such an assumption implies that large, rapid, changes in employment come with increasingly high costs. In other words, firms that are productive enough to expand do so in a gradual manner.

To solve the model, I employ a standard solution technique for heterogeneous agent models proposed in Krusell and Smith (1998a). However, the properties of the model are such that one quickly runs into the curse of dimensionality. For this reason, I propose to solve an approximate maximization problem which does not have this unfavorable property. Accuracy checks show that this is a valid procedure for the model at hand.

The calibrated model is consistent with the established empirical findings concerning firm age characteristics. The key to understanding the performance of the model is the inherent selection mechanism of successful firms. Relatively less productive firms shut down early in their lives leaving only the more successful businesses to expand and grow old. In other words, the risk of going out of business as well as job destruction rates are high for younger firms. Moreover, conditional on survival, young firms tend to grow faster than older businesses. The reason behind this comes back to the fact that given an efficiency level a small and a large firm will hire the same number of workers. However, smaller firms experience lower worker turnover (in absolute terms) leaving them with relatively more resources for expansion compared to older businesses. Such a quicksilver nature of young firms results in them creating relatively more (net) jobs, a large part of which is also due to business startups.

The model is not only consistent with firm-level dynamics, but it also generates realistic business cycle statistics of aggregate labor market variables, both in terms of co-movement and volatility. Moreover, firm heterogeneity together with the endogenous process of firm entry, growth and exit, create two new propagation mechanisms that

---

are not present in a standard model assuming a representative firm. Their interaction generates greater persistence and richer dynamics than those predicted by a standard model.

The framework in this paper stresses the importance of firm heterogeneity and especially the dimension of firm age providing a natural setting in which to analyze policy measures aimed at supporting young firms. Such measures have been proposed for instance under the recent "Startup America" initiative of the White House. In the presented model there is a role for government intervention as the model is characterized by firms getting only a small fraction of output potentially resulting in too little entry and too much firm exit. The fact that the majority of firm selection happens early in the life of a firm (young firms have higher exit rates) justifies the consideration of a subsidy aimed at young firms.

The results suggest that subsidizing firm entry increases welfare as more new firms help reduce unemployment and increase output. However, if the government focuses its resources only on existing young firms aggregate output actually falls. The reason is that a subsidy for existing firms enables relatively less productive firms to remain in business, crowding out potential entrants. As a result, average firm productivity declines and the re-allocation of workers from relatively less productive firms to more efficient businesses is hampered. Thus, the model suggests that policy measures should focus on reducing barriers to entry and thereafter they should quickly be withdrawn so that the economy can pick its own winners.

While many studies have focused on the link between firm size and firm growth, recent findings suggest that this paradigm might be misleading and point to firm age as an important determinant of business growth rates. So far there are relatively few studies empirically investigating the link between a firm’s growth and its age. The first contribution of this paper is to extend the current empirical findings by examining further the role of firm age over the business cycle. Furthermore, while there are many theoretical models of firm growth, in my understanding of the literature, none of them explicitly focuses on firm age as an important factor for business growth and the implications young firms have for aggregate labor market dynamics. Thus, the second, main, contribution of this paper is to provide a novel framework capturing the
cross-sectional facts related to firm age. The presented model can thus serve as a tool for answering policy questions related to firm age and job creation. Such questions have gained in importance as unemployment rates remain stubbornly high in many countries affected by the crisis.

The paper is structured as follows. In Section 4.2, I provide a short overview of studies related to this paper. Section 4.3 establishes empirical facts about the firm age characteristics in the U.S. economy. Section 4.4 then builds a general equilibrium model aiming at explaining the established facts. Section 4.5 describes the solution method and Section 4.6 provides the calibration of the model. In Section 4.7, I summarize the model predictions for firm age characteristics and aggregate labor market dynamics. Section 4.8 uses the model to analyze the impacts of a government subsidy supporting young firms and finally, Section 4.9 gives some concluding comments.

### 4.2 Related research

In this section, I provide a short overview of three areas of research that are related to this paper. The topics include the relationship between firm size and firm growth (while not accounting for firm age), the link between firm age and growth of businesses and finally, theoretical models incorporating firm heterogeneity.

The contributions of Birch (1981) and Birch (1987) sparked interest in the link between firm size and firm growth. The central message of Birch’s research was that small firms are the most important source of job creation in the U.S. economy, since they tend to grow faster than older businesses. Davis, Haltiwanger, and Schuh (1996) criticized the analysis on the basis of it being subject to the "regression fallacy" or "regression to mean effect". When using a given base year to classify firms into size categories, two types of errors can occur. Businesses that are not small can be earmarked as small either because of measurement error or a transitory negative shock to their employment levels. In both cases the firm will tend to "grow faster" in the following periods as their employment levels revert back to their mean. The opposite logic can be used for large firms leading to a downward bias in the relationship between firm size and growth.
Davis, Haltiwanger, and Schuh (1996) propose a different firm size definition to weaken this statistical pitfall. Using manufacturing plant data, they then conclude that the regression fallacy fully accounts for the negative relationship between size and growth. On the other hand, using the National Establishment Time Series (NETS) database covering the entire U.S. economy and avoiding the above caveats, Neumark, Wall, and Zhang (2011) document an inverse relationship between net growth rates and firm size.

Similar studies have also been conducted on data from other countries. Davidsson, Lindmark, and Olofsson (1998) use Swedish data and conclude that small firms contribute more to net job creation than large ones. Baldwin and Picot (1995), Barnes and Haskel (2002), Broersma and Gautier (1997), and Voulgaris, Papadogonas, and Agiomirgianakis (2005) come to similar conclusions using data from Canada, the United Kingdom, The Netherlands and Greece, respectively. However, none of these papers took into account the effect of firm age.

The above studies are a starting point of more recent research that did draw its attention to firm age as the driver of growth of businesses. Using the BDS database, Haltiwanger, Jarmin, and Miranda (2000) document the high job creation and destruction rates of young businesses in the period between 1992 and 2005 in the U.S. economy. Neumark, Wall, and Zhang (2006) find an important role of firm startups in the high job creation share of small firms in the NETS database. Moreover, Haltiwanger, Jarmin, and Miranda (2010) argue that once one controls for firm age, the negative relationship between firm size and growth vanishes. A similar conclusion is made in Ibsen and Westergaard-Nielsen (2011) for Danish data. Halabisky (2006) shows that in Canada young firms account for the bulk of net job creation. Using a cross-section of 99 countries, Ayyagari, Demirguc-Kunt, and Maksimovic (2011) conclude that small and young firms have higher job creation and destruction rates.

Finally, the theoretical model of firm growth, entry and exit, presented in this paper, is closely related to several papers, starting with the early study of Jovanovic (1982). In his model, new firms grow faster and are more likely to fail than older ones, as they learn about their efficiency level. Hopenhayn and Rogerson (1993) present a general equilibrium model including rich firm dynamics (without aggregate uncertainty) and
analyze the welfare impacts of firing taxes. More recently, Acemoglu and Hawkins (2010), Elsby and Michaels (2010), Kaas and Kircher (2011), Moscarini and Postel-Vinay (2010b) and Schaal (2010) extend the Mortensen-Pissarides model to include multiworker firms. While Acemoglu and Hawkins focus on the implications of the firm size distribution for unemployment and vacancy persistence, Elsby and Michaels show that a model with endogenous separations and a role for firm size can account for the business cycle features of aggregate labor market variables. Kaas and Kircher address the question of efficiency in search and matching models with multiworker firms and Moscarini and Postel-Vinay provide a theoretical underpinning for their empirical finding that larger employers fluctuate more than smaller ones. Finally, Schaal uses a rich heterogeneous firm model to analyze the interaction of aggregate productivity shocks and uncertainty shocks to explain the simultaneous and persistent increase in unemployment and a sharp rise in labor productivity observed in the recent recession.

4.3 Firm age and job creation in the U.S. economy

This section uses the most recent version of the Business Dynamics Statistics database, which includes the latest recession, to confirm the findings of previous studies related to average firm age characteristics across business cycles. Then, I extend these findings and document two new facts concerning young firms and their business cycle behavior.

The BDS is a publicly available database constructed by the Census Bureau. It covers approximately 98% of U.S. private employment and contains information on employment stocks and flows. The data is annual and runs from 1977 to 2009 and is broken down by location, industry, firm size and firm age.

I report the empirical findings in two blocks. First, facts regarding averages over the sample period between 1992 and 2009. This is a compromise between analyzing the data over a longer sample and including a rich enough age structure. Second, empirical findings related to business cycles. For this second block, I choose to differentiate only between firms younger than six years and the rest. This enables me to extend the sample period such that it starts in 1982. While the first block is an extension to previous studies based on an earlier version of the BDS database that ended in 2005,
the second block of empirical facts is to my knowledge new. Before analyzing the data, I define the concepts of job creation, job destruction, firm size as well as the age categories that will be used throughout the paper.

4.3.1 Definitions

Following Davis, Haltiwanger, and Schuh (1996), for all the age groups one can define

- **gross job creation** as the sum of employment gains over all businesses whose employment level has increased during the last year
- **gross job destruction** as the sum of employment losses over all businesses whose employment level has decreased during the last year
- **net job creation** as the difference between gross job creation and gross job destruction
- **firm size** is the simple average of firm employment in year \( t \) and \( t - 1 \).

In order to ease the exposition, I choose to define the following groups of firms based on their age

- **new firms** are those younger than 1 year
- **young firms** are those 5 years of age and younger (hence, including new firms)
- **old firms** are those older than 5 years

4.3.2 Firm age characteristics across business cycles

In this section, I first describe the general characteristics of the firm age distribution. Then I document the characteristics of firm dynamics, which are related to the findings of Haltiwanger, Jarmin, and Miranda (2010) about the determinants of firm growth. Finally, I summarize facts about job creation and job destruction according to firm age.

\[ \text{This firm size definition is also known as the "current" or "average" firm size. An alternative, perhaps more natural firm size concept is simply firm employment in period } t \ ("base year" \text{ firm size definition). Davis, Haltiwanger, and Schuh (1996) introduced the "current" firm size definition to diminish the "regression fallacy" as mentioned in the literature overview.} \]
Figure 4.1: Firm shares according to size and age

(a) Age shares according to size

(b) Size shares according to age

Notes: Panel (a) plots the share of each firm age group in the total number of firms in a given size category. The shares add up to 100 within each size dimension. Panel (b) plots the share of each firm size group in the total number of firms in a given age category. The shares add up to 100 within each age dimension. BDS data, averages between 1992 and 2009.
The distribution of firms

On average 40% of all firms are younger than 6 years. New firms, a subset of young businesses, account for about 10% of all firms. Firms in the oldest category (16 years and older) account for almost 30% of all businesses.

The BDS database also reports statistics according to the joint break-down into firm age and size. Panel (a) of Figure 4.1 shows the firm age shares in a given size category, while panel (b) of Figure 4.1 depicts the opposite ordering and shows firm size shares in a given age category. Panel (b) documents that in all age categories the size shares fall monotonically as firm size increases. On the other hand, this is not apparent in panel (a) where the age shares for small firms do not display a decline as firm age increases. Hence, it is the case that young firms are mostly small (panel (b)), but small firms are not necessarily young (panel (a)).

Size, age and firm growth

The U.S. economy is highly dynamic with on average 10% of all firms shutting down and 10% of all firms being startups each period. Table 4.1 documents that younger firms have higher exit rates than older ones, with young firms accounting for almost 1/2 of all firm closures.

Not only do younger firms have higher exit rates but, conditional on survival, they also tend to grow faster than more mature firms resulting in a strong "up-or-out" tendency. Haltiwanger, Jarmin, and Miranda (2010) document the negative relationship between firm age and firm growth. Using a non-parametric specification and controlling for detailed industry and year fixed effects, they regress net employment growth rates on firm age classes and find that young firms grow faster than older ones with new firms being crucial for this result. Furthermore, they also investigate the case when size is added as an additional explanatory variable. In this case, they no longer find that small firms have systematically larger net employment growth rates than larger businesses, as previously documented in the literature.
Table 4.1: Firm exit/entry rates and shares according to firm age, 1992-2009

<table>
<thead>
<tr>
<th>age category</th>
<th>firm share</th>
<th>Exit rate</th>
<th>share</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7.9</td>
<td>25.1</td>
<td>17.4</td>
</tr>
<tr>
<td>2</td>
<td>6.7</td>
<td>18.6</td>
<td>10.4</td>
</tr>
<tr>
<td>3</td>
<td>5.9</td>
<td>15.9</td>
<td>7.7</td>
</tr>
<tr>
<td>4</td>
<td>5.2</td>
<td>14.1</td>
<td>6.1</td>
</tr>
<tr>
<td>5</td>
<td>4.7</td>
<td>13.0</td>
<td>5.1</td>
</tr>
<tr>
<td>6 – 10</td>
<td>18.5</td>
<td>10.7</td>
<td>16.5</td>
</tr>
<tr>
<td>11 – 15</td>
<td>12.9</td>
<td>8.6</td>
<td>9.7</td>
</tr>
<tr>
<td>16+</td>
<td>28.0</td>
<td>7.0</td>
<td>27.1</td>
</tr>
<tr>
<td>all</td>
<td>1</td>
<td>10.5</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: "Firm share" gives the share of firms in a given category relative to the total number of firms. "Exit rate" is the fraction of firms that shut down in a given category relative to the number of firms in that age category. "Exit share" gives the fraction of firms shutting down in a given category relative to the total number of firms shutting down. Reported values are in percent.

Job creation and destruction according to firm age

More than 1.5 million net new jobs are created each year on average. This number, however, hides a large amount of churning, since annually 17.5 million jobs are created and 16 million jobs are destroyed on average. Put differently, almost 30% of all jobs are either destroyed or newly created.

Rates. Table 4.2 shows the job creation and destruction rates according to firm age.5 Both job creation and destruction rates drop gradually with firm age. The stunning feature of Table 4.2 is that only new firms have positive net job creation. However, gross job creation is large in all age categories and thus one should not conclude that it is only new firms that account for all job creation in the economy.

Shares. The above suggests an important role of young businesses in job creation. However, young firms are mostly small and thus it is not clear whether high job creation rates also translate into a large fraction of newly created jobs. Table 4.3 shows the shares of gross job creation and destruction of given firm age groups in the total. The

---

5The 200% job creation rate of new firms is an artifact of the firm size definition. Job creation and destruction rates are calculated as the total number employment gains (losses) over all businesses whose employment level has increased (decreased) during the last year divided by firm size. The firm size definition is based on the simple average of period $t$ and $t - 1$ employment levels. In the case of new firms this results in firm size being half of period $t$ employment, since these firms did not exist in period $t - 1$. 
Table 4.2: Job creation and destruction rates by firm age, 1992-2009

<table>
<thead>
<tr>
<th>age category</th>
<th>net JC</th>
<th>gross JC</th>
<th>gross JD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>200</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>-2.2</td>
<td>28.0</td>
<td>30.2</td>
</tr>
<tr>
<td>2</td>
<td>-6.5</td>
<td>20.9</td>
<td>27.5</td>
</tr>
<tr>
<td>3</td>
<td>-5.1</td>
<td>18.9</td>
<td>24.0</td>
</tr>
<tr>
<td>4</td>
<td>-4.3</td>
<td>17.8</td>
<td>22.1</td>
</tr>
<tr>
<td>5</td>
<td>-3.7</td>
<td>16.9</td>
<td>20.6</td>
</tr>
<tr>
<td>6 – 10</td>
<td>-3.0</td>
<td>15.0</td>
<td>17.9</td>
</tr>
<tr>
<td>11 – 15</td>
<td>-1.9</td>
<td>13.7</td>
<td>15.6</td>
</tr>
<tr>
<td>16+</td>
<td>-0.5</td>
<td>12.0</td>
<td>12.5</td>
</tr>
<tr>
<td>all</td>
<td>1.5</td>
<td>16.2</td>
<td>14.8</td>
</tr>
</tbody>
</table>

Notes: Net and gross job creation rates and gross job destruction rates for different firm age groups.

Table 4.3: Job creation and destruction shares by firm age, 1992-2009

<table>
<thead>
<tr>
<th>age category</th>
<th>gross JC</th>
<th>gross JD</th>
<th>employment share</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>17.4</td>
<td>0</td>
<td>1.4</td>
</tr>
<tr>
<td>1</td>
<td>4.9</td>
<td>5.8</td>
<td>2.8</td>
</tr>
<tr>
<td>2</td>
<td>3.4</td>
<td>5.0</td>
<td>2.7</td>
</tr>
<tr>
<td>3</td>
<td>2.9</td>
<td>4.0</td>
<td>2.5</td>
</tr>
<tr>
<td>4</td>
<td>2.6</td>
<td>3.5</td>
<td>2.4</td>
</tr>
<tr>
<td>5</td>
<td>2.4</td>
<td>3.2</td>
<td>2.3</td>
</tr>
<tr>
<td>6 – 10</td>
<td>9.5</td>
<td>12.6</td>
<td>10.3</td>
</tr>
<tr>
<td>11 – 15</td>
<td>7.7</td>
<td>9.6</td>
<td>9.1</td>
</tr>
<tr>
<td>16+</td>
<td>49.2</td>
<td>56.3</td>
<td>66.5</td>
</tr>
</tbody>
</table>

Notes: Shares of gross job creation, destruction and employment of different firm age groups in the total.

last column reports the respective employment shares. The clear "outlier" is the group of new firms that accounts for 17.4% of all jobs created in expanding firms, while their employment share is only 1.4%. Young firms create 34.3% of all jobs in expanding businesses and they destroy 21.5% of all jobs in contracting firms, even though they account for only 15.1% of total employment. Thus, young firms are important for aggregate job creation as they create a disproportionately large amount of jobs.

---

6 Using the base year size definition the employment share would be twice as large, 2.8%.
7 One should note that while the number of young firms is relatively stable over time, the total number of firms is growing. However, as indicated in the paragraphs above, it is important to put the job creation and destruction shares of young firms in relation to the respective employment shares, which were also declining over the given sample. Therefore, the overall picture that young firms create a disproportionately large amount of net jobs still holds.
4.3.3 Business cycle characteristics

The previous paragraphs spoke about the average importance of young firms for job creation across business cycles. In this section, I first examine the cyclical properties of young firms’ employment growth rates. Second, I conduct two counter-factual scenarios to highlight the importance of young firms for aggregate labor market dynamics.

Employment growth rates of young and old firms

Employment growth rates of both young and old firms are procyclical, with that of young businesses being 2.5 times more volatile than that of old firms. The correlation coefficients between the cyclical component of the unemployment rate (HP-filtered with smoothing coefficient of 6.23) and the two employment growth rates are $-0.57$ and $-0.63$ for young and old firms, respectively.\(^8\)

To further understand this cyclical pattern, one can look at the business cycle properties of the difference between employment growth rates of young and old firms. This was proposed by Moscarini and Postel-Vinay (2010a) for the group of small and large firms. Figure 4.2 plots this differential growth rate together with the unemployment rate (both detrended with a quadratic trend). Young firms contract faster or expand slower than older firms (a low differential growth rate) during recessions and early in recoveries and they grow faster or contract slower than older ones (a high differential growth rate) later on in expansions. The correlation coefficient of the two series is $-0.62$.\(^9\)

Understanding the importance of young firms

To further analyze the impact of young firms’ dynamics on aggregate labor market outcomes, I consider two counterfactual scenarios. First, what would the unemployment rate look like if young firms’ job creation and destruction rates were the same as those of old firms? Second, I focus on the latest recession which was especially hard on young firms. I ask what would the aggregate unemployment rate be if young firms’ job

---

\(^8\)The choice of a smoothing coefficient of 6.23 for annual data is based on the recommendation in Ravn and Uhlig (1997).

\(^9\)Detrending with an HP filter instead delivers a slightly lower correlation coefficient of $-0.43$. 
creation and destruction behaved as it did on average during the previous recessions?

Job creation and destruction of young businesses as that of old firms. I construct a counterfactual aggregate employment level by replacing actual job creation and destruction rates of young firms by those of old businesses. The difference between actual employment and this counterfactual value is then added to the unemployment rate.

Not surprisingly, there is a level effect since young firms’ net job creation rates are higher than those of old ones. The counterfactual unemployment rate is roughly 2 percentage points higher.\(^{10}\) This difference is mainly due to firm entry (which does not have a counterpart in the group of old firms). Higher firm exit and job destruction on the part of young firms compared to older businesses then counteracts these effects partly.

More interestingly, there are also some differences in dynamics. Figure 4.3 shows the

\(^{10}\)Note that this level effect is likely to be underestimated as I leave the employment of old firms unchanged. Specifically, I do not account for the effect that lower net job creation rates of young firms translate into lower employment levels of old firms in the future.
Figure 4.3: Actual-counterfactual unemployment decomposition; old firms

Notes: The figure plots the difference between the actual unemployment rate and the (mean adjusted) unemployment rate that arises when job creation and destruction rates of young firms are the same as those of old firms. The bars indicate the contributions of new firms (entry), young firm closures (exit), job creation of young continuers (JC), and job destruction of young continuers (JD). Shaded areas are NBER recessions.
difference between the actual and mean adjusted unemployment rates together with its decomposition into the contributions of firm entry, exit, job creation and destruction of existing firms. A positive difference between the two unemployment rates means that the different dynamics of young firms’ job creation and destruction compared to those of old firms contributed to unemployment increases. Similarly, a negative difference indicates that young firms’ dynamics, relative to those of old firms, contributed to a decrease in the unemployment rate.

The figure shows that the different dynamics can account for up to 0.9 percentage points of unemployment. Looking at the decomposition, firm entry accounts for the largest share in the observed difference between actual and counterfactual unemployment. The 2001 recession is somewhat specific as young firms maintained their growth and contributed to a lower unemployment rate.

**Job creation and destruction of young firms as in an average recession.** I now zoom in on the latest recession only. I ask what would the unemployment rate look like if young firms behaved as they did on average during the other recessions in the sample? To answer this I detrend (log) real GDP, job creation and destruction with a linear trend. I then calculate the average response of young firms’ job creation and destruction to a 1 percentage point decrease in real GDP over the previous recessions. Next, I use this average response and the observed drop in real GDP to create counterfactual (un)employment in the most recent downturn. Based on the average response from previous recessions, employment in new firms would have fallen by about 5% in the latest recession, while it actually dropped by almost 30%.

Table 4.4 shows the actual and counterfactual unemployment rate for the latest recession. The difference in 2009 is more than 1 percentage point. Thus, only the fact that young firms were hit especially hard during the latest downturn accounts for 25% of the unemployment run-up during the latest recession.

---

11 The mean adjusted counterfactual unemployment rate is constructed by replacing young firms’ job creation and destruction rates by those of old businesses, but keeping the average levels unchanged. Simply subtracting the average unemployment rate delivers similar results.

12 Note that old firms have zero entry rates by definition. The difference thus arises because business startups are above or below their average value.

13 Using a quadratic trend instead changes little.

14 The average decline of 5% may seem small. However, employment in new firms typically continues in its decline after recessions, further impacting the unemployment rate. The most recent crisis is therefore unique in the sense that employment in new firms plummeted already during the recession.
Table 4.4: Unemployment rate decomposition; 2008-2009

<table>
<thead>
<tr>
<th></th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>unemployment</td>
<td>5.81</td>
<td>9.29</td>
</tr>
<tr>
<td>unemployment\textsubscript{count}</td>
<td>5.48</td>
<td>8.24</td>
</tr>
<tr>
<td>difference</td>
<td>0.33</td>
<td>1.05</td>
</tr>
<tr>
<td>- entry</td>
<td>0.26</td>
<td>0.60</td>
</tr>
<tr>
<td>- JC</td>
<td>0.13</td>
<td>0.26</td>
</tr>
<tr>
<td>- exit</td>
<td>0.04</td>
<td>-0.00</td>
</tr>
<tr>
<td>- JD</td>
<td>-0.10</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Notes: The first two lines show the actual and the counterfactual unemployment rate based on young firms behaving as they would have based on their average response during the previous recessions. The bottom rows show the difference between the two and its decomposition into entry, exit, job creation and destruction of continuing firms.

One can again decompose this difference into contributions of entry, exit, job creation and destruction of continuers. Once again, the lion’s share of the unemployment differential is because the level of start-ups is lower than it usually is in recessions. The second most important contributor was lower job creation of continuing firms. Higher job destruction (either due to firm exits or firing in continuing firms) contributes little.\textsuperscript{15} Thus, the underlying message of these counterfactual scenarios is that young firms, and especially entrants, are important for aggregate labor market dynamics.

### 4.3.4 Summary of empirical facts

- **Importance of age for firm growth**: size does not matter for firm growth once age is taken into account (Haltiwanger, Jarmin, and Miranda, 2010). Young firms have higher exit rates and conditional on survival tend to grow faster than older businesses.

- **Firm age/size distribution**: 40% of firms are young and they are mainly small, while small firms are not always young.

- **Job creation and destruction across business cycles**: job creation and destruction rates fall with firm age. Young firms create a disproportionately large number of (net) jobs compared to their employment shares.

\textsuperscript{15}In 2008 job destruction in continuing firms was even lower than the historical average in the previous recessions.
• **Cyclicality of young and old firms**: employment growth rates of both young
and old firms are procyclical. Young firms are more volatile than older businesses
contributing to unemployment increases during and right after recessions and
boosting employment growth later on in expansions.

• **Importance of entry**: firm entry plays a crucial role in the effect of young
firms on aggregate labor market dynamics accounting for large shares of total
employment declines during recessions, especially in the latest downturn.

### 4.4 Model

I turn now to building a general equilibrium model aimed at capturing the above
empirical facts. The economy is populated by a large number of *heterogeneous*
firm that differ in their productivity levels which evolve (persistently) over time. Each firm
operates a production technology that uses labor as its only input. Firms obtain labor
by hiring workers on a frictional labor market.

In each period, based on aggregate and firm specific conditions, existing firms decide
whether or not to stay in the economy. At the same time, new firms endogenously enter
the market upon which they obtain an initial productivity draw from a wide range of
"business ideas". All firms that remain active in the economy are subject to aggregate
and firm specific productivity shocks. Based on their development, they produce and
decide on whether to expand or shrink their workforce. Hence, even when aggregate
productivity is fixed, the model still generates rich firm dynamics. Some firms start up
and some firms that shut down, some businesses will expand, while others will contract.
Next, I turn to explain the model more formally.

#### 4.4.1 Timing

The timing of events in this economy is depicted in Figure 4.4. At the beginning of the
period, *before* any shocks are revealed, new firms enter the market. At the same time,
incumbent firms choose whether to continue production or shut down. After entry and
exit decisions have been made, the innovations to both the aggregate and idiosyncratic
productivity levels are revealed. Active firms pay an operational cost, produce output
and pay their workers. They also decide how many vacancies to post and pay the appropriate cost. At the end of the period, a fraction $\delta$ of all firms exogenously shuts down and all their employees enter the unemployment pool. In addition, a fraction $\rho_x$ of workers in employment relationships with existing firms exogenously loose their jobs. All workers in the unemployment pool, including those entering at the end of the period, are ready to find a job in the next period.

### 4.4.2 Household behavior

Households are assumed to be risk neutral. The household consists of a continuum of *ex-ante homogeneous* workers of unit mass. The members of the household pool their incomes from firm ownership, employment and non-employment activities and spend it on consumption. The model abstracts from any investment or labor force participation decisions. The household thus maximizes expected life-time utility subject to the

---

16 Assumption risk averse households would not change the results qualitatively, but it would increase the computational burden of the solution method as firms would have to keep track of aggregate consumption in order to determine their stochastic discount factor.
following budget constraint
\[ c_t = W_t + bu_t + P_t, \]  
where \( c_t \) is aggregate consumption, \( W_t \) is aggregate wage income, \( b \) captures home production and the value of leisure, \( u_t \) is the mass of unemployed workers and \( P_t \) are aggregate firm profits. Costs of posting vacancies are assumed to be redistributed back to the household.

### 4.4.3 Individual firm behavior

Let us now consider the decision problem of a firm that chose to stay in the market in period \( t \).

**Maximization problem**

Active firms solve the following maximization problem

\[
V^F(z_t, p_{i,t}, n_{i,t}) = \max_{n_{i,t+1}, v_{i,t}} \left\{ z_t p_{i,t} n_{i,t} - w_{i,t} n_{i,t} - \xi - \frac{\kappa_0 v_{i,t}}{\kappa_1} + \beta (1 - \delta) \max \left\{ 0; E_t V^F(z_{t+1}, p_{i,t+1}, n_{i,t+1}) \right\} \right\}
\]

s.t.
\[ n_{i,t+1} = (1 - \rho_x) (n_{i,t} + q_{i,t} v_{i,t}), \]

where \( V^F(z_t, p_{i,t}, n_{i,t}) \) is the period \( t \) value of an (active) firm with idiosyncratic productivity \( p_{i,t} \) and employment level \( n_{i,t} \). Aggregate productivity is given by \( z_t \), \( n_{i,t} \) is firm level employment and \( w_{i,t} \) is the wage rate. \( \xi \) is an operational cost that needs to be paid at the beginning of the period and \( \kappa_0 v_{i,t}^\kappa_1 / \kappa_1 \) are the vacancy posting costs with \( v_{i,t} \) being the number of posted vacancies. Finally, \( \beta \) is the discount factor, \( \delta \) is an exogenous exit-inducing shock and \( E_t \) is the expectations operator taken over both aggregate and idiosyncratic productivity.

Firm value is composed of current profits (the difference between firm output and costs consisting of wages, vacancy posting costs and the operational cost) and the continuation value of remaining in operation. Each existing firm chooses whether to
shut down or not at the end of period $t$ (prior to observing period $t+1$ shocks). Therefore, the continuation value cannot fall below zero as the firm chooses to shut down once expected firm value turns negative.

(4.3) is the law of motion for individual firm employment, where $q_{i,t}v_{i,t}$ is the number of newly hired workers ($q_{i,t}$ is the probability with which a posted vacancy gets filled). At the end of each period a fraction $\rho_x$ of employed and newly hired workers get dismissed for exogenous reasons and they enter the unemployment pool.\footnote{The assumption that new matches can get separated prior to production is made for convenience. It is straightforward to assume that new matches cannot separate prior to production (unless their employer shuts down), but the expressions get messier.}

**Firm output and vacancy posting costs**

The production technology assumed in (4.2) uses labor as its only input and features constant returns to scale. The assumption of constant returns to scale production is common in the literature when capital is also considered as a production factor. However, many labor market models abstracting from capital also assume a linear production technology (a prominent example is the classic paper Mortensen and Pissarides (1994) and many others that build on it). Moreover, assuming a decreasing returns to labor technology would create an immediate link between firm growth, firm size and age. Young firms would tend to grow faster *automatically* only because they are born small, hence further away from their optimal size, implying higher marginal products of labor. This would contradict the findings of Haltiwanger, Jarmin, and Miranda (2010) that within a given age group smaller firms do not systematically grow faster than larger businesses. In the setting adopted in this paper, a small and a large firm with the same level of productivity will post the *same* number of vacancies.

A constant returns to labor technology, however, means that the concept of an optimal firm size vanishes. Nevertheless, the characteristics of the firm specific efficiency levels together with the presence of exogenous worker dismissals result in a well-defined firm size distribution.

Vacancy posting costs are assumed to be convex. Assumptions of increasingly costly factor adjustment are commonly made in many models (as for example in Acemoglu and Hawkins (2010), Bloom (2009) Kaas and Kircher (2011), Merz and Yashiv (2007)).
In the current context such an assumption implies that large, rapid, changes in employment are increasingly costly. In other words, firms that are productive enough to expand do so in a gradual manner.

**Wages**

Different models present different theories motivating wage setting. Some models base their wages on Nash bargaining, some on social norms, models with multi-worker firms often use the Stole-Zwiebel framework, while Kaas and Kircher (2011) introduce a new bargaining framework. In addition to the multiple theories that can stand behind wage setting, it is also an open question how to calibrate the bargaining parameters, which often lack clear empirical counterparts.\(^{18}\)

In this paper, I choose a reduced form approach and propose a simple wage setting rule. The wage rule parameters can then be calibrated such that the model delivers empirically plausible aggregate wage dynamics. The wage bill of a firm with productivity \(p_{i,t}\) is defined as

\[
w_{i,t} = \bar{\omega} z_{i,t} p_{i,t} + \omega p_{i,t}(1 - z_{i,t}). \tag{4.4}
\]

In the steady state, wages are a fraction \(\bar{\omega}\) of firm output. The parameter \(\omega\) controls the stickiness of wages with respect to aggregate productivity. The higher the \(\omega\), the lower the positive response of wages to aggregate productivity. In this way, workers are rewarded relatively more in response to an increase in firm specific efficiency compared to an increase in aggregate productivity common to all firms. The reason why the second term is not just a constant, as is the case other wage rule specifications, is that in the framework with idiosyncratic productivity shocks a fixed wage term would disadvantage relatively less productive firms.\(^{19}\) For these businesses it would be relatively costlier to pay wages, while the wage rule presented here puts all firms on the same footing.

---

\(^{18}\)For instance Hagedorn and Manovskii (2008) show that an alternative calibration of the Nash bargaining framework can help resolve the “volatility puzzle” of standard matching models.

\(^{19}\)For an example of a wage rule with a fixed wage term see den Haan and Lozej (2010)
Exogenous shocks

Aggregate productivity and firm specific productivity of existing firms have the following processes

\[
\log(z_t) = \rho_z \log(z_{t-1}) + \epsilon_t, \quad \epsilon_t \sim G_{\epsilon}, \quad (4.5)
\]

\[
\log(p_{i,t}) = \rho_p \log(p_{i,t-1}) + \eta_{i,t}, \quad \eta_t \sim G_{\eta}, \quad (4.6)
\]

where \(\rho_z\) with \(\epsilon_t\) and \(\rho_p\) with \(\eta_t\) are autocorrelation coefficients and innovations of aggregate and idiosyncratic productivity, respectively. \(G_{\epsilon}\) and \(G_{\eta}\) are assumed normal with zero mean and standard deviations of \(\sigma_{\epsilon}\) and \(\sigma_{\eta}\), respectively.

### 4.4.4 Vacancy posting

A firm posts vacancies until the costs of doing so are equal to the expected benefits. Define the value of a job at a firm with idiosyncratic productivity \(p_{i,t}\) as

\[
J(z_t, p_{i,t}) = \frac{\partial V^F(z_t, p_{i,t}, n_{i,t})}{\partial n_{i,t}} = z_t p_{i,t} - w_{i,t} + \beta(1 - \delta)(1 - \rho_x) \max\{0; E_t J(z_{t+1}, p_{i,t+1})\}. \quad (4.7)
\]

Given the functional form of vacancy posting costs the free entry condition, which assumes that the value of posting a vacancy is pushed down to zero, reads

\[
k_0 q_{i,t}^{n_{i,t}} = q_{i,t} \beta(1 - \delta)(1 - \rho_x) \max\{0; E_t J(z_{t+1}, p_{i,t+1})\}. \quad (4.8)
\]

### 4.4.5 Firm entry

At the beginning of each period, prior to observing any period \(t\) shocks, potential firms can enter the economy. The cost of entering is given by \(\psi\) and it represents all the administrative, financial and utility costs related to starting up a new business. Paying this cost gives potential firms the chance to startup a business. Potential firms weigh this cost with the expected benefits of entry.

Upon entry new firms start with an employment level \(n_0\) and they obtain an initial
idiosyncratic productivity draw $p_{i,0} \sim G_{p,0}$. Thereafter, the idiosyncratic productivity evolves (persistently) over time according to the law of motion in (4.6). $G_{p,0}$ is assumed to be normal with mean 0 and standard deviation $\sigma_{p,0}$ which is allowed to be different from $\sigma_{\eta}$. The initial productivity distribution is supposed to represent a wide range of "business ideas". The distribution $G_{\eta}$ then specifies how these business ideas evolve over time. It is therefore natural to think of the initial productivity distribution as being much wider than that of the innovations to firm specific productivity. As becomes clear in the calibration section, this is actually the case.

The condition for firm entry can then be written as

$$\psi \leq \lambda_t E_{t-1} V^F(z_t, p_{i,0}, n_0), \quad (4.9)$$

where $E_{t-1} V^F(z_t, p_0, n_0)$ is the expected value of a new firm (taken over both the aggregate and the idiosyncratic productivity level). $\lambda_t$ is the probability of actually starting up a firm once the entry costs are paid. Entry occurs until (4.9) holds with equality.

The entry probability is assumed to depend on the mass of firms entering the economy ($N_{New}^t$), reflecting two main effects. First, the possibility that not all potential firms that begin the process of starting up a new business actually finish or succeed. Firms may not start up because of bad luck, discouragement, inability to obtain the appropriate documents or funds, etc. The reason why it depends on the number of entering firms is supposed to capture the notion of competition for funds needed to start a business, queues at the offices dealing with business startups, etc. Second, $\lambda_t$ is also meant to capture the matching probability related to hiring an initial number of $n_0$ workers. The higher the number of new firms, the lower the probability each firm has of hiring the desired number of workers. The entry probability is assumed to take on the following functional form $\lambda_t = \alpha_0 (N_{New}^t)^{\alpha_1}$.

---

20 This does not relate to the fact that some startups are not productive enough. That is captured by endogenous exit of new firms.
4.4.6 Firm exit

At the beginning of each period, prior to observing period $t$ shocks, incumbent firms decide whether to continue operating or to shut down. The firm bases its decision on expected firm value.\textsuperscript{21} If the firm chooses to exit, all its workers fall into unemployment. If, on the other hand, the firm chooses to continue operating, it faces the optimization problem in (4.2).

Formally, one can define a cutoff point, $\tilde{p}(z_t, n_{i,t})$ for firm specific productivity, below which firms choose to exit. As aggregate productivity decreases, relatively more productive firms become so unprofitable that it no longer makes sense to stay in the market and they shut down. Similarly, a smaller firm will have a harder time generating enough revenue to be able to pay the operational cost. Hence, the cutoff point is inversely related to both the aggregate productivity shock and firm size. The cutoff point summarizing the firms exit decision is implicitly defined by the following equation

$$E_{t-1}[V^F(z_t, \rho \tilde{p}(z_t, n_{i,t}) + \eta_{i,t}, n_{i,t})] = 0.$$  

(4.10)

4.4.7 The labor market and other aggregate variables

Let $u_t$ be the mass of workers that are unemployed and let $v_t$ be the mass of vacancies posted by all active firms in period $t$. Unemployed workers and vacancies match randomly on the labor market according to an aggregate matching function

$$M_t = \gamma u_t^\mu v_t^{1-\mu},$$

(4.11)

where $\gamma$ is match efficiency and $\mu$ is the elasticity of matches with respect to unemployment. The choice of a Cobb-Douglas matching function with constant returns to scale follows common practice in the literature. The probability of a given unemployed worker finding a job is given by $f_t = M_t / u_t$. Similarly, the average probability of a firm filling its vacancy is given by $q_t = M_t / v_t$.

The total number of matches, $M_t$, is then ”distributed” proportionally to individual

\textsuperscript{21}Note that firm profits can be negative if the value of staying in operation is large enough. Since the household is the owner of all firms, such losses then show up in aggregate firm profits in the household budget constraint.
firms based on their relative share in aggregate vacancies. This implies that all firms face the same probability of filling a vacancy. Define the number of matches of an individual firm as $m_{i,t} = v_{i,t}q_{i,t}$. One can then write

$$q_{i,t} = \frac{m_{i,t}}{v_{i,t}} = \frac{M_{t}v_{i,t}}{v_{i,t}} = \frac{M_{t}}{v_{i}} = q_{t}. \quad (4.12)$$

Let $n_{t}$ be the mass of workers that are employed (producing) in period $t$. Then, the law of motion for aggregate unemployment can be written as

$$u_{t} = 1 - n_{t} = u_{t-1} + \int_{i \in X_{t}^{exit}} \tilde{n}_{i,t-1} dt + \int_{j \in X_{t}^{stay}} \rho_{x} \tilde{n}_{j,t-1} dj - M_{t-1} - N_{t}^{new} n_{0}, \quad (4.13)$$

where $X_{t}^{exit}$ is the set of firms that were active in period $t - 1$ but shut down in period $t$, $X_{t}^{stay}$ is the set of firms that are active in both period $t - 1$ and $t$. $\tilde{n}_{i,t}$ is the number of employment relationships in firm $i$ at the end of the period and thus includes not only workers employed in period $t$, but also newly hired workers, $\tilde{n}_{i,t} = n_{i,t} + m_{i,t}$ ($n_{j,t}$ is defined analogously). Hence, the change in unemployment is given by the difference between the number of workers who exogenously separated or who were employed in firms that shut down and the number of unemployed workers who found jobs in existing or new firms.

Finally, aggregate vacancies and aggregate output are sums of the respective individual levels in active firms, $v_{t} = \int_{j \in X_{t}^{stay}} v_{j,t} dj$ and $y_{t} = \int_{j \in X_{t}^{stay}} n_{j,t} z_{j} p_{j,t} dj$.

### 4.4.8 Equilibrium

The individual state variables of each firm are its productivity and employment, $s_{i,t} = (p_{i,t}, n_{i,t})$. The aggregate state is given by aggregate productivity and the cross-sectional distribution of firm specific productivity and employment levels, $F_{t}(p_{i}, n_{i})$. The reason why the cross-sectional firm distribution is a state variable is because firms need to know the aggregate probability of filling a vacancy in order to be able to solve their maximization problem. This depends on the number of aggregate vacancies and unemployment. These variables are in turn determined by individual vacancy, employment, entry and exit decisions of all the firms in the economy. Thus, each firm needs
to know the entire distribution of firm specific productivity and employment levels in order to be able to predict the value of the probability of filling a vacancy.

Let \( S_t = (z_t, F_t(p_i, n_i)) \) denote the aggregate state. A competitive equilibrium is defined by

- (i) individual firm policy functions for employment, \( n(s_{i,t}, S_t) \), vacancies, \( v(s_{i,t}, S_t) \), and the exit decision, \( \bar{p}(s_{i,t}, S_t) \), that solve the individual firm problem in (4.2) and are consistent with the exit condition in (4.10),

- (ii) a mass of new entrants \( N_{New}(s_{i,t}, S_t) \) that satisfies the entry condition in (4.9),

- (iii) a firm specific productivity distribution, \( F_t(p_i, n_i) \), that is determined by the interaction of the (aggregate and idiosyncratic) exogenous productivity shocks and the employment, vacancy posting, exit and entry decision rules,

- (iv) and exogenous driving processes for aggregate and all individual productivity levels that are given in (4.5) and (4.6).

Finally, note that individual firm wages and output are determined by the above defined policy rules. Aggregate employment, vacancies, wages and output are given by the respective sums of individual firm variables over all active firms. Aggregate profits are the difference between total firm output and total costs consisting of wage, vacancy posting and operational costs, and aggregate consumption satisfies the budget constraint in (4.1).

### 4.5 Solution method

Maximization of the value function in (4.2) is not trivial, because among other things, one of the state variables is the cross-sectional firm distribution. This is a high-dimensional object and, more importantly, in the presence of aggregate productivity shocks, it varies over time. To deal with this issue, I follow Krusell and Smith (1998a) and assume that firms track only a few moments of the idiosyncratic productivity distribution. In the next paragraphs, assuming that I can solve the individual firm
problem, I describe an iterative scheme that solves for the equilibrium. I refer to this scheme as the Krusell-Smith (KS) algorithm.

Next, I show that given the characteristics of the KS algorithm one can easily run into the curse of dimensionality. At the same time, a perturbation solution of the individual firm problem is not possible because of a non-differentiability in the continuation value. I thus propose to solve an *approximate* maximization problem that allows for a perturbation solution for which higher dimensions of the problem pose no extra computational burden.

### 4.5.1 Krusell-Smith algorithm

The general idea of the KS algorithm is that instead of tracking the entire firm specific productivity distribution, firms follow only a few of its moments. The resulting equilibrium is thus an approximate one, since some relevant information is left out. However, the choice of this set of moments is such that the resulting forecast errors from omitting other information are very small.

Denote the moments of interest as $\mathcal{M}_t = (\mathcal{M}_{1,t}, \mathcal{M}_{2,t}, ... \mathcal{M}_{I,t})$, where $I$ is the number of moments firms follow. Firms perceive a law of motion for these moments given by $H_m$, so that $\mathcal{M}_t = H_m(\mathcal{M}_{t-1}, z_t, Q_{m,t})$, where $Q_{m,t}$ is a collection of past aggregate productivity shocks and/or further lags of $\mathcal{M}$.

As explained earlier, the knowledge of $H_m$ allows firms to solve for their optimal behavior. The resulting decision rules can be used to simulate the economy, which delivers time series of simulated moments of the individual productivity distribution. These can then be compared to the perceived laws of motion in $H_m$. The resulting approximate equilibrium must be such that the goodness of fit is high. In other words the law of motion, $H_m$, must track the evolution of simulated moments accurately.

As is clear from the maximization problem, firms are ultimately interested in the aggregate vacancy filling probability, $q_t$. Thus, I let firms track the evolution of $q_t$ *directly*, rather than letting them follow moments of $\mathcal{F}_t(p_t, n_t)$ and then relating these to $q_t$. The perceived law of motion for the aggregate vacancy filling probability is $q_t = H(q_{t-1}, z_t, Q_t)$, where $Q_t$ is a collection of past aggregate productivity shocks and/or further lags of $q$. Further details, such as a detailed description of the iterative
procedure, the exact functional form of $H$, the composition of $Q_t$, stopping criteria as well as accuracy tests are in Appendix 4.A.

4.5.2 An approximate maximization problem

From the exposition in the previous subsection it is clear that the dimensionality of the maximization problem can be quite high, if $Q_t$ includes many variables. As is described in detail in Appendix 4.A, $Q_t$ includes 5 variables and one therefore runs into the curse of dimensionality. Solving the individual firm problem with, for instance, value function iteration thus becomes exceedingly computationally expensive.

One would thus like to solve the individual firm problem with perturbation techniques, in which case the high dimensionality of the maximization problem poses no extra computational costs. However, perturbing the value function in (4.2) is not possible, because the option to exit introduces a non-differentiability. To overcome this problem, I propose to solve an approximate maximization problem, which is smooth and can therefore be solved using perturbation. Specifically, I replace the firms’ objective function with the following Bellman equation

$$
\tilde{V}^F(z_t, p_{i,t}, n_{i,t}) = \max_{n_{i,t+1}, v_{i,t}} [y_{i,t}(n) - w_{i,t}(n) - \xi - \kappa(v_{i,t}) + \beta (1-\delta) E_t \tilde{V}^F(z_{t+1}, p_{i,t+1}, n_{i,t+1})].
$$

The above equation basically ignores the non-differentiability in the continuation value. Hence, the firm behaves as if it does not account for the option value of shutting down.\footnote{Any losses incurred by the firm are transferred to the household as the owner of all firms. Furthermore, under risk neutrality of both firms and households and in the absence of a savings technology there are no precautionary motives that could alter agents’ behavior.}

Figure 4.5 illustrates the problem at hand. The true continuation value exhibits a non-differentiability where the zero lower bound kicks in. The perturbation solution to the approximate maximization problem, however, just extrapolates into the negative part of the state-space. Remember that the area of the state-space to the left of the "kink" is essentially irrelevant for continuing firms’ vacancy posting decisions. The reason is that if they find themselves in that part of the state-space, they choose to
exit the market. What will be distorted, however, is the average value of the policy function. This value is lower in the case of the perturbation solution, since in some parts of the state-space the perturbation policy rule is negative, while the true policy rule is zero. However, as long as the dynamics of the respective decisions are not affected, the level difference poses no issues since the calibration targets realistic average values of firm closures and aggregate vacancies.

I will now look at a simple case which does not run into the curse of dimensionality and which I can therefore solve accurately with, for instance, value function iteration. In particular, I assume that the true law of motion for $q_t$ depends only on $z_t$. I then compare this accurate solution of the true problem (including the non-differentiability) to that obtained using perturbation techniques to solve the approximate maximization. Figure 4.6 shows demeaned simulated exit and vacancy posting decisions based on the accurate solution of the true maximization problem (“true”) and those based on the perturbation solution to the approximate maximization problem (“approximate”). The

\[\text{Notes: The figure shows a diagram of a fictional true value function that includes a non-differentiability together with its perturbation approximation that ignores it.}\]
Notes: Panel (a) compares individual firm exit decisions for a simulated series of aggregate productivity (keeping firm employment at its steady state) under the true and approximate maximization problem. Panel (b) compares individual firm vacancy posting decisions for a simulated series of aggregate productivity (keeping firm employment at its steady state) under the true and approximate maximization problem.
two solutions yield very similar dynamics suggesting that using perturbation techniques on the approximate maximization problem is a valid procedure. Further details on the approximate maximization problem are presented in Appendix 4.B.

4.6 Calibration

To facilitate the calibration procedure, I divide the model parameters into two groups. First, parameters that are relatively standard or can be determined according to other studies in the literature, and second, parameters that are calibrated such that the model fits certain statistics in the data. Table 4.5 summarizes all the parameter values.

4.6.1 Parameters taken from the literature

The first group of parameters contains the discount factor, $\beta$, the standard deviation of aggregate productivity innovations, $\sigma_z$, the autocorrelation coefficient of aggregate productivity, $\rho_z$, match elasticity, $\mu$, the exogenous separation rate, $\rho_x$, the exogenous firm exit probability, $\delta$, the value of home production and leisure, $b$, and the power in the vacancy posting cost function, $\kappa_1$.

$\beta$ is set to 0.99 as the model period is assumed to be one quarter, $\sigma_z$ is set to 0.007 and $\rho_z$ is fixed at 0.95 as is standard in the literature. $\mu$ is set to 0.72 following the recommendation of Shimer (2007). Helfand, Sadeghi, and Talan (2007) report that in the BDS database through years 1990 to 2005 17.2% of gross job loss was due to firm closures. This is similar to the value of roughly 1/6 reported in Davis, Haltiwanger, and Schuh (1996), which is, however, related only to manufacturing firms. $\rho_x$, is thus set to $(1 - 0.172)\rho_{total}$, where $\rho_{total}$ is the total separation rate. The calibration related to the total separation rate is explained in the next subsection. Helfand, Sadeghi, and Talan (2007) also document that firm closures account for 1.7% of gross job loss at firms with more than 1,000 employees. The model in this paper predicts that the endogenous exit probability of large firms is virtually zero (large firms are also more productive and thus they first go through periods of contraction before they exit). Therefore, $\delta$ is set such that $\delta = 0.017(\rho_x + \delta)$. Without loss of generality $b$ is normalized to zero. Because of the structure of the adopted wage rule in (4.4) the outside option of workers does not
Table 4.5: Model parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Target/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discount factor</td>
<td>$\beta$</td>
<td>0.99</td>
</tr>
<tr>
<td>Autocorr. coef., agg. shock</td>
<td>$\rho_z$</td>
<td>0.95</td>
</tr>
<tr>
<td>St. dev, agg. shock</td>
<td>$\sigma_z$</td>
<td>0.007</td>
</tr>
<tr>
<td>Exog. separation rate</td>
<td>$\rho_x$</td>
<td>0.828</td>
</tr>
<tr>
<td>Exog. exit probability</td>
<td>$\delta$</td>
<td>0.0014</td>
</tr>
<tr>
<td>Match elasticity</td>
<td>$\mu$</td>
<td>0.72</td>
</tr>
<tr>
<td>Home production</td>
<td>$b$</td>
<td>0</td>
</tr>
<tr>
<td>Match efficiency</td>
<td>$\gamma$</td>
<td>0.741</td>
</tr>
<tr>
<td>Operational cost</td>
<td>$\xi$</td>
<td>0.562</td>
</tr>
<tr>
<td>Scale vacancy cost</td>
<td>$\kappa_0$</td>
<td>0.159</td>
</tr>
<tr>
<td>Scale entry prob.</td>
<td>$\alpha_0$</td>
<td>6.1e3</td>
</tr>
<tr>
<td>Entry cost</td>
<td>$\Psi$</td>
<td>1</td>
</tr>
<tr>
<td>Power entry prob. par.</td>
<td>$\alpha_1$</td>
<td>-2.92</td>
</tr>
<tr>
<td>1st wage parameter</td>
<td>$\tilde{\omega}$</td>
<td>0.943</td>
</tr>
<tr>
<td>2nd wage parameter</td>
<td>$\omega$</td>
<td>0.25</td>
</tr>
<tr>
<td>Idio. shock st. dev.</td>
<td>$\sigma_p$</td>
<td>0.073</td>
</tr>
<tr>
<td>Idio. shock persistence</td>
<td>$\rho_p$</td>
<td>0.976</td>
</tr>
<tr>
<td>Initial idio. shock st. dev.</td>
<td>$\sigma_{p,0}$</td>
<td>0.251</td>
</tr>
<tr>
<td>Initial employment</td>
<td>$n_0$</td>
<td>5.793</td>
</tr>
</tbody>
</table>

Notes: $\rho_{total}$ is the total separation rate, $n_{entry}$ is total employment of new firms, $\sigma(g)$ is the dispersion of employment growth rates (including entry and exit) and job persistence is defined according to Davis, Haltiwanger, and Schuh (1996) as the fraction of new jobs that survive into the next quarter.

enter wages. In the current setup it only affects the scale of aggregate output. Finally, the vacancy cost function is assumed to be quadratic in vacancies, $\kappa_1 = 2$, following Acemoglu and Hawkins (2010). Given this value (and the value of $\kappa_0$, the calibration of which is described in the next section) costs of a newly hired worker in an average firm are 8.7% of output per worker.\footnote{The total vacancy posting costs amount to roughly 17% of output per worker in an average firm. Each period there are $qv$ workers hired. This means that it takes $17/(vq) \approx 8.7\%$ of output per worker to hire one new employee in an average firm.} Nevertheless, the choice of quadratic vacancy posting costs is somewhat arbitrary and therefore I also investigate the case when $\kappa_1 = 5$.\footnote{Results available upon request.} The main results of the paper still hold under this alternative calibration.
4.6.2 Parameters chosen to match statistics in the data

The second group of parameters that are calibrated to match statistics in the data consists of match efficiency, $\gamma$, the scale parameter in the vacancy cost function, $\kappa_0$, the operational cost, $\xi$, the entry cost, $\Psi$, the scale parameter in the entry probability function, $\alpha_0$, the power parameter in the entry probability function, $\alpha_1$, the two wage rule parameters $\tilde{\omega}$ and $\omega$, the standard deviation of individual firm productivity innovations, $\sigma_p$, the autocorrelation coefficient of individual firm productivity, $\rho_p$, the standard deviation of the initial firm productivity draw, $\sigma_{p,0}$, and initial employment size, $n_0$. These nine parameters are set such that the model is able to match nine statistics in the data. To ease the exposition, I group the statistics into three categories: first order moments of aggregate variables, second order moments of aggregate variables, and firm level statistics.

The first category of parameters consists of $\gamma$, $\kappa_0$, $\xi$, $\Psi$ and $\alpha_0$. These parameters are set such that the model delivers a steady state unemployment rate of 12% common in the literature. Following den Haan, Ramey, and Watson (2000) and van Ours and Ridder (1992) the model targets a vacancy filling probability of 71%. The model further targets a total separation rate of 10% which is typically done in the literature. This value is based on evidence by Hall (1995) and Davis, Haltiwanger, and Schuh (1996) and is used for example by den Haan, Ramey, and Watson (2000) and Krause and Lubik (2007). Without loss of generality, $\Psi$ is normalized to 1 and $\alpha_0$ is calibrated such that in the steady state entry equals exit.

The second category of parameters includes $\alpha_1$, $\tilde{\omega}$ and $\omega$. These parameters are set such that they fit three second order moments of aggregate variables relative to the standard deviation of (log) labor productivity. The three statistics are the relative volatility of the share of new firm employment in total employment equal to 0.24, the relative employment rate volatility of 0.44 and the relative volatility of (log) wages equal to 0.76.

Finally, $\sigma_p$, $\rho_p$, $\sigma_{p,0}$ and $n_0$ are set such that the model matches four cross-sectional statistics. First, the average dispersion of firm employment growth rates of roughly

---

26 This implies a value of 8.28% for the exogenous separation rate $\rho_x$.

27 The entry condition in (4.9) shows that $\alpha_0$ and $\Psi$ cannot be identified separately.
0.56 as reported in Davis, Faberman, Haltiwanger, Jarmin, and Miranda (2008). This dispersion measure is the average cross-sectional standard deviation of firm employment growth rates, including entry and exit, over the period between 1992 and 2007. Second, the average persistence of new jobs is targeted to be 0.68 as documented in Davis, Haltiwanger, and Schuh (1996). This value gives the fraction of newly created jobs that survive into the next year.\footnote{It does not refer to individual jobs, but rather to an increase in employment that persists for one year.} Finally, the model targets the average firm size and the average size of new firms to be 21.3 and 6.1, respectively, as found in the BDS database.

### 4.7 Model results

The main goal of this paper is to build a theoretical framework that would help us understand the role of firm age in determining firm growth and its importance for aggregate labor market dynamics. In this section, I document that the model in this paper correctly predicts that both firm age and firm size alone are negatively related to the growth rate of firms as in the data. However, once age and size are taken into account together, the negative coefficient on firm size looses the majority of its statistical significance as documented by Haltiwanger, Jarmin, and Miranda (2010).

Furthermore, this section shows that the model is consistent with the other empirical findings presented in Section 4.3 and that it predicts realistic aggregate labor market dynamics, both in terms of comovement and volatility.

The reported results are based on values from 1,000 simulations. Each simulation has 1,108 quarters, where the first 1,000 are dropped to obtain 108 quarters as a counterpart to the sample used in the empirical part.

#### 4.7.1 Size, age and firm growth

Table 4.6 reports model based regression results of firm level employment growth rates on age, size and age and size together. The upper panel, comparable to Table 4 in Haltiwanger, Jarmin, and Miranda (2010), shows results for a specification which uses
Table 4.6: Firm growth regression on size and age

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Age only</th>
<th>Size only</th>
<th>Age and size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>2.000 (0.010)</td>
<td>2.552 (0.011)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.025 (0.010)</td>
<td>0.044 (0.017)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.008 (0.010)</td>
<td>0.129 (0.016)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>−0.005 (0.010)</td>
<td>0.125 (0.013)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>−0.013 (0.011)</td>
<td>0.089 (0.012)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>−0.017 (0.011)</td>
<td>0.050 (0.019)</td>
</tr>
<tr>
<td></td>
<td>6-10</td>
<td>−0.026 (0.005)</td>
<td>0.004 (0.006)</td>
</tr>
<tr>
<td></td>
<td>11-15</td>
<td>−0.029 (0.006)</td>
<td>−0.014 (0.007)</td>
</tr>
<tr>
<td></td>
<td>16+</td>
<td>−0.031 (0.004)</td>
<td>−0.016 (0.005)</td>
</tr>
</tbody>
</table>

|           | 0-8       | 0.194 (0.008) | −0.552 (0.006) |
|           | 8-12      | 0.040 (0.014) | 0.008 (0.015)  |
|           | 12-16     | 0.015 (0.019) | −0.059 (0.019) |
|           | 16-20     | −0.002 (0.017) | −0.121 (0.018) |
|           | 20-24     | −0.064 (0.009) | −0.149 (0.010) |
|           | 24-30     | −0.063 (0.005) | −0.067 (0.007) |
|           | 30-36     | −0.009 (0.006) | −0.005 (0.006) |
|           | 36-42     | −0.003 (0.011) | 0.005 (0.007)  |
|           | 42+       | −0.003 (0.007) |              |

Notes: The table reports coefficients of size and age in a regression on firm employment growth. "Regressor" indicates either a dummy variable for a certain age/size category, or actual age and size values. Standard errors are reported in brackets. The average firm size definition is used (base year size definition delivers similar results).

dummy variables for age and size groups. The coefficients can then be interpreted as averages in the respective age/size category.

Employment growth rates monotonically decline with firm age (when firm size is not considered) as documented in the data. Furthermore, a negative relationship can also be seen between firm size and business growth (when firm age is not considered), although to a lesser extent, since the largest firms are basically stagnant.

The most interesting case, however, is when both age and size are used together as regressors. In this case, age still retains its negative relationship with respect to firm employment growth rates. However, the inverse relationship between firm size and

\footnote{In the specification where age and size are used together, the category of the largest firms is excluded.}
its growth is considerably weakened. The clearest example is the group of very small firms for which the average growth rate (compared to the largest group) is significantly negative.

Although informative, the age and size categories are still arbitrary. Therefore, the lower panel reports results from a regression where age and size are not grouped into categories. In this case I regress firm employment growth rates on their age and size directly, and not on the respective dummy variables. The above-mentioned patterns are confirmed by these regression results. Both age and size alone display a negative, statistically significant, coefficient. However, when considering age and size together, the coefficient on size loses most of its statistical significance (the t-statistic falls from $-10.9$ when only size is considered to $-2.0$ when both age and size are taken into account). The model thus performs very well in capturing the recent finding that it is mainly age that matters for firm growth, not size.

The negative relationship between firm age and firm growth is driven by the effect of firm entry and the evolution of firm specific productivity. As firms age and expand the mean-reverting character of individual productivity kicks in lowering further incentives for expansion. Furthermore, since new firms start up small they can also benefit from their lower worker turnover (in absolute terms). In other words, a larger firm needs to use more of its resources to cover the gap after a greater number of exogenously dismissed workers before it can expand.

When controlling for firm age, the group of small firms is “cleansed” of the effect of firm startups and the effects of firm specific productivity developments. For this reason the negative coefficient on firm size loses considerably on statistical significance. However, the effect of smaller firms having to use less resources for covering the gap after exogenously dismissed workers before expanding is still present.

### 4.7.2 Other model predictions related to firm age

In the following paragraphs I document that the model is also consistent with the other empirical findings presented in Section 4.3.
Figure 4.7: Firm shares according to age for small firms, model and data

![Figure 4.7: Firm shares according to age for small firms, model and data](image)

Notes: The figure plots the firm share of each firm age group in the total number of firms conditioning on small firms. Small firms are defined as the smallest 43%, which corresponds to firms with 1 to 4 employees in the data.

Figure 4.8: Firm shares according to size for young firms

![Figure 4.8: Firm shares according to size for young firms](image)

Notes: The figure plots the firm share of each firm size group in the total number of firms conditioning on young firms.
Firm age/size distribution

Even though the calibration does not target firm age shares, the nature of the firm growth process is such that the model predicts that 38% of all firms are young. This is very close to the 41% found in the data.

Furthermore, in reality the majority of young firms are small, while the converse is less true. Figure 4.7 shows the firm age shares in a given group of small firms. The model overpredicts the age shares for the very young at the expense of old firms, but the shares do not die out with age. Hence, it is not the case that small firms are mainly young. Figure 4.8 plots the size shares of new entrants, firms younger than 2 years and firms younger than 6 years. All three groups are predominantly small as in the data.

The results are driven by the fact that young firms are born small and conditional on survival they gradually expand. However, the linear production technology does not restrict older (larger) firms from contracting if their firm specific productivity falters. Thus, while young firms are mainly small, small firms are not always young.

Exit rates, job creation and destruction according to firm age

Exit rates. Figure 4.9 depicts the empirical and model-based exit rates as a function of firm age. The negative relationship between firm exit rates and age is present, but it is weaker, especially after the first year of a firm’s life. The model exaggerates the exit rate of new firms while it underpredicts the probability of shutting down for all the other age groups. This is driven by the selection process of successful firms. The relatively less productive businesses shut down early in their lives leaving only the more efficient ones to grow old. In the model, this process is relatively strong and therefore exit rates fall sharply after the first periods of a firms’ life.

Job creation and destruction rates. Figures 4.10 and 4.11 show the empirical and model-based job creation and destruction rates according to firm age. The model

30 In the data firms with 1 to 4 workers account for about 43%. The model size distribution is not entirely comparable with the empirical one due to the absence of very large firms. Thus, the figure compares the empirical age shares of firms with 1 to 4 employees, to the age shares of the smallest 43% of all firms in the model.

31 Empirical counterparts are not shown because, as mentioned, the model size distribution is not entirely comparable to that in the data.
Figure 4.9: Firm exit rates according to age, model and data

Notes: The figure plots firm exit rates of each firm age group. The exit rate is defined as the fraction of firms shutting down in the total number of firms in a given group.

Figure 4.10: Job creation rates according to age, model and data

Notes: The figure plots job creation rates of each firm age group both for the model and the data.
Figure 4.11: Job destruction rates according to age, model and data

Notes: The figure plots job creation rates of each firm age group both for the mode and the data.

qualitatively captures the declining job creation and destruction rates according to age, but it underpredicts their levels (except for new firms).

The declining nature of job creation and destruction rates with firm age is again driven by the selection process of successful firms. This, in combination with the effect of firm entry results in young firms having higher (net) job creation rates than older businesses. The relatively low job destruction rates of older firms are driven by their low exit rates.

**Job creation and destruction shares.** In the data, 15% of all workers are employed at young firms. At the same time young firms account for 34.3% and 21.5% of all gross job creation and destruction, respectively. In the model, young firms account for 18.5% of employment, 54.5% of all gross job creation and 19.6% of all gross job destruction. Hence, as in the data, young firms play an important role in job creation since they create a disproportionately large amount of (net) jobs compared to their size. The model exaggerates the share of young firms in job creation which is related to the underpredicted level of job creation rates of older firms. Thus, although the model underpredicts the extent of job reallocation (the sum of gross job creation and destruction), it performs well in capturing the importance of young firms in this process.
Cyclicality of firm-level employment

The model is consistent with the procyclical behavior of employment growth rates of young firms (correlation coefficients with respect to the cyclical component of the unemployment rate of $-0.35$). As aggregate productivity increases, the incentives to post vacancies rise and the risk of shutting down falls. Moreover, these incentives are not dwarfed by the costs of replacing exogenously separated workers as young firms are mainly small.

The correlation coefficient of unemployment and the employment growth rate of old firms, however, is virtually zero at $-0.06$. The reason for this is twofold. A large part of old firms are productive and therefore also large. For these firms the additional incentives for hiring brought up by an increase in aggregate productivity are dwarfed by their large costs of replacing exogenously dismissed workers. Those old firms that are relatively small and could benefit more from an increase in aggregate productivity, however, also also relatively unproductive and face a high risk of shutting down. Hence, only if old firms are not too large and are productive enough to expand does a boom provide extra incentives to hire.

Because young firms have higher job creation and destruction rates on average, young businesses are also more volatile than older ones. The standard deviation of the employment growth rate of young firms is 3.5 larger than that of old firms. This is slightly higher than in the data. Moreover, the model is consistent with the negative (albeit a slightly weaker) correlation between the differential employment growth rate of young and old firms with the unemployment rate (correlation coefficient of $-0.32$ compared to $-0.62$ in the data).

Importance of entry

Section 4.3 documented that firm entry is important for aggregate dynamics. One can also zoom in on new firms in the model.\textsuperscript{32} Starting at the steady state Figure 4.12

\textsuperscript{32}The typical way to simulate is to use a large (finite) number of firms. However, the group of new firms constitutes only a small fraction of the total number of firms and thus one needs to be concerned with sampling uncertainty even if the total number of firms is large. To this end I use a non-stochastic cross-section simulation method. This grid technique does not feature cross-sectional sampling uncertainty and is thus suitable for this purpose (details are provided in Appendix 4.C)
Figure 4.12: Decomposition of IRF of the unemployment rate

Notes: The figure plots the IRF of the unemployment rate to a negative one-standard-deviation shock to aggregate productivity. It further decomposes it into contribution of "entry" and a combined contribution of exit, job creation and destruction of continuing firms ("rest").

shows the impulse response function (IRF) of the unemployment rate to a negative one-standard-deviation shock to aggregate productivity. The IRF is then decomposed into the contribution of firm entry and the combined contribution of exit, job creation and destruction of continuing firms ("rest").

Upon impact lower firm entry accounts for almost 40% of the unemployment increase. Moreover, firm entry remains below its steady state level long after the initial hit because of persistently lower expected benefits of starting up a business. Remember that new firms enter based on the expected value of individual firm productivity (not a realization) and they obtain a "business idea" only after entry. During periods of lower aggregate productivity, there is an increase in the chance of the initial draw not being high enough in order for the firm to survive in the market, reducing the incentives for entry.

The combined contribution of exit, and net job creation of continuing firms reverts within a few periods and starts to push unemployment down to its steady state. This is because the larger unemployment pool makes it easier for existing firms to hire workers. Note that potential firms also take this into account as firm value depends on
the aggregate probability of finding a worker. However, this effect is not strong enough to overturn the negative impact on incentives for entry caused by the higher expected exit rates.\footnote{Remember that the initial productivity distribution is wide and therefore changes in aggregate productivity affect a relatively large mass of new firms.}

under the current specification this effect is not strong enough to overturn the above-mentioned channel reducing incentives to entry.

4.7.3 Model predictions related to aggregate variables

In this section, I analyze the implications of the model for aggregate variables. First, I examine impulse response functions to an aggregate productivity shock. The IRFs of the presented heterogeneous firm (HF) model are compared to those obtained from a representative firm (RF) matching model using the same calibrated parameters. Second, I examine the business cycle properties of aggregate labor market variables predicted by the HF model and compare them to those observed in the data.

Impulse response functions

Heterogeneity in firm specific productivity levels together with the endogenous process of firm entry, growth and exit create two new propagation channels that are not present in a standard matching model. In the HF model the number of firms varies procyclically and average firm productivity fluctuates countercyclically. These two opposing effects result in new interesting dynamics as well as greater propagation.

Figure 4.13 shows the impulse responses of unemployment, vacancies, the probability of filling a vacancy and output to a positive one-standard-deviation shock to aggregate productivity generated by both the HF and RF model. The response of output in the HF model is further decomposed into the effect of employment, average firm productivity and aggregate productivity. The middle right panel shows impulse responses of the number of active firms and the cutoff value \( \tilde{p}(z_t, n_0) \), representing the risk of shutting down (these do not have counterparts in the RF model).\footnote{Remember that the cutoff value depends on both aggregate productivity and employment. The IRFs of cutoff values for different firm sizes are qualitatively similar.}

First, the HF model is characterized by greater propagation as the responses of
Figure 4.13: IRFs to an aggregate productivity shock

Notes: The figure plots the IRFs of vacancies, unemployment, the probability of filling a vacancy and output generated by the heterogeneous firm (HF) model and a comparable representative firm (RF) version. The output response of the HF model is decomposed into contributions of the aggregate productivity shock, average firm productivity and employment. It further shows the IRFs of the number of all active firms and the cutoff value $\tilde{p}(z_t, n_0)$ (these do not have counterparts in the RF model).
unemployment, vacancies and the probability \( q \) die out more gradually than in the RF case. The response of unemployment and the probability of filling a vacancy displays a larger drop in the HF model compared to the RF one. This is due to endogenous firm entry and exit. As aggregate productivity improves less firms exit and more firms enter the economy pushing down unemployment further. This directly translates into a stronger decline of the probability of finding a worker as the pool of available unemployed is relatively smaller than in the RF case. This stronger drop in the chances of hiring new workers provides relatively less incentives to post vacancies and thus their response is weaker compared to the RF model.

Second, the response of vacancies does not mimic that of the exogenous shock in the HF case. Instead, it displays a hump-shaped response as is the case in the data.\(^{35}\) The reason for this is the hump-shaped response of the total number of active firms in the economy. As survival rates increase in reaction to a positive aggregate productivity shock the new entering firms thus meet with more and more existing firms which have survived from previous periods. Hence, the total number of firms gradually cumulates before returning to its steady state level.

Third, while the response of output in the RF model gradually declines with aggregate productivity, in the HF model it has intriguing dynamics. After roughly 4 years, output actually increases slightly before returning to its steady state. There are two forces at play. As aggregate productivity increases the risk of shutting down declines. This results in relatively less productive firms being able to stay in the market reducing average firm productivity. At the same time, more firms enter the economy, a large part of which are highly productive. However, since the cumulation of new productive firms takes time, initially the negative effect of higher survival rates of relatively less efficient firms dominates and average firm productivity falls. After exit rates have returned back to their steady states the positive effects of a larger fraction of (highly productive) new firms takes over and average firm productivity rises. This leads to an increase in output which eventually returns back to steady state together with the total number of firms.

\(^{35}\)The response of vacancies mimicking the dynamics of the exogenous shock is counterfactual to the data but stubbornly robust in standard matching models as pointed out by for instance Fujita and Ramey (2006).
Table 4.7: Standard deviations and cross-correlations of selected variables; U.S. data

<table>
<thead>
<tr>
<th></th>
<th>u</th>
<th>v</th>
<th>v/u</th>
<th>n</th>
<th>f</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>standard deviation</td>
<td>0.54</td>
<td>0.37</td>
<td>12.25</td>
<td>0.53</td>
<td>3.18</td>
<td>1.14</td>
</tr>
<tr>
<td>autocorrelation</td>
<td>0.87</td>
<td>0.89</td>
<td>0.92</td>
<td>0.87</td>
<td>0.81</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>u</th>
<th>v</th>
<th>v/u</th>
<th>n</th>
<th>f</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>-0.85</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v/u</td>
<td>-0.88</td>
<td>0.93</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>-0.99</td>
<td>0.85</td>
<td>0.88</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>-0.88</td>
<td>0.77</td>
<td>0.97</td>
<td>0.88</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>-0.81</td>
<td>0.82</td>
<td>0.81</td>
<td>0.81</td>
<td>0.74</td>
<td>1</td>
</tr>
</tbody>
</table>

Notes: The data used are the following: the unemployment rate is taken from the BLS, the vacancy rate is taken from Barnichon (2010), the employment to labor force ratio ("n") is taken from the BLS, "f" is the job finding rate taken from Shimer (2007) and "y" is real GDP published by the BEA. The data are quarterly and run from 1982Q1 to 2007Q1 (the end of the sample is dictated by availability of the job finding rate data). Real GDP is logged and all variables are detrended with an HP filter (smoothing coefficient 1,600). The reported standard deviations are relative to the standard deviation of labor productivity (output per worker in the non-farm business sector).

Business cycle properties

Table 4.7 summarizes the standard deviations and correlations of selected labor market variables in the U.S. economy. Table 4.8 reports the same statistics generated by the HF model. For comparison, Table 4.9 also reports the business cycle statistics generated by the standard matching model assuming a representative firm. This time, however, the RF model is recalibrated to fit the same statistics as the HF model.

The HF model captures the observed volatility and autocorrelations of labor market variables well. The key reason behind the ability to capture the respective volatilities is that the surplus share of the production relationship is small for the firm. This is the result of the calibration of the wage setting rule. A small firm surplus share implies that aggregate productivity shocks have a larger effect on the value of jobs for a firm and thus on the vacancy posting incentives as pointed out in Hornstein, Krusell, and Violante (2005b). As explained in the previous section, firm heterogeneity together with the endogenous process of firm entry and exit generate greater persistence allowing the model to capture well the observed autocorrelation coefficients.

The only exception where the model noticeably underpredicts volatility is labor market tightness (and hence the job finding probability). This is related to the weaker
Table 4.8: Standard deviations and cross-correlations of selected variables; HF model

<table>
<thead>
<tr>
<th></th>
<th>u</th>
<th>v</th>
<th>v/u</th>
<th>n</th>
<th>f</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>standard deviation</td>
<td>0.54</td>
<td>0.30</td>
<td>6.87</td>
<td>0.53</td>
<td>1.39</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.011)</td>
<td>(0.37)</td>
<td>(0.024)</td>
<td>(0.056)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>autocorrelation</td>
<td>0.88</td>
<td>0.72</td>
<td>0.82</td>
<td>0.88</td>
<td>0.83</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.012)</td>
<td>(0.021)</td>
<td>(0.014)</td>
<td>(0.022)</td>
<td>(0.017)</td>
</tr>
</tbody>
</table>

correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>u</th>
<th>v</th>
<th>v/u</th>
<th>n</th>
<th>f</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>-0.55</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v/u</td>
<td>-0.94</td>
<td>0.79</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>-1.00</td>
<td>0.55</td>
<td>0.94</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.079)</td>
<td>(0.039)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>-0.95</td>
<td>0.79</td>
<td>1.00</td>
<td>0.95</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.015)</td>
<td>(0.00)</td>
<td>(0.039)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>-0.91</td>
<td>0.61</td>
<td>0.89</td>
<td>0.91</td>
<td>0.90</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.027)</td>
<td>(0.021)</td>
<td>(0.033)</td>
<td>(0.021)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The heterogeneous firm model was simulated 1,000 times, each simulation lasted for 1,108 quarters where the first 1,000 were dropped to obtain a sample of 27 years as in the data. The reported statistics are averages over the 1,000 simulations (standard deviations across simulations are in brackets). The simulated data were treated in the same way as their empirical counterparts described in the note of Table 4.7.

(negative) correlation between vacancies and unemployment (the Beveridge curve). The reason is that in a boom less firms are forced to shut down and more firms enter, which reduces the pool of unemployment available for hiring. This in turn lowers the probability a given vacancy is filled and thus the incentives to post vacancies are diminished. Similar logic applies to recessions. The model captures the rest of the correlation structure well. Hence, overall the model does a good job in generating realistic aggregate labor market dynamics, both in terms of co-movement and volatility.

The RF model strongly overstates the volatility of vacancies (and thus also labor market tightness and the job finding rate) for reasons explained in the previous section. Moreover, it performs relatively worse in capturing the autocorrelation coefficients compared to the HF model.
Chapter 4

Table 4.9: Standard deviations and cross-correlations of selected variables; RF model

<table>
<thead>
<tr>
<th></th>
<th>u</th>
<th>v</th>
<th>v/u</th>
<th>n</th>
<th>f</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>standard deviation</td>
<td>0.54</td>
<td>1.82</td>
<td>18.44</td>
<td>0.53</td>
<td>4.34</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.054)</td>
<td>(0.520)</td>
<td>(0.014)</td>
<td>(0.165)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>autocorrelation</td>
<td>0.78</td>
<td>0.60</td>
<td>0.74</td>
<td>0.78</td>
<td>0.73</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>(0.050)</td>
<td>(0.082)</td>
<td>(0.062)</td>
<td>(0.050)</td>
<td>(0.062)</td>
<td>(0.064)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>u</th>
<th></th>
<th></th>
<th>v</th>
<th></th>
<th></th>
<th>v/u</th>
<th></th>
<th></th>
<th>n</th>
<th></th>
<th></th>
<th>f</th>
<th></th>
<th></th>
<th>y</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>correlation matrix</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>u</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>-0.56</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v/u</td>
<td>-0.73</td>
<td>0.97</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>-1.00</td>
<td>0.56</td>
<td>0.73</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.086)</td>
<td>(0.063)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>-0.73</td>
<td>0.98</td>
<td>0.99</td>
<td>0.73</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.005)</td>
<td>(0.002)</td>
<td>(0.063)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>-0.71</td>
<td>0.98</td>
<td>0.99</td>
<td>0.71</td>
<td>0.99</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.065)</td>
<td>(0.001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The reported business cycle statistics were generated by the representative firm model. For further details see note of Table 4.8.

4.8 Government policy supporting young firms

The presented framework stresses the importance of firm heterogeneity for aggregate outcomes, especially the dimension of firm age. It thus provides an ideal laboratory in which to analyze questions relating to the role of young firms. Such questions are gaining on importance as unemployment rates remain stubbornly high in many countries affected by the most recent crisis and governments struggle to boost employment growth. For instance, the recent "Startup America" initiative of the White House is partly aimed at supporting young firms and consists of four main points: (i) easing access to capital, (ii) reducing regulatory barriers, (iii) providing mentoring and advice and (iv) tax relief and incentives. In this section I first argue that there is a role for government intervention in the presented model. I then show the effects of a government policy aimed at supporting young firms, similar to the Startup America program.

Under the presented model calibration, the wage setting rule stipulates output

36For more information see http://www.whitehouse.gov/issues/startup-america.
shares for workers and firms that are likely to be far away from the efficient ones dictated by the Hosios condition (in the steady state the firm gets 3% of output). In such a case, the economy would be characterized by "underinvestment" resulting in too little firm entry and too much firm exit.

The above suggests that there is a role for government intervention in the form of supporting business operations. For instance, subsidizing the operational costs $\xi$, which does not show up in the wage setting rule, raises firm profits, increases the resources for hiring new workers and lowers the risk of a subsidized firm shutting down. The majority of firm selection takes place early on in a firm’s life (young firms exhibit high exit rates) justifying the consideration of a subsidy for young firms.

Specifically, I assume that firms pay only a fraction $1 - \tau$ of their operational cost, where $\tau$ is the government subsidy which is assumed to decay monotonically with the firm’s age. I consider two different implementations of the government policy. First, only firm entry is subsidized leaving the behavior of existing firms unchanged. Second, only firms older than one quarter are subsidized with the subsidy dying out monotonically until the firm reaches 3 or 5 years of age. For comparability, in all cases the total amount spent on the subsidy is equal to 1% of steady state aggregate output in the case of no subsidy.

Table 4.10 summarizes the results, where the reported values are relative to the case with no government intervention. In the case when only firm entry is subsidized (first column) the government policy leads to higher output (the reported values for output are net of the costs of the subsidy) and thus increases welfare. With more firms entering the economy unemployment decreases and average firm productivity rises as many of the new firms are highly productive. At the same time, however, the average exit rate increases, since a larger fraction of new firms comes also with relatively more firm closures.

Under the second scenario (second and third column) only existing firms are subsidized. For both considered durations of the subsidy unemployment declines, but

---

37 Further details and issues related to the technical implementation of the subsidy in the solution of the model are in Appendix 4.D.

38 One does not need to be concerned by fluctuations when examining the welfare in the presented model, since all agents are assumed to be risk neutral.
output actually falls. The reason behind this lower output level is a fall in average firm productivity. The subsidy on existing firms lowers the exit rates of active firms enabling the relatively less productive ones to remain in business. Unemployment falls as less workers lose their jobs due to firm closures lowering the probability of filling a vacancy. This reduces incentives for entry and less businesses startup.\(^{39}\) Lower firm entry comes with a relatively lower mass of highly productive firms. Thus, the combination of higher survival rates of relatively less productive businesses and lower firm entry results in a fall in average firm productivity.

Moreover, firm closures release workers from relatively unproductive firms making them available to be hired by more productive businesses. This reallocation process is less efficient when the government subsidizes existing firms. Such an adverse effect is not present when subsidizing only firm entry as there is no firm selection at that stage of a firm’s life. Only upon entry do firms obtain their initial productivity draws.

The model thus predicts that the most important part of policies such as the Startup America program is getting rid of barriers to entry. Once new firms can easily startup, government support should be swiftly withdrawn as relatively inefficient incumbent firms crowd out potentially highly productive entrants and the worker reallocation process is disrupted.

\(^{39}\) Note that new firms also take into account the increased survival rates of older firms. This effect dampens the decline in firm entry, but does not overturn it.
4.9 Conclusion

Recent research has shown that it is not small firms that grow faster and are important for job creation in the aggregate economy, but rather young businesses. This questions the current way of thinking that has treated firm size as an important determinant of business growth. This paper turns its attention to firm age as the driver of firm growth and examines the role of young businesses in determining aggregate labor market dynamics.

It has been documented that young businesses have higher exit rates and grow faster than older ones, job creation and destruction rates fall with firm age and that young firms create relatively more (net) jobs. I extend these findings by documenting that young businesses are also more volatile than older ones and that firm entry is important for unemployment rate developments.

To further understand these relationships, I build a novel general equilibrium model with labor market frictions and heterogeneous firms that is consistent with the above facts and produces realistic dynamics of aggregate labor market variables. Firm heterogeneity together with the endogenous process of business start-ups, firm growth and firm closures create a new propagation mechanism resulting in greater persistence and more complex dynamics than otherwise present in a standard matching model. The presented framework also enables us to study policy questions related to firm age. The results suggests that governments should mainly focus on easing barriers to entry. Supporting existing firms disrupts the selection process of successful firms leading to lower average firm productivity and in turn lower output.

Although this framework performs well in terms of capturing the firm level characteristics present in the data, and provides us with a laboratory to study new policy questions, some aspects of the model still deserve more attention. In particular it would be especially interesting to allow workers to search for new jobs while being employed or to let them quit. This would then alter the worker reallocation process, since in the presented model workers can change employers only following an exogenous dismissal, or when their current employer shuts down.

Furthermore, the model indicates that the number of business start-ups is a crucial
aspect to overall job creation. Moreover, the policy exercise documented that new firms
are key in not only creating jobs, but also increasing productivity. Further analysis
of the entry decision supported by empirical evidence related to characteristics of new
firms could sharpen the policy implications of the model.

Finally, there is a potentially important issue concerning young firms that has been
left out completely in this paper. Namely, the importance of financial frictions for
young businesses. As a new firm, it can be much harder to obtain external funding,
for instance because entrants cannot prove themselves with a successful credit history.
Incorporating financial frictions could strengthen the results of the presented model,
as the firms that are responsible for a large part of job creation in the economy could
also be those that are hit hardest by financial frictions. Large propagation effects could
arise as the worsened conditions for obtaining external funds during recessions would
further discourage firm entry. These and other extensions, however, are left for future
research.
4.A Krusell-Smith algorithm - details and accuracy checks

In this section I provide details on the exact specification of the aggregate law of motion of the aggregate vacancy filling probability, $H$. Also, details of the simulation, stopping criteria and accuracy checks are shown.

4.A.1 Details on solution algorithm and simulation

The KS algorithm consists of the following iterative procedure

- 1. select $Q_t$
- 2. guess a functional form for $H$ and values of its coefficients
- 3. given $H$ solve the individual firm problem
- 4. given the decision rules from step 3, simulate the economy and obtain a simulated time series of the vacancy filling probability, $q_{sim}$
- 5. use the simulated time-series from step 4 and estimate parameters ($H_{new}$) of the aggregate law of motion for $q$
- 6. compare $H_{new}$ and $H$. If the parameters are not close to each other, update the guess of $H$ using $H_{new}$ and go back to step 3. If the parameters have converged, but the goodness of fit is low, increase the number of lags of $z$ and/or the number of lags of $q$ in $Q_t$ and go back to step 3. Alternatively, assume a different functional form for $H$ and go back to step 3. If the parameters have converged and the goodness of fit is high, stop

As mentioned in the main text, the aggregate law of motion is summarized as $q_t = H(q_{t-1}, z_t, Q_t)$, where $Q_t$ is a collection of (past) aggregate productivity shocks and/or moments of the idiosyncratic productivity distribution and further lags of the aggregate probability of filling a vacancy. The choice of $Q_t$ is a balance between a parsimonious specification and accuracy.
Chapter 4

In the current setup $H$ is assumed to be log linear and $Q_t = (z_{t-1}, z_{t-2}, q_{t-2}, q_{t-3})$.

The reason why lagged values of the aggregate productivity shock are useful for predicting aggregate labor market tightness (and thus the aggregate probability of filling a vacancy) is that in the end it is aggregate productivity that drives fluctuations in the distribution of individual productivities. Alternatively, one could also use moments of the firm productivity distributions, such as the mass of firms at certain quantiles, the dispersion etc. However, these moments will necessarily be arbitrary choices and they will themselves become state variables that will likely depend on lagged values of aggregate productivity.

Given coefficient values in the aggregate law of motion, $H$, the individual firm problem can be solved and the model can be simulated. The simulation is done with 200,000 firms (increasing the number of firms does little to the results). The model is then simulated for 5,000 periods, where the first 1,000 are dropped. The remaining 4,000 periods are used to update the coefficients in $H$.

The stopping rule is based on the maximum absolute percentage difference between the coefficients in $H$ used to solve the firm problem and those that come out of the regression using simulated time-series. The stopping criterion is $10^{-6}$. Finally, new coefficients are updated with a dampening factor of 0.25. This means that the new coefficients in $H$ are a weighted average of regression coefficients and the coefficients from the previous iteration, where the weight on the regression coefficients is 0.25.

### 4.A.2 Accuracy checks

Figure 4.14 shows an accuracy plot which compares the simulated time path of $q$ and the simulate path based on the aggregate law of motion. Note that the true simulated values of $q$ are not used as an input in the aggregate law of motion. The two are very close to each other.

Looking at the average percentage difference between the two time series shows only a very small average difference of 0.04%. However, this could still hide large difference. Therefore, I also consider the maximum percentage difference, which in this case is 0.26%. This value occurs during a very sharp drop in aggregate productivity (a fall from 0.02 to $-0.005$ which is a change corresponding to roughly 1.5 times the standard
Notes: The figure plots the simulated probability of filling a vacancy ("actual") and the ones based on the aggregate law of motion used in by the firms in their maximization problems ("aggregate law of motion"). This series does not use at any point the true simulated value of $q$.

deviation of $z$). Excluding this extreme drop the maximum percentage difference falls to 0.14% a very low value.

Finally, the impulse responses of aggregate variables based on using the simulated value for $q$ and that based on the aggregate law of motion are virtually identical to each other.\textsuperscript{40} The above evidence suggests that the aggregate law of motion for $q$ does a very good job at approximating the actual (simulated) behavior of $q$.

\textbf{4.B An approximate maximization problem}

As is documented in the previous section, the dimensionality of the individual firm maximization problem is quite high due to the KS algorithm. Apart from the "typical" state variables (individual firm productivity, firm employment and aggregate productivity) firms also take into account 5 additional state variables. The latter help predict the value of the aggregate vacancy filling probability needed in the firms’ maximization problem.

\textsuperscript{40}Results upon request.
In this section I compare two solutions to the individual firm problem presented in this paper. For convenience I replicate the equations of the maximization problem below. An individual firm maximizes its value subject to the law of motion for firm employment

\[ V^F(z_t, p_{i,t}, n_{i,t}) = \max_{n_{i,t+1}, v_{i,t}} \left[ y_{i,t}(n_{i,t}) - w_{i,t}(n_{i,t}) - \xi - \kappa(v_{i,t}) + \beta(1 - \delta) \max \{ 0; E_t V^F(z_{t+1}, p_{i,t+1}, n_{i,t+1}) \} \right] \]

s.t.

\[ n_{i,t+1} = (1 - \rho_x)(n_{i,t} + q_{i,t} v_{i,t}). \] (4.16)

In the main text the firms do not actually maximize (4.15), but rather an approximate objective function of the following form

\[ \tilde{V}^F(z_t, p_{i,t}, n_{i,t}) = \max_{n_{i,t+1}, v_{i,t}} \left[ y_{i,t}(n) - w_{i,t}(n) - \xi - \kappa(v_{i,t}) + \beta(1 - \delta) E_t \tilde{V}^F(z_{t+1}, p_{i,t+1}, n_{i,t+1}) \right]. \] (4.17)

As mentioned in the main text, the policy function based on the perturbation solution to the approximate maximization problem will have a lower mean. However, as long as the dynamics of the firm decision rules are not affected, the level difference poses no issues, since the calibration targets realistic average values of firm closures and aggregate vacancies.

The goal of this section is to document how close are the solutions to the approximate and the true maximization problems in terms of dynamics. In order not to run into the curse of dimensionality, which is the original reason to use perturbation techniques, I consider a simplified version of the aggregate law of motion for \( q \). Namely, \( H \) only consists of only the current aggregate productivity state. All other parameter values are as in the main text. Note that this is a valid procedure, since I am interested only in the accuracy of the individual firm problem and I am not solving for the general
equilibrium.

A discrete approximation was made for the exogenous Markov processes for individual and aggregate productivity. While individual productivity was approximated using 100 grid points, only 20 grid points were used for aggregate productivity. A higher number of grid points for individual productivity ensures enough accuracy for the firm exit decision. Furthermore, a grid of 250 points was created for firm level employment with the maximum employment level being 4 times the average firm size. Sensitivity analysis indicated that 250 grid points for employment were enough to ensure that the discreteness of the grid does not affect the results. Finally, values implied by the perturbation solution were used as a starting point in the value function iteration.

Figures 4.6 and 4.6 show the demeaned simulated paths of the individual firm productivity cutoffs (exit decisions) and vacancy posting decisions for a given realization of the aggregate productivity shock.\footnote{For the value function iteration solution linear interpolation between grid points was used.} The two are very similar. For simplicity, in both cases, firm employment is held fixed at its steady state level. Repeating the exercise for a different value of firm employment yields similar results. The figures show that the solution to the approximate maximization problem follows the true one very well suggesting that the proposed procedure is a valid one for the problem at hand.

4.C Non-stochastic cross-section simulation method

Instead of simulating a large number of firms characterized by their productivity and employment levels \((p, n)\), this method works with the \emph{mass} of firms at grid points representing productivity and employment levels \(p\) and \(n\). A first step is thus to create a fine grid for both individual firm productivity and employment levels. Denote the number of productivity grid points \(N_p\) and the number of employment grid points \(N_n\).

Each grid point \([p_i, n_j]\), where \(i \in [1, 2, \ldots, N_p]\) and \(j \in [1, 2, \ldots, N_n]\), thus gives the mass of firms with firm specific productivity \(p_i\) and employment level \(n_j\). The simulation can be described by the following iterative scheme

- 1. at each grid point firms decide whether or not to exit (based on past values of aggregate and firm specific conditions).
• 2. new firms enter the economy based on the entry condition in (4.9) with employment \( n_0 \) and a draw from the firm specific productivity distribution \( G_0 \).

• 3. update aggregate productivity according to its law of motion in (4.5).

• 4. distribute all the mass at each grid point \([p_i, n_j]\) to the grid points \([p_k, n_j]\), where \( k \in [1, 2, ..., N_p] \), according to the individual firm productivity law of motion in (4.6).

• 5. transfer all the mass at each grid point \([p_i, n_j]\) to the grid point \([p_i, n_j']\), where \( n_j' = (1 - \rho_x)n_j + v(p_i)q_t \), with \( v(p_i) \) being the vacancy posting policy function. Go back to step 1.

The simulations in the current paper use 400 grid points for productivity and 400 grid points for employment. The grid points for (log) firm productivity run between \(-1.1\) to \(1.5\) and for the employment they are between \(0\) and \(100\). Widening the ranges did not change the results. If the employment choice falls outside the grid range, the respective mass is assigned to the appropriate corner grid point.

4.D Implementation of government subsidy for young firms - details

In this section I describe the technical implementation of the government subsidy to young firms. As mentioned in the main text, the government subsidy decays over time. This character of the subsidy is dictated by the solution method, since one needs to express it in recursive form. Let the initial subsidy a firm that has survived the first quarter be \( \tau_0 \). Then, the subsidy for older firms is assumed to be give by

\[
\tau_j = \rho_{\tau} \tau_{j-1},
\]

(4.18)

where \( j \) indicates the age of a firm in quarters. \( \rho_{\tau} \) is determined such that firms of a given age (in the main text I consider 3 and 5 years) obtain 5% of \( \tau_0 \).
Chapter 5

Conclusion

This thesis examines the costs of business cycles and the way they are shaped by frictions on the labor market. A better understanding of the ease (or difficulty) with which unemployed people find jobs or the way different firms behave over the business cycle can help us evaluate alternative policy measures aimed at boosting employment or smoothing out the negative effects of business fluctuations. These are the issues that have gained in importance after the most recent crisis as unemployment rates remain high in many countries and governments struggle to increase job creation. In the paragraphs that follow, I discuss the main findings of this thesis and possible avenues for future research.

Chapter 2 deals with the question of why business cycles could be detrimental for the economy. Even though there is an underlying feeling that economic fluctuations lower the wellbeing of society, economic theory has struggled to show that. If recessions are followed by times of increased economic activity, it is not clear why the economy should be worse off overall, compared to a world without business cycles. Together with my advisor, Wouter den Haan, we present a framework, in which the presence of business cycles results in permanent losses of output. The reason is an interaction of costs to entry and inefficiencies that force some productive projects to shut down. With inefficiencies in continuation and startup decisions, the presence of business cycles reduces the expected duration (and benefits) of some projects that would otherwise be able to operate in a world without economic fluctuations. Combined with costs to entry, this means that for some of these projects it no longer pays to startup. In other words, the interaction between entry costs and inefficiencies in our model results in a
permanent drop in the level of output in the presence of business cycles. Our estimates show that this permanent drop is in the range of 2 percent and in some cases much higher, which is several orders of magnitude larger than is typically found in other studies.

The model presented in Chapter 2 is based on the assumption that entrepreneurs correctly understand the working of the economy. Specifically, it is assumed that they know how often booms and recession occur allowing them to correctly evaluate the expected benefits of starting a business. However, in reality it is more likely that entrepreneurs do not have such information and instead they need to learn it or form beliefs about it. One can imagine that after a severe recession, such as the one we have just experienced, entrepreneurs form relatively more pessimistic beliefs about the frequency, or severity, of recessions. This would in turn lower the expected benefits of their projects. The result would be that recessions are followed by periods of protracted lower firm entry, only because beliefs about future economic conditions turn relatively more pessimistic after downturns. This would result in even more severe business cycle costs than predicted by our model.

Having established that business cycles can be detrimental for the economy, Chapter 3 examines the role of the severity of search and matching frictions, "match efficiency", on the unemployment rate over the business cycle. Using data from the U.S. economy and employing econometric techniques, I show that match efficiency is estimated to be moving together with the cycle. This means that recessions are times when unemployed workers have a harder time finding a job, not only because there are less vacancies and more unemployed competing for them, but also because the severity of frictions on the labor market is larger. Estimates suggest that up to 1/5 of the unemployment run-up during the most severe recessions can be attributed to a deterioration in match efficiency alone.

To further analyze the role of match efficiency in determining aggregate labor market dynamics, Chapter 3 also examines a simple theoretical model. I show that measured match efficiency moves together with the business cycle in a model incorporating search and matching frictions, workers with different skills and firms which are free to flexibly hire and dismiss employees.
The theoretical model predicts that taking into account the changes in job requirements of firms and/or the skill characteristics of the unemployed over the business cycle should result in estimates of match efficiency that are less volatile or even close to constant. An interesting extension of the presented empirical work would be to include such measures of skill characteristics of jobs and/or unemployed workers. If the match efficiency estimates would then turn out to be less volatile, it would provide support for the specific channel predicted by the model.

Finally, Chapter 4 studies the role of young businesses in determining aggregate outcomes both on the labor market and in the economy as a whole. Recent studies have shown that the well established negative relationship between a firm’s size and its growth vanishes once the age of firms is taken into account. Furthermore, recent studies show that young firms are mainly small, but not all small firms are young. Rates of job creation and job destruction fall with firm age and young firms create relatively more jobs. Finally, young firms are found to have a higher risk of going out of business, but if they survive they tend to grow faster than older firms. Using the latest version of the Business Dynamics Statistics database, I extend these findings by showing that young firms are also more volatile than older ones and that business startups are important for determining the developments of the unemployment rate.

Then, I build a novel general equilibrium model incorporating labor market frictions and heterogeneous firms which differ in their productivity. Firms in this model are free to enter the economy and if individual business conditions are so low that it no longer pays to operate, they choose to shut down. Furthermore, firms choose whether to expand or shrink their workforce depending on aggregate and firm specific conditions.

This model is shown to be consistent with the above-mentioned empirical facts relating to firm age. The key to understanding the models’ performance is an inherent selection process of successful firms. Only the relatively more efficient firms are able to survive, expand and grow old. Therefore, younger firms exhibit a higher risk of shutting down and thus also a higher rate of job destruction. Moreover, younger firms are mainly small and as such they can take advantage of lower worker turnover and the associated costs (in absolute terms) resulting in them having relatively more resources for expansion compared to older businesses. The model is also consistent with the
dynamics of aggregate labor market variables, such as the unemployment rate, the
vacancy rate and the probability of finding a job. The model correctly replicates the
co-movement of these variables over the business cycle and also predicts realistic sizes
of their fluctuations.

The model is then used to analyze the impact of a government policy aimed at
supporting young firms as drivers of job creation. Such a measure was recently proposed
under the "Startup America" initiative of the White House and this experiment thus
addresses a highly relevant question. The results suggest that subsidizing firm entry
does increase welfare as more new firms lower unemployment and increase aggregate
output. However, if the government focuses its resources only on existing young firms
welfare decreases. The reason is that subsidizing existing firms enables relatively less
productive firms to remain in business crowding out entrants. These two effects result in
lower average firm productivity. Moreover, the reallocation of workers from relatively
less productive firms to more efficient ones is hampered. The overall effect on the
economy is that aggregate output falls and thus welfare decreases.

Although the presented model performs well in capturing the empirical facts re-
lated to firm age and also delivers realistic aggregate labor market dynamics, it can
be extended in several dimensions. Specifically, enabling workers to search for new
jobs while being employed or allowing them to quit would influence the worker reallo-
cation process since in the presented model workers can change employers only after
exogenous separations or if their current employer shuts down. Moreover, consider-
ing a wider range of frictions, such as financial frictions which are likely to have a
greater impact on young firms, could have the potential of strengthening the presented
results. Incorporating financial frictions could magnify the effects of aggregate pro-
ductivity shocks even more as worsened conditions for obtaining external funds during
downturns would further discourage firm entry. Finally, the model indicates that the
number of business start-ups is a crucial aspect to overall job creation. Moreover, the
policy exercise documented that new firms are key in not only creating jobs, but also
increasing productivity. Further analysis of the entry decision supported by empirical
evidence related to characteristics of new firms could sharpen the policy implications
of the model.
We leven in een wereld waar economische fluctuaties, conjunctuurcycli, een gegeven zijn. De meest recente recessie is een duidelijk voorbeeld van hoe dergelijke verschijnselen een ieder van ons benvloeden. De impact van de crisis op de arbeidsmarkt was bijzonder hard. In veel landen, die getroffen zijn door de crisis, blijft de werkloosheid hardnekkig hoog. Het analyseren van de rol die de arbeidsmarkt speelt bij het veroorzaken van economische fluctuaties en de kosten van conjunctuurcycli, zijn precies de onderwerpen waarop ik mij in mijn proefschrift focus.

Hoe is de arbeidsmarkt van invloed op de totale economie? Net zoals vele andere markten, werkt ook de arbeidsmarkt niet vlekkeloos. Het kost werknemers tijd en moeite om een baan te vinden en op eenzelfde manier is het kostbaar voor bedrijven om een geschikte werknemer te vinden. Werknemers verschillen in vaardigheid en opleidingsniveau, terwijl bedrijven gedifferentieerde goederen produceren en zich in verschillende gebieden bevinden. Een ervaren bouwvakker is niet erg bruikbaar voor een ziekenhuis op zoek naar een hersenchirurg. Op eenzelfde manier is een bouwvakker in Frankrijk niet zo handig voor een bouwbedrijf in Italië. Deze "zoek- en matchingfriceties" zijn de reden waarom werklozen en vacatures naast elkaar kunnen bestaan. De wisselwerking tussen deze fricties en de economische omstandigheden bepaalt vervolgens hoe de werkgelegenheid, en daarmee dus ook de productie, zich ontwikkelen over de cyclus.

In het tweede hoofdstuk van mijn proefschrift (gezamenlijk werk is met mijn begeleider Wouter den Haan) stellen we de vraag waarom conjunctuurcycli kostbaar kunnen zijn voor de economie. Hoewel er een breed gedragen gevoel bestaat dat economische fluctuaties een samenleving schaden, is dit moeilijk aan te tonen in economische mod-

*This summary owes its life to Tim, Siert and Mathilde.
ellen. Waarom loont het de moeite voor centrale banken en overheden om te proberen fluctuaties glad te strijken, wanneer slechte tijden gevolgd worden door goede? Om onze gedachtegang duidelijk te maken, vergelijken we de typische economie waarin wij leven, met een hypothetische wereld waarin er geen fluctuaties bestaan. In beide economien is een scala aan projecten beschikbaar (hetzij banen dan wel complete bedrijven) welke productief zijn. Elk van deze projecten verschilt echter in haar productiviteit. Sommige projecten kunnen veel produceren, terwijl de productie van andere veel lager ligt.

Een ander belangrijk ingredint in ons model is dat de beslissing om een project te starten of voort te zetten onderhevig is aan ”inefficiencies”. Ondernemers moeten bijvoorbeeld vaak een beroep doen op externe financiering. Het niet in staat zijn om de vereiste rente te betalen, kan tot een faillissement leiden, zelfs wanneer het project in principe winstgevend genoeg is om de lening in de toekomst terug te kunnen betalen. Tijdens een recessie is het voor een ondernemer moeilijk om de bank ervan te overtuigen dat de zaken beter zullen gaan in de toekomst en dat de lening uiteindelijk zal worden terugbetaald. Met andere woorden: in sommige gevallen zullen winstgevende projecten gedwongen worden te sluiten. Een voorbeeld van een inefficiency is een bank die een ondernemer een winstgevend project niet toevertrouwt, bijvoorbeeld omdat hij geen kredietgeschiedenis heeft, of omdat de bank zijn beloftes simpelweg niet gelooft.

In een wereld zonder fluctuaties zouden projecten die niet productief genoeg zijn om de inefficiencies te overwinnen niet eens starten, aangezien ze onmiddellijk failliet zouden gaan. In een economie met conjunctuurncycli zijn inefficiencies moeilijker te overwinnen in slechte tijden. Recessies zijn peroides waarin bijvoorbeeld de vraag van consumenten laag is en ondernemers minder verdienen, waardoor het voor hen moeilijker is om de reguliere rente aan de bank te voldoen. De aanwezigheid van recessies betekent in feite dat een aantal projecten, die prima zouden opereren in een wereld zonder conjunctuurncycli, nu gedwongen worden te sluiten tijdens recessies. Met andere woorden, de levensverwachting van dergelijke projecten neemt af. Wat dan nog? Aan de ene kant dwingen recessies enkele projecten die kunnen overleven in een wereld zonder fluctuaties te sluiten, maar aan de andere kant stellen expansies projecten die niet zouden starten in een wereld zonder conjunctuurncycli, in staat op te starten aangezien de inefficiencies dan gemakkelijker te overwinnen zijn. Opnieuw
lijken de verschillende krachten elkaar uit te wisselen, waardoor conjunctuurfrequenties niet relevant zijn voor het welzijn van een samenleving.

Echter, tot dusverre heb ik n cruciaal ingrediënt uit ons verhaal achterwege gelaten. In werkelijkheid zijn er kosten verbonden aan het opstarten van een project. Het bouwen van een fabriek kost geld, net zoals het vinden van een geschikte werknemer voor een vacante positie niet kosteloos is. Dit verandert het verhaal dramatisch. Vergeet niet dat sommige projecten die kunnen overleven in een wereld zonder fluctuaties, door recessies gedwongen worden te sluiten, waardoor hun verwachte levensduur afneemt. Gezien het feit dat een ondernemer de verwachte inkomsten van het opstarten van een project zal vergelijken met de verwachte kosten, heeft de lagere levensverwachting voor een aantal van deze projecten fatale gevolgen. Een verkorting van de verwachte levensduur vermindert de verwachte voordelen drastisch. Het maakt nogal uit of je denkt dat je een bedrijf voor de rest van je leven kunt managen, of slechts tot de volgende recessie. Hierdoor zullen sommige projecten niet langer de moeite van het opstarten waard zijn, omdat hun verwachte inkomsten te laag zijn om de opstartkosten te dekken. Deze projecten, die zouden kunnen overleven als er geen conjunctuurfrequenties zouden zijn, worden niet eens opgestart in een wereld met fluctuaties, waardoor de productie permanent geschaad wordt. Onze schattingen suggereren dat deze permanente daling van de productie ruim enkele procenten kan bedragen en in sommige gevallen zelfs enkele tientallen procenten.

In hoofdstuk 3 wordt bekeken hoe makkelijk (of moeilijk) werkloze mensen een baan vinden en hoe het gemak van het vinden van een baan en het werkloosheidspercentage benauwd worden door de efficiëntie waarmee vraag en aanbod op de arbeidsmarkt aan elkaar gekoppeld worden (match-efficiëntie). Op basis van data voor de VS gebruik ik econometrische technieken om te schatten hoe de kans op het vinden van een baan voor een werkloze afhangt van de match-efficiëntie. Hoewel deze niet direct observeerbaar is kan ik met behulp van een econometrische techniek een schatting voor de match-efficiëntie uit de data halen. Uit de schatting blijkt dat de match-efficiëntie meebeweegt met de conjunctuur van een economie. Gedurende recessies is het voor werklozen moeilijker om een baan te vinden, niet alleen omdat er minder banen zijn en meer mensen op zoek zijn naar een baan, maar ook omdat arbeidsfricties die met
het zoekproces samenhangen (zoekfricties en matchingfricties) in tijden van een laag-conjunctuur sterker worden, waardoor de match-efficiëntie omlaag gaat. Kwantitatief gezien wordt ongeveer 25% van de kans van een werkloze om een baan te vinden gedreven door de match-efficiëntie. Vanuit een ander perspectief bekeken: tijdens de zwaarste recessies gaat het werkloosheidspercentage met gemiddeld 3.5 procentpunt omhoog, waarvan 1/5 wordt veroorzaakt door de afnamen van match-efficiëntie tijdens deze recessies.

Om de mogelijke oorzaken van fluctuaties in de gemeten match-efficiëntie beter te begrijpen wordt hierna in Hoofdstuk 3 een simpel theoretisch model, met zoek- en matchingfricties, werknemers met verschillende productiviteitsniveaus en bedrijven die makkelijk werknemers kunnen aannemen en ontslaan, geanalyseerd. In dit model beheert match-efficiëntie net als in de data met de conjunctuur op en neer. De drijvende kracht hierachter in het model is dat tijdens recessies alleen de productievere werknemers hun baan kunnen behouden. De eisen, die bedrijven aan baanzoekende werklozen stellen, gaan tijdens een recessie omhoog, en daarmee wordt de kans op een baan voor deze werklozen kleiner. Immers, een groter percentage van de werkloze populatie voldoet niet aan de eisen voor de vacante posities, waardoor de match-efficiëntie daalt.

Hoofdstuk 4 bestudeert tot slot de rol van nieuwe ondernemingen bij het tot stand komen van geaggreerde totalen op zowel de arbeidsmarkt als in de economie als geheel. Gebruikmakend van de Business Dynamics Statistics (BDS) Database constateer ik 5 empirische bevindingen: (i) jonge bedrijven zijn voornamelijk klein, maar niet alle kleine bedrijven zijn jong, (ii) de mate waarin banen geschapen en vernietigd worden daalt naargelang een bedrijf langer bestaat en jonge bedrijven creëren relatief meer banen dan oudere, (iii) jonge bedrijven hebben een grotere kans om failliet te gaan, maar degenen die dat niet doen, groeien sneller dan oudere bedrijven, (iv) jonge bedrijven zijn volatieler dan oudere en (v) nieuwe bedrijven spelen een belangrijke rol in het verloop van de werkloosheid.

Vervolgens wordt in Hoofdstuk 4 een theoretisch model ontwikkeld met daarin heterogene bedrijven die verschillen in hun productiviteit, wat ons in staat stelt om de rol die de leeftijd van bedrijven speelt binnen een economie beter te begrijpen. Bedrijven kunnen in dit model vrij toetreden tot de markt en als het met een bedrijf zo
slecht gaat dat doorgaan met de productie niets oplevert, dan kiezen bedrijven ervoor om te sluiten. Verder wordt aangenomen dat het aannemen van veel werknemers tegelijk relatief duurder is dan het aannemen van enkele werknemers, waardoor bedrijven die productief genoeg zijn om uit te breiden dit op geleidelijke schaal zullen doen. Dit model levert resultaten op die consistent zijn met de bovengenoemde empirische bevindingen. Essentieel om de uitkomsten van het model te begrijpen is de selectie van succesvolle bedrijven in dit model. Alleen de relatief beter presterende bedrijven blijven operationeel, breiden uit en worden oud. Door dit selectie proces hebben jongere bedrijven een hogere kans om failliet te gaan en dus ook een hogere mate van baanverlies. Bovendien zijn jongere bedrijven voornamelijk klein en kunnen ze dientengevolge profiteren van de lagere kosten die gepaard gaan met aannemen en ontslaan van hun werknemers, wat uitmond in een snellere groei dan bij oudere bedrijven. Het model is eveneens consistent met de dynamiek van geaggregeerde arbeidsmarkt variabelen in de data zoals de werkloosheid, het aantal vacatures en de kans om een baan te vinden.

Hierna wordt het model gebruikt om de impact van overheidsbeleid gericht op het ondersteunen van jonge bedrijven als drijvende kracht achter de creatie van banen te analyseren. Zo’n maatregel is kortgeleden voorgesteld in het "Startup America" initiatief van het Witte Huis. Dit model experiment analyseert dus een uiterst relevante vraag. Het blijkt dat het verlenen van subsidie om nieuwe bedrijven te beginnen, de werkloosheid verlaagt en het nationaal product verhoogt. Het succes van deze maatregel staat of valt echter volledig met de creatie van nieuwe bedrijven. Als reeds bestaande bedrijven subsidie ontvangen gaat het nationaal product juist omlaag, omdat de natuurlijke selectie van succesvolle bedrijven dan verstoord wordt. Hierdoor daalt de gemiddelde productiviteit van bedrijven. De aanbeveling die hieruit volgt is dat een beleid, dat erop gericht is jonge bedrijven op gang te helpen, vooral moet proberen de toetredingsbarrières voor nieuwe bedrijven weg te nemen, maar dat de overheid daarna direct een stap terug moet doen.


Po úvodu v první kapitole se druhá kapitola této disertační práce, která je napsána ve spolupráci s mým vedoucím Wouter den Haanem, snaží odpovědět na otázku, proč jsou výkyvy v ekonomické aktivitě nepříznivé pro blahobyt společnosti. Přestože existuje obecné podvědomí, že hospodářské cykly nejsou pro ekonomiku prospěšné, ekonomická teorie má problémy toto dokázat. Selským rozumem se zdá, že pokud jsou

---

*I'd like to thank my grandfather and father for checking my deteriorating Czech spelling.*
recese následovány konjunkturami, čili obdobími se zvýšenou ekonomickou aktivitou, tak se tyto výkyvy přibližně "vynulují". Proč jsou tedy vlády a centrální banky tolik posedlé "vyhlazováním cyklu"?


Ve světě bez ekonomických výkyvů pak firmy, které nejsou dostatečně produktivní, aby překonal tyto nedokonalosti, nejsou vůbec započaté. Kdyby je někdo nastartoval, tak by byly stejně ihned nuceny zastavit svou činnost. V ekonomice s hospodářskými cykly je těžší překonat tyto nedokonalosti ve špatných časech. Recese jsou období, kdy poptávka je nízká, podnikatelé nemají vysoké tržby a tím pádem mají více problémů se splátkami úvěrů. Přítomnost recesí v podstatě znamená, že některé firmy, které mohou spokojeně vyrábět ve světě bez výkyvů, jsou nyní nuceny zastavit činnost. Jinými slovy, očekávaná životaschopnost, či doba trvání, takovýchto firem je snížena ve světě s ekonomickými cykly.

No a co? Na jedné straně recese donutí uzavřít některé firmy, které mohou vyrábět ve světě bez výkyvů. Na straně druhé ale konjunktury zas umožní výrobu jiným firmám, které ve světě bez cyklů vyrábět nemohou. Čili opět se zdá, že se vše jen "vynuluje" a negativní dopad hospodářských cyklů na blahobyt společnosti je mizivý.

Doposud jsem však tajil poslední, velmi důležitou, součást našeho modelu. Život v ekonomikách, které jsem takto popsal, je až příliš snadný. V realitě založení firmy
stojí peníze. Zohlednění nákladů na započetí projektu však značně změní naší situaci. Vzpomeňte si, že některým firmám, které jsou schopny vyrábět ve světě bez výkyvů, se v přítomnosti hospodářských cyklů výrazně sníží jejich očekávaná životaschopnost. Tato omezená doba trvání pak dramaticky snižuje očekávané přínosy těchto projektů. Je velkým rozdílem, pokud si myslíte, že svou firmu budete moci provozovat do konce svého života, nebo pokud jste přesvědčení, že přežijete jen do přístí recess. A skutečně některé projekty už nebudou stát za to je zahájit, poněvadž jejich přínosy budou natolik nízké, že nepokryjí počáteční náklady. Tyto projekty, které jsou schopny vyrábět ve světě bez výkyvů, ale v přítomnosti hospodářských cyklů nejsou vůbec započaty, pak trvale snižují úroveň HDP.

Naše odhady naznačují, že tento trvalý pokles v úrovní HDP je v rozmezí 0,2 až 2,1 procentních bodů. V některých případech však může být i mnohem vyšší dosahující až 30-ti procent HDP. Tato čísla jsou řádově vyšší než jaká se nacházejí v ostatních studiích, kde se pohybují pod úrovní 0,1% HDP.

Třetí kapitola se zabývá závažností frikcí na trhu práce, "efektivitou hledání", a jaký mají dopad na nezaměstnanost a tím pádem také jak ovlivňují průběh hospodářského cyklu. Příkladem může být pozoruhodný vývoj na trhu práce v USA během poslední ekonomické krize, který ekonomům zamotal hlavu. Stejně jako v předchozích recessích, míra nezaměstnanosti vzrostla a počet pracovních míst poklesl. Po běžné recessi se počet volných pracovních míst začne zvyšovat a tím stlačí míru nezaměstnanosti, protože si lidé snáze hledají práci. To se však po nedávné krizi nestalo. Počet volných pracovních míst sice narostl, ale míra nezaměstnanosti zůstala tvrdohlavě vysoká. Jinými slovy, přestože se v ekonomice nacházel mnoho volných pracovních míst a mnoho nezaměstnaných, z nějakého důvodu se tato pracovní místa nezaplnila. Jednín z možných důvodů vysvětlující tento vývoj je zhoršení frikcí, efektivity hledání. Příkladem může být velké množství pracovníků ve stavebnictví, kterí ztratili zaměstnání. Zároveň se otevřelo mnoho volných pracovních míst ve zdravotnictví a školství, ale samozřejmě tato pracovní místa nebyla vhodná pro nezaměstnané stavbaře. Dalším důvodem mohly být problémy na trhu nemovitostí. V období propadu na tomto trhu bylo relativně težší se stěhovat za prací, protože přefinancování hypotéky bylo mnohem obtížnější než dříve. Toto jsou příklady, proč
se může efektivita hledání měnit během hospodářských cyklů a jaký může mít dopad na zbytek ekonomiky.

Ve třetí kapitole se zaměřuji na lehkost (či obtížnost), s jakou si nezaměstnaní lidé jsou schopni nalézt práci a jaký vliv na to má efektivita hledání, či závažnost frikcí na pracovním trhu. Na datech USA s pomocí ekonometrických metod odhaduji jak je pravděpodobnost, s jakou si nezaměstnaní jsou schopni nalézt práci, ovlivněna efektivitou hledání. Ačkoliv není efektivita hledání přímo pozorovatelná, použitá ekonometrická metoda je přesto schopna odhadnout její vliv. Odhady ukazují, že se efektivita hledání pohybuje spolu s hospodářským cyklem. To znamená, že recessi jsou období, kdy nezaměstnaní lidé si hůře hledají práci nejen proto, že je méně pracovních míst a více nezaměstnaných, které o ně soutěží, ale také proto, že frikce na trhu práce jsou závažnější. Pohyby v pravděpodobnosti, se kterou si nezaměstnaní nacházejí práci jsou zhruba z jedné čtvrtiny důsledkem výkyvů v efektivitě hledání. Řečeno jinými slovy, v obdobích recessi míra nezaměstnanosti průměrně stoupla o 3.5 procentních bodů. Zhruba jednu pětinu tohoto nárůstu má na starosti pokles v efektivitě hledání.

Za účelem dalšího pochopení zdrojů, které možnou stát za výkyvy v efektivitě hledání, používám ve třetí kapitole rovněž teoretický model. Tento model zahrnuje frikce na trhu práce, pracovníky s odlišnými schopnostmi a firmy, které mohou pružně najímat a propouštět pracovníky. V tomto jednoduchém modelu lze rovněž pozorovat, jak se efektivita hledání pohybuje s cyklem. Důvodem je, že jen ti relativně schopnější pracovníci si mohou udržet zaměstnání v obdobích recessi. Jinými slovy, období recessi jsou časy, kdy firmy mají vyšší nároky na schopnosti svých zaměstnanců a lidí, které najímat. To má pak za následek pokles v průměrné pravděpodobnosti nalezení si práce. Čili, stejně jako v příkladu uvedeném dříve, v ekonomice se nachází mnoho volných pracovních míst a mnoho nezaměstnaných. Avšak to, že firmy požadují v recesích lepší schopnosti od lidí, které najímat, znamená, že relativně větší část nezaměstnaných je pro tato volná pracovní místa nevhodná. To se pak projevuje v poklesu efektivity hledání.

Nakonec čtvrtá kapitola se zabývá otázkou jaký vliv mají mladé firmy na průběh hospodářských cyklů. Poslední ekonomická krize jasně ukázala, že recessi mají na různé firmy odlišné dopady. Zaměstnanost v malých a mladých firmách byla obzvláště
zasážena během posledního ekonomického propadu. Počet pracovních míst v malých a mladých firmách v USA poklesl o 10,4%, respektive 14,8% zatímco ve velkých a starších podnicích tento pokles byl 7,5%, respektive 2,8%. Čtvrtá kapitola se snaží odpovědět na otázku, jak důležité jsou mladé podniky pro celou ekonomiku a jakou roli může hrát stát při podporování mladých firem za účelem zvýšení celkové zaměstnanosti v ekonomice.

Z dat databáze "Business Dynamics Statistics" vyplývá pět empirických faktů: (1) mladé firmy jsou většinou malé, ale ne všechny malé firmy jsou mladé, (2) míry tvorby a destrukce pracovních míst klesají s věkem firem a mladé firmy vytvářejí relativně více pracovních míst, (3) mladé podniky vykazují vyšší riziko odchodu z trhu, ale pokud přežijí, tak rostou rychleji než starší firmy, (4) mladé firmy jsou více volatilní než staré a (5) nové firmy jsou důležitým faktorem určujícím vývoj míry nezaměstnanosti.

Čtvrtá kapitola pak buduje teoretický model s heterogenními firmami, které se liší ve své produktivitě, címně se nám otevřou dveře k podrobnější analýze role mladých firem na celkovou ekonomiku. Model rovněž představuje ideální laboratoř, ve které lze zkoumat dopady vládních zásahů podporujících mladé firmy. Podobná politika byla nedávno zavedena Bílým Domem v rámci programu "Startup America" a tudíž analýza těchto jevů představuje vysoce relevantní otázku. V uvedeném modelu mohou firmy vstupovat na trh a pokud jsou jejich individuální podmínky natolik špatné, že už nestojí za to dále firmu provozovat, mohou i z trhu odejít. Zároveň se předpokládá, že najímání většího množství pracovníků je spojeno s rostoucími náklady, címně pádem růstu produktivních firem je postupný.

Tento model je pak schopen replikovat empirická fakta ustanovená v předchozí části. Klíčovým důvodem tohoto úspěchu je přirozený výběr úspěšných firem. Jen ty relativně produktivní firmy jsou schopny přežít, růst a stárnout. Mladé firmy tím pádem vykazují vyšší rizika odchodu z trhu a také vyšší míry destrukce pracovních míst. Navíc většina mladých firem je také malá, což jim umožňuje využít nižší náklady obratu zaměstnanců a mladé firmy tím pádem i rostou rychleji než starší podniky. Model je navíc také schopen správně předpovědět společný pohyb agregátních veličin na trhu práce, jakými jsou například míra nezaměstnanosti, volných pracovních míst, nebo pravděpodobnost, se kterou si nezaměstnaní lidé jsou schopní najít práci.
Shrnutí

Tento model pak slouží jako laboratoř, ve které zkoumáme dopad vládní politiky, která podporuje mladé podniky za účelem zvýšení zaměstnanosti a tím i HDP. Podobné návrhy se objevily v nedávném programu Bílého Domu "Startup America". Model ukazuje, že podobná politika je skutečně úspěšná co se snížení nezaměstnanosti a navýšení HDP týče. Celý úspěch těchto opatření však stojí na bedrech nových firem. Podpora existujících podniků naopak vede k poklesu HDP, poněvadž dochází k narušení přirozeného procesu výběru úspěšných firem, čímž se snižuje průměrná produktivita firem. Podobné vládní politiky by se tedy měly zaměřit především na podporu přístupu nových firem k trhu, ale pak by jejich pomoc měla rychle odeznít.
Bibliography


ity,” unpublished manuscript, University of Mannheim.

Market,” IZA discussion paper no. 5452.

105, 211–248.

the New Keynesian Model with Search Frictions,” Journal of Monetary Economics,
54(3), 706–727.


the Welfare Effects of Eliminating Business Cycles,” Review of Economic Dynamics,
12, 393–404.

Krusell, P., and A. Smith (1998a): “Income and Wealth Heterogeneity in the

Krusell, P., and A. A. Smith, Jr. (1998b): “Income and Wealth Heterogeneity in

at the Industry Level: The Real Effects to Financial Liberalization,” Journal of
Development Economics, 89, 210–222.

Loayza, N. V., R. Ranciere, L. Serven, and J. Ventura (2007): “Macroe-
conomic Volatility and Welfare in Developing Countries: An Introduction,” World


93, 1–14.

Martin, P., and C. A. Rogers (2000): “Long-Term Growth and Short-Term Eco-
nomic Instability,” European Economic Review, 44, 359–381.


