University of Zurich[UZH]

**LEVERAGE, BUSINESS CYCLES AND HUMAN CAPITAL ACCUMULATION**

Dissertation submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktor der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Luca Mazzone
Italy

approved in July 2020 at the request of

Prof. Dr. Felix Kübler
Prof. Dr. Dirk Krueger
Prof. Dr. Simon Scheidegger

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zürich, 15.07.2020

The Chairman of the Doctoral Board: Prof. Dr. Steven Ongena

# ACKNOWLEDGMENTS

I wish to express all my gratitude to my advisor, Felix Kübler for his time, unique feedback and support over the years. My gratitude goes as well to the other members of my committee. To Dirk Krueger, for having been an authentic mentor from day one - he represents all the good things I can say about our profession. And finally, to Simon Scheidegger, for pushing me beyond what I thought was wizardry. He has been a guide when needed, a teacher when required, and always a friend. Furthermore, I want to thank Karl Schmedders, Jesús Fernández-Villaverde, José Víctor Ríos Rull and Harold Cole for fruitful conversations on my work. I also express my gratitude to all the participants to the Advances in Computation in Economics and Finance Seminar and to the Brown Bag Lunch Seminar in Zürich, to the Macro Lunch and to the Wharton PhD Student Seminar at Penn. The intellectual stimulation and the quality of feedback have been invaluable.

A special thanks goes to my coauthors. To Marc, for the shared work and the patience bearing that coauthoring a job market paper involves. To Edoardo and Andrea, for bringing me on board and then again to Edoardo, for enduring the pandemic roommate experience with me and making it full of good memories.

I wish to express my gratitude for people who provided support, pushed me to do better, and were there when I needed advice. Johannes Brumm and Daniel Harenberg, as long as they have been in town. My officemates in Plattenstrasse, Adriano and Marlon. And my friends around Switzerland involved in the same adventure, Luciano and Virginia.

I also want to spend a few words for saying thank you to everyone involved in my experience visiting the University of Pennsylvania. Moving to Philadelphia gave me new enthusiasm, another place to call home, and shaped the economist I am now. In particular, I am grateful for having met André, Jason, Marco, Margaux, Paolo, Sara, and Stefano. There is a future where Jihed is still open for us, and John is ready with our coffee.

Finally, I owe amounts of gratitude that are too large to be included to people in my private life that are not strictly related to the work included here. Each of them deserves a separate, and private, acknowledgment.

During these years I was alternatively supported by the Swiss Finance Institute; the University of Zürich and the Swiss National Science Foundation.I want to express my gratitude for their investment and trust, for which I will always be honored.

# ABSTRACT

LEVERAGE, BUSINESS CYCLES AND HUMAN CAPITAL ACCUMULATION

Luca Mazzone

This thesis collects three essays that deal with important methodological, theoretical or empirical issues within macroeconomics. The importance of distributional channels for the aggregate dynamics of economies with heterogeneous agents (first chapter); the role of financing frictions and initial conditions in shaping human capital and wealth accumulation decisions during the life cycle (second chapter); the interplay of labor market rigidities and business cycles that leave some workers behind in the job ladder.

The first chapter introduces a global solution method for the computation of infinite-horizon, heterogeneous agent macroeconomic models with aggregate uncertainty. Details of the algorithm are illustrated by presenting its application to a an example model: in it, aggregate dynamics depends explicitly on firm entry and exit, and individual choices are often constrained by a form of market incompleteness. The proposed strategy thus combines adaptive sparse grids with a cross-sectional density approximation, and introduces a framework for solving the more general class of dynamic models with firm or household heterogeneity accurately.

The second chapter studies the impact of student loans on post baccalaureate choices, using within-college variations in financial aid policies. We find that higher levels of debt induce a front loading of earnings, an underinvestment in human capital and an anticipation in home ownership. We then estimate a life-cycle model using a representative panel of college graduates and analyze the mechanisms behind the interaction between student debt, career choices and housing. Our results indicate that lower net wealth generates a trade-off between career and housing choices for college graduates.

The third chapter describes a structural model of the labor market that features worker and firm heterogeneity, where workers accumulate human capital and can search on the job. We analyze the optimal provision of insurance within the firm through an optimal dynamic contract. In particular, we show that limited liability on the firm side generates downward wage rigidity. In addition,aggregate fluctuations alter the sorting between workers and firms and distort incentives to accumulate human capital.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# Chapter 1

# ON THE SOLUTION OF HIGH-DIMENSIONAL MACRO MODELS WITH DISTRIBUTIONAL CHANNELS

By Luca Mazzone †

## 1.1 Introduction

In this paper I propose a global solution method for solving infinite-horizon, discrete time, heterogeneous agent models with aggregate uncertainty. I illustrate how it works by presenting a model economy which, in its relative simplicity, contains numerous computational challenges that make most existing methods inapplicable or unfeasible.

The framework I explore builds on decades of work on dynamic models, which only recently have been adapted to account explicitly for heterogeneity. As Yellen 2016 stated, "*it is important for policymakers to understand and monitor the effects of macroeconomic developments on different groups within society*", and yet " *various linkages between heterogeneity and aggregate demand are not yet well understood, either empirically or theoretically*". In fact, despite considerable amount of effort aimed at incorporating agent heterogeneity and distributional effects into macroeconomics, their inclusion in the standard policy toolbox is far from widespread. A relevant obstacle is the computation of equilibria in models where heterogeneity plays a crucial role, whether the heterogeneity is *ex ante* (agents' characteristics, like time discounting, differ) or *ex post* (agents' characteristics are identical, but idiosyncratic shocks differ). I will concentrate on the second class of models. Such economies typically contain two sources of risk: agents are subject to idiosyncratic shocks as well as some form of aggregate uncertainty, and risk can be non-diversifiable. It is known that economies where all idiosyncratic uncertainty cancels out

†University of Zürich and Swiss Finance Institute.

in the aggregate are characterized by a steady state cross-sectional distribution which can be computed fairly easily. With aggregate uncertainty, however, a steady state wealth distribution will not exist in general. Also, as the effects do not cancel out in the aggregate, so the cross-sectional distribution changes with the stochastic aggregate shock. As a consequence, the set of state variables has to include the time-varying cross-sectional distribution of agents' characteristics. In models with a continuum of agents, the object that characterizes the cross-sectional distribution becomes infinite-dimensional, and thus intractable.

The development of computational methods in economics of the 1990s (see Judd 1998 and Ríos-Rull 1997) has led to an increased interest in the computation of equilibrium. Contemporaneous increased availability of computational tools and resources allowed solution procedures to exploit parallelization - see Dongarra and Steen 2012.

A seminal contribution on the solution of distributional models with aggregate uncertainty came by Krusell and Smith 1998, who proposed a simple and versatile solution method. Their model is based on Aiyagari 1994: ex-ante identical agents face non-insurable unemployment risk and have to decide over consumption and savings. The presence of non-insurable risk implies individual histories of shocks of all agents are relevant for determining current choices: their distribution affects prices, hence the cross-section of choices enters the individual decision problem.

In practice, however, in the model economy of Aiyagari 1994 only a rather small fraction of agents is constrained, and this small fraction of agents hold an even smaller fraction of aggregate wealth. Therefore, the aggregate behavior of the economy depends almost entirely on the actions of non-constrained agents, and it can be summarized by the first moment of the distribution - a property known as approximate aggregation[1]. This feature drives the main result of their paper, that is the possibility to use the mean of individual choices as a sufficient statistic to predict future aggregates. When approximate aggregation holds, the aggregate law of motion influences the distribution of individual choices, but the solution of each individual problem can rely on aggregates as a set of sufficient statistics for the whole distribution. As Krusell and Smith 1998 themselves argue in their article, a more involved interplay between the cross-sectional distribution of agents' choices and individual policies requires different solution methods.

Algan, Allais, and Den Haan 2008 propose an algorithm that is aimed specifically at solving models where distributional channels have an important role in individual choices: their global solution method uses a smooth cross-sectional density approximation to compute agents' expectation of future aggregates. However, this approach has seen limited use,

[1] A note of caution: approximate aggregation does not imply that the economy can be equivalently described by a representative agent model, nor that there will be zero dispersion among agents.

mostly because it requires a full discretization of the aggregate state space that becomes rapidly unfeasible as the dimensionality of the state space increases - a classic issue known as the curse of dimensionality[2].

To address these issues with the help of a practical example, I introduce a model economy where aggregate dynamics depends explicitly on firm entry and exit, and individual choices are often constrained by a form of market incompleteness. A simpler version of the model has been proposed by arellano2012financial in order to study the effects of uncertainty shocks on the business cycle. The baseline model includes heterogeneous firms whose choices are characterized by a timing structure that forces them to choose labor with a period delay before its use in production. Timing structure makes uncertainty on individual productivity relevant for profit maximization[3]. Although market incompleteness does distort firm choices, entry and exit play no explicit role in the model, as firms default exogenously and the measure of active firms is restricted to be constant over time. In my own extension, idiosyncratic risk generates revenue risk that cannot be hedged and might force the firm to default depending on their past choices, so that exit as well as entry are endogenous, and they are not restricted to cancel each other. Time-varying net entry means the measure of active firms is itself time-varying. Default risk, in turn, implies that firms will become more cautious when uncertainty increases, generating a potential aggregate first-order effect from shocks to the second moment of the distribution of productivity.

In this framework, a departure from approximate aggregation is more likely, as every agent is affected by default risk with the intensity of this distortion depending on its past choices, so that higher order moments of the distribution of individual choices will influence individual policies. In presence of such micro-level dynamics, the strategy in Krusell and Smith 1998 is expected to fail in approximating the aggregate law of motion accurately, and global solution methods can be required; however, given the multidimensional nature of the aggregate state space, those methods would be computationally burdensome. The computational challenge is twofold: the aggregate law of motion has to be determined at many points in the state space, and each point involves an expensive computation.

The problem of approximating an aggregate law of motion can be seen from the perspective of the interpolation of a high-dimensional function: if the function is sufficiently smooth, sparse grids perform this task with small losses in terms of accuracy (see Bungartz and Griebel 2004); if this is not the case, sparse grids are of little help. Brumm and Scheidegger 2017 is the first application of using sparse grids with another layer of

[2] see Bellman 1961

[3] A similar timing structure can be traced back to the capital investment technology in kydland1982time

sparsity to economic problems: their solution algorithm "learns" which areas of state space need a finer grid by relying on so-called adaptive sparse grids (see Pflüger 2012 ). I follow their approach and show how the use of adaptive sparse grids makes it feasible to compute global solutions of heterogeneous agents models even with a reasonably high-dimensional state space.

The present paper combines adaptive sparse grids with a global solution method for economies populated by a continuum of agents, and proposes a general framework for solving dynamic models with firm or household heterogeneity. This approach manages to obtain accurate solutions in settings where common methods have limited application: non-linear or even discontinuous individual policies, big aggregate shocks, multidimensional state spaces, and relevant distributional channels. It also alleviates the curse-of-dimensionality problem discussed above. However, even if adaptive sparse grids reduce significantly the number of points to be evaluated, obtaining a solution in a reasonable time can still be an impossible task. I discuss how my approach can be implemented on a HPC cluster, and describe a suitable parallelization scheme that allows a fast computation of the solution. As the main computation tasks in my algorithm are fully independent of each other, they can be solved in parallel by distributing them via Message Passing Interface among different computer nodes (see Dalcín, Paz, and Storti 2005 and Dalcín et al. 2008 ). The model is solved on the Swiss National Supercomputing Centre (CSCS) HPC Cluster "Piz Daint", using up to 72 MPI ranks. I also show that my strategy is very scalable and adaptable to various modification of the model.

I compare the relative performance of my algorithm to an adapted version of the Krusell and Smith 1998 method, after illustrating different suitable measures of accuracy for heterogeneous agent models with aggregate uncertainty, and their application to the example economy of my paper.

The proposed method is successful in solving for the model equilibrium, using various accuracy measures based on training sets over the whole state space, or on simulation outcomes. The latter case is particularly relevant, because it allows a fair comparison with the accuracy results of the competing approaches; in particular, Krusell and Smith 1998 fails in obtaining equally accurate approximations of the aggregate law of motion. Also, by analyzing the equilibrium solutions of the model, interesting dynamics that could not be observed without a global solution method emerge. Specifically, higher uncertainty induces a misallocation of labor caused by a real option effect on firms that are constrained by default risk; at the same time, net entry can increase because of a combination of general equilibrium effects and increased dispersion of productivity which make high productivity projects even more promising. Uncertainty thus creates two "option" effects that go in different directions, and so the sign of business cycles responses to a second-moment shock can vary depending not only on being paired with other shocks, but also on their duration.

The method proposed in this paper opens the way for analyzing models where the economy is hit by shocks that combine first moment and second moment shifts, and in general for giving a more precise definition of business uncertainty, policy uncertainty, and so forth. Future work on solution methods can instead improve the performance of the algorithm discussed here, in particular by combining adaptive approaches with strategies that concentrate computational efforts on areas of the state space that are part of the ergodic set.

## 1.2 Related Literature

The present work is mostly related to two strands of literature: models that aims at exploring connections between macroeconomic aggregates and idiosyncratic risk, and papers that aim at providing useful solution methods for models of the mentioned type.

In particular, this paper relates to the literature that analyzes the role of firm heterogeneity in the business cycle, as in Khan and Thomas 2008 and Jermann and Quadrini 2012. A more general survey of the inclusion of heterogeneity in macroeconomics is provided by Guvenen 2011 and Krueger, Mitman, and Perri 2016. Firm heterogeneity is a natural environment for analyzing the effects of aggregate shocks, which in turn points to recent efforts in understanding the effects of changes in uncertainty. In a seminal paper, Bloom 2009 reports that uncertainty shocks produce a sharp but temporary decline in aggregate output and employment. This occurs because higher uncertainty causes firms to temporarily pause their investment and hiring . Once macroeconomic relevance of policy and business uncertainty is recognized, a natural step forward is to model explicitly the impact of shocks to higher moments of the processes driving the economy, as in Bloom et al. 2013, in order to understand which channels are more important and why.

The model in this paper includes uncertainty shocks as a source of aggregate risk, and draws mostly from the economy in Arellano, Bai, and Kehoe 2016. It differs in making default depending on the idiosyncratic state of the firm, rather than on some exogenous process, and entry to depend on current business cycle conditions, rather than calibrated in order to have constant measure of firms. Besides the methodological relevance, such an explicit dynamics is interesting in light of the role of net firm entry in the business cycle. Evidence in, among others, Davis, Faberman, and Haltiwanger 2012, points to a greater sensitivity of small businesses to aggregate conditions and a fundamental role in entry and exit in the business cycle: in bad times, small firms tend to contract employment more and to default more than big firms. Also, Haltiwanger 2012 shows for the US that, between 1980 and 2009, 18% of gross job creation is accounted for by new firms, while 19% of gross job creation by the opening of new establishments of existing firms. In terms of gross job destruction, 17% is accounted for by firm exit, 14% by the closing of establishments of existing firms. Those numbers, especially relative to firm exit, are obviously even higher during recessions.

The main contribution of the paper lies in the literature whose aim is to provide suitable solution methods to macroeconomic models with heterogeneous agents and aggregate uncertainty. As discussed above, this features makes it difficult to approximate equilibria, because history-dependent policies in economies with a continuum of agents generate an infinite-dimensional state space in individual agents' problem. Investigating the conditions under which equilibrium exists in these economies has been the subject of an extensive theoretical literature - see Miao 2006, Cheridito and Sagredo 2016 and Brumm, Kryczka, and Kubler 2017.

Turning to the problem of computation, it is possible to identify three main methodological approaches. The seminal approach proposed by Krusell and Smith 1998 is based on a combination of projection and simulation techniques[4]. Their algorithm guesses an aggregate law of motion and simulates a time series of the economy, then re-estimates the law of motion itself from the simulated series using least squares, and proceeds until the set of coefficients has converged following some criteria. I describe the application of this strategy to my model economy in detail in Appendix B. On the same vein, an algorithm based on a combination of projection and simulation is proposed by Judd, Maliar, and Maliar 2011, while Young 2010 shows that the same approach can be used more efficiently by replacing simulation with a histogram-based non-stochastic algorithm. Another strand of literature puts together projection techniques with flexible functional forms to approximate the cross-sectional distribution of agents - see Den Haan and Rendahl 2010 and Algan, Allais, and Den Haan 2008.

Finally, a third approach relies on linear perturbations around a steady state, as it happens in most macroeconomic applications without heterogeneity - the software package DYNARE has long been an important tool for economists working on dynamic models (see Adjemian et al. 2011). In heterogeneous agent models, solution methods on this vein have been advanced in this context by Reiter 2009 and, more recently, by Winberry et al. 2016 .

A comparison of the methodologies listed above has been performed by Algan et al. 2013, and by Terry 2015 using respectively the Aiyagari economy and the Kahn and Thomas economy as benchmark models. Testing the performance of competing methodologies in terms of accuracy and speed, Terry finds the algorithm by Krusell and Smith to be superior to its competitors. The result is to be expected, given that in both

[4]By "projection", here, I refer to the notion that aggregate variables enter explicitly into agents' individual problems, which are computed on different points of the (aggregate) state space: in the case of Krusell and Smith, the points of the state space are the outcome of a simulation.

models aggregate dynamics is driven at large by unconstrained agents. More recently, Pröhl 2017 shows this class of models can be solved with a solution method that obtains the fully rational equilibrium depending on the whole cross-sectional distribution, using polynomial chaos expansions. A relevant finding that follows from this approach is that idiosyncratic risk does not aggregate in equilibrium, and approximate aggregation does not hold anymore.

Finally, the method proposed in this paper relies on the use of adaptive sparse grids. Krueger and Kubler 2004 noticed that, in problems that are sufficiently well behaved, one can address the curse of dimensionality by using sparse grids. For mappings that can be represented as vector functions with bounded second order derivatives, i.e. for models that imply a sufficiently smooth law of motion, classical sparse grids define an optimal (a priori) selection of grid points. However, many economic applications have features such that these prerequisites are not met, either at the individual level or at the aggregate.

Brumm and Scheidegger 2017 show that classical sparse grids are not sufficient to obtain an accurate approximation of the state space using two examples (an international real business cycle model with irreversible investment and a firm price-setting problem with menu costs), and propose to use adaptive sparse grids instead. In their algorithm, the dynamic programming problem of the agent is affected by the *curse of dimensionality* because it has to take into account of the action of each other agent in equilibrium, so that adaptive sparse grids are used as a tool to perform higher-dimensional dynamic programming.

## 1.3 From Classic to Adaptive Sparse Grids

Obtaining the solution to a dynamic programming problem, the aggregate law of motion of a model economy, or even the coefficients of a statistical model over some training set, can be seen as performing the interpolation of a function, say $f : \Omega \to \mathbb{R}$ that is known only algorithmically. In those cases, even if $f$ is not known, it can be evaluated at points in its domain using a numerical procedure. A discretization of the domain of interest $\Omega$ to obtain those points is trivial and obviously not problematic for lower dimensional problems.

In the simplest case, the first step for building an interpolant $u$ of $f$ is to construct a Cartesian grid $G$ over $\Omega$ - i.e., a grid constituted of equidistant grid points with mesh width $h_n = 2^{-n}$ - then evaluate $f$ at $G$. A piece-wise linear interpolant can then be obtained as the weighted sum of piecewise linear basis functions $\varphi_i(x)$:

$$f(x) \approx u(x) = \sum_i \alpha_i \varphi_i(x)$$

The coefficients $\alpha_i$ will be obtained as solutions to a (linear) system of equations generated by function evaluations at $G$. It is easy to see that this approach will face the curse of dimensionality rather soon. Moving to the d-dimensional case, basis functions can be obtained via tensor products, so that:

$$\varphi_i(x) = \Pi_{j=1}^{d} \varphi_{i_j}(x_j)$$

with $i$ being the index for each dimension. Define $V_n$ as the space of piecewise d-linear functions with mesh width $h_n$, so that $V_n = \bigoplus W$, with $W = \text{span}\{\varphi_i(x)\}$.
The asymptotic interpolation error decay is:

$$\|f(x) - u(x)\| \in \mathcal{O}\left(h_n^2\right) \tag{1.1}$$

The function can thus be satisfactorily approximated if it is sufficiently well behaved, but this requires $\mathcal{O}\left(h_n^{-d}\right) = \mathcal{O}\left(2^{nd}\right)$ function evaluations - and here lies the curse of dimensionality.

A well-known strategy to overcome this issue relies on the hierarchical decomposition of the approximation spaces which allows to build sparse grids. I will illustrate the the univariate case for illustrative purposes, as the same reasoning carries over to the multidimensional case with just heavier notation. Consider a standard hat function:

$$\varphi(x) = \max\{1 - |x|, 0\} \tag{1.2}$$

from which one-dimensional hat basis functions at level $l$ can be derived as

$$\varphi_{l,i}(x) = \varphi\left(2^l x - i\right) \tag{1.3}$$

with $0 < i < 2^l$. The set of hierarchical subspaces is obtained as:

$$W_l = \text{span}\{\varphi_{l,i}(x) : i \in I_l\} \tag{1.4}$$

The space of piecewise linear functions for a given level $l$ can then be expressed as a direct sum of the hierarchical subspaces, i.e. $V_n = \bigoplus_{l \leq n} W_l$. The basic intuition behind sparse grids is to exclude ex ante those subspaces that contribute little to the interpolant as they have many basis functions and small support. The procedure yields the sparse grids space

$$V_n^{(1)} = \bigoplus_{|l|_1 \leq n+d-1} W_l \tag{1.5}$$

Under some regularity conditions (namely, boundedness of mixed second derivatives), the resulting sparse grid is optimal with respect to the $L_2$–norm and the maximum-norm. Different choices of error measurement would lead to different optimal grids. The advantage of sparse grids is evident: while asymptotic error decay is now

$$\|f(x) - u(x)\| \in \mathcal{O}\left(h_n^2 (\log h_n^{-1})^{d-1}\right) \tag{1.6}$$

the number of function evaluations decreases sharply to $\mathcal{O}\left(h_n^{-1}(\log h_n^{-1})^{d-1}\right)$.

If the function of interest exhibits sharp local behavior - discontinuities, kinks, or even steep regions - classic sparse grids will not provide an accurate approximation. Those limitations, however, can be addressed if adaptivity is used: in this case, refinement depends on local error estimation, and suits the problem at hand. In adaptive sparse grids implied by local rules, interpolation is based again on hierarchical piecewise polynomials with local support and varying order. In contrast to classical sparse grids, however, adaptive sparse grids use functions with support restricted to a neighborhood of each point and adopt a iterative refinement strategy. This implies that, as the hierarchy of points in the grid advances to lower levels, points are added based on function evaluations and error estimation. In the context of my example economy, this implies concentrating in areas of the aggregate state space $S_t$ where the aggregate law of motion, call it $\Gamma$, has more curvature or sharper behavior. The solution is then approximated as

$$\Gamma(S_t) \approx \sum_{k=1}^{l} \sum_{i \in I_l} \alpha_{k,i} \varphi_{k,i}(S_t) \tag{1.7}$$

where $\varphi$ are basis functions of arbitrary order evaluated at grid points on the aggregate, high-dimensional, state space. What is important to highlight, here, is that coefficients $\alpha_{k,i}$, called *refinement surpluses*, are generated directly from the refinement of the grid. Different refinement strategies are discussed in the next subsection.

### On refinement strategies

Consider the single-dimensional case, first. The hierarchical structure in sparse grids can be highlighted by ordering all points in the grid $G$ according to an index-to-level map $g(j) : \mathbb{N}_0 \to \mathbb{N}_0$, so that point $x_j$ is associated with level $l = g(j)$. Define level index sets $D^l$ as the indexes of points in level $l$. Cumulative level sets $V^l$ are defined as all points at levels less than or equal to $l$, which implies $V^0 = D^0$, and $V^l = V^{l-1} \bigcup D^l$. Each point $x_j$ is associated to a hierarchical family structure, so that:

$$P_j = \left\{ i \in \mathbb{N}_0 : x_i \text{ is a parent of } x_j \right\} \subset D^{g(j)-1} \tag{1.8}$$

$$O_j = \left\{ i \in \mathbb{N}_0 : x_i \text{ is a child of } x_j \right\} \subset D^{g(j)+1} \tag{1.9}$$

A graphical illustration of the parent-children structure is offered in Figure 1, taken from ma2010adaptive :

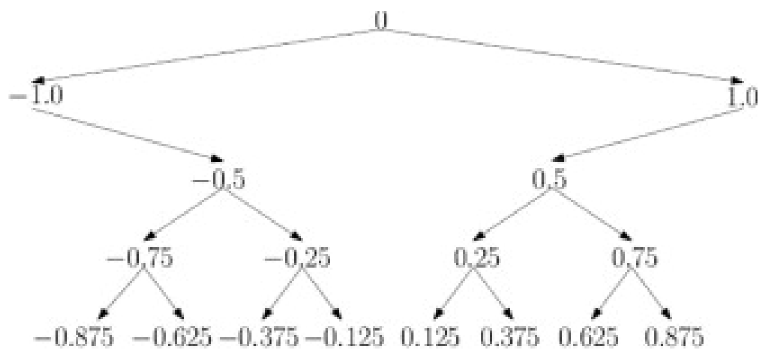


Figure 1.1: Data structure

The first point, at 0, belongs to $D^0$ and obviously has no parents. Point $-0.5$ is in $D^2$, has $-1.0$ as a parent (with $-1.0 \in D^1$) and $\{-0.75, -0.25\}$ as children. Refining the grid under this framework imposes no methodological issues: given basis functions $\left\{\varphi_j\right\}_{j \in \mathbb{N}_0}$, coefficients $\alpha$ for the interpolant in (1.7) can be obtained by matching the value of the function at grid points. This already suggests the refinement strategy: one includes points from the next level of the hierarchy only around points associated with large surplus coefficients. Formally, define the set of large surpluses as:

$$B^l = \left\{ j \in V^l : \frac{|\alpha_j|}{f_{\max}} > \varepsilon \right\} \tag{1.10}$$

for desired tolerance $\varepsilon > 0$. The set of large surpluses yields the indexes of points that are candidates for belonging to the refinement set $R_j^l$. A way to determine the convergence of the interpolation algorithm is to reach a level for which $V^{l+1} = V^l$, or $R_j^{l+1} = \emptyset$. A graphical illustration of the construction of adaptive grids in this context can be found in Appendix C.

In the multidimensional case, however, the situation is not so simple. As Bungartz and Dirnstorfer 2003 point out, refinement in multiple dimensions might lead to unbalanced grids and children nodes being generated with an empty parent sets (so called "orphans"), a feature that can undermine accuracy, or lead to unstable results. Formally, define the set of large surpluses in the multidimensional context as:

$$\mathbf{B}^l = \left\{ \mathbf{j} \in \mathbf{V}^l : \frac{|\alpha_{\mathbf{j}}|}{f_{\max}} > \varepsilon \right\} \tag{1.11}$$

The refinement set $\mathbf{R}_j^l$ is not determined mechanically now. For instance, it is possible that some point $x_j$ is associated with a large coefficient, but that some parents of $x_j$ are not included in the interpolant. Also, the researcher has to decide whether to refine isotropically (i.e., in all directions) or anisotropically (i.e., in a selected set of directions).

The first approach to refinement of adaptive sparse grids (hence the label "classic" refinement) is isotropic, and adds only the children of points with large coefficients. Refinement sets is then defined as:

$$\mathbf{R}_j^l = \bigcup_{\alpha=1}^{d} O_j^\alpha \tag{1.12}$$

The classic refinement approach is typically working fine for most problems. However, adding points in every direction can be inefficient in higher-dimensional problems, increasing the computational burden. In other cases, failure to include the full family of a point can lead to instability around *orphans*.

To deal with the latter issue, the family-selective refinement approach checks whether parents of every point in $\mathbf{B}^l$. If they are not part of the grid, it adds them first, otherwise it adds children. The refinement set is thus:

$$\mathbf{R}_j^l = \left( \bigcup_{\alpha \in \Delta_j^l} P_j^\alpha \right) \bigcup \left( \bigcup_{\alpha \notin \Delta_j^l} O_j^\alpha \right) \tag{1.13}$$

where $\Delta_j^l$ is the set of *orphan* points at level $l$.

On the other hand, the problem with the classic approach can be that too many points per level are added on straight lines like the main axis, the boundaries, and so on. The direction-selective refinement approach aims at limiting this tendency by adding points only on $\Lambda_j^l$, defined as the set of directions associated with a large one-directional coefficient. Hence the refinement set becomes:

$$\mathbf{R}_j^l = \bigcup_{\alpha \in \Lambda_j^l} O_j^\alpha \tag{1.14}$$

As for the classic approach, this refinement strategy can be fragile when failure to add parents leads to instability.

For problems where instability can be a problem but it is limited to a small portion of the function domain, a combination of the advantages of the direction-selective and the family-selective refinement can be desirable. This is the case of the family-direction-selective refinement approach: each direction is considered separately, but children are added only if parents are not missing. Finally, the refinement set results in:

$$\mathbf{R}_j^l = \left( \bigcup_{\alpha \in \Lambda_j^l \cap \Delta_j^l} P_j^\alpha \right) \bigcup \left( \bigcup_{\alpha \notin \Lambda_j^l \setminus \Delta_j^l} O_j^\alpha \right) \tag{1.15}$$

To illustrate how different refinement schemes work, take the function:

$$f(x,y) = \frac{1}{|0.5 - x^4 - y^4| + 0.1} \tag{1.16}$$

with $x, y \in [0,1] \times [0,1]$. Brumm and Scheidegger 2017 show adaptive sparse grids are a suitable interpolation algorithm for $f$, while classic grids and sparse grids end up performing considerable worse[5]. The function is plotted in Figure 2 using points from the last level of the adaptive sparse grid. As refinement proceeds, some areas of the state space become "denser", while other are covered by very few points: computation resources are thus invested where they are more needed to obtain an accurate solution. The final grid, after just seven refinements, is at the bottom right of Figure 3. Tolerance parameter, which determines refinement, is constant and set at $1.E-4$ (more efficient interpolation would obtain allowing the tolerance parameter to adjust for different levels). The grid interpolates the function with a maximum absolute error of $3.63e-03$ on a training set of 1000 points scattered over the state space.

Since refinement is an iterative process (and points are added in "cohorts", from parents to children), the threshold need not be constant across refinement levels. Instead, there is room for adapting the criterion by taking into consideration the specificity of the problem. Clearly, this implies there is a tension between the degree of accuracy one can obtain: too high refinement threshold at the first stages, i.e. a coarse grid at the beginning, could miss important action in some regions of the state space; conversely, a too low refinement threshold, hence a finer grid, could invest too many resources in areas of low curvature and dampen the benefits of the whole strategy. As discussed above, some refinement strategies can loosen the efficiency-accuracy trade-off: the desirability of each strategy depends on the specificity of the problem, with parents-first being preferable for functions with local sharp behavior and family-direction-search over-performing alternatives as dimensionality of the problem grows.

Figure 4 compares the adaptive grid after the sixth refinement level for the same function. Points that would be present both under classic and FDS refinement are

[5] Regarding classic sparse grids, notice that the regularity conditions are not satisfied by $f$: namely, the function is non-differentiable in a segment of its domain

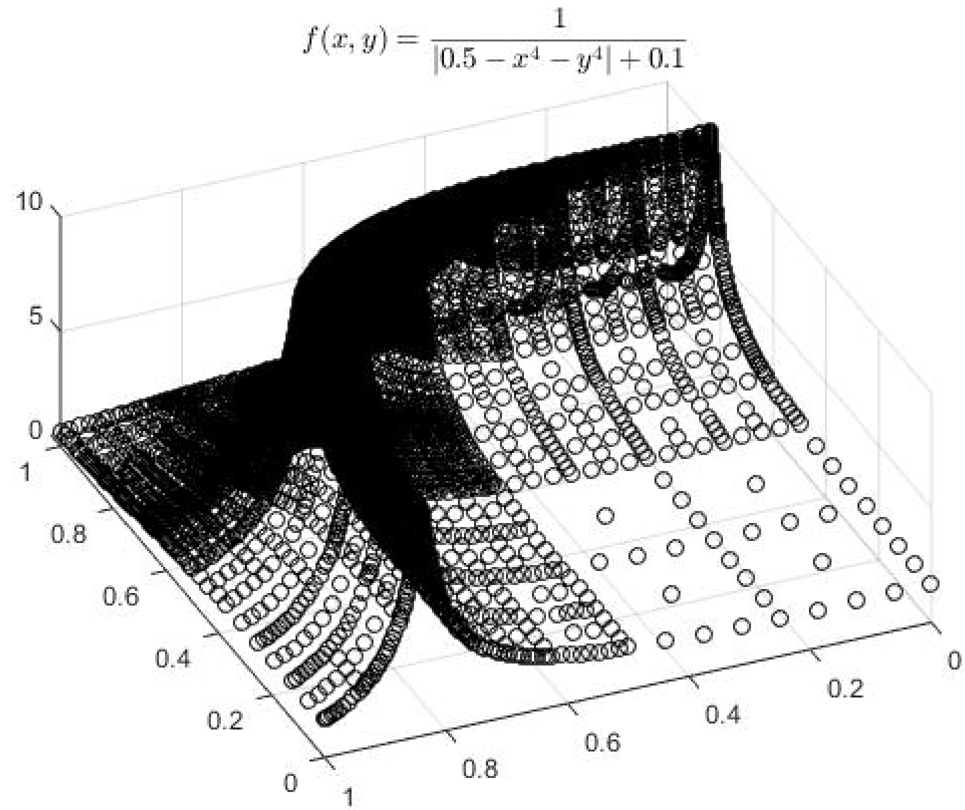



Figure 1.2: The example function

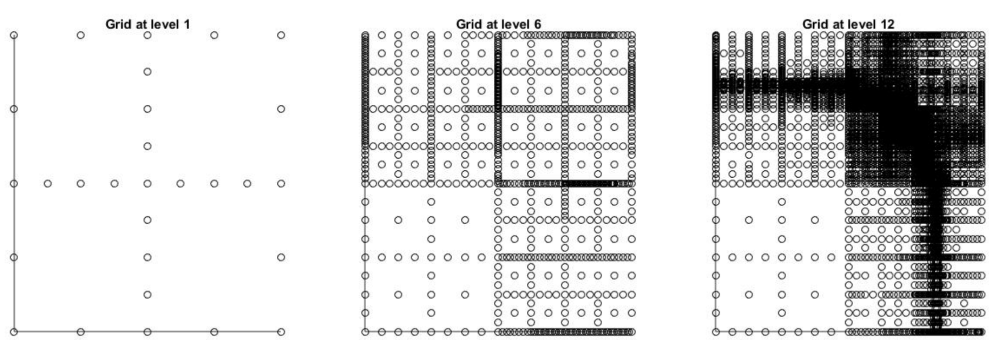



Figure 1.3: Refinement on the example function using FDS

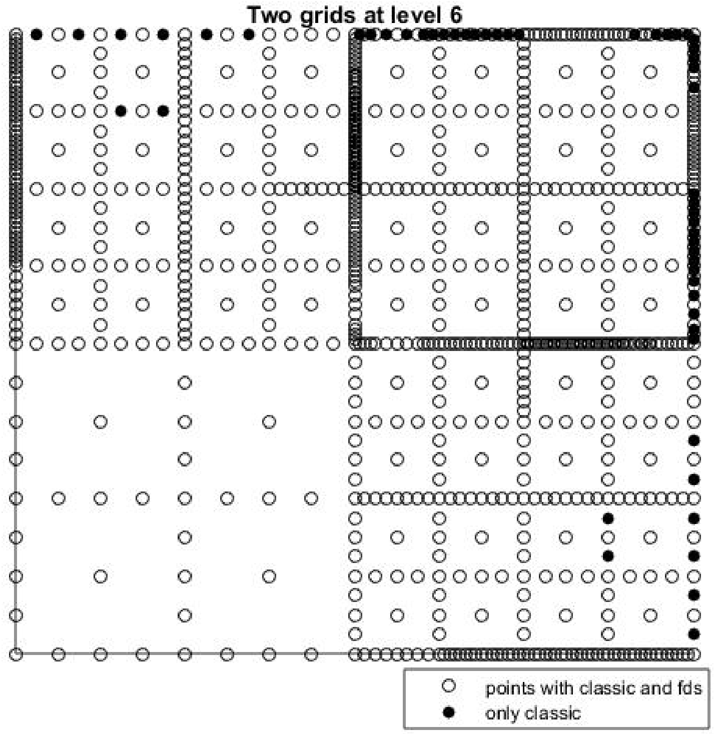


Figure 1.4: Refinement on the example function

represented by empty dots, whereas I use full dots to highlight the position of points that are added only by the classic refinement strategy. Classic refinement starts adding more points, especially at the boundaries, already at such lower refinement levels. Notice that this happens even if, in principle, FDS could end up adding more points because of the "parents-first" component of the strategy.

## 1.4 Example Economy

In order to illustrate the algorithm, I introduce a model whose solution poses a series of computational challenges. For the sake of clarity, the explanation of the algorithm will often refer to components of this economy. Still, the method has broader validity, and I chose the model to which it is applied only because, despite its small size, it contains a good number of interesting methodological features.

Time is indexed by $t \in \mathbb{N}$. The economy is populated by a unit mass of identical households, by a unit mass of identical final firms, and by a unit mass of intermediate firms subject to idiosyncratic demand shocks.

Market incompleteness in the form of non-insurable idiosyncratic risk and aggregate uncertainty in the form of time-varying distribution of idiosyncratic shocks imply that intermediate firms are subject to default (and exit) risk; at the same time, households can finance the creation of new firms.

## Exogenous Processes

The idiosyncratic and aggregate shocks follow two autoregressive processes, given respectively by:

$$\log(z_t) = \mu_t + \rho_z \log(z_{t-1}) + \sigma_{t-1}\varepsilon_t \tag{1.17}$$

$$\log(\sigma_t) = (1-\rho_\sigma)\log(\mu_\sigma) + \rho_\sigma \log(\sigma_{t-1}) + v_t \tag{1.18}$$

where $\varepsilon_t \sim N(0,1)$, and $v_t \sim N(0,\varphi^2)$.

In this application, the two auto-regressive processes are approximated, using the method developed by Tauchen 1986, with Markov chains. The discretization generates $n_s$ states for the aggregate shock ($n_s = 2$ implies there is a high and a low uncertainty state), and $n_z \cdot n_s$ states for the idiosyncratic shocks. Transition probabilities among idiosyncratic states, i.e. probability of $z'$ given $z$, will be denoted as $\pi_z(z'|z)$.

## Households

A unit mass continuum of identical households supplies labor to firms, consumes non-durables produced by final good firms, owns a unit mass distribution of firms. There is a representative household, and we will refer to it as a single agent, and it has flow utility given by:

$$U(C_t, N_t) = \frac{C_t^{1-\eta}}{1-\eta} - \theta \frac{N_t^{1+\chi}}{1+\chi} \tag{1.19}$$

where $C_t$ and $N_t$ are, respectively, aggregate consumption and hours. Define the aggregate state variable $S_t$ as $S_t = (\sigma_{t-1}, \sigma_t, N_t, C_t, Y_t)$ at time $t$, where $\sigma_t$ is the aggregate shock known at time $t$ [6].

Households have access to a saving and borrowing technology, which can be interpreted as a riskless bond, so that the representative household's budget constraint can be written as:

$$C_t + Q(\sigma_t|S_t)B_{t+1} = w_t(S_t)N_t + B_t(\sigma_{t-1}) + D_t(S_t) - T_t(S_t) \tag{1.20}$$

[6] Timing structure is an important feature of this model: the index $t$ refers to the moment in which agents learn about the realization of $\sigma$, but idiosyncratic shock at $t$ depends on the previous, and not the current, realization of it: this way, agents know that whether next periods' idiosyncratic shocks will be more, less or equally volatile than current shocks

where $Q(\sigma_t|S_t)$ is the security price, $B_t$ is aggregate borrowing, $w_t(S_t)$ is the hourly wage, $D_t(S_t)$ are dividends accruing to households from firms, and $T_t(S_t)$ is a lump-sum tax levied by the government to finance a subsidy handed to workers of defaulting firms. Solving the household problem yields the usual stochastic discount factor $Q$, and the wage $w_t$. Finally, consumption is determined by market clearing, using the condition $Y = C + I$, where $I$ denotes the resources used to activate new firms, following the dynamics described below.

## Firms

Perfectly competitive final good firms purchase inputs $y_t$ at prices $p_t(x_t)$ from intermediate good firms, then produce $Y_t$ according to the technology:

$$Y_t = \left( \int z_t y_t(n_t, b_t, z_t)^{\frac{\gamma-1}{\gamma}} d\Upsilon_t(n_t, b_t, z_t) \right)^{\frac{\gamma}{\gamma-1}} \tag{1.21}$$

where $\Upsilon_t$ is the measure of active intermediate firms, indexed by their characteristics: $z_t$ is the idiosyncratic shock to demand of their good, $n_t$ the choice of hours for period $t$, and $b_t$ the amount of debt due in period $t$. Good $y_t$ is produced by monopolistic intermediate firms using technology $y_t = n_t^\alpha$ and sold to final good firms at price $p_t(n_t, b_t, z_t)$, so that demand for intermediate goods is given by:

$$y_t(n_t, b_t, z_t) = \left( \frac{z_t}{p_t(n_t, b_t, z_t)} \right)^\gamma Y_t \tag{1.22}$$

Intermediate goods firms are subject to idiosyncratic demand shocks that can generate temporary losses, which can be covered with non-contingent debt issuance. I abstract from agency problems and assume the manager maximizes the expected value for shareholders. The timing of their problem can be described as follows: at the beginning of time $t$, each firm observes current prices $Q$, $p$ and $w$, the current uncertainty state $\sigma_t$, its past choices $n_t, b_t$, and its demand shock $z_t$. Each firm then has to make its choices on how much labor to purchase (it will be employed and paid in the following period), and how much to borrow, depending on their forecasted prices $p$ and $w$. Risk-neutral financial intermediaries observe firm choices, assess the probability of default of each firm and set the price of credit accordingly.

The main friction comes into play here: firms may - and often will - be subject to a form of debt overhang, as described by Myers 1977, in that past debt choices affects current choices of labor. This, in turn, will make the cross-sectional distribution relevant for equilibrium prices.

The problem of incumbent firms can thus be written as:

$$V(n_t, b_t, z_t, S_t) = \max_{n_{t+1}, b_{t+1}} \kappa d_t + (1-\kappa)\beta \mathbb{E}[Q(\sigma_t|S_t)V(n_{t+1}, b_{t+1}, z_{t+1}, S_{t+1})] \tag{1.23}$$

subject to

$$d_t = p_t n_t^{\alpha} - w_t n_t - b_t + q_t b_{t+1} \geq 0 \tag{1.24}$$

Since firms have a credit line consisting of the (endogenously determined) maximum amount of resources that can be borrowed in each period, they can be forced to default when a very low realization of $z_t$ occurs and there is no policy $\{n_{t+1}, b_{t+1}\}$ such that $d_t \geq 0$ [7].

If funds generated and borrowed in the current period are sufficient to cover existing debt, the firm continues; otherwise, it defaults and pays wages first, then the rest is claimed by debtholders, who get $y_t p_t - w_t n_t$. If wages cannot be paid in full, the proceeds of lump-sum taxes levied on households are used to pay workers. The defaulting firm then exits. The model also has a continuum of potential entrants, characterized by the same demand shock process as active firms. Each firm can be "activated" by paying a fixed cost $\xi$ (the cost is sustained by households), and enters the economy with zero debt. The potential entrant problem is thus:

$$V^e(z_t, S_t) = \max_{n_{t+1}} \kappa \xi + (1-\kappa)\beta \mathbb{E}[Q(\sigma_t|S_t)V(n_{t+1}, 0, z_{t+1}, S_{t+1})] \tag{1.25}$$

Clearly, a firm will be started only if $V^e(z_t, S_t) > 0$.

Modeling choices in the firm problem may seem a bit ad hoc, but they reflect an important feature of business cycles: the constraint on dividends can be seen as if the firm is only able to access debt to finance its obligations, whereas other sources of funding (most importantly, equity) are unavailable. Also, as discussed above, small firms tend to increase their default rate more than big firms during recessions. They are, in general, more sensitive to shocks during bad times, consistently with evidence summarized in Fort et al. 2013. In sum, intermediate good producers in this model can be seen as small firms, subject to a series of frictions in their access to capital and risk-sharing[8].

## Financial Intermediaries

Financial intermediaries operate in a perfectly competitive market, borrowing funds from households and lending to firms with an endogenous discount schedule á la Eaton and

[7]Notice that the most relevant friction in this case is the absence of state-contingent debt, as firms cannot insure against a specific realization of the idiosyncratic shock $z$

[8]More in general, the combination of endogenous entry dynamics, financial frictions that force firms to rely on debt only and production where labor is the most important input point to a model that can be interpreted as an economy of heterogeneous *small* firms.

Gersovitz 1981.

Since competition prevents intermediaries from achieving positive profits, discount rates $q_t(S_t, z_t, n_{t+1}, b_{t+1})$ will simply equal the expected value of the amount that each firm pledges to repay to the intermediary. Borrowers will obtain the discounted amount of the face value of the bond, multiplied for the probability of the firm being active next period, plus the expected value of what is left in the firm in case of default.
Bond rates will then be denoted as:

$$q_t(S_t, z_t, n_{t+1}, b_{t+1})b_{t+1} = \mathbb{E}[\phi_{t+1}Q(\cdot)b_{t+1} + (1-\phi_{t+1})Q(\cdot)\max\{p_{t+1}y_{t+1} - w_{t+1}n_{t+1}, 0\}] \tag{1.26}$$

where $\phi_{t+1} = 1$ indicates the firm is continuing, and $\phi = 0$ indicates the firms is defaulting. Risk-neutral bond pricing from intermediaries implies a time-varying, endogenous borrowing limit for firms, which becomes more stringent in bad times, inducing the typical pro-cyclical borrowing behavior that is seen in models of strategic default. Even if default is not strategic here, it still happens that, in good times, relatively less profitable firms are kept alive by buoyant financial markets; conversely, some relatively more productive firms find themselves *underwater* during downturns.

## Government

The government has a limited role in this economy: it runs a balanced budget, levying lump-sum taxes on all household to finance the difference between payments that defaulting firms owe to workers and the remaining resources within those firms.

## Parameters

Most calibration choices are reflecting commonly used values. Table 2 reports parameter values in three blocks: households, firms and exogenous processes. Labor share $\alpha$ does target the actual labor share in the US, which averages around 60% in US postwar data. Discount rate $\beta$ targets a (riskless) rate of return close to $1,5\%$, risk aversion of 1 means household has log-preferences, and the choice of $\chi$ implies a labor elasticity of 2, which is a standard calibration.

The parameter $\kappa$ is introduced in the firm problem to capture a tension between shareholders and managers as in Jensen 1986: incomplete markets and non-negativity of dividends would give firms an incentive to build up savings, the parameter is instead chosen as summarizing a contractual agreement that induces managers to borrow. Elasticity of substitution between goods $\gamma$ is parametrized in order to give intermediate firms a markup $\frac{\gamma}{\gamma-1}$ (see equation (2.2)) that is around 15%. Entry costs target U.S. capacity utilization in manifacturing, which fluctuates around 75%.

Calibration of the exogenous shocks reflects estimates in Stanfield, Haltiwanger, and Syverson 2008, that are used in Arellano, Bai, and Kehoe 2016 to target moments in the distribution of firms based on Compustat data. Finally, the two autoregressive processes are discretized into a Markov chain with $n_z$ states using the technique proposed by Tauchen 1986 and setting $n_z = 10$.

| Parameter | Symbol | Value |
|---|---|---|
| Discount factor | $\beta$ | 0.985 |
| Risk aversion | $\eta$ | 1 |
| (Inverse) labor elasticity | $\chi$ | 0.5 |
| Labor share | $\alpha$ | 0.6 |
| Jensen effect | $\kappa$ | 0.4 |
| Entry cost | $\xi$ | 1.0 |
| Elasticity of substitution | $\gamma$ | 7.7 |
| persistence of idiosyncratic shock | $\rho_z$ | 0.7 |
| persistence of aggregate shock | $\rho_\sigma$ | 0.75 |
| std of of aggregate shock | $\varphi$ | 0.1 |
| mean of uncertainty process | $\mu_\sigma$ | 0.09 |
| number of idiosyncratic states | $n_z$ | 10 |

Table 1.1: Parameter calibration

## 1.5 Solving the Model

The present model, in its relative simplicity, contains a collection of methodological difficulties. Starting from the firm problem, notice how the presence of default risk induces a discontinuity of the value function in (1.24), as some combination of choices for hours and debt imply $d_t < 0$. The non-negativity constraint then binds, so the firm exits and thus its value function *jumps* to zero. Moreover, solving the firm problem requires knowledge of some mapping between current state space and future aggregates and prices. Obtaining this forecasts is the actual challenge of this model. The firm problem is not standard, but a simple approach for solving it, i.e. iterating on the value function, is still an option once future prices are at hand. Other options, however, are on the table: in Appendix A I discuss how to implement a version of the Howard acceleration based on Ljungqvist and Sargent 2012, that I use in the paper. The next subsection discusses the role of market incompleteness in making it hard to obtain an accurate forecast of future prices.

### Addressing the importance of distributional channels

In principle, firms should also be aware of the full cross-sectional distribution of other firms' choices, which is an infinite-dimensional object, and which determines future aggregates. In fact, to obtain the expected values for $N, Y, C$, i.e. $\hat{C}_{t+1} = E_t(C_{t+1}|\mathscr{F}_t)$, $\hat{Y}_{t+1} = E_t(Y_{t+1}|\mathscr{F}_t)$ and $\hat{N}_{t+1} = E_t(N_{t+1}|\mathscr{F}_t)$, where $\mathscr{F}_t$ is the information set of the firm at time t, one has to calculate an expectation over policy functions that takes into account the whole history of shocks occurred to each agent, in an economy where the number of agent is infinite. Such integral is clearly intractable, representing an obstacle to a solution that involves rational expectations. Different solution strategies are typically characterized by how they deal with this specific issue.

Most of the algorithms in the literature, as discussed above, are extremely successful in obtaining an accurate solution (in some cases, even for models with high-dimensional state spaces, large shocks and discontinuities in the agents' value and policy functions) under particular conditions. The most important is that aggregate dynamics are mainly driven by the fraction of agents whose choice are unconstrained - and thus independent of their own idiosyncratic state or history. In other circumstances, departures from approximate aggregation can be quite severe. To see this in my model economy, take labor choice by firms. In an unconstrained environment firm problem has a closed form solution, and optimality implies that the marginal product of labor equates a constant markup over wage, i.e.:

$$\mathbb{E}\left(p(z_t)\right)\alpha n_t^{\alpha-1,*} = \frac{\gamma}{\gamma-1} w_t \tag{1.27}$$

As pointed out by Arellano, Bai, and Kehoe 2016, when financial markets are incomplete and there is aggregate risk, the firm problem changes. Default risk implies there is some threshold value for the idiosyncratic shock, call it $\hat{z}$, below which the firm with past choices $n_t, b_t$ will always default. So the optimality condition becomes:

$$\mathbb{E}\left(p(z_t)\right) z_t \alpha n_t^{\alpha-1,*} = \frac{\gamma}{\gamma-1}\left[w_t + V \frac{\pi_z(\hat{z})}{1-\Pi(\hat{z})}\frac{d\hat{z}}{dn_t^*}\right] \tag{1.28}$$

where $\Pi(\hat{z})$ is the cumulative distribution function associated with the density $\pi_z$. Labor choices impact the threshold value $\hat{z}$, and through it, the hazard rate $\frac{\pi_z(\hat{z})}{1-\Pi(\hat{z})}$: therefore, all firms' choices are to some extent distorted by market incompleteness, and distributional channels become relevant. Notice that the effect of this distortion of the individual firm does not have necessarily the same sign for all levels of the idiosyncratic shock: indeed, as we will see, firms with high $z_t$ will produce more in a high-uncertainty environment, because of general equilibrium effects. At the same time, inactive firms that draw a higher

$z_t$ enter more often in high uncertainty times, counterbalancing reduced output by existing firms whose shock is low-to-medium.

To address this additional complexity, a strand of literature starting from Den Haan and Marcet 1990, argues for approximating expectations with the use of polynomials; for heterogeneous agents' models, Algan, Allais, and Den Haan 2008 propose a polynomial approach for approximating cross-sectional distribution of agents in a global solution scheme with a flexible functional form. A similar strategy would directly take into account the importance of higher order moments, and thus take into account distributional channels directly, by using an approximation of the cross-sectional density in the forecasting of aggregates.

## Cross-Sectional Density Approximation

Global solution algorithms typically use quadrature methods to integrate individual policies into aggregates. I build on the contribution of Algan, Allais, and Den Haan 2008 to the latter approach, which is to use two sets of moments in the solution algorithm: one characterizes the aggregate state space, and determines the grid over which we have to approximate an aggregate law of motion; the other is composed of "reference" moments, which pin down the cross-sectional distribution through a flexible functional form, but are not used as state variables. The separation into variables that enter directly into agents' problems and reference moments is what makes the approach feasible: if one had to take higher moments into account explicitly, the dimensionality of the problem would quickly explode.

The calculation of reference moments, as the formulation of the functional form for approximated density, implies choices on the solution strategy which depend on the characteristics of the model. In this case, I make the simplifying assumption that the distributions of individual choices in debt and labor are independent of each other. Independence allows to characterize the relevant distribution for obtaining aggregate hours and production in one dimension, which greatly improves the stability and convergence properties of the algorithm. Reference moments and coefficients can be updated in the solution by using simulation outcomes: as in Den Haan and Rendahl 2010, the algorithm starts from the computation of reference moments from a steady state solution with no aggregate uncertainty. Simulation then supplies information on the behavior of the model in the ergodic set. Another option is to adjust the first reference moments using information from the approximate law of motion.

A classical choice for the flexible form is the following:

$$P(n,\rho^{z_i}) = \rho_0^{z_i} exp\left\{\rho_1^{z_i}\left(n - m_1^{z_i}\right) + \rho_2^{z_i}\left[\left(n - m_1^{z_i}\right)^2 - m_2^{z_i}\right]\right\} \times \cdots$$
$$\cdots \times exp\left\{\rho_{n_M}^{z_i}\left[\left(n - m_1^{z_i}\right)^{n_M} - m_{n_M}^{z_i}\right]\right\}$$

where $\left\{m_1^{z_i},...,m_{n_M}^{z_i}\right\}$ are the $n_z n_M$ reference moments, initially calculated from a steady state solution without aggregate uncertainty[9], $n_z$ is the number of points in which the idiosyncratic process is discretized, and $n_M$ is the polynomial degree of the functional form. Once the reference moments are calculated for each aggregate state, one obtains a convex minimizazion problem, in that coefficients $\rho_1^{z_i},...,\rho_{n_M}^{z_i}$ solve

$$\min_{\rho_j^{z_i},j=1,...,n_M} \sum_{i=1}^{n_z} \int_{\underline{n}}^{\bar{n}} P(n,\rho^{z_i})dn \tag{1.29}$$

and the solution immediately follows from finding the zeros of the system of first order conditions of (1.29). Notice that $\rho_0^{z_i}$ is calibrated in order to have $\int_{\underline{n}}^{\bar{n}} P(n,\rho^{z_i})dn = 1$ .

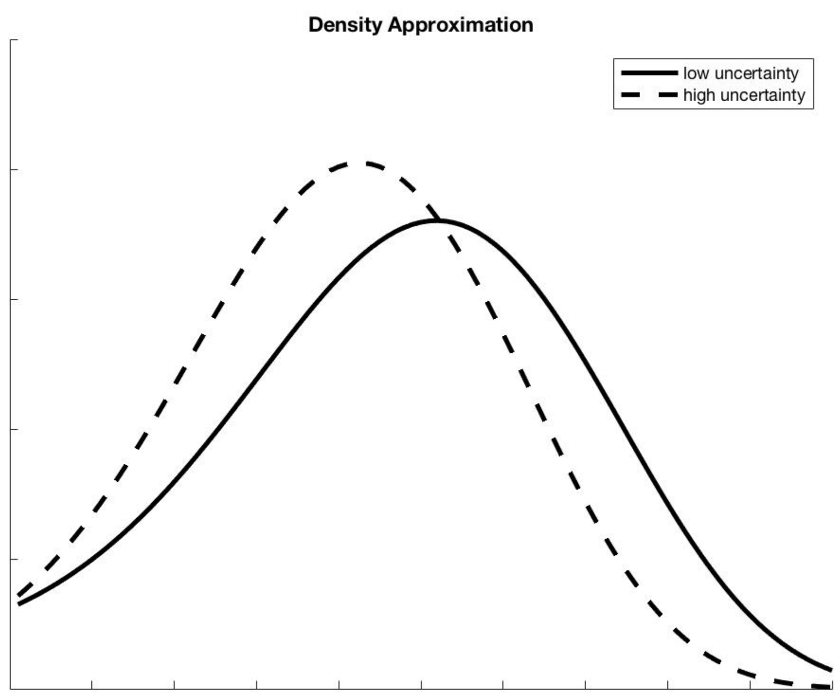



Figure 1.5: $P(n,\rho^z)$, with reference moments and coefficients from steady state

The integral is computed on a given number of Simpson nodes, where the values of individual policies are obtained through interpolation, using splines. Once reference

[9]the steady state solution yields reference moments following the histogram approach of Young 2010

moments and coefficients are at hand, one can obtain aggregate variables using the measure obtained from the flexible functional form on the Simpson integral. This allows to compute aggregate hours and production, as:

$$N_t = \sum_{i=1}^{n_z} \tilde{\pi}_i^z \int n_{t+1}^*(n_t, b_t, z_t, S_t) P(n, \rho_\sigma^{z_i}) dn \tag{1.30}$$

$$Y_t = \left[ \sum_{i=1}^{n_z} \tilde{\pi}_i^z \int z_t y_t(n_t, b_t, z_t)^{\frac{\gamma-1}{\gamma}} P(n, \rho_\sigma^{z_i}) dn \right]^{\frac{\gamma-1}{\gamma}} \tag{1.31}$$

where $\tilde{\pi}_i^z$ is the asymptotic distribution of the Markov chain that discretizes the exogenous process for idiosyncratic shocks. Using flexible functional forms allows another use of simulated data: the one-step-ahead forecast can be used to adjust the first reference moments of the parametrized distributions, in order to follow the time-varying cross section of agents' choices. This, however, doesn't come at no cost: to get the one-step ahead forecast of higher moments, one has to include them explicitly in the law of motion to keep track of.

## Using Adaptive Sparse Grids

Appropriately interpolating on a discretized grid over the aggregate state space would, in principle, yield a sufficiently accurate approximation of the law of motion of the economy. However, a discretization of the full aggregate state space is computationally burdensome even in a relatively small model like the one presented in this work, where the importance local dynamics requires a fine grid in each dimension. As discretization proceeds with adding points, finer Cartesian grids see dimensionality growing exponentially, in a classic display of the *curse of dimensionality* problem.

To understand why this matters here, notice that elements that enter directly into the firm problem at $t$ are three future prices: equilibrium wage, prices of intermediate goods, and the stochastic discount factor. Those are, in turn, determined in equilibrium by three continuous variables (consumption, hours, output) and one discrete variable (aggregate uncertainty). Hence, a minimal representation of the the economy at time $t$ has to include at least these three aggregates. In solving the benchmark model of Aiyagari 1994, where heterogeneous agents face non-insurable unemployment risk, Algan et al. 2013 include the measure of unemployed agents in the set of state space variables. This is done in order to better keep track of changes in the cross-section of agents. Similarily, I include the measure of active firms $\mu_t$. This brings the dimension of the state space to four (continuous) plus two (discrete) states - already problematic if one wants to use Cartesian grids.
The point is illustrated in Table 2. First column reports the dimension; second column

reports the number of points for a full Cartesian grid; third column reports local polynomial grids that constitute the starting point of adaptive sparse grids before refinement; fourth column reports classical (global) sparse grid induced by Lagrange polynomials

| $d$ | $\|V_4\|$ | $\|V_4^L\|$ | $\|V_4^G\|$ |
|---|---|---|---|
| 1 | 17 | 17 | 9 |
| 2 | 289 | 65 | 28 |
| 3 | 4913 | 177 | 84 |
| 4 | 83521 | 400 | 210 |
| 5 | $1419 \cdot 10^3$ | 801 | 462 |
| 10 | $2015 \cdot 10^9$ | 8801 | 8008 |

Table 1.2: Number of gridpoints for different grid types of level 4

In the model economy presented in this work, the main obstacle to an accurate approximation of the law of motion is the relevance of important local sharp behavior in given areas of the aggregate state space. In particular, the behavior of the solution depends crucially on the extent to which there is firm entry or not in that particular point[10] . The algorithm will adaptively add points in regions of the state space with more curvature - this implies the need to choose a refinement threshold below which the approximation is considered sufficient and points are not added. However, as argued above, different refinement strategies are not equivalent, and choosing one against the other is a possible source of stark differences in algorithm performance.
In implementing the solution algorithm, I postulate the existence of a box that contains the endogenous state; sparse grids will be defined in this box $B \subset \mathbb{R}^N$. However, there is no guarantee that resources in approximating the law of motion will be spent in the most relevant areas of the state space - i.e. at values visited by the solution. Solutions will all be inside the box, if it is suitably large, but they could lie in a rather small fraction of its total area. Moreover, even if information about the ergodic set can be learned from simulating the model, the use of this information then depends on the algorithm chosen. In the case of my method, the researcher can update the boundaries of set **B**, and the flexible form for the cross-sectional density approximation can be updated with new reference moments and coefficients. However, unfortunately, the method cannot yet account for irregular geometries. For a discussion on why the geometry of state spaces is relevant, and on how to deal with possible irregular shapes, see Scheidegger and Bilionis 2017.

[10]The reason why entry matters has to do with the presence of mass points induced by the discretization of the idiosyncratic shock process. It is useful to consider the entrant problem in my example economy: optimal hours chosen by entrants depend only on $z_t$ and on forecasted aggregates. Therefore, if $z_t$ can take only finite many values, at every discretized value of the shock corresponds a positive mass point of firms choosing the same employment level, making the cross-sectional approximation less accurate.

## The Algorithm

Before running the main iteration, one needs to obtain the Markov process - again using Tauchen 1986 - to obtain the chain for aggregate shocks to be used for the simulation later. The algorithm will then need to use steady state reference moments and distributions; these can be obtained by solving the model without aggregate uncertainty. I then obtain coefficients by solving (1.29); notice that this can be done by taking the first order condition, and then using a Broyden–Fletcher–Goldfarb–Shanno method. With everything at hand, one can finally start the solution algorithm.

After the algorithm has reached convergence, the model is solved, it can be simulated and used to perform impulse response analysis. However, as discussed above, other useful information on the behavior of the model can be obtained via simulation. In particular, new reference moments and coefficients can be obtained from simulated series. If one wants to follow this path, the solution algorithm adds another (outer) loop on reference moments and coefficients, which stops once it reaches some convergence criterion.

As discussed above, choice of threshold value $\bar{\varepsilon}$ is crucial: it turns out that, in this model, choosing "wrong" values either worsens accuracy or ends up producing unnecessarily dense grids in the first stages of iteration. However, as the starting value is not too low from the beginning, it is possible calibrate it dynamically, by having it adjusting to lower values with higher levels of refinement. Regarding the refinement itself, a reasonable prior would be to assume that family-direction search is the strategy that delivers the best performance. Relevant non-linearities around the endogenous threshold that separates areas of the state space where there is no firm entry and areas with positive firm entry, in fact, could prevent the algorithm from converging because of the instability around *orphan* points discussed in Section 2. I solve the model using classical refinement for robustness, and results confirm this intuition.

**Data:** Level of initial (classical, unrefined) sparse grid, $\tilde{l}$. Initial guess $\Gamma_0(\cdot)$ for the law of motion at every point of initial grid $G_{\tilde{l},0}$. Maximum number of refinement steps $l_{\max}$. Approximation accuracy $\epsilon$ for obtaining guess mapping and threshold value $\bar{\epsilon}$ for hierarchical surplus in adaptive sparse grids interpolation, with $\epsilon > \bar{\epsilon}$. Coefficients $\rho_{z,m}$ and reference moments $m_z$.

**Result:** The (approximate) equilibrium aggregate law of motion $\Gamma_{\tilde{l}}$, the corresponding policy functions $\sigma^*(\cdot,\cdot)$ for every firm type indexed by $\{n_t, b_t, z_t\}$, and entry decisions $I_t^*(z_t)$.

**while** $e > \epsilon$ **do**

  **for** $g \in G_{\tilde{l},0}$ **do**

    With $\Gamma_0(g)$ at hand:
    obtain policy functions $\sigma^*(\cdot,\cdot)$ for active firms, and
    determine entry choices $I_t^*(z_t)$ of potential entrants.
    Clear markets and obtain aggregates, hence obtain $\hat{\Gamma}_0(g)$.

  **end**

  Obtain approximation error $e = \Gamma_0(G_{\tilde{l},0}) - \hat{\Gamma}_0(G_{\tilde{l},0})$.
  Update $\Gamma_0(G_{\tilde{l},0})$ using $\hat{\Gamma}_0(G_{\tilde{l},0})$ with dampened fixed point iteration.

**end**

Now, randomly generate $n$ training inputs $X = \{x_i\}_{i=1}^n \subset B$.

**while** $\bar{e} > \bar{\epsilon}$ *and* $\tilde{l} \le \tilde{l}_{max}$ **do**

  Update $l = l + 1$
  Use $\hat{\Gamma}_{l-1}(\cdot)$ to refine the grid, obtain $G_{\tilde{l},l}$
  Interpolate to get law of motion $\Gamma_l(G_{\tilde{l},l})$

  **for** $g \in G_{\tilde{l},l} \setminus G_{\tilde{l},l-1}$ **do**

    With $\Gamma_l(g)$ at hand:
    obtain policy functions $\sigma^*(\cdot,\cdot)$ for active firms, and
    determine entry choices $I_t^*(z_t)$ of potential entrants.
    Clear markets and obtain aggregates, hence obtain $\hat{\Gamma}_l(g)$.

  **end**

  Obtain approximation error $\bar{e} = \Gamma_l(X) - \hat{\Gamma}_l(X)$.

  Update $\Gamma_l(G_{\tilde{l},l})$ using $\hat{\Gamma}_l(G_{\tilde{l},l})$ with dampened fixed point iteration.

**end**

Algorithm 1.1: Overview of the critical steps of the solution algorithm.

The key of the present solution method lies in the way individual policy functions are aggregated across the state space. I obtain firms' policy functions using an adapted version of time iteration based on Ljungqvist and Sargent 2012. The algorithm for solving the firm problem is illustrated in the Appendix. Value functions of incumbent firms at $b = 0$ are then used to evaluate entry decisions. Consumption is obtained via market clearing equation $C_t = Y_t - I_t$ - this implies that general equilibrium effects come from changes in value functions. Notice that current and forecasted consumption determine the stochastic discount factor, then firms' value functions and hence decision by potential entrants. Since $Y_t$ is a state, the inner loop starts by guessing an equilibrium value of $\tilde{C}_t$ determined by entry decisions, and iterates using bisection until guessed consumption converges to consumption implied by entry choices.

## Parallelization

A useful feature of the solution method presented above is the possibility to exploit high-performance computing systems. The algorithm is based on Message Passing Interface (MPI), a communication protocol for parallel computers, based on a distributed shared memory concept. A detailed outline of MPI is offered in Gropp, Lusk, and Skjellum 1999. My parallelization strategy starts from splitting computation among MPI groups defined by the number of discrete states in the aggregate state space, as described by Brumm, Kubler, and Scheidegger 2016 and illustrated in Figure 3. In this application, the model needs only two MPI groups, as $\sigma = \{\sigma_L, \sigma_H\}$. For each group, the algorithm creates the corresponding grid, then interpolates to obtain an approximate law of motion at each level following the steps summarized in Algorithm 1.1. Each point $g$ in the grid $G_{\tilde{l},l}$ is then endowed with independent firm problem, entry decision and market clearing. Gridpoints are split among multiple MPI processes, so that each process is receiving an independent problem to solve.

Notice that, as the number of gridpoints is not constant but changes at every level $i$, one needs to create a queue and let MPI processes be assigned gridpoints gradually as they proceed solving for equilibrium. An example can clarify what happens here. Suppose $m$ ranks are available, and the grid at level $i$ consists of $n > m$ points. Then a queue is created, and the first $m$ points are assigned to different ranks. For the preservation of strong scaling properties, it is important that $n > m$ always holds, but when using adaptive sparse grids this is not always guaranteed: especially at initial steps, when starting from a good guess, the refinement procedure might reduce more than increase the number of new points added to the grid. This induces the choice of a starting grid that could be "finer"

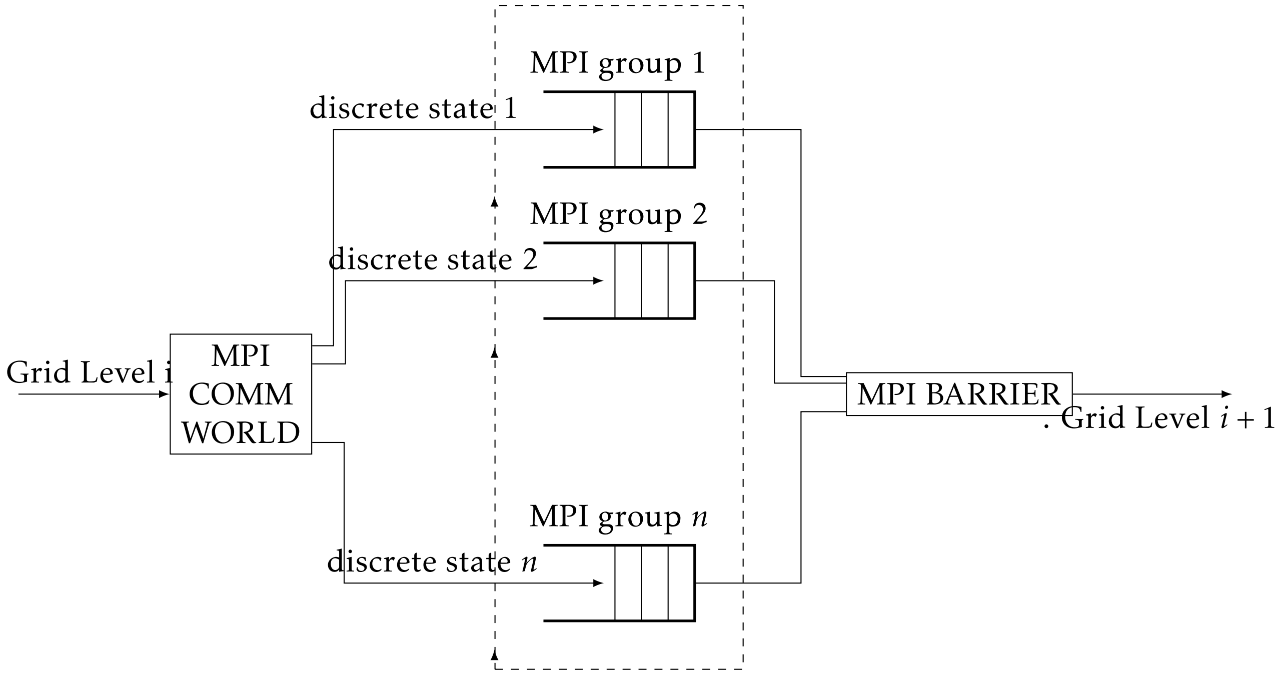


Figure 1.6: Schematic representation of the parallelization of Algorithm 1

than otherwise obtained by efficiency considerations in a non-parallel environment.
As an equilibrium is found in any of the gridpoints $j \in m$, the *freed* rank is now going to pick the next gridpoint from the queue. [11]. Once the queue is exhausted, solutions are merged, and a new refinement level is attainable by gathering points from all processes, updating the adaptive sparse grid interpolation, and finally obtaining new forecasts.

To measure of how the solution method benefits from parallelization, I show the algorithm exhibits strong scaling properties. To do so, I report on the single node performance[12] on different refinement levels (i.e. the computation of the whole grid $G_{\bar{l},l}$ at levels $l = 1, 2, 3...$) with increasingly larger numbers of MPI ranks. The first level of the adaptive grid in the low uncertainty state has 114 points - since this is a relatively small number of jobs, the maximum efficiency is reached at an already relatively low number of ranks. In general, parallel efficiency improves in the size of the grid, as the workload would be better distributed among different MPI processes. Since adaptive sparse grids do not increase monotonically in size, parallel performance doesn't necessarily increase as one proceeds to higher levels. Still, it is possible to say that scaling efficiencies will be even stronger in higher-dimensional models.

[11] I owe Philipp Eisenhauer for my understanding on how to implement this on a HPC cluster

[12] In this section I always refer to the architecture of the "Piz Daint" system installed at Swiss National Supercomputing Centre (CSCS).

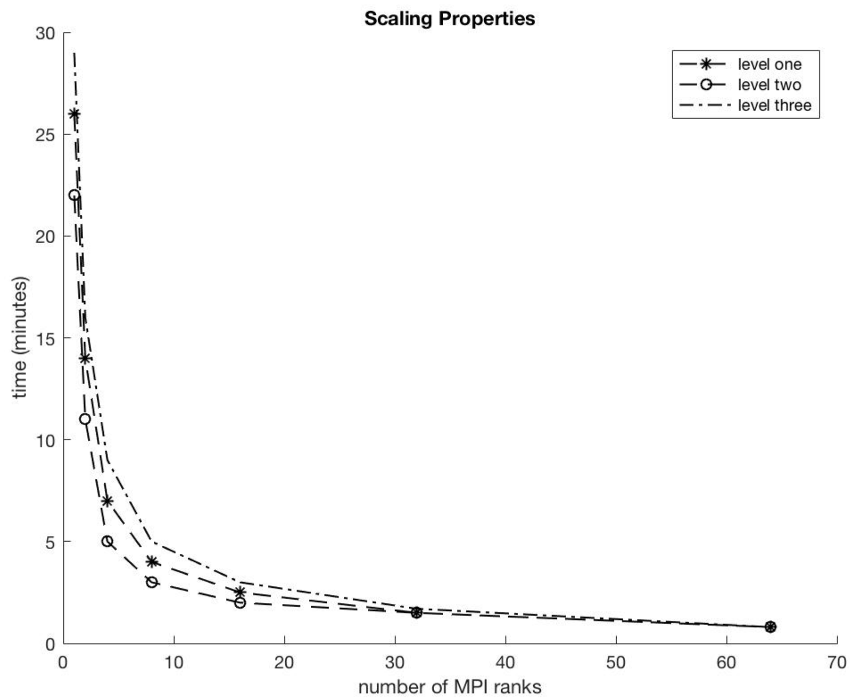



Figure 1.7: Performance for levels of refinement and number of MPI ranks

## 1.6 Results

### Measuring Accuracy

Once the model is solved, evaluating the performance of the algorithm depends on its ability to obtain accurate approximations. In most dynamic models with heterogeneous agents, heterogeneity takes place at the household level. Hence a straightforward measure of accuracy is given by some function of errors in the approximation of Euler equations over a set of points. In the model of this paper, the crucial task is the approximation of the aggregates law of motion, that firms use to forecast prices. It is thus tempting, when comparing the proposed method to more established solution strategies like Krusell and Smith, to report the R-squared for evaluating the accuracy of the latter. The analysis of this chapter will show that it would be misleading, consistently with issues highlighted by, among others, Den Haan 2010. Table 3 reports the maximum and mean percentage absolute differences between realized data and static forecasts for all of the relevant variables as accuracy statistics. Both are relevant measures: an approximation that tracks the dynamics nicely most of the time but is subject to enormous, albeit rare, deviations will not do a good job; an approximation that always stays close to realized data but noisily tracks the process will not be ideal either.

| | Hours | Output | Consumption |
|---|---|---|---|
| Max | 2.77 | 2.99 | 2.75 |
| Mean | 1.00 | 0.76 | 0.63 |

Table 1.3: Maximum and mean percentage differences between realized data and forecasts in the training set

It is an established tenet of the literature that such measures are not informative enough: they do not concentrate on the ergodic set, but consider all of the state space equally, and do not take into account of the extent to which forecast errors can cumulate over time if the actual law of motion is never observed by agents. While the second issue depends on specific model assumptions, the first is extremely relevant in a multi-dimensional setting, where areas of the state space are included in the state space hypercube but are very unlikely to be ever visited by the solution. In turn, this implies that using accuracy measures based on dynamic forecasts can in general show an improved performance in multi-dimensional settings when compared to static forecast accuracy measures. A first diagnostic can be made by looking at how forecasting accuracy is distributed over the training set $\mathbf{X} \subset \mathbf{B}$, as in Figure 4. Errors are concentrated in high-output, low-hours areas, which suggests that they might not affect heavily accuracy in simulations.

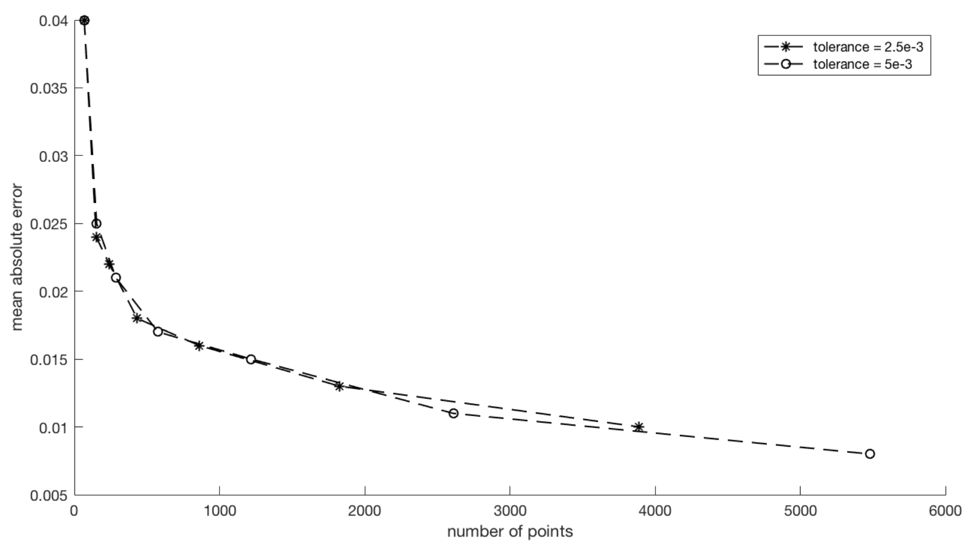



Figure 1.8: Accuracy over $\mathbf{B}$ for different refinement levels at $\sigma = \sigma_H$

The set of statistics from Table 4 provide an account of how accurate the solution is on the ergodic set: dynamic forecasts are computed from the absolute percentage differences of one-step and two-steps-ahead forecasts versus realized data for the relevant variables. The third line reports, for the sake of comparison, one-step ahead forecast errors from the Krusell and Smith solution. Properties of approximation seem to be preserved, if not

improved, in the ergodic set. A comparison with one-step errors from the Krusell and Smith simulation shows the improvement in accuracy, that is massive for hours, and still quite relevant for consumption. Figure 9 gives an example of simulated series and their approximation, which show that the law of motion is followed rather closely.
Another issue arises when it is to be established whether approximation errors are systematic: in other words, if the approximate law of motion is constantly leading to higher or lower expectations of realized values. A way to address this issue has been proposed by Den Haan 2010: if the law of motion generates systematic errors, then n-steps ahead forecasts ($n > 1$) will progressively perform worse.

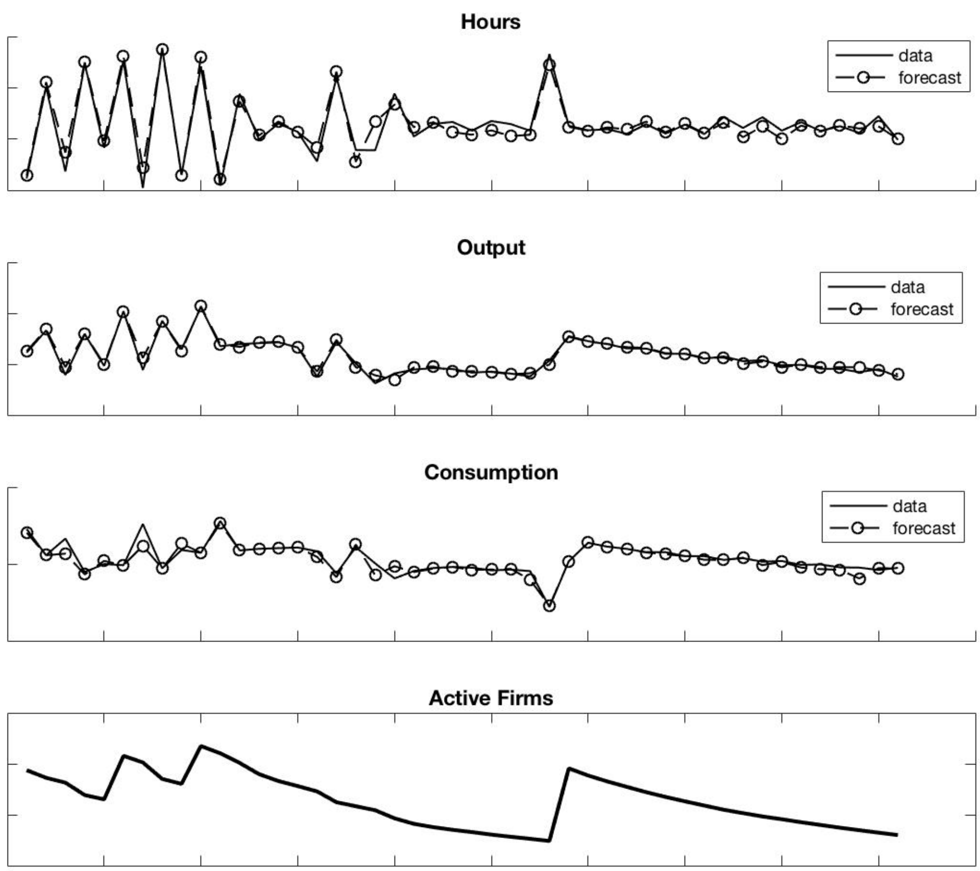
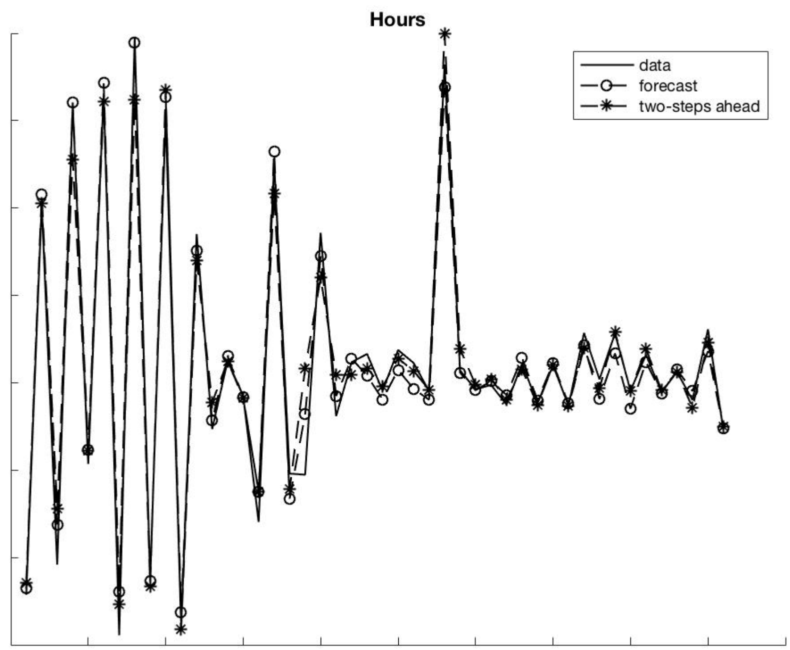
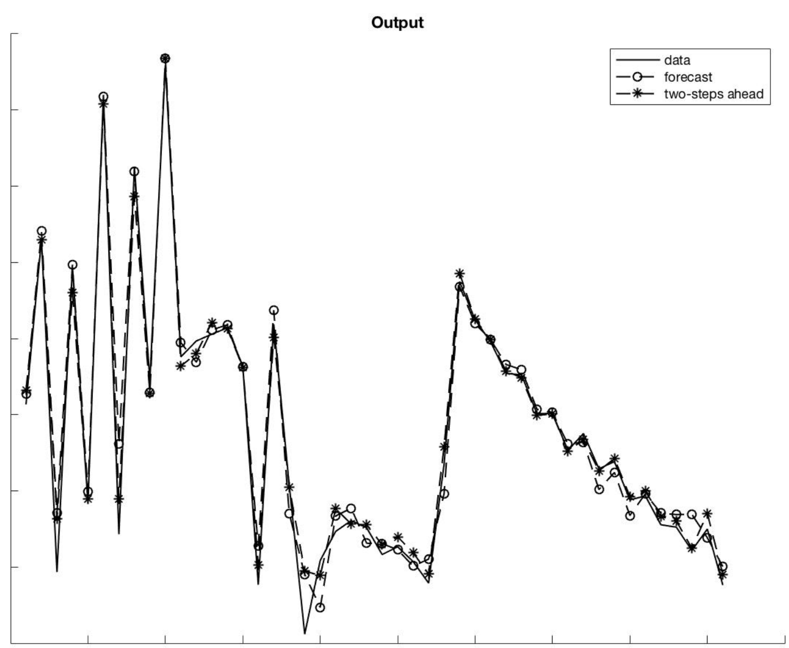



Figure 1.9: A series of 50 periods of simulated data


Hours
data
forecast
two-steps ahead


Output
data
forecast
two-steps ahead

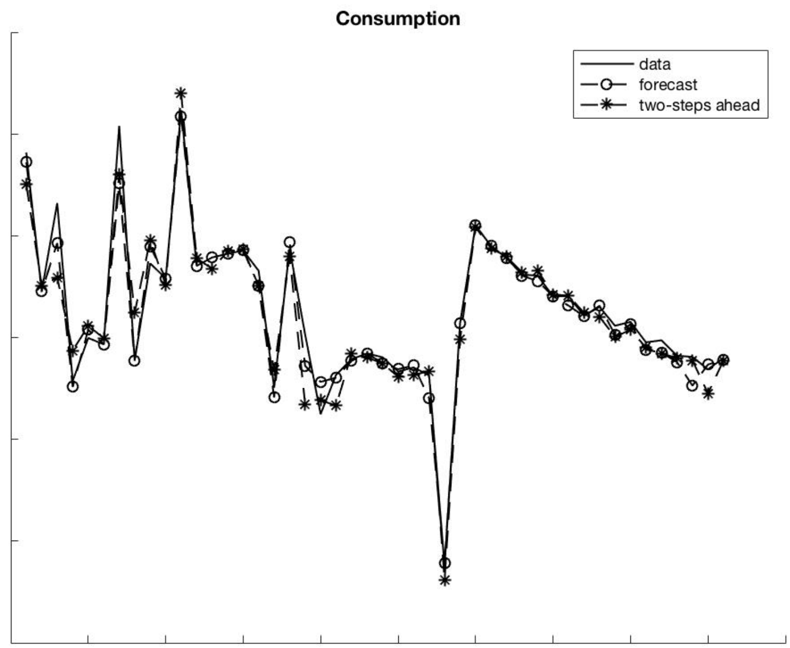



Figure 1.10: One-step vs. two-steps ahead forecasts

| | | Hours | Output | Consumption |
|---|---|---|---|---|
| One-step ahead | Mean | 0.62 | 0.42 | 0.4 |
| | Max | 2.77 | 2.36 | 2.84 |
| Two-steps ahead | Mean | 0.78 | 0.29 | 0.68 |
| | Max | 2.43 | 1.66 | 3.65 |
| One-step in KS | Mean | 6.19 | 2.6 | 1.2 |
| | Max | 13.4 | 8.15 | 4.79 |

Table 1.4: Maximum and mean percentage differences between realized data and forecasts

A formal testing procedure for numerical solutions of dynamic models has been advanced by Den Haan and Marcet 1994. They observe that any solution method implies that approximation error $e_t = \hat{\Gamma}_t(S_t) - \Gamma_t(S_t)$ should satisfy

$$\mathbb{E}_t\left[e_{t+1} \otimes h(S_t)\right] = 0 \tag{1.32}$$

for any function $h : \mathbb{R}^k \to \mathbb{R}^k$. It is then possible to construct a test statistic, $C = TB_T^{'} A_T^{-1} B_T$ such that $C \longrightarrow \chi^2_k$, where:

$$B_T \equiv \sum_{t=1}^{T} \frac{e_{t+1} \otimes h(x_T)}{T} \quad \text{and} \quad A_T \equiv \sum_{t=1}^{T} \frac{e_{t+1}^2 h(x_T) h^{'}(x_T)}{T} \tag{1.33}$$

It is known that any given numerical solution will fail the accuracy test for sufficiently large T. However, for reasonable number of simulated periods, the solution passes the test: critical values for a $\chi^2_k$ with $k = 3$ are 0.115 at 1% and 0.352 at 5% .

| T | 200 | 250 | 500 | 1000 |
|---|---|---|---|---|
| Test statistic | 0.0257 | 0.0276 | 0.0317 | 0.0621 |

Table 1.5: Application of the test by Den Haan and Marcet (1994): the simulation passes the test at the 1% significance level for all lengths

## Impulse Responses

The model economy used in this paper is constructed to analyze business cycle effects of uncertainty shocks; however, lacking a linear representation $X_t = AX_{t-1} + B\varepsilon_{\sigma,t}$ classical impulse response functions $\hat{x} = A^{t-1}B$ are not available. Following Koop, Pesaran, and Potter 1996, I adopt the following procedure to compute impulse responses:

1. Fix large $N$ simulation samples of $T_{\text{IRF}}$ periods, and fix shock period $T_{\text{shock}}$, then simulate the Markov chain of the aggregate shock for the $T_{\text{IRF}}$ periods of each sample;

2. For each sample $i \in N$, draw $u_i \sim U(0,1)$; if $u_i > \bar{u}$, then label the series $i$ as belonging to the "shock" group, and impose high uncertainty in $T_{\text{shock}}$, then let the process transition as usual[13]. Otherwise, simply label the series as "no shock";

3. Simulate the series, so to obtain $X^{\text{shock}}$ and $X^{\text{no shock}}$

4. Obtain the Impulse Response Function as:

$$IRF = 100\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{X_{i,t}^{\text{shock}}}{X_{i,t}^{\text{no shock}}}\right) \tag{1.34}$$

The impulse responses are represented in Figure 11 for the most relevant variables of the model economy. They highlight an interesting and perhaps counter-intuitive effect of the interaction between endogenous firm entry and uncertainty shocks: constrained incumbent firms immediately respond by producing less output, as in the partial equilibrium. A general equilibrium effect, operating via decreasing wages, makes firms with a high idiosyncratic shock willing to use more labor, thus dampening the impact on aggregate hours. On impact, then, the effect of uncertainty operates primarily via a misallocation of labor: firms that are more constrained by default risk will reduce employment, while

[13]The number of periods after $T_{\text{shock}}$ can change: in this paper, I model the shock as lasting two periods, so that in practice the economy starts its transition back to "normal" times in $T_{\text{shock}} + 2$

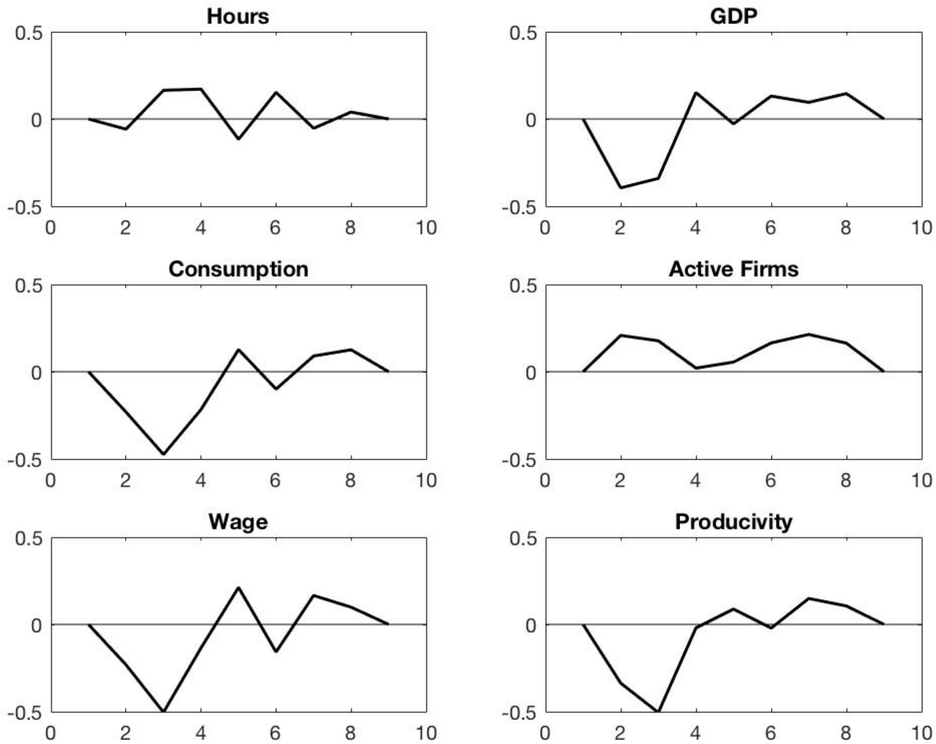



Figure 1.11: One standard deviation shock to uncertainty

others will take advantage of falling wages and increase their production. As a result of the two effects, aggregate output is lower. However, the real option effect on (some) incumbent firms is not the only relevant dynamics induced by increased uncertainty: as future productivity can take higher as well as lower values, the value of more promising projects increases, and this implies an increase in firm entry. This increase more than compensates any increase in firm exit, and thus boosts output out of the small uncertainty-induced contraction. Longer periods of uncertainty can sustain longer (and deeper) contractions; however, the interpretation of such long periods of high uncertainty is less immediate.
Segal, Shaliastovich, and Yaron 2015 discuss the impact of uncertainty shocks that are not mean-preserving, allowing to identify and treat separately increased policy uncertainty (bad shocks) or increased uncertainty induced by a flow of innovation (good shocks). Such a framework can contribute with important insights for business cycle analysis - and it implies an extension of the presented model that would be simple to model under the solution method proposed in this paper.

## 1.7 Conclusions

In this paper I propose a global solution method for solving infinite-horizon, discrete time, heterogeneous agent models with aggregate uncertainty. I illustrate how it works by presenting a model economy which, in its relative simplicity, contains numerous computational challenges that make most existing methods inapplicable or unfeasible.
The approach is successful in solving the model with considerable accuracy; by treating distributional channels explicitly, it is able to analyze models where cross-sectional dynamics play a major role. The algorithm is designed for efficient implementation on HPC clusters, a feature that will become increasingly important in quantitative economics.
Future research will be needed to deal with an even larger state space, where adaptive sparse grids become unfeasible. Another promising avenue is the combination of adaptive approaches with strategies that concentrate computational efforts on areas of the state space that are part of the ergodic set.

## References


Adjemian, Stéphane et al. (2011). “Dynare: Reference manual, version 4”. In:

Aiyagari, S Rao (1994). “Uninsured idiosyncratic risk and aggregate saving”. In: *The Quarterly Journal of Economics* 109.3, pp. 659–684.

Algan, Yann, Oliver Allais, and Wouter J Den Haan (2008). “Solving Heterogeneous-agent Models with Parameterized Cross-sectional Distributions”. In: *Journal of Economic Dynamics and Control* 32.3, p. 875.

Algan, Yann et al. (2013). “Solving and simulating models with heterogeneous agents and aggregate uncertainty”. In: *Handbook of Computational Economics* 3, p. 277.

Arellano, Cristina, Yan Bai, and Patrick J Kehoe (2016). *Financial frictions and fluctuations in volatility*. Tech. rep. National Bureau of Economic Research.

Bellman, Richard E (1961). *Adaptive control processes: a guided tour*. Princeton university press.

Bloom, Nicholas (2009). “The impact of uncertainty shocks”. In: *Econometrica: Journal of the Econometric Society* 77.3, pp. 623–685.

Bloom, Nicholas et al. (2013). “Really uncertain business cycles”. In:

Brumm, Johannes, Dominika Kryczka, and Felix Kubler (2017). “Recursive equilibria in dynamic economies with stochastic production”. In: *Econometrica* 85.5, pp. 1467–1499.

Brumm, Johannes, Felix Kubler, and Simon Scheidegger (2016). “Computing equilibria in dynamic stochastic macro-models with heterogeneous agents”. In:

Brumm, Johannes and Simon Scheidegger (2017). “Using Adaptive Sparse Grids to Solve High-Dimensional Dynamic Models”. In: *Econometrica* 85.5, pp. 1575–1612.

Bungartz, H-J and Stefan Dirnstorfer (2003). “Multivariate quadrature on adaptive sparse grids”. In: *Computing* 71.1, pp. 89–114.

Bungartz, Hans-Joachim and Michael Griebel (2004). “Sparse grids”. In: *Acta numerica* 13, pp. 147–269.

Cheridito, Patrick and Juan Sagredo (2016). “Existence of Sequential Competitive Equilibrium in Krusell–Smith Type Economies”. In:

Dalcín, Lisandro, Rodrigo Paz, and Mario Storti (2005). “MPI for Python”. In: *Journal of Parallel and Distributed Computing* 65.9, pp. 1108–1115.

Dalcín, Lisandro et al. (2008). “MPI for Python: Performance improvements and MPI-2 extensions”. In: *Journal of Parallel and Distributed Computing* 68.5, pp. 655–662.

Davis, Steven J, R Jason Faberman, and John C Haltiwanger (2012). “Recruiting intensity during and after the Great Recession: National and industry evidence”. In: *The American Economic Review* 102.3, pp. 584–588.

Den Haan, Wouter J (2010). "Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents". In: *Journal of Economic Dynamics and Control* 34.1, pp. 79–99.

Den Haan, Wouter J and Albert Marcet (1990). "Solving the stochastic growth model by parameterizing expectations". In: *Journal of Business & Economic Statistics* 8.1, pp. 31–34.

— (1994). "Accuracy in simulations". In: *The Review of Economic Studies* 61.1, pp. 3–17.

Den Haan, Wouter J and Pontus Rendahl (2010). "Solving the incomplete markets model with aggregate uncertainty using explicit aggregation". In: *Journal of Economic Dynamics and Control* 34.1, pp. 69–78.

Dongarra, J Jack and Aad J van der Steen (2012). "High-performance computing systems: Status and outlook". In: *Acta Numerica* 21, pp. 379–474.

Eaton, Jonathan and Mark Gersovitz (1981). "Debt with potential repudiation: Theoretical and empirical analysis". In: *The Review of Economic Studies* 48.2, pp. 289–309.

Fort, Teresa C et al. (2013). "How firms respond to business cycles: The role of firm age and firm size". In: *IMF Economic Review* 61.3, pp. 520–559.

Gropp, William, Ewing Lusk, and Anthony Skjellum (1999). *Using MPI: portable parallel programming with the message-passing interface*. Vol. 1. MIT press.

Guvenen, Fatih (2011). "Macroeconomics with hetereogeneity: a practical guide". In: *Economic Quarterly* 3Q, pp. 255–326.

Haltiwanger, John (2012). "Job creation and firm dynamics in the United States". In: *Innovation policy and the economy* 12.1, pp. 17–38.

Jensen, Michael C (1986). "Agency costs of free cash flow, corporate finance, and takeovers". In: *The American economic review* 76.2, pp. 323–329.

Jermann, Urban and Vincenzo Quadrini (2012). "Macroeconomic effects of financial shocks". In: *The American Economic Review* 102.1, pp. 238–271.

Judd, Kenneth L (1998). *Numerical methods in economics*. MIT press.

Judd, Kenneth L, Lilia Maliar, and Serguei Maliar (2011). "Numerically stable and accurate stochastic simulation approaches for solving dynamic economic models". In: *Quantitative Economics* 2.2, pp. 173–210.

Khan, Aubhik and Julia K Thomas (2008). "Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics". In: *Econometrica: Journal of the Econometric Society* 76.2, pp. 395–436.

Koop, Gary, M Hashem Pesaran, and Simon M Potter (1996). "Impulse response analysis in nonlinear multivariate models". In: *Journal of econometrics* 74.1, pp. 119–147.

Krueger, Dirk and Felix Kubler (2004). "Computing equilibrium in OLG models with stochastic production". In: *Journal of Economic Dynamics and Control* 28.7, pp. 1411–1436.

Krueger, Dirk, Kurt Mitman, and Fabrizio Perri (2016). "Macroeconomics and Household Heterogeneity". In: *Handbook of Macroeconomics* 2, pp. 843–921.

Krusell, Per and Anthony Smith (1998). "Income and wealth heterogeneity in the macroeconomy". In: *Journal of Political Economy* 106.5, pp. 867–896.

Ljungqvist, Lars and Thomas J Sargent (2012). *Recursive macroeconomic theory*. MIT press.

Miao, Jianjun (2006). "Competitive equilibria of economies with a continuum of consumers and aggregate shocks". In: *Journal of Economic Theory* 128.1, pp. 274–298.

Myers, Stewart C (1977). "Determinants of corporate borrowing". In: *Journal of Financial Economics* 5.2, pp. 147–175.

Pflüger, Dirk (2012). "Spatially adaptive refinement". In: *Sparse grids and applications*. Springer, pp. 243–262.

Pröhl, Elisabeth (2017). "Computing the Cross-Sectional Distribution to Approximate Stationary Markov Equilibria with Heterogeneous Agents and Incomplete Markets". In: *Working Paper*.

Reiter, Michael (2009). "Solving heterogeneous-agent models by projection and perturbation". In: *Journal of Economic Dynamics and Control* 33.3, pp. 649–665.

Ríos-Rull, José-Víctor (1997). "Computation of equilibria in heterogeneous agent models". In: *Federal Reserve Bank of Minneapolis Staff Report*.

Scheidegger, Simon and Ilias Bilionis (2017). "Machine learning for high-dimensional dynamic stochastic economies". In:

Scheidegger, Simon and Adrien Treccani (2016). "Pricing American Options Under High-Dimensional Models with Recursive Adaptive Sparse Expectations". In:

Segal, Gill, Ivan Shaliastovich, and Amir Yaron (2015). "Good and bad uncertainty: Macroeconomic and financial market implications". In: *Journal of Financial Economics* 117.2, pp. 369–397.

Stanfield, Brian, John Haltiwanger, and Chad Syverson (2008). "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?" In: *The American economic review* 98.1, pp. 394–425.

Stoyanov, M (2015). *User Manual: TASMANIAN Sparse Grids*. Tech. rep. ORNL/TM-2015/596. One Bethel Valley Road, Oak Ridge, TN: Oak Ridge National Laboratory.

Tauchen, George (1986). "Finite state markov-chain approximations to univariate and vector autoregressions". In: *Economics letters* 20.2, pp. 177–181.

Terry, Stephen (2015). "Alternative Methods for Solving Heterogeneous Firm Models". In:

Winberry, Thomas et al. (2016). "No More Excuses! A Toolbox for Solving Heterogeneous Agent Models with Aggregate Shocks". In: *2016 Meeting Papers*. 1335. Society for Economic Dynamics.

Yellen, Janet (Oct. 2016). *Macroeconomic Research After the Crisis*. "The Elusive 'Great' Recovery: Causes and Implications for Future Business Cycle Dynamics" 60th annual economic conference sponsored by the Federal Reserve Bank of Boston. URL: https://www.federalreserve.gov/newsevents/speech/yellen20161014a.htm.

Young, Eric R (2010). "Solving the incomplete markets model with aggregate uncertainty using the Krusell–Smith algorithm and non-stochastic simulations". In: *Journal of Economic Dynamics and Control* 34.1, pp. 36–41.

## 1.A Appendix

### Solving the Firm Problem using Howard Acceleration

The (intermediate) firm is defined by: past choices over labor and hours, $l_t$ and $b_t$, and current level of idiosyncratic shock to demand, $z_t$. These define the idiosyncratic state $x_t$. The economy is characterized by the aggregate hours and debt, $N_t$ and $B_t$, and the shock to volatility $\sigma_t$. These define the aggregate state as $S_t$. Moreover, aggregate production $Y_t$ is obtained via Dixit-Stiglitz aggregation of each firm's production $y_t = l_t^\alpha$
The firm problem for the choice of hours and debt is then:

$$V(x_t, S_t) = \max\nolimits_{l,b}\left\{\kappa d_t + \beta(1-\kappa)\Sigma_{z_{t+1}}\pi(z_{t+1}|z_t)Q(\sigma_{t+1}|\sigma_t)V(x_{t+1}, S_{t+1})\right\}$$
$$\text{sub } d_t = p_t(x_t)l_t^\alpha - w_t l_t - b_t + q_t(x_t, S_t)b_{t+1}$$

As an additional constraint, dividends are restricted to be non-negative: when a firm can do nothing but distribute negative dividends, it defaults. The firm takes all aggregates in $S_t$ and wages $w_t$ as given. The price of the intermediate good is

$$p_t = z_t\left(\frac{Y_t}{y_t}\right)^{\frac{1}{\gamma}}$$

whereas intermediaries fix the price on the bond $q_t$ in a way such that $q_t(x_t, S_t)b_{t+1}$ equals the expected payment.

1. **Starting values**
   Since $S_t$ is exogenous for the firm, values of $Y_t$, $N_t$, as well as the law of motion that determines $\hat{C}_{t+1} = E_t(C_{t+1}|\mathscr{F}_t)$, $\hat{Y}_{t+1} = E_t(Y_{t+1}|\mathscr{F}_t)$ and $\hat{N}_{t+1} = E_t(N_{t+1}|\mathscr{F}_t)$, where $\mathscr{F}_t$ is the information set of the firm at time t, are determined outside of this problem. Given these values, one chooses a $n \times n$ grid in choice variables $l_{t+1}$, $b_{t+1}$, as well as a starting guess $g_0$ for the policy function $(l^*_{t+1}, n^*_{t+1}) = g_0(x_t, S_t)$.

2. **Obtain V from g**
   Consider that, if $g_0(x_t, S_t)$ is optimal, then

$$V_{i+1} = kd_t(g_0(x_t)) + \beta(1-\kappa)\Sigma_{z_{t+1}}\pi(z_{t+1}|z_t)Q(\sigma_{t+1}|\sigma_t)V_i \tag{1.35}$$

   with $V_i = V_{i+1}$. Then (1.35) can be rearranged as

$$V = (I - \beta(1-\kappa)Q(\sigma_{t+1}|\sigma_t)T)^{-1}\kappa d_t(g_0(x_t))$$

where

$$T = \begin{pmatrix} \pi_{1,1}J_1 & \pi_{1,2}J_1 & \cdots & \pi_{1,5}J_1 \\ \pi_{2,1}J_2 & \pi_{2,2}J_2 & \cdots & \pi_{2,5}J_2 \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{5,1}J_5 & \pi_{5,2}J_5 & \cdots & \pi_{5,5}J_5 \end{pmatrix}$$

and $J_i$, i=1,...,5 is a $n^2 \times n^2$ sparse matrix whose $j,k$-th entry equals 1 if $\text{pos}(l_t, b_t) = j$, $\text{pos}(g(x_t)) = k$ [14].

3. **Obtain the new policy function**
   With $V$ in hand, one can compute the expected continuation value $C(g_i(x_t)) = E\Big[V(g_i(x_t), S_{t+1})\Big]$, so to get the objective function as

$$OBJ(g_i(x_t)) = \kappa d_t + \beta(1-\kappa)C(g_i(x_t)) \tag{1.36}$$

   Clearly, $g_{i+1}(x_t) = \text{argmax}_{l_{t+1}, b_{t+1}} OBJ(g_i(x_t))$.

4. **Check convergence**
   Obtain $V_{i+1}$ as in **1**. Stop if $\frac{g_{i+1}-g_i}{g_i}$ < some criterion, otherwise go back to **2**.

[14] pos refers to the correspondent entry in the $(l_t, b_t)$ grid

## Applying the Solution Method of Krusell and Smith

First, choose a functional form for calculating conditional expectations. This implies a forecasting system:

$$log(\hat{X}_{t+1}) = \alpha\delta_\sigma + \beta log(X_t) + \gamma\Upsilon_t \tag{1.37}$$

where $X_t = [N_t, Y_t, C_t]$, and $\delta_\sigma$ is an aggregate state dummy, which can take on four values, depending on present and future uncertainty state. Guess coefficients $\Gamma_0 = [\alpha, \beta, \gamma]$, draw a $T$- period long time series for idiosyncratic and aggregate shocks from the discretized Markov Chain of $\{z_t\}_{t=0}^T$ and $\{\sigma_t\}_{t=0}^T$. Choose a tolerance value $\varepsilon$.

In each iteration $s$, and period $t$:

1. Given $\Gamma_{s-1}$, and existing aggregate state $S_t$, guess a consumption level $\tilde{C}$. For aggregate state $S_t$ and law of motion $\Gamma_{s-1}$, solve for individual policy functions $n_{t+1}(x_t, S_t)$, $b_{t+1}(x_t, S_t)$ and obtain the expected value of entrants: with this and the free-entry condition $V^e(x_t, S_t, \xi_t) > 0$, one obtains $\int \xi d\mu I_t$ ;

2. Update $\tilde{C}$ using bisection;

3. If guessed $\tilde{C}$ equals consumption implied by $C = Y - I$, market clears. Take next period's idiosyncratic shocks to obtain aggregate variables $S_{t+1}$;

4. After having obtained a full time series from the $T$ iterations between 1 and 3, update the forecasting system using least squares, and obtain $\Gamma_s$. If $|\Gamma_s - \Gamma_{s-1}| < \varepsilon$, for some tolerance value $\varepsilon$, the algorithm stops; otherwise, go back to point 1 with the new forecasting system at hand;

## Results from Krusell and Smith solution

The following table reports estimation results from the application of the Krusell and Smith algorithm to the model economy of the paper.

Table 1.6: Krusell and Smith Regression

| | Hours | | Output | | Consumption | |
|---|---|---|---|---|---|---|
| Constant* | -1.373 | | -1.1912 | | -1.2416 | |
| Hours* | 0.1029 | | -0.1030 | | -0.248 | |
| Output* | -0.779 | | 0.0181 | | 0.158 | |
| Consumption* | -0.915 | | -0.4421 | | -0.3861 | |
| | | -1.386 | | -1.1125 | | -1.1663 |
| | | -0.232 | | -0.3496 | | -0.465 |
| | | -0.2806 | | 0.3438 | | 0.6433 |
| | | -0.8982 | | -0.503 | | -0.4898 |
| Aggregate Shock Dummy | -0.0009 | | 0.0071 | | -0.0083 | |
| Utilization | 0.1006 | | 0.2857 | | 0.4048 | |
| $R^2$ | 0.999 | | 0.989 | | 0.999 | |

* = Left Column is for $\sigma = \sigma_L$, Right Column for $\sigma = \sigma_H$

## Building Adaptive Sparse Grids

Be $\Omega = [0,1]^d$ the domain of the problem, where $d$ is its dimensionality[15]. Let $\vec{x}_{\vec{l},\vec{i}}$ be a d-dimensional gridpoint in our discretized state space: $\vec{l} = (l_1, ...., l_d) \in \mathbb{N}^d$ are the indices representing the refinement level and $\vec{i} = (i_1, ..., i_d) \in \mathbb{N}^d$ represents the spatial position of the gridpoint. A full grid $\Omega_l$ on $\Omega$ will then have mesh size defined as:

$$h_{\vec{l}} = (h_{l_1}, ..., h_{l_d}) = 2^{-\vec{l}} \tag{1.38}$$

The grid is such that, along each dimension, gridpoints are equidistant. However, the mesh size $h_l$ may differ along dimensions.
Let us now assume we want to interpolate a function $f : \Omega \to \mathbb{R}$; for the sake of simplicity we will illustrate the one-dimensional case, as the multimensional extension is straightforward[16]. The set of points on the grid is characterized by a hierarchy, and at each refinement level we can associate functions of arbitrary order $p$:

$$\varphi_{l,j}^{(p)}(x) = \begin{cases} \Pi_{i \in F_l} \frac{x - x_i}{x_j - x_i}, & \text{if } x \in \left\{x_{l,j} - h_l, x_{l,j} + h_l\right\} \\ 0, & \text{otherwise} \end{cases}$$

Notice that a function can have order $p$ only if there are $l$ level of *ancestor* points of $x_j$; otherwise, the largest possible $p$ is taken (eventually, the function will be linear). To order basis functions, define increment spaces as

[15] The domain can clearly be adapted to the nature of the problem: in applications it is a common choice to tailor it around the ergodic set

[16] Stoyanov 2015 provides with a careful explanation on how to extend this reasoning to a multidimensional environment

$$W_l := \text{span}\left\{\varphi^{(p)}_{l,i:i\in I_l}\right\},\ I_l = \left\{i \in \mathbb{N}, 1 \leq i \leq 2^l - 1, i \text{ odd}\right\} \tag{1.39}$$

This implies that different basis functions have mutually disjoint support within a level and that the support of each function on level $l$ is the support of two functions on the next refinement level $l+1$. A function $f$ is then approximated as

$$f(x) \approx \sum_{k=1}^{l} \sum_{i\in I_l} \alpha_{k,i}\varphi_{k,i}(x) \tag{1.40}$$

The coefficients $\alpha_{k,i}$ are called *hierarchical surpluses*. To understand how they are determined and their crucial role in the determination of grids, we will introduce an analytical example. In Figure 1.12, a univariate normal distribution is displayed.

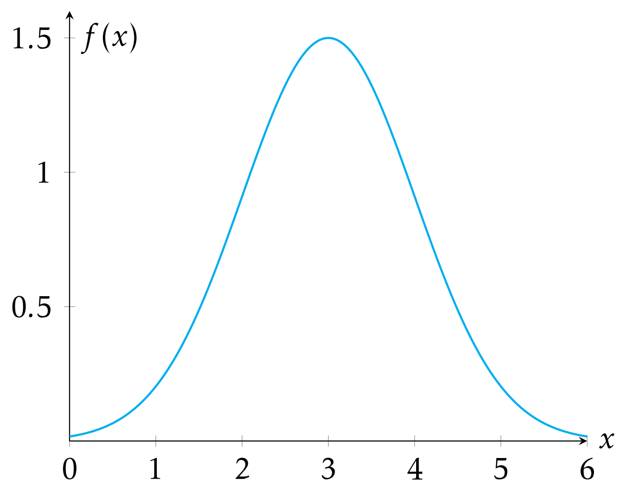



Figure 1.12: Normal distribution with $\mu = 2$, $\sigma^2 = 2$

To approximate it, we first may want to rescale $\Omega$ so that at least it covers the interval between 0 and 6. The first level approximation involves hierarchical function to be $\varphi_{1,1}(3) = 1$; to calculate the surplus, simply notice that $\alpha_{1,1} = f(x_1) - \varphi_{1,1}(3)$. This is highlighted in Figure 1.13. At this point, an evaluation should be made: if $\alpha_{1,1} > \varepsilon$, where $\varepsilon$ is some tolerance value, two *children* points are added - and by the reasoning made above they will be 1.5 and 4.5.

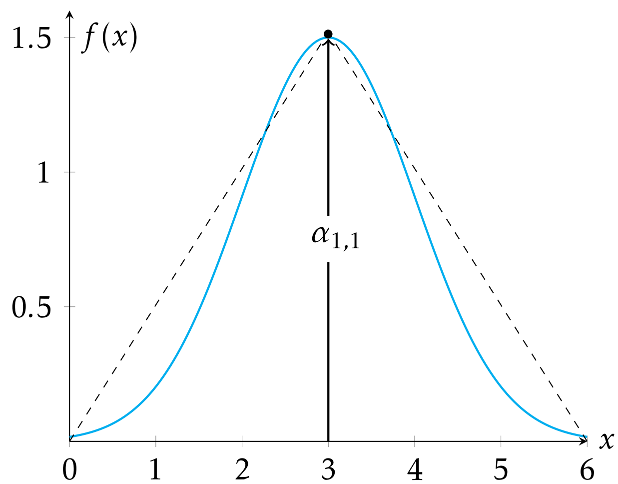


Figure 1.13: Level 1 hierarchical function and coefficients

In 1.14, the two points are added, and the relative surpluses are calculated, as above, from the difference between the approximate function at the previous level and the function evaluated in the two points. The resulting approximation for $l = 2$ is given by the sum of the black dashed line and the red dashed line, which is precisely $\sum_{k=1}^{l} \sum_{i \in I_l} \alpha_{k,i} \varphi_{k,i}(x)$

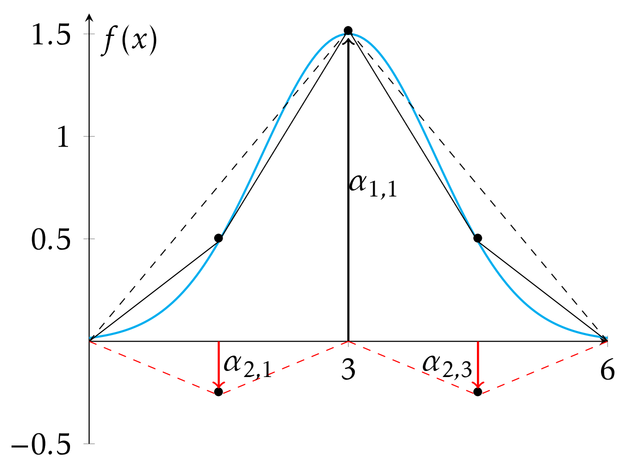


Figure 1.14: Level 2 hierarchical function and coefficients

Notice how the process can go on until it reaches arbitrary accuracy, limited by the tolerance value $\varepsilon$, by adding points in each level in the area of the domain where the function exhibits more curvature. As we will see, this will allow us to deal with the aggregate state space of the model, by approximating agents' expectations with high accuracy without spending too much time in areas of the state space where the mapping is more linear. For a discussion on properties of adaptive sparse grids in approximating functions that are known only algorithmically, see Scheidegger and Treccani 2016.

# Chapter 2

# GO BIG OR BUY A HOME: STUDENT DEBT, CAREER CHOICES, AND WEALTH ACCUMULATION

By Marc Folch† and Luca Mazzone ‡

## 2.1 Introduction

[table,figure]style=Plaintop
We study the effects of student debt on career and housing choices of young workers, and introduce a theoretical framework to look at their interaction. As of December 2018, there were 44 million student loan borrowers who owed $1.46 trillion in total. This, however, is a relatively new feature of US labor markets. Between the years 2004 and 2018, the outstanding stock of student debt more than tripled in the United States (**Figure ??**). Student borrowing is now more likely to be a burden for a higher percentage of college graduates and a relevant factor they take into account in their economic and financial decisions[17].
In presence of financial constraints, student debt affects post graduation choices by making further borrowing more difficult. As the relative value of current consumption grows, and workers postpone additional human capital investment, a series of life cycle decisions are consequently affected. We highlight one often overlooked cost of additional investment in human capital, that is the postponing of household formation. When this channel is taken into account, the initial impact of financial constraints on human capital accumulation is amplified, as the relative value of additional human capital investment decreases

†University of Pennsylvania.
‡University of Zürich and Swiss Finance Institute.
17 Most of the increase in student debt has been attributed to the substantial rising cost of college over the last decade (see **Looney and Yannelis 2015** for a comprehensive account.). Since 2004, tuition at four year colleges increased at an average rate of 3% per year, and student debt surely helped to moderate the impact of higher costs on college enrollment, but more students leave now from college with student debt and borrowers graduate with higher balances.

throughout the life cycle, due to stronger horizon effects induced by mortgage repayment and retirement.

The persistent effects of workers' initial conditions on labor market outcomes have been subject to extensive research. Business cycles have strong effects on long term outcomes for earnings and wealth accumulation of those who start their career from a less favorable position (**Kahn** 2010 and **Oreopoulos, Von Wachter, and Heisz** 2012), while initial wealth is an important determinant of long term wage growth (**Griffy** 2019). Student debt, being the second largest form of household debt after mortgages, is the most natural choice for understanding the role of financial constraints on young workers.

We use the Baccalaureate and Beyond Longitudinal Study (**B&B**), a restricted access dataset compiled by the National Center for Education Statistics. The **B&B** surveys cover a representative sample of U.S. college graduates interviewed on successive waves, starting in 1991. In order to empirically examine how college borrowing affects career choices, earnings, and wealth accumulation in the years after graduation, we need to overcome notorious identification problems, as the amount borrowed may be determined by unobserved individual ability or different expectations, which in turn would affect all post graduation choices.

We address the identification problem by introducing an instrument based on variations in colleges' financial aid. Composition of aid at the college level is calculated using public access data from the Integrated Postsecondary Education Data System (**IPEDS**). We focus on institutional grants, which are funded from private sources and net assets of the institution and experience significant variations year-by-year. We use these supply changes in financial aid during college enrollment to extract variation in student debt that is not correlated with post bachelor choices through unobserved characteristics. We also show that institutional grant changes are not correlated with college measures of quality.

Our empirical results suggest that a negative net wealth position induces a trade-off between career and housing choices for young workers. Higher levels of student debt cause a front loading of earnings, while significantly and persistently deterring additional human capital investment. This, in turn, contributes to lower earnings growth. Indebted graduates earn 0.2% more for each percentage increase in student debt in the first year. Over time, this effect is compensated by wage growth being 0.1% lower. Borrowing also generates an anticipation in home ownership and first-time marriage, although more indebted graduates end up buying less expensive homes.

We develop a model with endogenous (risky) human capital accumulation enriched by career choices and housing decisions to rationalize the empirical evidence and understand the importance of life cycle forces in shaping post graduation outcomes. After graduating from college, individuals enter the labor market and are heterogeneous in ability, student

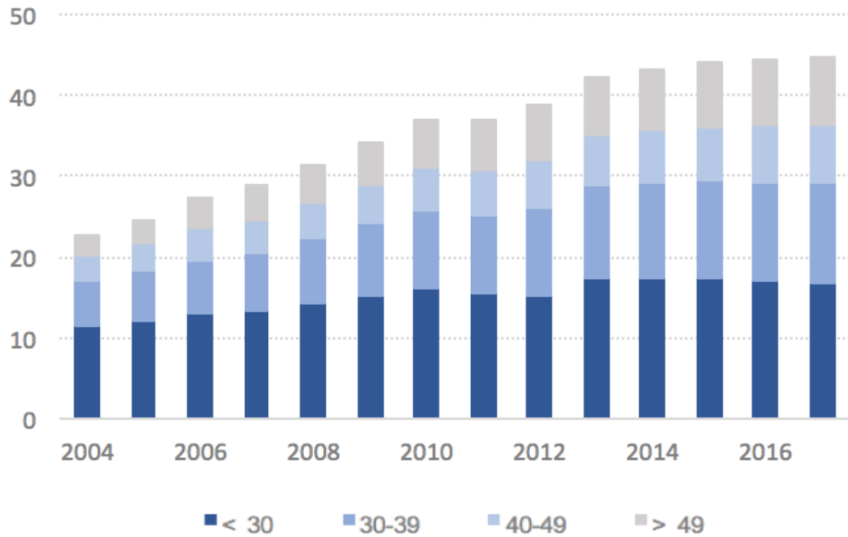


Figure 2.1: Student Loan Borrowers (millions)

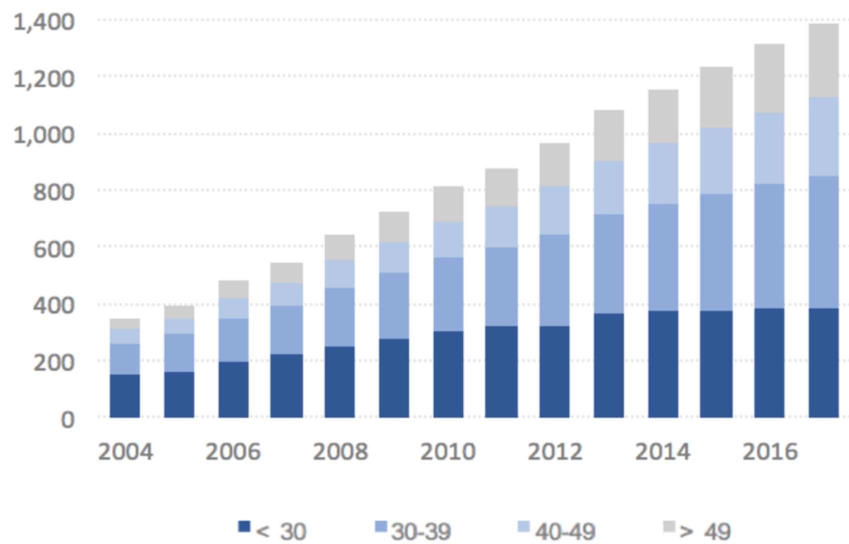


Figure 2.2: Student Loan Balances (billions of dollars)
Source: FRBNY Consumer Credit Panel/Equifax

debt, human capital and initial liquid wealth. They sequentially decide on human capital investment, savings, non housing and housing consumption while they pay for student debt. At any point, they can enroll in a post bachelor program and, if they do, take on additional student debt. Available careers differ in the way productivity, and thus compensation, is linked to accumulated human capital. Workers with a post bachelor degree gain access to a career path that carries a positive skill premium.

The model is estimated by Simulated Method of Moments using a combination of data from **B&B** and Current Population Survey (**CPS**). Our theoretical framework implies that, because of introducing career heterogeneity and post schooling credit constraints, student debt plays a key role in explaining inequality throughout the life cycle. College graduates with higher student debt sort into careers with lower compensation for human capital accumulation, and then experience lower earnings growth. A key factor in this sorting result is the increased preference for earnings front loading induced by student debt. This effect is stronger for low ability individuals.

Structural estimation highlights substantial non monetary returns to post bachelor education, that yields consumption-equivalent utility consisting of more than \$30.000 per year. On the other hand, skill premium for the post bachelor degree educated workers corresponds to less than 40% of the earnings differential with respect to workers with just a bachelor degree. Individuals with relatively higher ability are thus able to afford the cost of higher education given the mentioned composition of returns, while others postpone or choose the alternative career path. Two frictions are crucial in determining these results: binding credit constraints for leveraged households and limited ability to transfer human capital across careers. The second friction also helps explaining why indebted graduates do not simply enroll in graduate school after large part of their debt is repaid, since an implied cost of leaving their career is given by the destruction of part of their human capital accumulated on the job.

The model has also speaks to the effects of borrowing on home ownership. Unconstrained graduates who choose a career with a steeper earnings path (as is the case for those who enroll in graduate programs) are at the same time more likely to postpone their investment in housing. On the other hand, workers that choose to remain in a career that implies lower earnings growth consider housing a relatively more attractive investment. A first, intuitive channel, would point at higher borrowing causing reduced home ownership. On the other hand, career choice induced by the initial debt position is able to counterbalance this effect. Home ownership is relatively higher for young graduates that started their career with more student debt, a finding consistent with our empirical results. Student debt has the apparently counter intuitive effect of anticipating entry into home ownership. As workers age, those into careers characterized by a steeper income path eventually catch up on housing - both in the data and in the model, the two groups have the same rate of home ownership around age 37.

These results point to an easy counterfactual exercise that helps highlighting the way in which, conversely, housing affects enrolment patterns. In a model without home ownership, distortions to human capital accumulation induced by student debt are smaller, enrolment in post bachelor programs is higher, and income inequality decreases. The main reason for this effect is that postponing household formation is costly: higher debt forces graduates to postpone additional education, or to invest less in human capital to accumulate higher savings. While doing so, workers realize that enrolling at a later age would mean a further postponement of household formation, as the downpayment constraint will bind for an even longer period of time, and choose give up on additional education or choose careers with a *flatter* income profile.

Finally, we use the model to evaluate the impact of a widespread adoption of an income based repayment plans (IRP) and compare it to a more radical forgiveness plan, as the one proposed by Sen. Elizabeth Warren in her presidential campaign. We find that the introduction of the IRP provides the foundation for reducing the unintended consequences of student loan debt. By lowering an individual's monthly payments, IRP provides a consumption smoothing mechanism that reduces the need to choose a higher paying job. More surprisingly, the implementation of a forgiveness plan and the widespread adoption IRP yield similar outcomes, both increasing enrollment in graduate programs and earnings over the life cycle. In both cases, however, alleviating the debt burden does imply an increase in housing demand of younger workers. This is particularly true in the case of IBR, due to a combination of increased post bachelor degree attendance, longer time horizon for debt repayment, and higher overall repayments for borrowers with higher balances and low returns on their human capital investments.

The paper is organized as follows. **Section** 2.2 summarizes the literature, **Section** 2.3 describes the data and presents the empirical results, **Section** 2.4 provides an overview of the model and the life cycle choices of individuals, **Section** 2.5 calibrates and estimates the model to observed data patterns, **Section** 2.6 presents the main results of the model, **Section** 2.7 concludes with some policy discussion and future work.

## 2.2 Related Literature

This paper relates to the strand of research that aims at assessing the extent to which initial labor market conditions have strong effects on long term outcomes for earnings and wealth accumulation, as in **Kahn** 2010 and **Oreopoulos, Von Wachter, and Heisz** 2012. In particular, it places itself inside the literature that attempts to identify the impact of student loans on labor market outcomes after college. While research on student loans mostly relied on reduced form estimates, the broader literature on long term consequences of early career decisions is often based on structural models. In this work, we aim at bringing two branches of this literature together.

Isolating the effects of student debt on post graduate choices is made complex by student debt being typically negatively selected, as pointed out by **Looney and Yannelis** 2015. The empirical evidence on how student debt affects earnings mostly points to a positive relationship, at least in the short run[18]. Based on a natural experiment in an elite university,

[18]For empirical studies that conclude a negative or neutral effect of student debt on earnings see: **Weidner**

**Rothstein and Rouse** 2011 show that student debt causes college graduates to choose jobs with an initial higher salary and reduces the probability that they choose "public" low paid jobs. **Luo and Mongey** 2019 find that a version of these results generalizes to the cross section of the U.S. colleges. In particular, they find that higher student debt causes college graduates to take jobs with higher wages, lower job satisfaction, and more on the job search.

Using a difference-in-difference approach, **Gerald and Smythe** 2019 study the impact of student debt on various labor market outcomes (income, hourly wages, and hours worked). They conclude that indebted students have initial higher earnings due to higher work hours rather than higher wage rates. **Chapman** 2015 finds that exogenously increasing the loan burden of a college graduate by $1,000 increases their income by $400-$800 one year after graduation. **Field** 2009 shows that law students who were offered loans were more likely to accept jobs in higher paying corporate law rather than public interest law. Nonetheless, higher initial earnings may not necessarily lead to higher lifetime earnings if they are not followed by further human capital investment (**Becker** 1962**, Ben-Porath** 1967**, Hause** 1972 **and Mincer** 1974). In this line of thought, **fos2017debt** investigate the effects of student debt on additional human capital investment measured as graduate school enrollment. They find that a $4,000 increase in student debt reduces the likelihood of enrollment in graduate school by 1.5 percentage points. In this paper, we contribute to this literature by considering general labor market outcomes for a nationally representative sample of college graduates. We also provide a unified framework for analyzing the relationship between student debt, earnings and human capital investment after college.

Another set of empirical articles have analyzed the role of student loans on first time home ownership. Controlling for multiple factors, **Houle and Berger** 2015**, Cooper and Wang** 2014 **and Gicheva and Thompson** 2014 show that student debt reduces the likelihood of homeownership for young households. This negative relationship likely reflects the underwriting process of a mortgage contract. First, student loans are due when borrowers have the least capacity to pay, leaving borrowers with a lower disposable income and less room for savings towards the down payment of a house. Second, and specially after the financial crisis, the inclusion of student loan payments in the debt to income ratio implies that some agents may delay home purchase until they can qualify for a (larger) mortgage. Using administrative data and tuition induced variation in student debt, **Bleemer et al.** 2017 find that the recent increase in student debt could explain between 11 and 35 percent of the decline in young's homeownership over 2007-2015. Using a similar approach, **Mezza et al.** 2016 estimate that a $1,000 increase in student debt decreased first time homeownership by approximately 1.5 p.p. for public 4-year college graduates

2016, **Akers** 2012 , **Zhang** 2013.

who left school between 1997 and 2005. We contribute to this literature by providing evidence of the effects of student debt on first time homeownership using new data on college graduates. In addition, we rationalize this relationship in a quantitative life cycle framework.

Our analysis also relates to the literature that study student loan program design within a quantitative framework. For example, **Ionescu** 2009 finds that repayment flexibility increases college enrollment significantly, whereas relaxation of eligibility requirements has little effect on enrollment or default rates. In a similar framework, **Ionescu and Simpson** 2016 find that tuition subsides increase aggregate welfare by increasing college investment and reducing default rates in the private market. **Johnson** 2013 also shows that tuition subsidies provide larger increases in college enrollments than increasing borrowing limits. Compared to this literature, our model provides a more detailed characterization of college graduates career choices and post schooling wealth accumulation.

In a related paper, **Di Maggio, Kalda, and Yao** 2019 examine the effect of student debt forgiveness on individual credit and labor market outcomes. Using hand collected lawsuits filings matched with individual credit bureau information, they find that borrowers experiencing the debt relief shock reduce their overall indebtedness by 26%. They also find that borrowers' probability to change jobs increase after the discharge and this leads to an increase in earnings by more than $4000 over a three year period. We examine the effects of an hypothetical student debt forgiveness plan on both earnings and first time home ownership.

Finally, our paper relates to the literature that analyzes how initial conditions affect lifetime inequality. In particular, this literature focuses on the importance of initial conditions relative to shocks over the life cycle. **Huggett, Ventura, and Yaron** 2011 study how heterogeneity in initial wealth and human capital affect lifetime inequality by modelling earnings growth through a Ben-Porath production function. They find that initial conditions, as measured at age 23, determine more than 60 percent of variation in lifetime utility, and that the majority of this variation is determined by initial human capital differences. In a similar framework, **Heathcote, Storesletten, and Violante** 2014 use a model with heterogeneous preferences and productivity, and find instead that life cycle productivity shocks account for half of the cross sectional variance of wages.

The role of initial conditions in shaping long term human capital accumulation has been addressed in the search and matching literature as well. Using a model with directed search and heterogeneous asset holdings, **Griffy** 2019 finds that initial wealth plays a crucial role in determining life cycle inequality, and heterogeneity in skills has a relatively smaller impact. This difference is caused by the inclusion of frictional labor markets, which makes wealth have a first order effect on earnings. In a similar vein, **Eeckhout**

**and Sepahsalari** 2019 show that there is positive sorting between workers with net asset holdings and more productive firms. In this article, we focus on college graduates and also find that initial wealth (student debt) plays a crucial role in life cycle decisions. Differently from this strand of literature, we model the labor market as career paths with different additional human capital requirements. We also include housing as a mechanism through which career choices could interact with financial constraints and affect lifetime wealth. Finally, our model is related to the one in **Athreya et al.** 2019, as they build a life cycle model of education choice where agents are assumed to be heterogeneous in ability, liquid assets and human capital to understand the returns to college attendance for various segments of the population.

## 2.3 Data and Empirical Analysis

### Description of Data

Our main source of data comes from the restricted use dataset from the National Center for Education Statistics (NCES) Baccalaureate and Beyond Survey (**B&B**). The survey follows several cohorts of bachelor's degree recipients over time and contains a mix of administrative and self reported data about their income, student debt, occupation, graduate school enrollment and homeownership (among other variables).

B&B draws its cohorts from the National Postsecondary Student Aid Study (NPSAS), which collects data from large, nationally representative samples of postsecondary students and institutions to examine how students pay for postsecondary education. B&B samples are representative of graduating seniors in all majors and colleges. Our analysis focuses on the most recently available cohort (2007/08), which was followed up one and four years after graduation and was interviewed again in 2018 (forthcoming). We also use as robustness students that graduated from college in 2016 and were followed one and four years after graduation (forthcoming).

We restrict the sample to traditional college students: students who attended only one college, enrolled between 2002 and 2004, and graduated at age 21-23. In terms of colleges, we focus on four year public and private non-profit colleges, excluding private for-profit and special focus institutions. After imposing these restrictions, we also remove all colleges for which we do not have more than 5 students - this is necessary since we use an instrument that is based on college level variation and we need enough students per college for the sample to be representative.

**Table** 2.1 provides the main statistics for the whole sample and for the restricted sample. The table also provides CPS statistics for individuals with at least a BA degree and aged

Table 2.1: Summary Statistics

| | B&B 08/09/12 | | CPS |
|---|---|---|---|
| | Full Panel | Restricted Panel | Restricted |
| **Outcomes** | | | |
| **2009** | | | |
| Current primary job salary | 29,007 | 27,082 | 29,153 |
| **2012** | | | |
| Current primary job salary | 41,869 | 42,409 | 43,886 |
| With a Graduate Degree | 22% | 22% | 22% |
| Home ownership | 37% | 39% | 40% |
| **Debt** | | | |
| % Indebted | 65% | 66% | |
| Percentile 25 (d>0) | 12,482 | 12,000 | |
| Percentile 50 (d>0) | 20,784 | 20,125 | |
| Percentile 75 (d>0) | 33,500 | 33,000 | |
| College Obs. | 1,442 | 512 | |
| Individual Obs. | 14,409 | 7,026 | |

Source: Baccalaureate and Beyond Longitudinal Study 2008/2012 and CPS 2009-2012.

22-24 in 2009 and 24-26 in 2012. Measures of earnings for Baccalaureate & Beyond for college graduates are similar to the ones in Census: the average earning for a college graduate in the restricted sample was $27,082 right after college, while $42,409 four years after graduation. Around 40 percent owned a house and 22 percent had a Graduate Degree by 2012.

## Earnings, Careers and Housing

We are interested in the effect of student loans on earnings and career choices. At the same time, we look at other post baccalaureate choices intrinsically linked to careers such as post baccalaureate education and households formation.

In order to understand the role of career sorting and its relationship with earnings and entry into home ownership, we characterize career paths based on their implied earnings profile. More specifically, we classify careers using the average earnings growth within the large occupation groups that constitute each of them. We merge B&B occupations with Current Population Survey (CPS) occupations and classify as steep careers those in the top quintile of earnings growth between age 25-30 and age 45-50 for those workers with at least a Bachelor's Degree.

Those that we define as *steep careers* contain occupations in healthcare (non-nurses), legal,

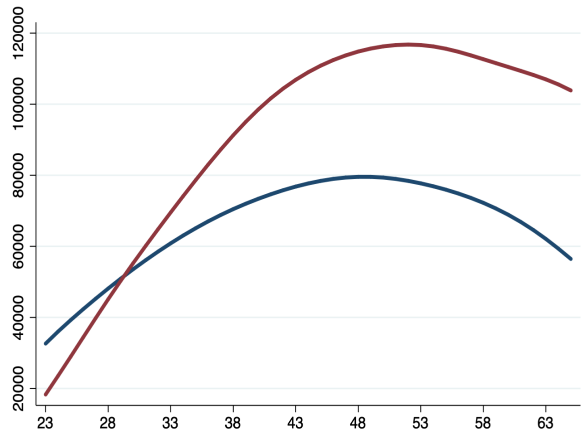


Figure 2.3: Median Earnings

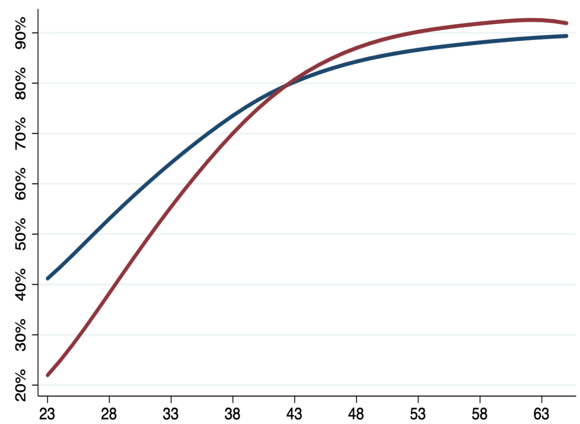


Figure 2.4: Homeownership

Figures 2.3 - 2.4 shows median annual earnings and average homeownership by age for workers in steep earnings career (in red) and other careers (in blue). Source: CPS (2010-2019), white males with at least a Bachelor's Degree. Occupations are classified as steep careers if they are in the top quintile of earnings growth between age 25-30 and age 45-50.

math, post-secondary educators, life scientists and social scientists. Most of these career paths require some post bachelor education or long-term training, adding more back-loading to the implied earnings profile. As we can see in **Figure** 2.3, careers with higher earnings growth typically have lower initial earnings. In addition, career choices are inextricably linked with housing choices: **Figure** 2.4 shows that steeper careers have significantly lower initial home ownership rates.

## Empirical Estimation

Isolating the causal effect of student debt is challenging, as borrowing is hardly an exogenous variable in students' decisions. The bias could go in either direction. On the one hand, if low ability students are less likely to receive grants, $\beta$ will reflect the latent negative correlation between ability and borrowing. On the other hand, high ability students with higher earnings expectations could be more willing to borrow, resulting in debt being positively selected. Using the same reasoning, if colleges with lower instructional expenditure per student have a higher incidence of debt, then $\beta$ will also capture lower college quality.

In order to obtain an unbiased estimate of the causal effect of student debt, the regression should include all of the college and individual characteristics that affect the amount borrowed during college and the post baccalaureate decision. To address this issue, we group colleges into six different categories based on their sector (public or private non-profit) and their Carnegie Classification (Doctorate granting Universities, Master's or Baccalaureate Colleges).

In addition, we also include a rich set of individual controls. We use individual characteristics that are included in the FAFSA financial aid application form (financial need and dependency status), the year they started college (2002, 2003 or 2004), gender and ethnicity. We also include the SAT score and the major of study in order to account for individual's ability[19]. Thus, the relationship between debt and post college outcomes can be expressed in the following reduced form Equation:

$$y_{i,t+\tau} = \alpha_{j(i),t} + \beta d_{i,t} + \Gamma w_{i,t} + \epsilon_{i,t} \tag{2.1}$$

where $y_{i,t+\tau}$ is the individual's post college outcome, $\alpha_{j(i),t}$ is the vector that captures college fixed effects clustered in 6 groups, $d_{i,t}$ is the log of the cumulative amount of loans (federal and private) borrowed for undergraduate degree at time of graduation, and $w_{i,t}$ is the set of individual controls.

Nonetheless, unobserved college and/or students' characteristics could still be relevant in determining access to different forms of aid and have a direct impact on students' post baccalaureate decisions; this makes $d_{i,t}$ a potential endogenous variable, and thus, the OLS estimate of Equation (1) could still be biased.

## Instrumental Variable: Institutional Grants

In this section, we show that supply side variations in the financial aid options faced by all students in a particular college offer a way to overcome the identification problem. In practice, students usually receive a year-by-year financial aid package that is determined by college financial aid officers, but is not known in advance at the time of application. It includes student loans, scholarships, and, grants from the government and the institution itself.

Differently from government grants and student loans, institutional grants are funded from private sources and net assets of the institution. Since institutional grants do not

[19] See Appendix **??** for more details about how these variables are defined.

require repayment, they are preferred to loans and are the first to be added into a financial aid package. Loans are therefore the marginal source of funds to most students. In order to capture this substitution between institutional grants and loans, we compute the ratio of the value of total institutional grants issued by the college to the sum of grants and student loans (grant to aid henceforth):

$$x_j = \frac{inst.grant_j}{(inst.grant_j + loan_j)}$$

**Figure** 2.5 shows how variations in the grant to aid ratio capture the substitution between the two measures of funding for both public and private non-profit colleges. Nevertheless, the exclusion restriction may be violated if the grant to aid ratio is correlated with other college or individual characteristics and those characteristics have a direct impact on students' post baccalaureate decisions. This may happen because students are not randomly assigned to a college and they choose college based on a bundle of college characteristics, which include financial aid availability.

In order to reduce this source of bias, we exploit variation in grant to aid policies during college enrollment. These changes are likely unexpected for the student at the enrollment stage and might come from surprise returns to university endowments, unexpected large donations and/or changes in college costs[20]. These variations in institutional grants are significantly correlated with changes in student debt levels (**Figure** 2.7), and they are uncorrelated with other college (**Table** **A1**) or student characteristics (**Table** **A2**).

We thus take as our instrument the average variation in grant to aid during college enrollment (where $t_0(i)$ represents the year when the individual first enrolled):

$$\begin{aligned} z_{j(i)} &= \hat{x}_{j(i)} \\ &= \bar{x}_{j(i)} - x_{j(i),t_0(i)} \\ &= \frac{\sum_{t=t_0(i)+1}^{t=T} x_{j(i),t}}{T - t_0(i) - 1} - x_{j(i),t_0(i)} \end{aligned}$$

The amount of college debt is modeled as an outcome of individual demand for debt and these supply side college variations in institutional grants, represented by:

$$d_{i,j,t} = \mu_{j(i),t} + \delta z_{j(i)} + \Pi w_{i,t} + u_{i,t} \tag{2.2}$$

We estimate the model by two stage regression. Therefore, it is important that there is significant variation of the instrument with student debt across institutions. **Table** 2.2

[20] For instance, Harvard University endowment value declined 29.5% as investment returns reached −27.3% during the financial crisis. On the other hand, Michael Bloomberg's donated $1.8 billion in support of financial aid at John Hopkins University in 2018, that eliminated the need to borrow for prospective and current students.

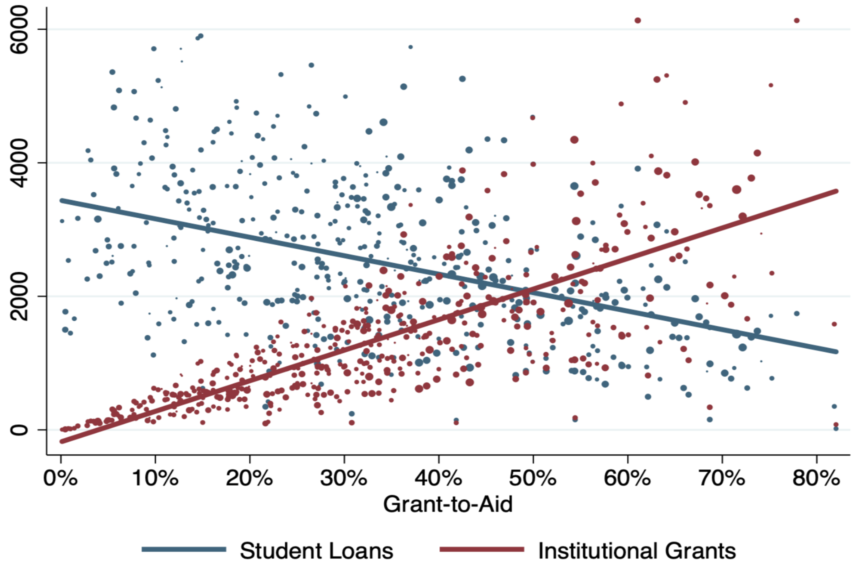


Figure 2.5: Institutional Grants and Student Loans : Public 4y Colleges

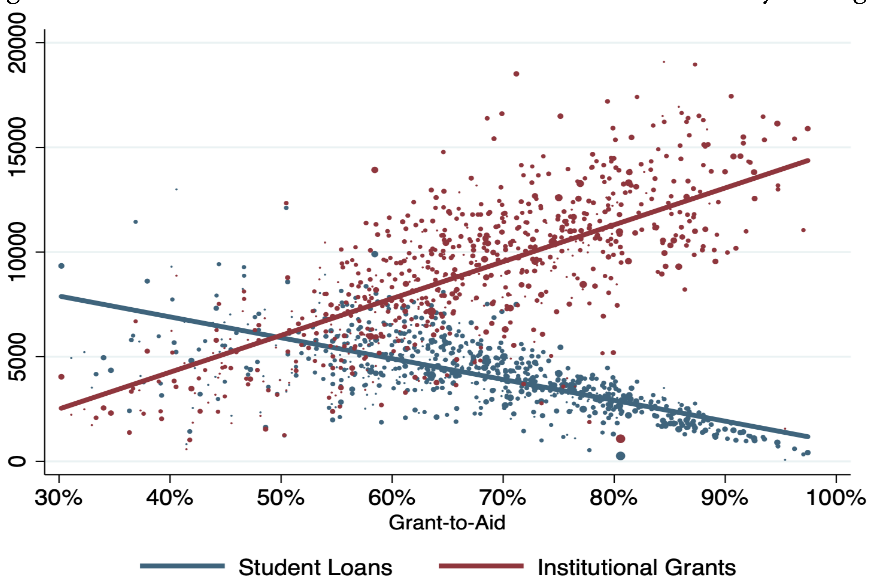


Figure 2.6: Institutional Grants and Student Loans : Private non-profit 4y Colleges

Figures 2.5 - 2.6 shows the average amount of institutional grants (in blue) and student debt (in red) by grant-to-aid for 4-year public and private non-profit colleges in the 2007/2008 academic year, respectively. Source: IPEDS (2007/08). Colleges are weighted by full-time first-time students enrollment.

shows that such condition is satisfied. The results imply that, on average, a change in 1 percentage point in grant to aid during college enrollment induces a corresponding 3% decrease in student debt, all else equal.

Table 2.2: First Stage Regression

| | $\log(\text{debt}_{2008})$ |
|---|---|
| | (1) |
| $\Delta$ Grant to Aid | -0.031*** |
| | [0.006] |
| Controls | Y |
| College FE | Y |
| Observations | 7,026 |

Clustered standard errors in brackets

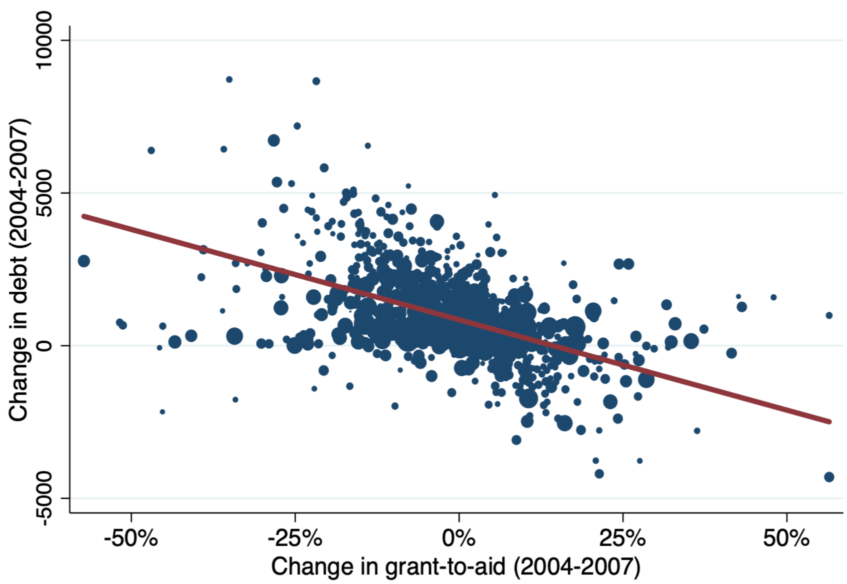



Figure 2.7: Change in Grant to Aid

Figure 2.7 shows the relationship between the change in grant-to-aid and the change in the average debt per student between 2004/05 and 2007/08 academic years. Colleges are weighted by the amount of full-time first-time students enrolled. Table 2.2 shows the regression output of the first-stage regression of cumulative debt at graduation on average change in grant-to-aid (with respect to enrollment year). Source: IPEDS (2007) and Baccalaureate and Beyond Survey (2008). Standard errors are clustered by college groups.

## Empirical Results

This section presents our main empirical results. We show the causal effect of student loans on earnings one, four and ten (forthcoming) years after college graduation. At the same time, we analyze the role of debt on other post baccalaureate choices intrinsically linked to earnings such as graduate school attendance, career choices and households formation. We use the estimation strategy proposed in the previous section, which produces a Local Average Treatment Effect (*LATE*), that is, the effect of student debt on the subset of compliers (colleges that changed grant to aid policies while our sample of students were enrolled).

Table 2.3: Earnings and Post Baccalaureate Education

| | Earnings | | Education (2012) | |
|---|---|---|---|---|
| | Log(wage) 2009 (1) | Growth 2009-2012 (2) | Post Ba Completion (3) | Post Ba Enrollment (4) |
| OLS / Probit | 0.003 | -0.001 | -0.002 | -0.005 |
| | [0.013] | [0.003] | [0.002] | [0.003] |
| IV | 0.298*** | -0.094*** | -0.045*** | -0.039*** |
| | [0.126] | [0.029] | [0.008] | [0.008] |
| Controls | Y | Y | Y | Y |
| College FE | Y | Y | Y | Y |
| Observations | 7,026 | 7,026 | 7,026 | 7,026 |

Standard errors, clustered by college groups, in brackets.

Results from the estimation of Equation (2.1) on earnings are given in the first two columns of **Table** 2.3. The first column shows the OLS and IV estimates for earnings one year after bachelor's degree completion. Column (1) implies that, on average, increasing a student's debt by 10% would lead to an increase in annual earnings of 2.98%. Column (2) runs the same equation on earnings growth 4 years after graduation and shows that the effect turns significantly negative: increasing a student's debt by 10% would lead to a reduction in earnings growth of 0.9%.

The front loading effect of borrowing on earnings is thus consistent with the hypothesis that highly indebted graduates need to boost their initial earnings to ease the burden of repaying their loans. However, the negative growth on earnings might point to indebted workers under-investing in additional human capital after college. Column (3) shows the average marginal effects of debt on post BA attainment 4 years after graduation. *Ceteris paribus*, a 10% increase in college debt reduces the probability of having a post baccalaureate degree four years after graduation by 0.45 percentage points. In addition, Column (4) shows the average marginal effect of debt on cumulative post BA enrollment since college graduation. As we can see, indebted students do not catch up on graduate enrollment even four years after graduation.

While part of the effect of debt on earnings is certainly coming from indebted students being more likely to delay further education (and then reap the benefits some years after), we cannot rule out another channel represented by the choice of different career paths: for instance, a career with a steeper earnings process could be also characterized by the need to do some internship work after graduation before getting a full time position.

Table 2.4: Career and Households Formation (2012)

| | Steep Occupation (1) | Homeownership (2) | First-time Marriage (3) | House Value (4) |
|---|---|---|---|---|
| OLS / Probit | -0.001 | 0.001 | 0.002 | -0.011 |
| | [0.001] | [0.001] | [0.001] | [0.081] |
| IV | -0.034*** | 0.052** | 0.027** | -0.481*** |
| | [0.015] | [0.025] | [0.013] | [0.031] |
| Controls | Y | Y | Y | Y |
| College FE | Y | Y | Y | Y |
| Observations | 7,026 | 7,026 | 7,026 | 1,901 |

Standard errors, clustered by college groups, in brackets.

**Table** 2.4 shows the relationships between student debt and career choices, first-time home ownership and marriage. Measuring the impact of student loans on first entry into home ownership is an important piece for validating our main hypothesis. Consistent with our previous results, a 10% increase in college debt reduces the probability of working in a steeper career four years after graduation by 0.34 percentage points, while increasing the probability of being home owner by 0.52 p.p. Such a positive relationship with housing is consistent with indebted students working more initially after graduating and under-investing in risky human capital. Column (3) shows a similar positive effect of debt on marital status.

Notice, however, that differential earning expectations play a role in determining what type of home is bought. In fact, as more indebted workers enter earlier into homeownership, they also tend to buy less expensive homes: the elasticity of house value to student debt, shown in column (4) is large: considering the median home purchase is worth around \$200.000, an increase in 10% in student debt reduces home value by almost \$10.000.

## 2.4 The Model

The model described in this section builds on important contributions to the human capital literature, as the career choice model of **Keane and Wolpin** 1997 and the **Ben-Porath** 1967 model presented in **Huggett, Ventura, and Yaron** 2011, extended to include student debt and housing.

A unit measure of finitely lived college graduates enter the labor market and are heterogeneous in student debt (d), human capital (h) and initial liquid wealth (k)[21]. Each household lives for $T$ periods deterministically. During working age, workers can decide to enroll in grad school: if they do, they access a different career path. Workers also sequentially decide labor and human capital investment within their career, savings and housing and non housing consumption while they pay for student debt (if any).

## Setting

**Preferences.** Each agent maximizes expected lifetime utility over non durable consumption (c) and housing services (s) (see **Kaplan, Mitman, and Violante** 2019):

$$u(c,s) = \frac{(c^{\zeta_1} s^{1-\zeta_1})^{1-\sigma}}{1-\sigma} \tag{2.1}$$

where c>0 and s=1+$\zeta_2$, where $\zeta_2$ is the housing service from owned housing.
**Labor Income.** When individuals work, hourly earnings are priced competitively to reflect their marginal productivity. Assuming a representative firm that uses human capital from workers in both careers and a linear production function, earnings are given by the human capital augmented number of hours worked multiplied by the equilibrium rental rate ($R_t$).

$$w_{j,t}(l_t, h_t) = R_t l_t \beta_j h_t \tag{2.2}$$

Workers are also exposed to unemployment risk: they can be separated from their job with probability $\rho$; while unemployed, they earn home production $b$, but cannot invest in human capital, so that $h_{t+1} = h_t$. When workers retire, they are assigned pension transfers that are proportional to their last earnings.
**Careers and human capital.** We restrict career choice to two different paths. In each career path, their compensation is equal to the marginal product of hours. Formally, normalizing rental rate $R_t = 1$, we get hourly wage $\tilde{w}_j = \beta_j h_j$, with $j = \{B, G\}$. The two paths differ in how workers' human capital accumulation translates into productive human capital. Human capital is less productive ($\beta_B < \beta_G$) for workers without graduate school education. Therefore, assuming workers make identical human capital investments, differences in earnings would grow as workers accumulate human capital.
After the career choice is made, individuals sequentially choose how much hours to work ($l_t$) and invest in further human capital $(1 - l_t)$. Human capital evolves according to the following Ben-Porath law of motion:

$$h_{t+1} = e^{z_{t+1}}(h_t + a((1-l_t)h_t)^\alpha), \quad z_{t+1} \sim N(\mu_z, \sigma_z^2) \tag{2.3}$$

[21] The distribution of initial liquid wealth is calibrated to match after college parental transfers documented in **Haider and McGarry** 2018

which depends on individual's ability (a) and with risk coming from human capital idiosyncratic shocks. The Ben-Porath formulation implies that switching to the "steeper" career path that follows graduate school has three contrasting effects on human capital investment decisions. On the one hand, since earnings in the steeper career path loads more on human capital, investments are riskier. Formally, comparing variances of hourly wages: $\text{Var}(\tilde{w}_G) = \beta_G^2 \text{Var}(h) > \beta_B^2 \text{Var}(h) = \text{Var}(\tilde{w}_B)$.

Additionally, higher marginal product of human capital gives weaker incentives for graduate school educated worker to invest in human capital because of a simple wealth effect. On the other hand, $\beta_G > \beta_B$ generates a strong substitution effect, in that every unit of consumption today that is foregone in order to invest in human capital generates higher returns in the future. The third effect seems to be dominant in the data, suggesting that difference in career paths are amplified by endogenous human capital investment.

**Graduate School.** Individuals can enroll in graduate school while in working age: if they do, they attend for two years, and then start to work in their new career. While enrolled, human capital grows in every period at rate $g_D$, and workers consume using a combination of their liquid savings and a fixed benefit $b_{grad}$. They also get non monetary utility $\xi$, which summarizes the amenity value of being in school as opposed to working.

Also, while they can switch careers at any point, they would lose all the human capital associated with it if they do. This friction implies that sorting choices made at the beginning of a worker's career can become hard to reverse as professional experience is accumulated, yielding longer term costs due to permanent underinvestment in human capital.

**Financial Markets.** Agents can save in liquid assets $k$. Workers are allowed to borrow short term, using the rate $r_-$, but they face a credit card borrowing constraint that can depend on their current income ($\phi$). If $k > 0$, savings yield a constant risk free rate $r_+$.

**Student Loans.** There are several options for repaying student loans, but the traditional and still most common is the 10-year fixed payment plan. Similar to a mortgage, the borrower makes constant payments over 120 months until the balance of principal and interest is paid off. Student loan payments ($P_\tau$) can be obtained as:

$$P_\tau = \frac{d_0}{\frac{(1+r_d)^\tau - 1}{r_d(1+r_d)^\tau}} \tag{2.4}$$

where $d_0$ is the student debt at the time of college graduation and $r_d$ is the gross interest rate on student loans. If a worker enrolls in graduate school, payments are suspended. Graduate school debt is added to the students' balance, debt is consolidated and a new standard repayment plan is started, giving the worker 120 months to repay the full amount.

**Housing.** Workers can buy a house at any moment of their life - except when they are enrolled in graduate school - as long as their life span is long enough that they can cover the 30-year mortgage and they have enough liquid assets to use as a downpayment. Workers are also subject to housing preference shocks, which capture shifts in life events (household formation or divorce). We model those shifts as taste shocks, i.e. additively separable choice specific random taste shocks, and assume they are i.i.d. Extreme Value type I distributed with scale parameter $\sigma_{\varepsilon}$. If a worker chooses not to own their house, she has to rent ($P_r$). The rental price is tied to the price of the house, $P_o$, and is set to match a given price to rent ratio. Individuals can ask for a 30-year fixed mortgage (m) to pay the price of the house ($P_o$).

There is no possibility of default or asking for a second mortgage. Home ownership is treated as an absorbing state, so if an individual is homeowner in a given year, then it will stay as homeowner at all future dates. Apart from mortgage payments, home ownership involves benefits that individuals can't get from renting (such as tax deductions) and additional expenses (insurance and manteinance). We include these expenses (and benefits) as $\delta$.

At the time of buying the house, individuals face two borrowing constraints: **(1)** they must make a downpayment (1-$\lambda$), **(2)** their monthly debt payments (student and mortgage debt) cannot exceed a proportion of their income ($\psi$). We assume that both constraints must be enforced at origination only.

Home owners must always pay the mortgage payment ($P_{\lambda}$) until mortgage balances are zero, following:

$$P_{\lambda} = \frac{(1-\lambda)P_o}{\frac{(1+r_d)^{30}-1}{r_d(1+r_d)^{30}}} \tag{2.5}$$

## Recursive formulation

We will illustrate the problem for agents of different stages of life, as the recursive formulation will differ according to it. The unit of time is two quarters. The choice is motivated by several facts: it corresponds to the length of the initial grace period (when student loan payments must not be made), it allows for a reasonable accounting of separation risk, and yet it reduces the time dimension enough so that we can solve and estimate the model.

We write future values in recursive expressions by adding a $'$ to them. The choice-specific value functions are denoted indicating the discrete state - for instance, $V^g$ indicates the value function of the worker with post-bachelor degree education.

**Retired workers:**

At retirement age $t = t_R$, workers are assigned pension transfers (p) that are proportional to their last earnings ($w_{t_R-1}$). Retired workers make consumption and saving decisions using their savings from working age ($k_{t_R-1}$). If they are home owners (o), they have to pay the residual parts of their mortgage ($m$) in equal payments ($P_\lambda$) until mortgage debt is fully paid off. Otherwise, if they are renters (r), they need to rent and pay $P_0$ every period. Retired workers cannot buy a house, as mortgage duration exceeds their life expectancy. We assume no bequests and terminal condition for liquid assets to be equal to zero. Finally, we impose a non-negativity constraint on consumption on all agents.

Recursive Problem for renters, for $t = t_R, ..., T$, is :

$$\begin{gathered} V_{a,r,t}(k,w) = \max_{k'} u(c,s) + \beta V_{a,r,t+1}(k',w) \\ c + k' + P_r = (1+r)\cdot k + pw_{t_R-1} \\ m_T = 0, k_T = 0,\ k' \geq \phi(pw_{t_R-1}),\ c \geq 0, \end{gathered} \tag{2.6}$$

The Problem for home owners for $t = t_R, ..., T$, with mortgage payment $P_\lambda$ is:

$$\begin{gathered} V^o_t(a,k,w,m) = \max_{k'} u(c,s) + \beta V^o_{t+1}(a,k',w,m') \\ c + k' + (P_\lambda + \delta) = (1+r)\cdot k + pw_{t_R-1} \\ m' = (1+r_d)m - P_\lambda \\ k_T = 0,\ k' \geq \phi(pw_{t_R-1}),\ c \geq 0 \\ \frac{P_\lambda + P_\tau}{w_j} \leq \psi \end{gathered} \tag{2.7}$$

In both cases, $r = r_+$ if $k \geq 0$, and $r = r_-$ otherwise. $P_\lambda$ is the mortgage payment as defined in equation (2.5) and depends on the downpayment the homeowner chose at the time the mortgage was originated.

**Workers (without student loans):**

Agents enter working age ($t = 1, ..., t_{R-1}$), and face two discrete choices every period: which career to pursue, i.e. whether to enroll in graduate school ($j = \{B,G\}$), and whether to buy a house or not ($H = \{r,o\}$). In both cases, workers are subject to preference shocks - respectively, denote the preference shock for the housing choice as $\sigma_\varepsilon \varepsilon_H$, and the preference shock for the schooling choice as $\sigma_\varepsilon \varepsilon_G$. Both preference shocks are i.i.d. Extreme Value

type I distributed with scale parameter $\sigma_\varepsilon$.
Workers' problem entails saving and choosing how much hours to work ($l$) and invest in further human capital ($1-l$) in every period. Human capital investment is risky and subject to an independent and identically distributed idiosyncratic shock every period ($z$). Earnings are given by the human capital augmented number of hours worked multiplied by the equilibrium rental rate as defined in (2.2).
Workers are also exposed to unemployment risk: while working ($u = 0$), they can be exogenously separated from their job with probability $\rho$; while unemployed ($u = 1$), they earn home production $b$, but cannot invest in human capital, so that $h' = h$. In order to apply for a mortgage and thus become a homeowner, workers have to satisfy the down-payment constraint (governed by the ratio $\lambda$) and at the same time satisfy the debt to income constraint (determined by the value $\psi$). Once the mortgage is approved, the payments ($P_\lambda$) are fixed for the next 30 years as defined in (2.5). Denote $a$ as workers' idiosyncratic ability. For notational convenience, we can collect shocks and exogenous states in $e = \{\varepsilon_H, \varepsilon_G, u\}$, and all the other idiosyncratic states in $x = \{a, h, k, d\}$, where $d$ indicates residual student debt balances, which in this case are equal to zero.
The recursive problem for renters without graduate school education, while employed, is thus:

$$\begin{gathered}
V_{r,t}(x,e) = \max_{k',l} \{u(c,s) + \beta \mathbb{E}[EV_{t+1}(x',e')]\} \qquad (2.8) \\
c + k' + P_r = (1+r)k + w_j(l,h) \\
h' = e^{z'}(h + a((1-l)h)^\alpha) \\
k' \geq \phi(w_j),\ c \geq 0
\end{gathered}$$

where:

$$EV_t(x,e) = \max\left\{V_{r,t}(x,e), V^g_{r,t}(x,e,s), V^o_t(x,e,m), V^{o,g}_t(x,e,s,m)\right\}$$

and where $V^g$ is the value function of a worker enrolled in grad school, and the state $s$ indicates the periods of attendance in the program. Unemployed workers' problem is analogous, with earnings replaced by $b$ and no human capital investment decision. Unemployed workers can find a job in the same career with probability $1-\rho$.
Home owners with housing payment $P_\lambda$ face the following problem:

$$\begin{gathered}
V^o_t(x,m,e) = \max_{k',l} u(c,s) + \beta \mathbb{E}\Big[EV^o_{t+1}(x',m',e')\Big] \qquad (2.9) \\
c + k' + (P_\lambda + \delta) = (1+r)\cdot k + w_j(l,h)
\end{gathered}$$

$$h' = e^{z'}(h + a((1-l)h)^{\alpha})$$

$$k' \geq \phi(w_j),\ c \geq 0$$

$$\frac{P_\lambda + P_\tau}{w_j} \leq \psi$$

$$m' = (1 + r_d)m - P_\lambda$$

$$P_\lambda = \begin{cases} \lambda P_0, & \text{if } m = P_0 \\ \frac{r_d(1+r_d)^{30}(1-\lambda)P_o}{(1+r_d)^{30}-1}, & \text{if } 0 \leq m < P_0 \end{cases}$$

where:

$$EV_{o,t}(x,e) = \max\left\{V_t^o(x,e,m), V_t^{o,g}(x,e,s,m)\right\}$$

If the worker is in the first period of home ownership, $P_\lambda$ equals to the downpayment required to buy the house. After that period, housing payments are determined by the mortgage equation (2.5), as before.

At this point we want to characterize the recursive problem of the individual attending graduate school. For simplicity, we will characterize only the problem of the renter. Define $\bar{S}$ as the number of periods required to get the degree. For $s \leq \bar{S}$:

$$V_{r,t}^g(x,e,s) = \max_{k'}\left\{u(c,s) + \beta\mathbb{E}[V_{r,t+1}^g(x',e',s')]\right\} \tag{2.10}$$

$$c + k' + P_r = b + (1+r)\cdot k$$

$$h' = h\cdot(1+g_D)$$

$$d' = (1+r_d)\cdot d + d_G\cdot \mathbb{1}_{s=1}$$

$$k' \geq 0,\ c \geq 0$$

We assume that, during graduate school, the borrowing constraint with liquid assets is tighter - since the individual is not working she has to keep her liquid assets positive. When $s > \bar{S}$, the recursive problem is analogous to the problem of a worker with student loans, conditional on career earnings' slope $\beta_j$, which is treated below.

**Workers (with student loans):**

Workers that enter the labor market with any positive amount of student debt ($d_0 > 0$) are by default enrolled in a 10-year fixed rate repayment plan, indicated by $\tau = 0$ [22]. Workers

[22] In this subsection both workers with undergraduate and graduate debt are treated together, assuming workers choose to consolidate their student loans at the day of graduation

don't have the option of defaulting or deferring on student loan payments. An employed renter would solve:

$$V_{r,t}(x,e) = \max_{k',l}\{u(c,s) + \beta\mathbb{E}[EV_{t+1}(x',e')]\} \tag{2.11}$$

$$c + k' + (P_\tau + P_r) = (1+r)\cdot k + w_j(l,h)$$

$$h' = e^{z'}(h + a((1-l)h)^\alpha)$$

$$d' = (1+r_d)d - P_\tau$$

$$k' \geq \phi(w_j),\ c \geq 0$$

where:

$$EV_t(x,e) = \max\left[V_{r,t}(x,e), V^g_{r,t}(x,e,s), V^o_t(x,e,m), V^{o,g}_t(x,e,s,m)\right]$$

as above. Home owners in working age with mortgage payment $P_\lambda$ face the following problem:

$$V^o_t(x,e,m) = \max_{k',l} u(c,s) + \beta\mathbb{E}\Big[EV^o_{t+1}(x',e',m')\Big] \tag{2.12}$$

$$c + k' + (P_\tau + P_\lambda + \delta) = (1+r)\cdot k + w_j(l,h)$$

$$h' = e^{z'}(h + a((1-l)h)^\alpha)$$

$$d' = (1+r_d)d + P_\tau \leq 0$$

$$k' \geq \phi(w_j),\ c \geq 0$$

$$\frac{P_\lambda + P_\tau}{w_j} \leq \psi$$

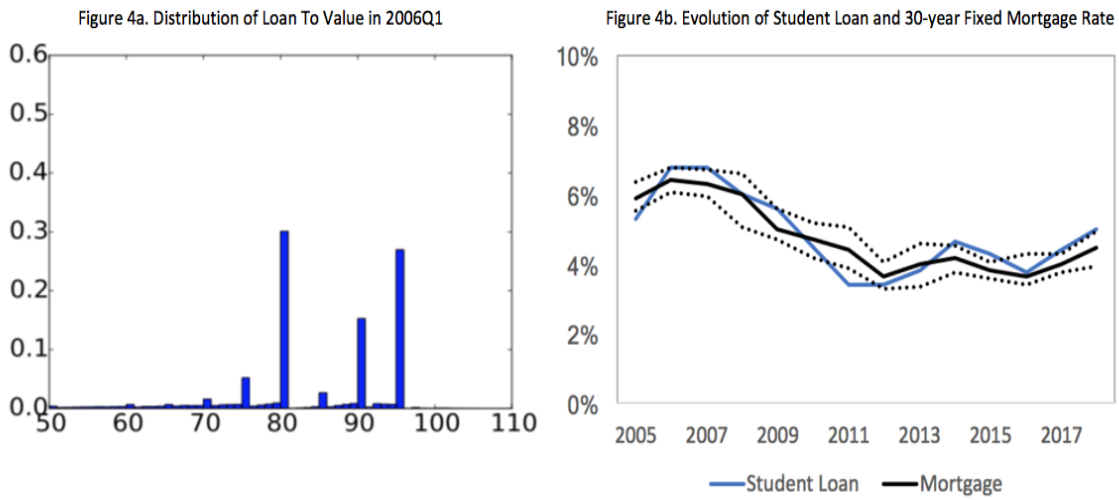
$$m' = (1+r_d)\,m - P_\lambda$$

$$P_\lambda = \begin{cases} \lambda P_0, & \text{if } m = P_0 \\ \frac{r_d(1+r_d)^{30}(1-\lambda)P_o}{(1+r_d)^{30}-1}, & \text{if } 0 \leq m < P_0 \end{cases}$$

Where $P_\tau$ is the student debt payment as defined in equation (2.4), and where

$$EV_{o,t}(x,e,m) = \max\Big\{V^o_t(x,e,m), V^{o,g}_t(x,e,s,m)\Big\}$$

Figure 2.8: Loan to Value and Interest Rate on Student loan and Mortgage debt




Note: Figure 4a shows the distribution of the Loan To Value at origination in 2006Q1 (taken from **Greenwald 2018**). Figure 4b shows the evolution of the federal student loan interest rate and the lowest, average and highest mortgage rate for a 30-year Fixed mortgage rate. Sources: Fannie Mae Single Family Dataset, Federal Student Aid, U.S. Department of Education, and, Freddie Mac's Primary Mortgage Market Survey (PMMS).

## 2.5 Calibration and Estimation

In this section, we discuss how we determine the parameters required for the analysis. We set these parameters in two ways. First, we set some parameters from elsewhere in the literature or by using data estimation (**Table** 2.5). The remaining parameters are estimated using indirect inference through the model.

### External Parameters

**Timing.** Each period time in the model represents two quarters. Individuals start making decisions when they graduate from college. After finishing college, they start working and repaying their student debt. Agents retire at the age of 65 and die when they are 80.
**Preferences.** Preferences are set using standard calibration in the macroeconomics literature. The yearly discount factor is set to be 0.99. We set the constant relative risk aversion in the utility function to 2.
**Career and Human Capital.** Following **Huggett, Ventura, and Yaron 2011**, we set the mean shock of human capital to 0, with 0.075 variance and the production function parameter $\alpha$ to 0.66. We assume that, when unemployed, worker gains access to unemployment benefits that sum up to $b$ calibrated to the Federal poverty threshold for an individual living alone in 2008 ($991 USD a month).
**Labor Income.** We set the rental rate to a yearly rate of 5% of the house price, and pension

Table 2.5: Calibrated Parameters

| | Parameter | Value | Description |
|---|---|---|---|
| **Preferences** | | | |
| | $\beta$ | 0.99 | Discount Rate |
| | $\sigma$ | 2 | Risk Aversion |
| **Careers, Human Capital** | | | |
| | a | 0.33 | Average Learning Ability |
| | $\alpha$ | 0.66 | Ben-Porath Production Function |
| | $\mu_z$ | 0.0 | Mean Shock to Human Capital |
| | $\sigma_z$ | 0.075 | Riskiness of Human Capital Investment |
| **Labor Income** | | | |
| | p | 0.45 | Pension Rate |
| | $\rho$ | $\{0.045, 0.06\}$ | Separation Probability |
| | b | \$991 | Home Production (monthly) |
| **Financial Markets, Student Debt** | | | |
| | $\phi$ | −\$ 5,000 | Credit Card Borrowing Limit |
| | $r^+, r^-$ | $\{2\%, 10.25\%\}$ | Interest on liquid assets |
| | $r_d$ | 7.1% | Interest on student loans and mortgages |
| | $\tau$ | 10 | Years for Fixed Repayment Plan |
| **Housing** | | | |
| | $P_o$ | \$ 250,000 | House price |
| | $\lambda$ | 0.15 | Downpayment (fraction of house price) |
| | $\psi$ | 0.43 | Debt to Income Ratio |

to be 45 percent of the last earned income. Finally, exogenous separation risk is set to 6 percent per year for bachelor-educated workers, and 4.5 percent for workers with a post-Bachelor degree, matching the average number of employment to unemployment transition of the two groups (see **Menzio, Telyukova, and Visschers** 2016).

**Financial Markets and student debt.** The annual interest rate for student loans and a 30-year fixed rate mortgage is calibrated to the 2004-2008 average rate of 6 percent (see **Figure** 2.7**b**). The risk free interest rate for savings is set at 0 following null real returns after 2008 and credit card borrowing rate is fixed at an annual 10 percent. We set a credit card borrowing limit of $-\$5,000$, targeting a median rate of credit limit to annual labor income for college graduates of 20 percent.

**Housing.** We set the price of the house at the average home price in the U.S. (\$250,000) during the years 2008-2012. The rental price per year is set at 5% of the house value to match the price to rent ratio (20). To calibrate the additional costs of homeownership, we compare 2015 ACS data for the median gross rent (rent and utilities) and median homeownership cost (mortgage payments, real estate taxes, insurance and utilities) in each state. We find that the median cost to own a home is 50% more than the median cost to rent each month.

The parameters that determine the LTV and DTI are chosen to match institutional features of the US mortgage market. For the LTV parameter, fix a downpayment constraint of $0.15 \cdot P_o$. This value is intended to reflect the distribution of the LTV in Freddie Mac data, which has two masses point around 80% and 90% (see **Figure 2.7a)**, where the first mass point is typically populated by younger buyers and thus seems more appropriate for pinning down the problem of first home ownership. In order to qualify for a Qualified Mortgage under CFPB guidelines, a borrower's total debt to income ratio, including the mortgage payment and all other recurring debt payments, cannot exceed 43 percent (Consumer Financial Protection Bureau 2013). Thus, we set the DTI parameter to 43%.

### Distribution of Initial Characteristics

In order to simulate the model, we have to make parameter choices regarding ability, starting values of liquid assets, human capital and student debt. We assume students leave college with zero liquid assets, but receive an exogenous transfer $\varepsilon_k$ from their parents, where $\log \varepsilon_k \sim \mathcal{N}(\mu_k, \sigma_k)$. Parameters of the log normal are calibrated to match parental transfers, as documented in **Haider and McGarry 2018**, that report an average transfer of $15,275 with the average being $27,247 conditional on considering only the 56% of graduates that receive a positive amount from their parents.

The distribution of other initial characteristics (ability, human capital, and student debt) is jointly log normally distributed. We determine these parameters in multiple steps. We calibrate the initial mean and standard deviations of human capital to match the mean and standard deviation of earnings after graduation from CPS data - respectively at $32.590 and $22.152. We match an average debt balance of $16.619, as reported by B&B in 2008.[23] We assume no correlation between initial human capital and student debt, and we take the joint distribution of human capital and ability from **Athreya et al. 2019**, who estimate a life cycle model of education choice on CPS data as well and report a correlation of 0.67. Finally, we need to determine the mean level of ability, $\mu_a$ and its correlation with initial cumulated student debt, $\rho_{a,d}$. We set the first parameter in order to match a yearly average growth rate of earnings of 2.9%. The second parameter has an important interpretation because, if correctly identified, it informs about the bias that an econometrician would be subject to when estimating equation (2.1) with least squares. We estimate the second parameter, jointly with other structural parameters, to match the key properties of the earnings and homeownership profiles on CPS and B&B data.

[23] The figure is composed by a percentage of 66% of borrowers, with cumulative average balances of $22.560 and a standard deviation of $11.070

## Estimation

Parameters $\Theta = \{\xi, g_s, \beta_G, \zeta_1, \zeta_2, \rho_{a,d}\}$ are jointly estimated by Simulated Method of Moments (see **Gourieroux, Monfort, and Renault** 1993 , **Smith Jr** 1993 and **Gallant and Tauchen** 1996 ). Let Let $x_i$ be an i.i.d. data vector, $i = 1, ..., n$, and $y_{is}(\Theta)$ be an i.i.d. simulated vector from simulation $s$, so that $i = 1, ..., N$ , and $s = 1, ..., S$. The goal is to estimate $\Theta$ by matching a set of simulated moments, denoted as $h(y_{i,s}(\Theta))$, with the corresponding set of actual data moments, denoted as $h(x_i)$. Define:

$$g_n(\Theta) = \frac{1}{n}\left[\sum_{i=1}^{n} h(x_i) - \frac{1}{S} h(y_{i,s}(\Theta))\right] \tag{2.13}$$

Building $g_n(\Theta)$ in this case faces an important challenge. In classic SMM estimation, exploration of the state space requires the model to be solved more than 10000 times. In the case of a model with a large state space like ours, this could be computationally expensive.[24]. To overcome the curse of dimensionality, we discretize the parameter space using sparse grids (see **Bungartz and Griebel** 2004) A similar approach in structural modelling as been using in the context of maximum likelihood estimation, see for instance **Heiss and Winschel** 2008.

By using functions with support restricted to a neighborhood of each point to build $h(y_{i,s}(\Theta))$, our approach is suitable for approximating the parameter-moment mapping even in cases of sharp behavior, like large fluctuations of the gradient (see **Stoyanov** 2013. Having $h(y_{i,s}(\Theta))$ at hand, we can construct an objective function that looks like:

$$\hat{\Theta} = \arg\min_{\Theta} g_n'(\Theta)\hat{W}_n g_n(\Theta) \tag{2.14}$$

where $\hat{W}_n$ is a positive definite matrix that converges in probability to a deterministic positive definite matrix $W$. There are many feasible choices for the covariance matrix, and it is common to simply rely on an identity matrix for $W$. To construct the optimal weight matrix, we use the influence function technique from **Erickson and Whited** 2002 (see also **Bazdresch, Kahn, and Whited** 2017 for an application closer to our case). The derivation is explained in detail in **Appendix 5**. Finding a solution to (2.14) faces the issue of the possible presence of many local minima: to make sure our solution is robust, we restart our optimization routine using multiple sets of starting values. Each routine solves its problem using a Nelder-Mead algorithm. Having an estimate of $h(y_{i,s}(\Theta))$ also allows us to obtain standard errors of parameter estimates, as they can be calculated knowing that

[24]Using a cluster with 144 CPUs, we manage to obtain a full solution of the model and simulate it in about 14 minutes.

$$\text{aVar}\left(\hat{\Theta}\right) = \left(1 + \frac{1}{S}\right)\left[\frac{\partial g_n(\Theta)}{\partial \Theta} W \frac{\partial g_n(\Theta)}{\partial \Theta'}\right]^{-1} \tag{2.15}$$

We want to match the empirical income profiles, the enrolment and the home ownership rates of individuals in working age. To do so, we take households with at least a BA degree, older than 23 years old from 2000-2018 Census data. We then separate the sample between those workers that obtained more than a bachelor degree at some point and those with only a bachelor degree. We use earnings in 2012 dollars, conditional on workers having a full time job, to calculate the income profiles. The six moments used in our estimation of the six parameters are computed as follows: we use the total student loan debt to income ratio at age 27, as we argue it proxies well both enrollment in additional education and the fact that it comes mostly from low indebted students. We then extract a constant and a linear trend from both the life cycle profiles of earnings and home ownership calculated from individuals aged 24-66 in Current Population Survey during years 2000-2018. In the first case, we use as a moment the ratio between constant and slope of earning profiles for workers with graduate degree and workers with only a bachelor degree. In the second case, we just aim at matching the overall life cycle profile of home ownership.

Table 2.6: Estimated Parameters

| Parameter | Description | Value | Standard Dev. |
|---|---|---|---|
| $\xi$ | Amenity Value of Grad School | \$55.171 | \$16.795 |
| $g_s$ | Grad School HC growth | 8.36% | 0.15% |
| $\beta_G/\beta_B$ | Skills Premium | 14.25% | 3.9% |
| $\zeta_1$ | Elasticity to Housing Service | 0.539 | 0.0069 |
| $\zeta_2$ | Housing Service | \$20.484 | \$695 |
| $\rho_{a,d}$ | Correlation (ability, debt) | −12.4% | 1.28% |

**Table A3** displays parameter estimates. [25] Standard errors, in the second column, tells us that estimates of parameters are precise - the only minor concern being represented by the objective function being less sensitive to changes in $\xi$. The model replicates well overall earnings dynamics, as in **Figure 2.3**. A better fit could be obtained by allowing a constant depreciation rate of human capital, which would induce a stronger concavity in the life cycle profile of earnings. However, the model matches pretty well average yearly income growth ( 1.9% in the data and 1.9% in our model), and earnings growth naturally slows because of income effects in the Ben-Porath problem. An extension of the model

[25] The amenity value of grad school is expressed in dollar terms, but does not correspond to $\xi$. To obtain it, we assume individuals in grad school are renters and have zero net liquid assets. Then the value is obtained by solving for the amount of consumption increase that would yield equivalent flow utility to grad school attendance.

Table 2.7: Target Moments

| Moments | Mean | |
|---|---|---|
| | Data | Model |
| A. Sample Means | | |
| Debt to Income at age 28[a] | 0.59 | 0.59 |
| Graduate to Bachelor Homeownership at age 38[c] | 1.0 | 0.77 |
| B. Regression Coefficients | | |
| Home ownership, constant[b] | 0.474 | 0.524 |
| Home ownership, slope[b] | 0.019 | 0.018 |
| Graduate to Bachelor earnings ratio, constant[b] | 1.10 | 1.04 |
| Graduate to Bachelor earnings ratio, slope[b] | 1.71 | 1.75 |

Sources:
a = B & B 2008 - 2012;
b = Current Population Survey, years 2000-2018, individuals with at least a bachelor degree, age 23-66, working full time
c = Current Population Survey, years 2000-2018, ratio between individuals with a bachelor degree and grad school education

Table 2.8: Untargeted Moments

| Moments | Mean | |
|---|---|---|
| | Data | Model |
| $\partial y_{t+1}/\partial d_t$ | 0.274 | 0.13 |
| $\partial \Delta y_{t,t+4}/\partial d_t$[a] | -0.185 | -0.216 |

$\Delta y_{t,t+4}$ is the 4-year growth in earnings after graduation

could perform the joint estimation of the Ben-Porath production function parameter and a linear depreciation rate of human capital.

While the model does not target anything but debt balances after graduation, it captures the front-loading incentives, as shown in **Table** 2.8: in the model, indebted graduates have 0.13% higher earnings for each 1% of additional student borrowing, but 0.26% lower earnings growth in the following four years.

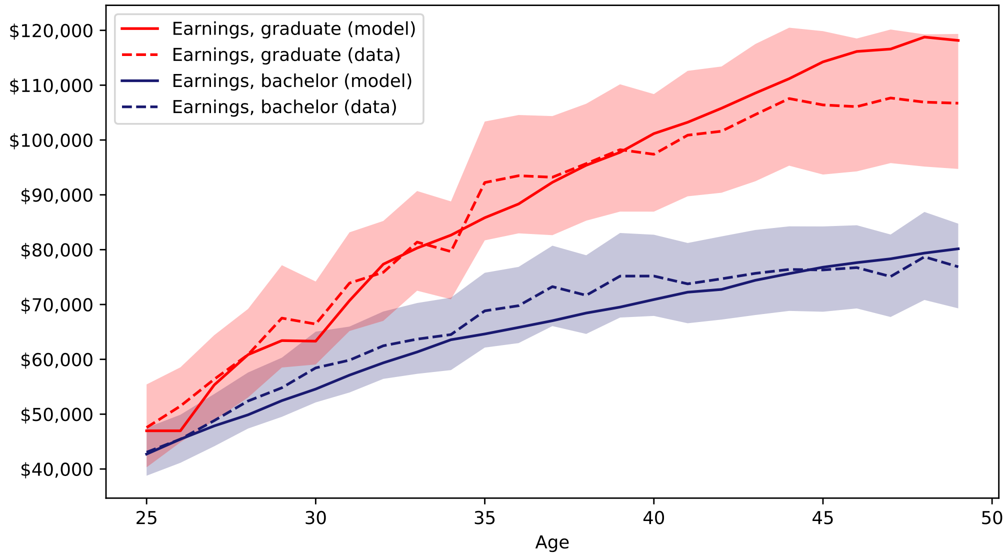



Figure 2.9: Life Cycle Profiles for Income, Model and Data
Data: Current Population Survey, years 2000-2018, population aged 24-66. Graduate School educated workers are all workers with an academic title higher than a 4-year college degree.

The patterns in enrollment, shown in **Figure 2.10**, replicates gradual entry into post graduate studies, and the level slightly more than a third of college educated workers pursuing further education. Because of the extreme assumption that human capital accumulated while working in one career is destroyed when switching[26], workers in the model tend to enroll slightly earlier than in the data. The slope in the life cycle pattern of home ownership in our model, as in **Figure 2.11**, is higher than in the data: especially in early years, home ownership is substantially lower, and then it catches up later in the working life. This can be explained with the choice of abstracting from bequest shocks in the model, as they would allow households to anticipate home ownership by relaxing their budget constraint. Decomposing the rate of home ownership by educational level, as done later in the text, we can see that the delay in purchases is almost entirely attributable to workers that pursue graduate studies.

[26]This choice is appropriate for some post-bachelor degrees, in particular the professional ones, where previous experience is hardly useful in the career implied by the degree. But it is clearly less appropriate to capture the role some other degrees, as MBAs and executive MBAs, play in the career of workers with some years of experience.

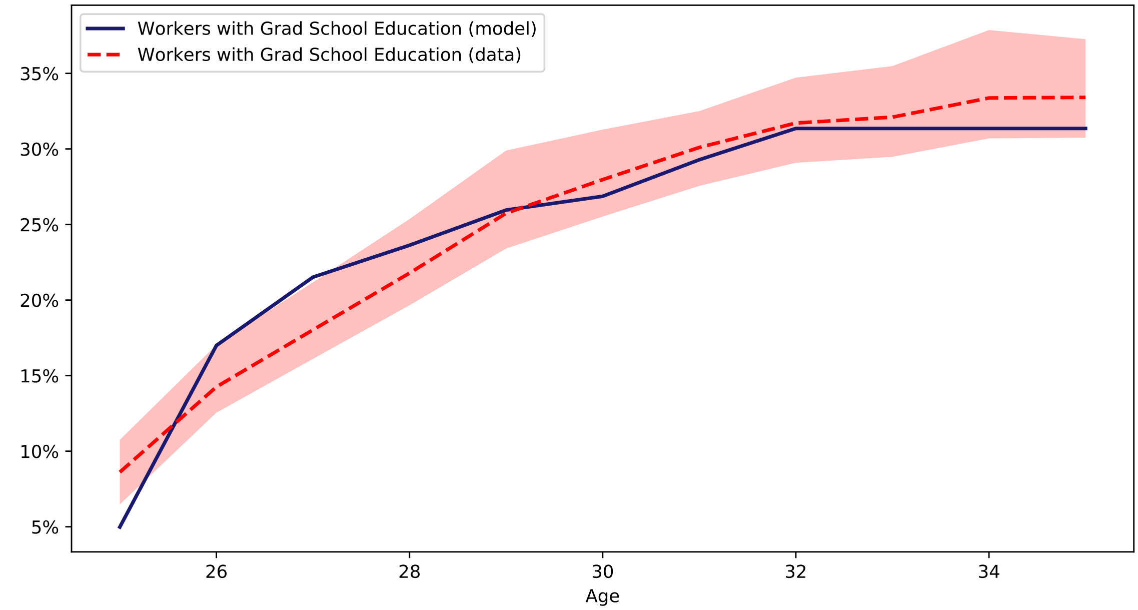



Figure 2.10: Life Cycle Profile of Enrollment in Graduate School
Data: Current Population Survey, years 2000-2018, population aged 23-66. Graduate School educated workers are all workers with an academic title higher than a 4-year college degree.

Another factor that limits earlier home purchases is our assumption of having just one size (and thus one price) available to workers. As results in the empirical section suggest, individuals who enter into home ownership earlier because college debt also purchase less expensive real estate. Extending the choice set by allowing individuals to differentiate their purchases with respect to price is the next step for improving the fit of our model to the data.

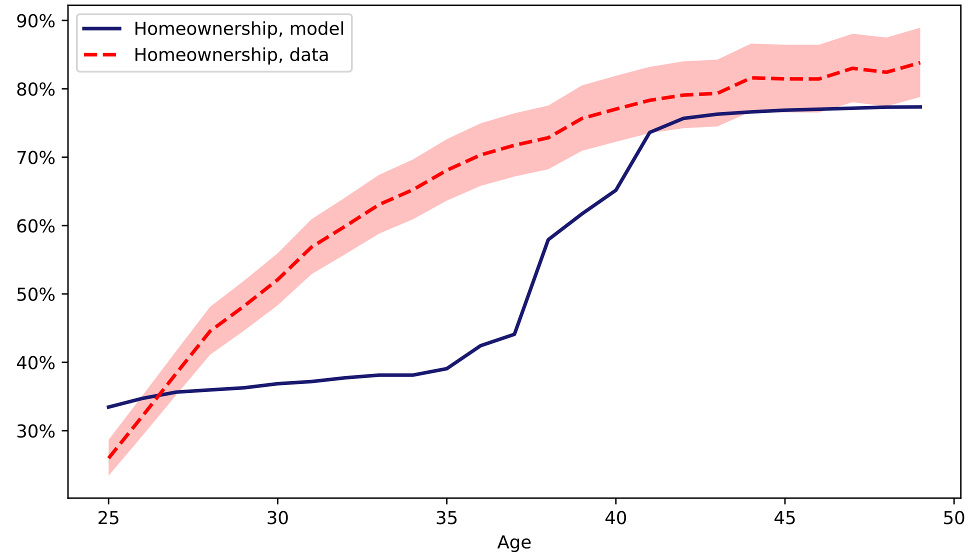



Figure 2.11: Life Cycle Profile of Entry into Home Ownership
Data: Current Population Survey, years 2000-2018, population aged 23-50.

## Identification

The model generates a large number of moments that can be used for estimation. Since interactions between each choice are quite complex, global identification is not possible even if one can attempt a one to one mapping between model parameters and empirical moments. Local identification, however, simply requires that the gradient of the model implied moments with respect to the parameter, $\partial h(y_{i,s}(\Theta))/\partial\Theta$, has full rank. This condition suggests that for a parameter to be identified, some subset of the vector of implied moments, must change when that particular parameter moves - see **Bazdresch, Kahn, and Whited** 2017.

We use the ratio of home ownership at age 37 by education groups, when the ratio is 1, as a target moment. The flow value of housing, $\zeta_1$ is identified by shifts in this ratio. The reason for this is that housing demand of workers with a post-bachelor degree is more sensitive to changes in the flow value: as it grows, not only less graduates enroll in post-bachelor programs, but workers with additional education try to enter into home ownership earlier, thus matching the home ownership of workers that only have a bachelor. A related parameter is the amenity value of graduate school: this, interestingly, is the value that mostly relates to the degree of sorting into additional education by ability, which in turn determines the ratio of the constants in income profiles. Hence, we can identify $\xi$ by looking at shifts in the relative level of incomes between post-bachelor and bachelor-only educated workers. We want to know about the two earnings parameters; $g_d$ ultimately determines the value of attending a post-bachelor degree. In the model, it also affects sorting, enrollment, and overall home ownership. However, only the relationship with the constant term in the home ownership life cycle profile is monotonous; as $g_d$ grows, it first increases debt to income ratios. Then it also starts to allow more indebted workers to postpone enrollment, and so the ratio decreases without affecting the ratio of the earnings profiles. The most straightforward identifying relationship is a result of higher $g_d$ simply increasing returns to graduate studies, thus allowing more workers to enter into home ownership at some point in time. More intuitively, the skill premium $\beta_g$ identifies the ratio of debt to income. The reason is that the skill premium, besides affecting earnings, is the main reason for increasing or decreasing early enrollment (remember the debt to income ratio is taken at age 25) in the model. Finally, we find that the correlation between debt and ability, $\rho_{a,d}$ is clearly identified by the ratio in the slopes of life cycle profiles of earnings. This is also intuitive: as the relationship between debt balances and ability becomes stronger (i.e. *more negative*), sorting into post bachelor degrees will unambiguously increase, other things not varying much. Hence, our model implies that growth in earnings differentials are mostly coming from increased borrowing of graduates with lower learning ability.

## 2.6 Results

In this section, we show the mechanisms behind the interaction between student debt, career choices and housing in our economy. We first analyze the performance of the baseline model in matching the empirical results presented in **Section** 2.3. We then provide some quantitative results that illustrate the contribution of each friction on the effects of debt on career choices and home ownership. Finally, we use the model to infer the effects of student debt on human capital and home ownership in the current environment.

### The Role of Student Debt on Earnings and Wealth

There are two main tradeoffs involved in the initial career choice. First, workers that not pursue additional education start with higher disposable income but then have lower income growth compared to the more human capital intensive careers. Second, income paths of bachelor educated workers are less volatile as human capital accumulation is a risky investment. This is immediate if looking at models' predictions for income, as in **Table** 2.9, where simulated results from the model are shown. Reported there is the ratio of income to average income for the same age. Workers whose undergraduate borrowing is above the median level start with higher earnings, because they are most likely to be working rather than being enrolled. After some years, the sorting effects of student loans start to affect earnings, and thus create a wide and persistent earnings gap.

Table 2.9: Earning Profiles by Debt Group

| Undergraduate Student Loans | Amount Borrowed | |
|---|---|---|
| | < \$22.560 | > \$22.560 |
| Age 25 | 0.90 | 1.11 |
| Age 28 | 1.036 | 0.94 |
| Age 30 | 1.06 | 0.92 |

In absence of frictions to borrowing or to the ability to transfer of human capital across careers, student loans should have no effect on career choices and human capital investment. In our model, the effects of student loans on career choices and lifetime earnings are ultimately the result of three main frictions. First, young workers face credit constraints that limit their ability to self insure against negative realizations of their human capital investment, or to smooth consumption through prolonged periods of unemployment. Second, human capital is not fully transferable across careers: we assume that any experience accumulated in one career path is lost when the worker transfers to the other career.

The assumption of limited human capital transferability has important consequences on the choice of using the career with a lower loading on human capital as an initial way to earn higher wages. If the worker wants to move on to the graduate school later in life, the decision will bear costs that increase in his (or her) tenure on the job. Finally, student loans follow a predetermined fixed repayment schedule and alternative repayment schemes are limited.[27]

Table 2.10: Sorting into Post-Bachelor Degrees

| Graduate School Enrollment | Student Debt | | Total |
|---|---|---|---|
| | < \$22.560 | > \$22.560 | |
| Graduate School Enrollment at Age 25 | | | |
| Low Ability | 2.8% | 2.33% | 2.64% |
| High Ability | 46.01% | 32.25% | 41.51% |
| Overall Graduate School Enrollment | | | |
| Low Ability | 13.52% | 11.20% | 12.74% |
| High Ability | 55.84% | 43.80% | 51.86% |

Skill Groups: below and above median ability level

[27] Our empirical analysis is focused on graduates that entered the labor market in 2008: during those years, less than 7% of borrowers enrolled in plans that allowed payments to be linked to earnings. After a series of reforms, enrollment in income based plans has increased substantially in the following decade.

**Table** 2.12 shows how entry into graduate school is affected by borrowing. More indebted students are significantly less likely to enroll. This happens for two reasons: on the one hand, while attending school allows to postpone payments, new debt is added to the existing one. Adding the burden of additional borrowing has compounding effects which put considerable pressure on future disposable consumption, thus discouraging enrollment. On the other hand, workers still have the possibility of starting to repay, while working, and then enrolling when their debt burden has reduced. The value of switching, however, decreases with tenure for two reasons: one is the mentioned nontransferability of human capital across careers. The other is a simpler horizon effect: as the worker gets older, and approaches the age where it would be optimal to start a mortgage, attending graduate school would imply a postponement of entry into home ownership because of the binding downpayment constraint, reducing the value of additional education.[28] Interestingly, the largest impact of borrowing on enrolment is on individuals with higher ability. A dampened sorting into careers points on the second effect being dominant: in fact, while relatively low ability individuals enroll smoothly over post graduation years, high ability individuals who would postpone getting additional education and switching career find the option becoming increasingly costly as they get older.

In order to understand the relative importance of different channels in affecting earnings over the life cycle, we decompose earnings growth differential between workers. To do so, we group workers based on different percentiles of student debt distribution. Notice one can obtain average earnings growth from **Equation** (2.3). Define $s_G$ as the share of post bachelor educated workers in a given group, $\bar{a}$ as the average ability and $F(h)$ as the distribution of human capital in that group. Average earnings growth will be defined as:

$$\Delta(w) = \int \underbrace{(s_G\beta_G + (1 - s_G)\beta_B)}_{\text{skill prem.}} \underbrace{\bar{a}}_{\text{ability}} \underbrace{((1-l)h)^{\alpha} dF(h)}_{\text{hum. capital}} \tag{2.16}$$

[28]We are not modelling household formation, and thus we are missing a potential counterbalancing effect, represented by adding a second income stream. However, as suggested by empirical evidence in the previous section, the impact of student debt on household formation goes in the same direction as the effects on home ownership. **Chang et al. 2019** points out that the recent decline in home ownership can be attributed to delayed household formation, providing additional support to the view that housing purchase and marriage can be considered as a joint choice.

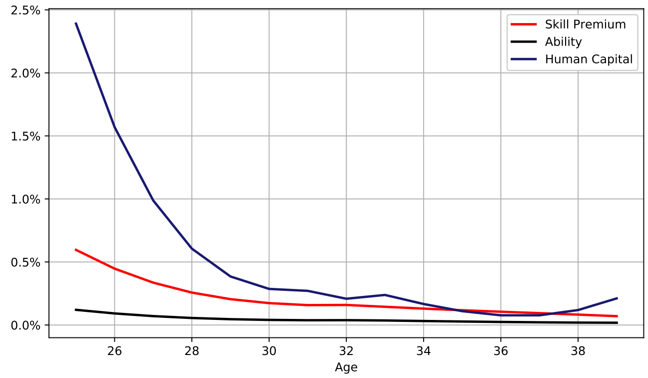



Figure 2.12: Decomposing Earnings Growth Differentials: Low versus High Debt

As argued above, highly indebted workers choose *flatter* earnings profiles. In **Figure** 2.12 we decompose the earnings growth differentials between the lowest and the highest tercile of workers ordered by undergraduate borrowing. Interestingly, ability plays a minor role in determining earnings growth differentials. This comes from two aspects. First, model estimates deliver small correlation between initial ability and debt. More importantly, though, there is ex post sorting that depends on the fact that human capital accumulation is risky at the individual level. High ability - high debt individuals with good human capital realizations experience both high initial wage growth and lower rate of enrollment in post bachelor degrees, as the option value of switching career decreases substantially after their human capital (and thus earnings) reach a higher level, hence they stay in the labor force while workers with lower ability and the same realizations find enrollment more valuable and thus enroll (disappearing temporarily from the workforce). Differences given by skill premium are coming from different post bachelor degree attendance patterns, as highlighted in **Table** 2.10. As highly indebted students catch up on enrollment, the contribution of skill premium decreases, but remains positive and eventually becomes the main factor driving earnings growth differentials as human capital investment behavior reaches a *plateau* for most workers.

Finally, we find that endogeonous human capital accumulation contributes for the lion share of earnings growth differentials. Two effects go in the same direction in determining this result. As one can see from the policy function of workers for the Ben-Porath human capital investment choice, highly indebted workers simply choose to invest less in order to have higher earnings in the current period. This is reinforced by career choices, as the same investment has higher returns for workers that enjoy a higher skill premium. This is consistent with earnings growth differentials being highest during the earlier years, as by **Figure** 2.3.

We now turn to housing. In our model, student loans affect home ownership through two main channels. On the one hand, highly indebted students are less likely likely to pursue extra education, which has lower returns to human capital, thus lower expected growth but also lower income risk. Thus, housing is a relatively more attractive investment at the start of the working career. On the other hand, student loan borrowers might face more difficulties in satisfying both the downpayment and the debt to income requirements for a mortgage. Since student loan payments reduce workers' disposable income, both investment in human capital and savings will be smaller. In addition, higher borrowing sorts workers into less human capital intensive careers, which negatively affects their lifetime earnings.

Table 2.11: Entry into Home Ownership

| Age of First Purchase | Non Borrowers | Borrowers | |
|---|---|---|---|
| | | < \$22.560 | > \$22.560 |
| Group | | | |
| All Workers | 30 | 29 | 29 |
| Only Bachelor[a] | 25 | 26 | 28 |

a= includes those who do not enroll in grad school at any point in time

As shown in **Table 2.11**, all those effects play a decisive role in determining the age at which households purchase their first home. From the second row it is possible to see that, for those workers who don't choose to enroll in graduate studies, borrowing affects home ownership mostly through the wealth effect. Hence, borrowers enter into home ownership later, with the delay growing nonlinearly in debt balances. In the aggregate, however, the role of post-bachelor enrollment dominates. As we can observe from the first row, the larger share of enrollment of non borrowers pushes home ownership to later in life. As balances grow, the two effects compensate each other - from **Table 2.10** we know that less than 27% of highly indebted workers undertake graduate studies, against an enrollment rate of 33% in the overall population.

There is a role of heterogeneity in ability, however, that dampens the effects of borrowing: once all workers share the same learning ability parameter $a$, the delaying effect of graduate school is stronger (see **Table A4** in Appendix). This happens because in the alternative model the population of workers that pursue additional education now has a *lower* average learning ability, and thus it takes more time for them on average to reap the benefits of additional education in terms of earnings. Also, notice this happens despite the model estimated with no ability heterogeneity features larger skill premium.

Non indebted workers initially invest in additional human capital and undertake riskier career paths. In going to graduate school, they face some periods of lower earnings, and subsequently some years of lower disposable income (because they borrow more to pay for graduate school tuition). There is also a consumption smoothing motive that explains later entry into homeownership. Workers with lower debt balances enter into housing market later because, as they sort into careers with higher income growth, they also find it optimal to delay home ownership until they can post the downpayment without impacting their disposable income in a substantial way. These two factors cause them to delay buying a house until they can afford it later in the life cycle. Before that, investment in human capital is more attractive . On the other hand, those who face a lower expected wage growth value housing as a more attractive investment, and then purchase as early as possible.

Looking at disposable income distributions in **Figure A.11** (in Appendix) helps understanding how the two effects play a role. Workers with post college education will have higher earnings, but facing the down payment will still force many of them to compress current consumption substantially. Postponing entry into home ownership is then consistent with willingness to smooth consumption over time, as their expected consumption growth is larger. On the other hand, workers with only a bachelor degree will have to compress their consumption anyway, through multiple periods of sustained savings or by accepting a period of lower consumption.

However, since their expected income growth is lower and more predictable, value of waiting is lower, and thus many opt into an early entry into home ownership. In the context of our two careers, the delay is particularly likely given that . This could lower the home ownership rate at the beginning, especially for the young who do not have much wealth. On the other hand, the increase in risk induces workers to increase precautionary savings, until the downpayment constraint is not binding, and inducing more transition from renting to home purchase.

## The Importance of Housing

To understand how relevant the housing channel is in determining education and career choices, we compare model predictions in a counterfactual scenario when workers are not allowed to access home ownership and remain renters during their whole life.[29] This way we are reducing available choices to workers compared to the baseline model, but we allow them to fully re-optimize given the new constraints they face. In this exercise, absent housing, agents can make different decisions about the timing of their investment in education (as well as about how much time to spend on human capital accumulation) than in the baseline.

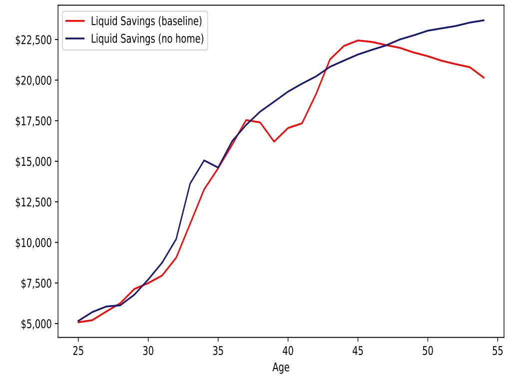
[29]An equivalent assumption is that we are imposing $\zeta_2 = 0$ while leaving all other parameters unchanged from the baseline estimation

Table 2.12: Enrollment with and without Housing

| Graduate School Enrollment | Student Debt | | Total |
|---|---|---|---|
| | < \$22.560 | > \$22.560 | |
| Graduate School Enrollment at Age 25 | | | |
| Baseline | 24.4% | 17.29% | 22.03% |
| No Housing | 52.83% | 29.4% | 45.02% |
| Overall Graduate School Enrollment | | | |
| Baseline | 34.68% | 27.50% | 32.29% |
| No Housing | 65.29% | 45.84% | 58.86% |

Two clear trends emerge: first, enrollment increases for both groups, and it does even more for those who borrowed less. Second, while highly indebted students still choose to postpone enrollment in order to reduce their debt balances, they do eventually enroll in the following years, while the baseline model suggests strong horizon effects. Switching costs (i.e. limited transferability of human capital) and borrowing constraints still matter, and determine the difference in enrollment patterns between graduates with different debt balances. Notice, however, that even in this context enrollment should not necessarily be identical along the debt distribution, to the extent that correlation with learning ability is different from zero - as turns out to be the case according to the estimates of our baseline model.

.5
Figure 2.13: Liquid Savings




.5
Figure 2.14: Earnings

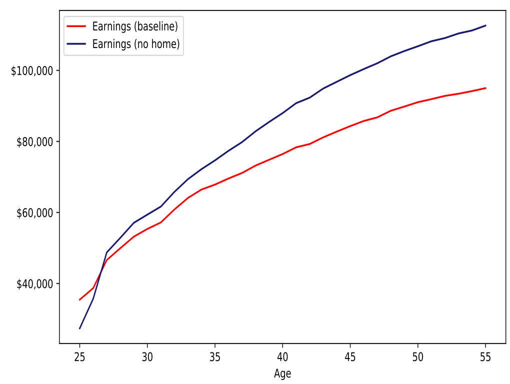
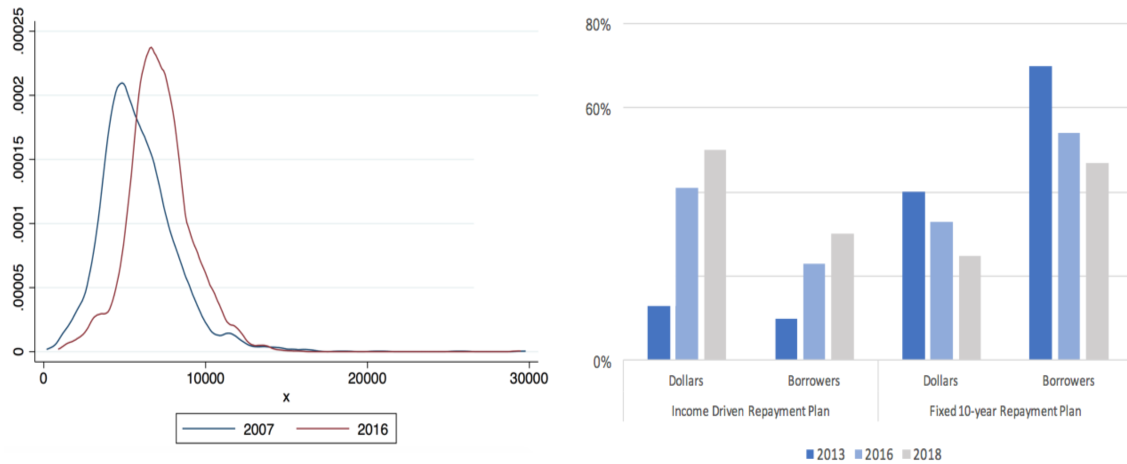



Figure 2.15: Baseline vs. no Homeownership

Increased enrollment in post-bachelor programs and the missing concern of savings in order to respect the downpayment constraint and then pay the mortgage have strong earnings effects, as shown in **Figure** 2.21. In this case, the change takes place mostly on the human capital investment side, as the pattern of savings is mostly unchanged - except for the later years, where a consumption smoothing motive drives workers in the counterfactual exercise into saving more.

## A "Debt to Equity Swap": Income Based Repayment

Figure 2.16: Evolution of Student Debt and Repayment Plans




Note: Figure 10a shows the distribution of yearly student loans awarded to full time first time undergraduates for 2007 and 2016. Figure 10b shows the percentage of student loan borrowers enrolled in repayment plans as well as the percentage amount of student debt each repayment plan represents. Sources: The Integrated Postsecondary Education Data System (IPEDS) and the Federal Student Aid Data.

Income Based Repayment plans are a popular solution to broadening access to higher education, as countries like Australia and Great Britain made them their baseline program for student finance (see **Chapman** 2016). They became available in the US to federal loan borrowers and depend on the borrower's discretionary income. Unlike fixed payment plans, there is no set horizon of loan repayment; instead, the borrower pays a percentage $\gamma$ of discretionary income each month until the loan is paid off or 20 to 25 years pass, in which case the remaining balance is forgiven (but included as taxable income). To be enrolled for these plans, borrowers have to report their income on an annual basis, and meet a series of eligibility criteria.

In this section, an income repayment plan in every period is introduced in the model as a baseline repayment scheme. The income repayment plan is defined to replicate the Pay As You Earn plan introduced in 2012: 10 percent of discretionary income for 20 years. At the end of the repayment period, remaining balances are forgiven and the forgiven amount is considered as additional income, to be taxed at a 25% rate. We rewrite the recursive problem in (2.11), as other problems are analogous:

$$V_{r,t}(x,e) = \max_{k',l}\{u(c,s) + \beta\mathbb{E}[EV_{t+1}(x',e')]\} \tag{2.17}$$

$$c + k' + P_r = (1+r)\cdot k + (1-\gamma)\cdot w_j(l,h)$$

$$h' = e^{z'}(h + a((1-l)h)^{\alpha})$$

$$d' = (1 + r_d)d - \gamma \cdot w_j(l, h)$$

$$k' \geq \phi(w_j),\ c \geq 0$$

where:

$$EV_t(x, e) = \max\left[V_{r,t}(x, e), V^g_{r,t}(x, e, s), V^o_t(x, e, m), V^{o,g}_t(x, e, s, m)\right]$$

A quantitative exercise is necessary to assess the extent to which income based repayment plans moderate the effects of initial student loan debt. On the one hand, enrollment in income driven repayment plans reduces the ratio of student loan payments to monthly wages, increasing disposable income. On the other hand, it can extend the repayment period significantly relative to a 10-year plan, thereby potentially increasing the total interest paid by the student loan borrower over the life of the loan.

The latter effect is the main reason why enrollment under IBR rises, but due mostly to higher enrollment by high ability graduates (see **Figure 2.20**). However, facing increasing payments during age 25-35, and a small risk of having to pay a lump sum tax in the late 30s because of residual balance forgiveness, workers under IBR delay entry into home ownership even more. After age 45, income effects start to dominate and overall home ownership grows compared to baseline.

A final remark on IBR connects to the increase in balances discussed in **Section 6.2**: as shown in this section, linking repayment to income does help alleviating financial constraints. Even if the program did not achieve full participation of graduates, the growth in IBR enrollment shown in **Figure 10b** can be credited with moderating the impact of the dramatic growth in undergraduate debt balances occurred between 2008 and 2016.

## Evaluating a Radical Policy: Debt Forgiveness for All

As student debt became a prominent issue in the public debate, various political actors have called for some sort of forgiveness plan. In particular, Senator Elizabeth Warren made student debt forgiveness a cornerstone of her political agenda.[30] In this chapter we introduce debt forgiveness under a balanced budget constraint, assuming the government can forgive all debt and then finance this program by spreading lump sum taxation over the life cycle of workers. This policy experiment that should serve as a benchmark for evaluating a more realistic forgiveness plan, which would most likely include some form of conditionality, and not be universal. Moreover, any forgiveness plan is going to be financed

[30]For details, see the Medium article (link here) where Warren articulates her proposal.

at least in part with some form of income, wealth or consumption taxes. In particular, Warren's proposal differs from ours in a few important aspects. An apparent major difference, yet quantitatively small in its effects, is the fact that debt cancellation would be capped at $50.000 - however, only 5% of borrowers as of 2018 has either higher balances or a household income that is high enough to exclude the borrowers from benefiting the plan. Another difference lies in the implementation: Warren proposes forgiveness of existing balances for graduates, and then the transition to a system with no fees charged to students of public schools. Our exercise is closer to the second scenario, as we wipe out all of undergraduate debt, and replace it with lump sum taxes levied over the life cycle to keep the reform under a balanced budget.

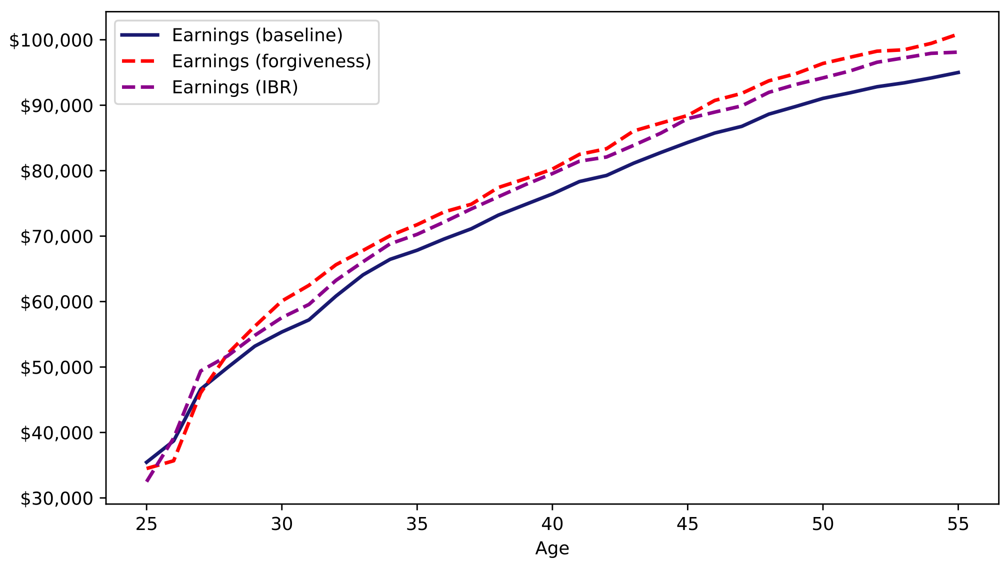



Figure 2.17: Baseline vs. Alternative Repayment Plans: Earnings

Figure 2.18: Enrollment

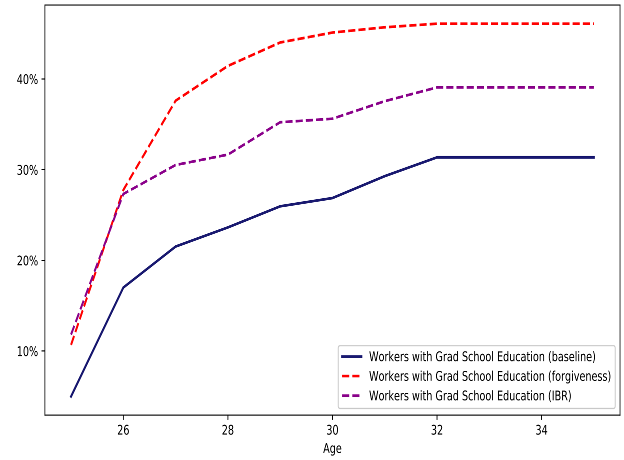



Figure 2.19: Home ownership

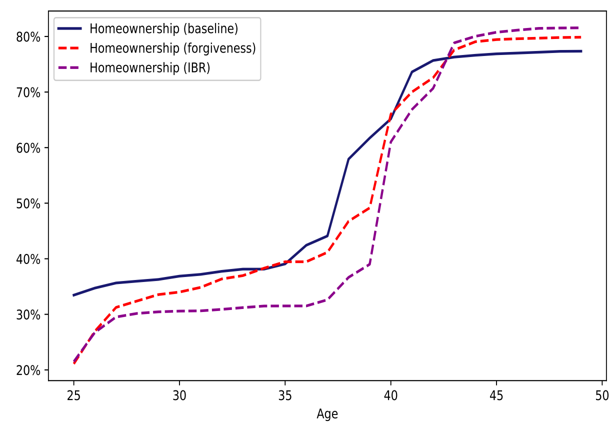



Figure 2.20: Baseline vs. Alternative Repayment Plans: Enrollment and Housing

According to our model, a forgiveness plan would have a large impact on post bachelor enrollment. While adoption of an IBR plan would increase enrollment in post-bachelor plans to 38%, forgiveness would bring it to 44%. It would both increase overall participation in graduate programs, and do it in particular doing early years. The second effect comes from the disappearance of the delaying motive that induces indebted graduates to postpone enrollment, while the first is a result of the relaxing of borrowing constraints on the same group. Given larger enrollment, it is not surprising that entry into home ownership is almost unchanged, as income effects move workers in the opposite direction. What happens, in fact, is that there is a small delay in access to home purchases, driven by increased enrollment.[31] The overall impact on earnings and later age home ownership, however, is not substantially larger than under the Income Based Repayment alternative

[31] A remark on Warren's plan is in order here. Senator Warren claims, citing **Mezza et al. 2016**, that student loans act as a drag on home ownership. She then goes on to suggest that a forgiveness plan would stimulate the housing market. As we argue here, student loans mostly affect the timing of access into home ownership rather than its overall level over the business cycle. Therefore, the plan should have little impact

plan. This comes from the differential impact the two plans have on sorting into graduate school: IBR achieves higher enrollment by a sharp increase in the enrollment of high ability individuals. On the other hand, forgiveness has negative effects on sorting, as it mostly increases the participation of workers with lower learning ability. This has a large impact on endogenous human capital accumulation which, as shown in **Section 6.2**, is the main driver of earnings growth. Lower ability workers enroll at a higher rate (and borrow for graduate studies), but their net monetary gain is small, and their endogenous human capital investment in age 30-35 is reduced compared to the baseline scenario where they had repaid their residual debt by that age.

## 2.7 Conclusions

What are the implications of higher levels of student debt on life cycle decisions? We find that graduating with higher levels of student debt causes higher earnings right after college, as well as earlier entry into home ownership, but lower income growth in the years after graduation. We then argue that this negative relationship is the result of student debt influencing career choices of college graduates. In particular, we find that individuals with higher levels of student debt are more likely to sort into careers that typically require less additional human capital after college, and specifically are less likely to enroll in post bachelor degree programs. We contribute to the existing literature by arguing that horizon effects determined by preferences for housing are an important channel for obtaining this result. While financial constraints are a necessary ingredient for initial financial conditions to affect life cycle outcomes, their interaction with a strong value attached to household formation is able to create a wide gap between outcomes of workers that start their careers with different debt balances.

Several policies have been advocated to help student loan borrowers. However, policy makers need guidance on the type of policies that are likely to be effective, from those that address liquidity constraints of borrowers to policies aimed to forgive a portion of student debt. We contribute to the policy debate by showing the merits of two alternative proposals. One, that is redistributive in nature, is to operate with a widespread forgiveness plan of all undergraduate debt, financed by lump sum taxes to be repayed over an extended period of time by the same cohort whose debt was forgiven. The other, that resembles closely the path chosen so far, aims at alleviating the burden of student debt by linking repayments to earnings. We show that an extension of existing policies is able to achieve results that

on housing and household formation. Our findings, however, are still consistent with **Mezza et al. 2016**, to the extent that they consider all student loans together, i.e. they include graduate debt. As we have shown, however, separating the two allows to highlight important channels and to better understand the impact of policy.

are quantitatively very similar to more ambitious forgiveness programs - namely, that the income based repayment plans that already attract a significant number of graduates are already an effective policy to reduce career and human capital accumulation distortions induced by student borrowing.

In future work, we plan to move in two directions. The first is to endogenize the college borrowing decision, by modeling undergraduate attendance, and nest our life cycle structure into a general equilibrium, overlapping generations framework. Those extensions will allow us to investigate the pattern of increased college attendance of the last decades, identifying its causes among shifts in technology, preferences and policy. After doing that, we will aim at comparing more comprehensive policies regarding education financing, human capital, and life cycle decisions. Another important question to address requires extending the housing decision part of the model to allow for location choice, and to make location choice relevant for career considerations. The decline of interstate migration in the U.S. has long been associated with reduced labor market dynamism, although recent research pointed at it resulting from a reduction of the component of occupation specific human capital. In presence of location choices, housing becomes not only an investment, but can also a drag or an obstacle to geographical and labor mobility. Changes in labor markets can thus have interactions with financial constraints, and generate interesting macroeconomic implications.

## References


Akers, B. (2012). *Excess Sensitivity of Labor Supply and Educational Attainment: Evidence from Variation in Student Loan Debt*. Tech. rep. Working Paper.

Athreya, Kartik et al. (2019). “Who Values Access to College?” In:

Bazdresch, Santiago, R Jay Kahn, and Toni M Whited (2017). “Estimating and testing dynamic corporate finance models”. In: *The Review of Financial Studies* 31.1, pp. 322–361.

Becker, Gary S (1962). “Investment in human capital: A theoretical analysis”. In: *Journal of political economy* 70.5, Part 2, pp. 9–49.

Ben-Porath, Yoram (1967). “The Production of Human Capital and the Life Cycle of Earnings”. In: *Journal of Political Economy* 75.4, pp. 352–365.

Bleemer, Zachary et al. (2017). “Echoes of Rising Tuition in Students’ Borrowing, Educational Attainment, and Homeownership in Post-Recession America”. In: *New York: Federal Reserve Bank of New York Staff* 820.

Bungartz, Hans-Joachim and Michael Griebel (2004). “Sparse grids”. In: *Acta numerica* 13, pp. 147–269.

Chang, Minsu et al. (2019). “A House Without a Ring: The Role of Changing Marital Transitions for Housing Decisions”. In: *2019 Meeting Papers*. 514. Society for Economic Dynamics.

Chapman, Bruce (2016). “Income contingent loans in higher education financing”. In: *IZA World of Labor*.

Chapman, S. (2015). *Student Loans and the Labor Market: Evidence from merit aid programs.* Tech. rep. Working Paper.

Cooper, Daniel and J. Christina Wang (2014). “Student Loan Debt and Economic Outcomes”. In: *Federal Reserve Bank of Boston, Current Policy Perspectives*.

Di Maggio, Marco, Ankit Kalda, and Vincent W. Yao (2019). “Second Chance: Life without Student Debt”. In: *Available at SSRN: https://ssrn.com/abstract=3376245*.

Eeckhout, Jan and Alireza Sepahsalari (2019). “The effect of asset holdings on worker productivity”. In: *Working Paper*.

Erickson, Timothy and Toni M Whited (2002). “Two-step GMM estimation of the errors-in-variables model using high-order moments”. In: *Econometric Theory* 18.3, pp. 776–799.

Field, E. (2009). *Educational debt burden and career choice: evidence from a financial aid experiment at nyu law school*. Tech. rep. AEJ: Applied Economics, 1(1):1-21.

Gallant, A Ronald and George Tauchen (1996). “Which moments to match?” In: *Econometric Theory* 12.4, pp. 657–681.

Garcia, CB and WI Zangwill (1981). *Pathways to solutions, fixed points, and equilibria*.

Gerald, Daniels and Andria Smythe (2019). *Student Debt and Labor Market Outcomes*. Tech. rep. AEA Papers and Proceedings, 109 : 171-75.

Gicheva, Dora and Jeffrey Thompson (2014). "The Effects of Student Loans on Long-Term Household Financial Stability". In: *UNCG Economics Working Papers* 14-2.

Gourieroux, Christian, Alain Monfort, and Eric Renault (1993). "Indirect inference". In: *Journal of applied econometrics* 8.S1, S85–S118.

Greenwald, Daniel (2018). *The Mortgage Credit Channel of Macroeconomic Transmission*. Tech. rep. MIT Sloan Research Paper No. 5184-16.

Griffy, Benjamin (2019). "Borrowing Constraints, Search, and Life-Cycle Inequality". In: *Discussion Papers 18-01, University at Albany, SUNY, Department of Economics.*

Haider, Steven J and Kathleen McGarry (2018). "Parental Investments in College and Later Cash Transfers". In: *Demography* 55.5, pp. 1705–1725.

Hause, J. (1972). "Earnings Profile: Ability and Schooling". In: *Journal of Political Economy* 80, S108–138.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2014). "Consumption and Labor Supply with Partial Insurance: An Analytical Framework". In: *The American Economic Review* 104.7, pp. 2075–2126.

Heiss, Florian and Viktor Winschel (2008). "Likelihood approximation by numerical integration on sparse grids". In: *journal of Econometrics* 144.1, pp. 62–80.

Houle, Jason and Lawrence Berger (2015). "Is Student Loan Debt Discouraging Home Buying Among Young Adults?" In: *Social Service Review* 89.4, S108–138.

Huggett, Mark, Gustavo Ventura, and Amir Yaron (2011). "Sources of Lifetime Inequality". In: *The American Economic Review* 101.7, pp. 2923–2954.

Ionescu, Felicia (2009). "The Federal Student Loan Program: Quantitative implications for college enrollment and default rates". In: *Review of Economic Dynamics* 12, pp. 205–231.

Ionescu, Felicia and Nicole Simpson (2016). "The Federal Student Loan Program: Quantitative implications for college enrollment and default rates". In: *Journal of Economic Dynamics and Control* 64, pp. 119–147.

Iskhakov, Fedor et al. (2017). "The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks". In: *Quantitative Economics* 8.2, pp. 317–365.

Johnson, Matthew T. (2013). "Borrowing Constraints, College Enrollment, and Delayed Entry". In: *Journal of Labor Economics* 31.4, pp. 669–725.

Kahn, Lisa B (2010). "The long-term labor market consequences of graduating from college in a bad economy". In: *Labour Economics* 17.2, pp. 303–316.

Kaplan, Greg, Kurt Mitman, and Giovanni L Violante (2019). "The Housing Boom and Bust: Model Meets Evidence". In:

Keane, Michael P and Kenneth I Wolpin (1997). “The career decisions of young men”. In: *Journal of political Economy* 105.3, pp. 473–522.

Looney, Adam and Constantine Yannelis (2015). “A crisis in student loans?: How changes in the characteristics of borrowers and in the institutions they attended contributed to rising loan defaults”. In: *Brookings Papers on Economic Activity* 2015.2, pp. 1–89.

Luo, Mi and Simon Mongey (2019). “Student debt and job choice: Wages vs. job satisfaction”. In: *New York University*.

McFadden, Daniel et al. (1973). “Conditional logit analysis of qualitative choice behavior”. In:

Menzio, Guido, Irina A Telyukova, and Ludo Visschers (2016). “Directed search over the life cycle”. In: *Review of Economic Dynamics* 19, pp. 38–62.

Mezza, Alvaro A. et al. (2016). “Student Loans and Homeownership”. In: *Finance and Economics Discussion Series. Washington: Board of Governors of the Federal Reserve System*.

Mincer, Jacob (1974). “Schooling, Experience, and Earnings”. In: *Columbia University Press*.

Oreopoulos, Philip, Till Von Wachter, and Andrew Heisz (2012). “The short-and long-term career effects of graduating in a recession”. In: *American Economic Journal: Applied Economics* 4.1, pp. 1–29.

Rothstein, Jesse and Cecilia Elena Rouse (2011). “Constrained after college: Student loans and early-career occupational choices”. In: *Journal of Public Economics* 95.1-2, pp. 149–163.

Smith Jr, Anthony A (1993). “Estimating nonlinear time-series models using simulated vector autoregressions”. In: *Journal of Applied Econometrics* 8.S1, S63–S84.

Stoyanov, M (2013). *User manual: Tasmanian sparse grids v2*. Tech. rep. 0. Technical report, Oak Ridge National Laboratory, Computer Science and Mathematics Division, One Bethel Valley Road, PO Box 2008, MS-6367, Oak Ridge, TN 37831-6164.

Weidner, J. (2016). *Does student debt reduce earnings?* Tech. rep. Working Paper.

Zhang, L. (2013). *Effects of college educational debt on graduate school attendance and early career and lifestyle choices*. Tech. rep. Education Economics, 21(2):154-175.

## 2.A Appendix

### B&B data

We use the restricted-use data and keep observations that have a positive value in the weight variable wte000, which represents the students who received a bachelor's degree in the 2007-08 academic year and responded to all interviews (2007-08, 2009, and 2012). The sample includes approximately 14,500 college graduates. We use and modify the following variables (available for public-use online in https://nces.ed.gov/datalab/powerstats):

Debt:

**Cumulative loan amount borrowed for undergraduate through 2007-08 (b1borat):** Indicates the cumulative amount borrowed from all sources for the respondent's undergraduate education through June 30, 2008. Does not include Parent PLUS loans. We log-transform this variable and deal with zero values by adding $1.

Post College outcomes:

**2012 Current Primary Job Salary (b2cjsal):** Indicates the respondent's annualized salary from their current or most recent primary job. Primary job is defined as the respondent's current or most recent job that lasted more than 3 months. We replace with a zero value the earnings of those who were not working at the time of the interview but reported the most recent earnings. We log-transform this variable and deal with zero values by adding $1.

**2009 Current Primary Job Salary (b1erninc):** Indicates the respondent's income from their current job as of the B&B:09 interview. For respondents with multiple jobs, salary is only for the primary job, the job at which the respondent worked the most hours. We log-transform this variable and deal with zero values by adding $1.

**2012 Primary Job: Occupation (b2cjocc33):** Indicates the occupation in which the respondent reported working in their current or most recent primary job as of the BB:12 interview, using 33 categories, based on the 2010 Standard Occupational Classification system developed by the Bureau of Labor Statistics. Primary job is defined as the respondent's current or most recent job that lasted more than 3 months; if more than one job meets these criteria, the job with the highest number of hours per week is selected. Variable categories are: Agriculture occupations; Air transportation professionals; Artists and designers; Business managers; Business occupations (non-management)...and Transport support occupations.

**2012 Current Value of Primary Residence (b2fhomval):** Indicates the approximate current value of the respondent's home(s), as reported by the respondent in the B&B:12 interview. We classify as home owners those observations with a value higher than zero.

For the value of the house, we consider houses with a value higher than \$100,000 and log-transform the variable.

**Highest degree attained since bachelor's as of 2012 (b2hideg):** Identifies the highest postsecondary degree or certificate the respondent had obtained after completing the 2007-08 bachelor's degree, as of the BB:12 interview. Variable categories are: Did not earn degree, Undergraduate certificate or diploma, Associate's degree, Additional bachelor's degree, Post-baccalaureate certificate...and Doctoral degree - other.

College Fixed Effects:

**Institution Sector in 2007/08 (sector4):** Indicates the sector of the 2007-08 bachelor's degree-granting institution, using five categories. Variable categories are: Public 4-year, Private nonprofit 4-year, Public 2-year, For-profit, and Others or attended more than one institution. WE keep Public 4-year and Private nonprofit 4-year colleges.

**Carnegie code (2005 basic, collapsed) for 2007-08 institution (cc2005c):** Indicates the Carnegie basic institution classification code, using collapsed categories, of the 2007-08 bachelor's degree-granting institution. Variable categories are: Associate's, Research and doctoral, Master's, Baccalaureate, and Special focus and other. We drop Associate's and Special focus and other institutions.

Individual Controls:

**Date of first postsecondary enrollment (pse_date):** Identifies the year and month, in YYYYMM format, when the respondent first enrolled in postsecondary education. We keep those students that enroll between 2002 and 2004.

**Student budget minus EFC in 2007-08 (sneed1):** Indicates the respondent's total need for need-based financial aid in 2007-08. We divide this variable by \$1,000.

**Dependency status in 2007/2008 (depend):** Indicates the respondent's dependency status during the 2007-08 academic year. Variable categories are: Dependent and Independent.

**SAT I score (tesatder):** Indicates the respondent's SAT I combined score, derived as either the sum of SAT I verbal and math scores or the ACT composite score converted to an estimated SAT I combined score using a concordance table from the following source: Dorans, N.J. (1999). Correspondences Between ACT and SAT I Scores (College Board Report No. 99-1).

**Field of Study (majors4y):** Indicates the respondent's major or field of study, using 10 categories, for the 2007-08 bachelor's degree. Variable categories are: Computer and information sciences; Engineering and engineering technology; Bio and phys science, sci tech, math, agriculture; General studies and other; Social Sciences, Humanities, Health-care, Business, Education and Other Applied. We classify them in three categories: STEM and health-care, Social Sciences and Business, Other.

**Race/Ethnicity (race):** Indicates the respondent's race/ethnicity with Hispanic or Latino origin as a separate category. Variable categories are: White, Black or African American, Hispanic or Latino, Asian, American Indian or Alaska Native, Native Hawaiian, other and More than one race. We classify them into four categories: White, Black, Latino, Asian, Other.
**Gender (gender):** Indicates the respondent's sex. Variable categories are: Male and Female.

## IPEDS data

Using harmonized college identifiers, we merge the B&B individual level data with institution level from the Institutional Post-Secondary Database (**IPEDS**). We use the IPEDS data in order to get information about the cost of attendance as well as the amount of grants and loans at the institutional level. We use the following variables for 2004-2007 from the IPEDS data center:

<u>College Student Debt:</u>

**Average amount of student loans awarded to full-time first-time undergraduates (loan):** Any monies that must be repaid to the lending institution for which the student is the designated borrower. Includes all Title IV subsidized and unsubsidized loans and all institutionally- and privately-sponsored loans. Does not include PLUS and other loans made directly to parents.
**Percent of full-time first-time undergraduates awarded student loans (ploan):** Percentage of full-time, first-time degree/certificate-seeking undergraduate students who were awarded student loans.

<u>Institutional Grants:</u>

**Average amount of institutional grant aid awarded to full-time first-time undergraduates (grant):** Scholarships and fellowships granted and funded by the institution and/or individual departments within the institution, (i.e., instruction, research, public service) that may contribute indirectly to the enhancement of these programs. Includes scholarships targeted to certain individuals (e.g., based on state of residence, major field of study, athletic team participation) for which the institution designates the recipient.
**Percent of full-time first-time undergraduates awarded institutional grant aid (pgrant):** Percentage of full-time, first-time degree/certificate-seeking undergraduate students who were awarded institutional grants (scholarships/fellowships).

<u>Grant-to-Aid:</u>

Some of the institutions have a missing value in grants or loans and at the same time the percentage of students who were awarded grants or loans is zero. We substitute these observations with a zero value in grants or loans. We then drop all colleges with a grant-to-aid of 0 or 100 in any of the six years (2002-2007).

Given that the average sum (and percent) of institutional grant and loan amounts are not available for 2002-2007, we construct the total institutional grant-to-aid ratio in the following way:

$$aid_{j,t} = ploan_{j,t}loan_{j,t}+pgrant_{j,t}grant_{j,t} = \left(\frac{TotalDebt_{j,t}}{Indebted_{j,t}}\right)\left(\frac{Indebted_{j,t}}{Students_{j,t}}\right)+\left(\frac{Grant_{j,t}}{Recipient_{j,t}}\right)\left(\frac{Recipient_{j,t}}{Students_{j,t}}\right)$$

$$x_{j,t} = \frac{\left(\frac{Grant_{j,t}}{Recipient_{j,t}}\right)\left(\frac{Recipient_{j,t}}{Students_{j,t}}\right)}{aid_{j,t}}$$

## Checking covariates

Table A1: Grant-to-Aid and College Characteristics

| | Stud/Faculty (1) | Grad. Rate (2) | Ret. Rate (3) | Tuition (4) | Gov Grant (5) |
|---|---|---|---|---|---|
| Δ Grant to Aid (2004-2007) | -0.05 | 0.13 | 0.15 | 57.3 | -10.1 |
| | [0.03] | [0.07] | [0.08] | [47.4] | [7.39] |
| Grant to Aid (2004) | -0.03 | 0.21** | 0.21 | 53.8 | -7.81 |
| | [0.02] | [0.07] | [0.16] | [64.6] | [7.19] |
| College FE | Y | Y | Y | Y | Y |
| Observations | 1,282 | 1,282 | 1,282 | 1,282 | 1,282 |

Clustered Standard errors in brackets.

Table A2: Grant-to-Aid and Undergraduate Students Characteristics

| | % Black (1) | %Age<25 (2) | %Full-time (3) | Avg. SAT (4) | % Income < 30,000 (5) |
|---|---|---|---|---|---|
| Δ Grant to Aid (2004-2007) | 0.01 | 0.21 | 0.09 | -0.08 | 0.68 |
| | [0.07] | [0.11] | [0.07] | [0.0.38] | [0.54] |
| Grant to Aid (2004) | -0.10* | 0.17 | -0.169 | 0.01 | 0.71 |
| | [0.04] | [0.11] | [0.11] | [0.10] | [0.64] |
| College FE | Y | Y | Y | Y | Y |
| Observations | 1,282 | 1,282 | 1,282 | 1,282 | 1,282 |

Clustered Standard errors in brackets.

## Home-Ownership Across Cohorts

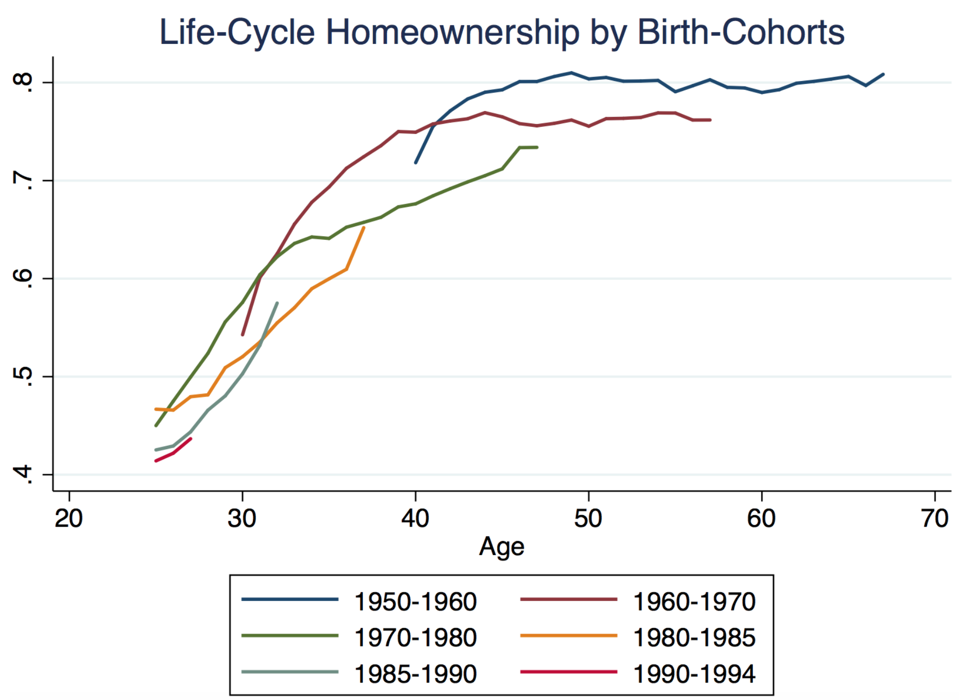



Figure 2.21: Evolution of First-Time Home-Ownership by Cohorts

## Solution Method

### Discrete-Continuous Choices

We illustrate how we take into account discrete choices with the problem of an employed renter with student loans, as in the Bellman Equation (2.11). For illustrative purposes only, we assume no borrowing constraints. If the worker had no discrete choices to make, the Bellman equation for the optimal consumption of a worker would satisfy the following first order condition known as the Euler equation:

$$0 = u_c'(c,s) - \beta(1+r)\mathbb{E}\left(u_c'(c',s')\right) \tag{2.1}$$

However, since at any period the renter worker can choose two discrete choices (to become a homeowner or switch career), the problem at the state vector point $\{a,h,j,d,e,t\}$ involves solving for all the possible combinations of available discrete choices.
Following **Iskhakov et al. 2017**, we assume instead that the discrete choices are affected by choice-specific taste shocks, $\sigma_e\varepsilon_t$, i.i.d. Extreme Value type I distributed with scale parameter $\sigma_\varepsilon$ as in **McFadden et al. 1973**.
Taking again the value function in (2.11). Abstracting from career and repayment choice, and focusing only on the home-ownership decision, the expected value of the future value function becomes:

$$\begin{aligned} \mathbb{E}[V'] = \max \mathbb{E}[V_r(k',h',j',m',d',e',t+1)], \mathbb{E}[V_{o,\lambda}(k',h',j',m',d',e',t+1)] = \\ = \max \mathbb{E}[V_r(\cdot,t+1) + \sigma_\varepsilon\varepsilon(o)], \mathbb{E}[V_{o,\lambda}(\cdot,t+1) + \sigma_\varepsilon\varepsilon(r)] = \\ = \sigma_\varepsilon \log\Big(\exp\{V_r(\cdot,t+1)/\sigma_\varepsilon\} + \exp\{V_{o,\lambda}(\cdot,t+1)/\sigma_\varepsilon\}\Big) \end{aligned} \tag{2.2}$$

Thus, the Euler equation for a renter can then be written as:

$$\begin{aligned} 0 = u_c'(c,s) - \beta(1+r)\mathbb{E}\Big[u_c'(c',s'>1)\cdot P(s'>1|k',h',j',m',d',e') \\ + u_c'(c',s'=1)\cdot P(s'=1|k',h',j',m',d',e')\Big] \end{aligned} \tag{2.3}$$

where $P(s'>1)$ and $P(s'=1)$ are conditional choice probabilities given by the binomial logit formula:

$$\begin{aligned} P(s'>1|k',h',j',m',d',e') = \frac{\exp\{V_{o,\lambda}(\cdot,t+1)/\sigma_\varepsilon\}}{\exp\{V_{o,\lambda_H}(\cdot,t+1)/\sigma_\varepsilon\} + \exp\{V_r(\cdot,t+1)/\sigma_\varepsilon\}} \\ P(s'=1|k',h',j',m',d',e') = \frac{\exp\{V_r(\cdot,t+1)/\sigma_\varepsilon\}}{\exp\{V_{o,\lambda}(\cdot,t+1)/\sigma_\varepsilon\} + \exp\{V_r(\cdot,t+1)/\sigma_\varepsilon\}} \end{aligned} \tag{2.4}$$

### Borrowing constraints

Solving (2.2) requires taking care of an additional issue. Formally, given the state $S$ and indicating the Euler equation as $\phi : S \times \mathbb{R}^m \to \mathbb{R}$, and the policy function as $k' : S \times \mathbb{R}^m \to \mathbb{R}$, one needs to find policy and multiplier $(k',\mu) \in \mathbb{R}\times\mathbb{R}$ s.t.

$$\phi(S,k',\mu) = 0\ ,\ k' \geq \phi \perp \mu \geq 0 \tag{2.5}$$

Following **Garcia and Zangwill 1981**, this problem can be transformed into a system of two equations, and can then be solved using standard solution algorithms for root finding. Define a variable $\alpha$ such that:

$$\alpha \equiv \begin{cases} \mu, & \text{if } \mu \geq 0,\, k' = \phi \\ -k', & \text{if } \mu = 0,\, k' \geq \phi \end{cases} \tag{2.6}$$

and

$$\begin{aligned} \alpha^{+} &= (\max(0, \alpha))^{k} \\ \alpha^{-} &= (\max(0, -\alpha))^{k} \end{aligned} \tag{2.7}$$

where $k \in \mathbb{N}^{+}$. The variable acts like a "penalty" when the constraint is violated, forcing the algorithm to search in the feasible set. The problem can be rewritten as finding policies and $\alpha$ such that:

$$\phi(S, k', \alpha^{+}) = 0 \;,\; k' - \alpha^{-} = 0 \tag{2.8}$$

## Optimal Weight Matrix for GMM

We follow **Erickson and Whited** 2002 in computing the optimal weight matrix $\hat{\Omega}^{-1}$ from the following formula for clustered covariance:

$$\hat{\Omega} = \frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\psi_{h(x_{i,t}))}\right)\left(\sum_{t=1}^{T}\psi_{h(x_{i,t}))}\right)' \tag{2.9}$$

in which $\psi_{h(x_{i,t}}$ is the vector of influence functions for the empirical moments $h(x_{i,t})$. Deriving the influence functions for choice of moments is relatively straightforward. Take any subset of $h(x_{i,t})$ and denote it as $\theta$. For those moments that are obtained from simple averages, i.e. $\hat{\theta} = \mathbb{E}(x_i)$, the influence function can be computed simply as:

$$\psi_{\hat{\theta}}(x) = x - \mathbb{E}(X) \tag{2.10}$$

In the case of linear regression coefficients, we need to get influence function for the slope and the constant. The slope is $\hat{\theta}(\beta) = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$. Then:

$$\begin{aligned}\psi_{\hat{\theta}(\beta)}(x,y) &= \frac{(x-\mathbb{E}(X))(y-\mathbb{E}(Y))-\text{Cov}(X,Y)}{\text{Var}(X)} - \frac{\left((x-\mathbb{E}(X))^2-\text{Var}(X)\right)\text{Cov}(X,Y)}{(\text{Var}(X))^2} = \\ &= \frac{(x-\mathbb{E}(X))(y-\mathbb{E}(Y))-\beta(x-\mathbb{E}(X))}{\text{Var}(X)} = \frac{(x-\mathbb{E}(X))}{\text{Var}(X)}\left[(y-\mathbb{E}(Y))-\beta(x-\mathbb{E}(X))\right]\end{aligned} \tag{2.11}$$

The constant is instead $\hat{\theta}(\alpha) = \mathbb{E}(y) - \frac{\text{Cov}(X,Y)}{\text{Var}(X)}\mathbb{E}(x) = \mathbb{E}(y) - \frac{\mathbb{E}(XY)\mathbb{E}(X)-(\mathbb{E}(X))^2\mathbb{E}(Y)}{\text{Var}(X)}$. Then:

$$\begin{aligned}\psi_{\hat{\theta}(\alpha)}(x,y) = -&\frac{(xy-\mathbb{E}(XY))\mathbb{E}(X)+(x-\mathbb{E}(X))\mathbb{E}(XY)-(y-\mathbb{E}(Y))(\mathbb{E}(X))^2-2(x-\mathbb{E}(X))\mathbb{E}(X)\mathbb{E}(Y)}{\text{Var}(X)}+ \\ &+y-\mathbb{E}(Y)+\frac{\left((x-\mathbb{E}(X))^2-\text{Var}(X)\right)\left(\mathbb{E}(XY)\mathbb{E}(X)-(\mathbb{E}(X))^2\mathbb{E}(Y)\right)}{(\text{Var}(X))^2} = y-\mathbb{E}(Y)- \\ &-\frac{(xy-y\mathbb{E}(X))\mathbb{E}(X)+(x-\mathbb{E}(X))(\text{Cov}(XY)-\mathbb{E}(X)\mathbb{E}(Y))}{\text{Var}(X)} - \frac{\left((x-\mathbb{E}(X))^2-\text{Var}(X)\right)}{(\text{Var}(X))^2}\end{aligned} \tag{2.12}$$

Finally, we use the ratio of regression coefficients. Take the ratio of slopes $\hat{\theta}(\beta_g/\beta_b)$. Then by the chain rule:

$$\psi_{\hat{\theta}(\beta_g/\beta_b)}(x_g,y_g,x_b,y_b) = \frac{\psi_{\hat{\theta}(\alpha)}(x_g,y_g)\beta_b - \psi_{\hat{\theta}(\alpha)}(x_b,y_b)\beta_g}{\beta_b^2} \tag{2.13}$$

And similarly, one can obtain the influence function for the ratio of constants.

## Model without heterogeneity in ability

We estimate the same model as in **Section 5**, assuming no heterogeneity in ability.

Table A3: Estimated Parameters

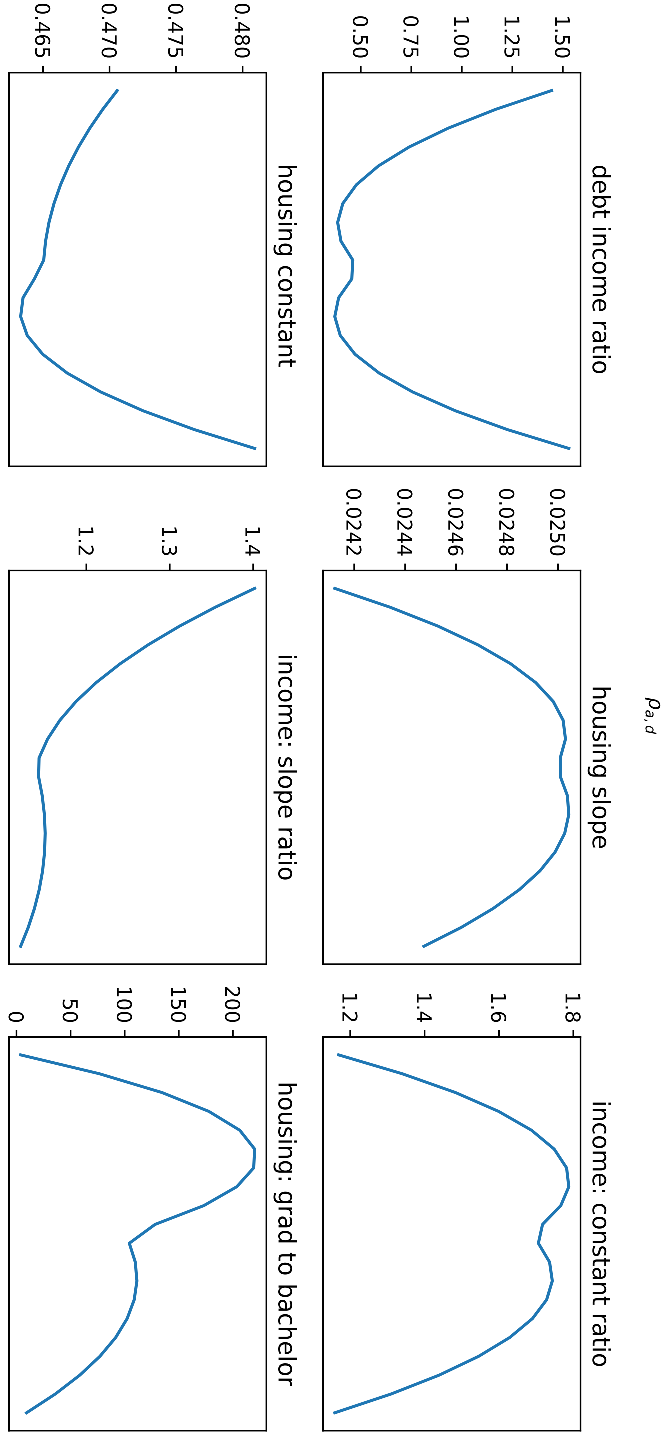
| Parameter | Description | Value | Standard Dev. |
|---|---|---|---|
| $\xi$ | Amenity Value of Grad School | \$74.080 | \$18.080 |
| $g_s$ | Grad School HC growth | 8.99% | 0.23% |
| $\beta_G$ | Skills Premium | 12.7% | 2.4% |
| $\zeta_1$ | Elasticity to Housing Service | 0.605 | 0.005 |
| $\zeta_2$ | Housing Service | \$22.760 | \$920 |

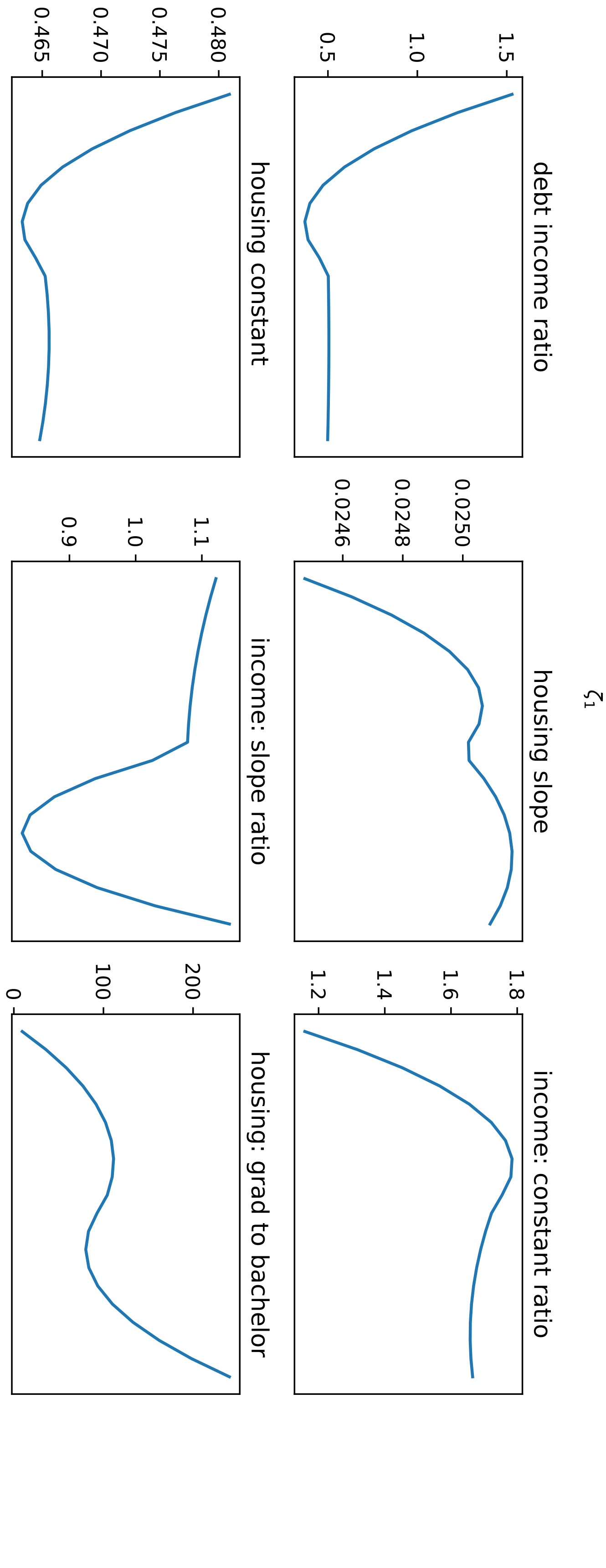
Table A4: Entry into Home Ownership

| Age of First Purchase | Non Borrowers | Borrowers | |
|---|---|---|---|
| | | < \$22.560 | > \$22.560 |
| Group | | | |
| All Workers | 34 | 26 | 29 |
| Only Bachelor[a] | 24 | 25 | 28 |

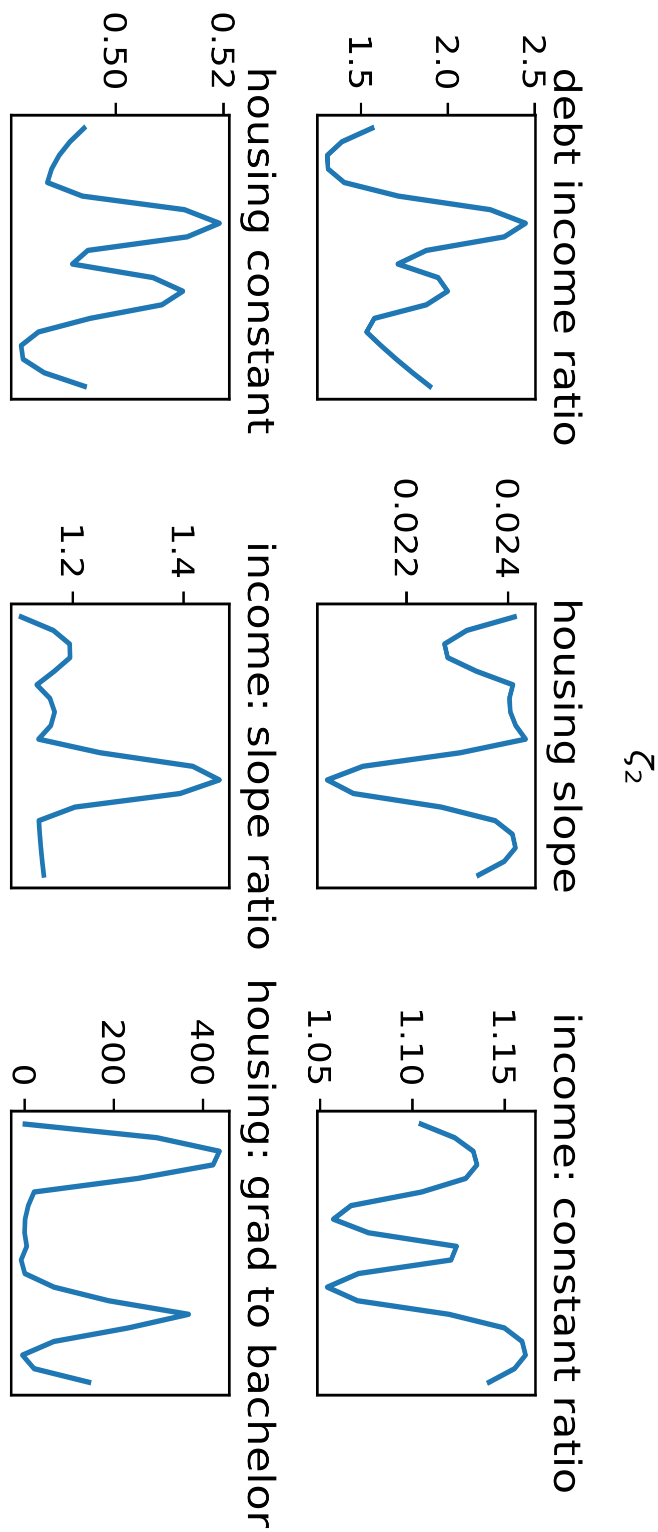
a= includes those who do not enroll in grad school at any point in time

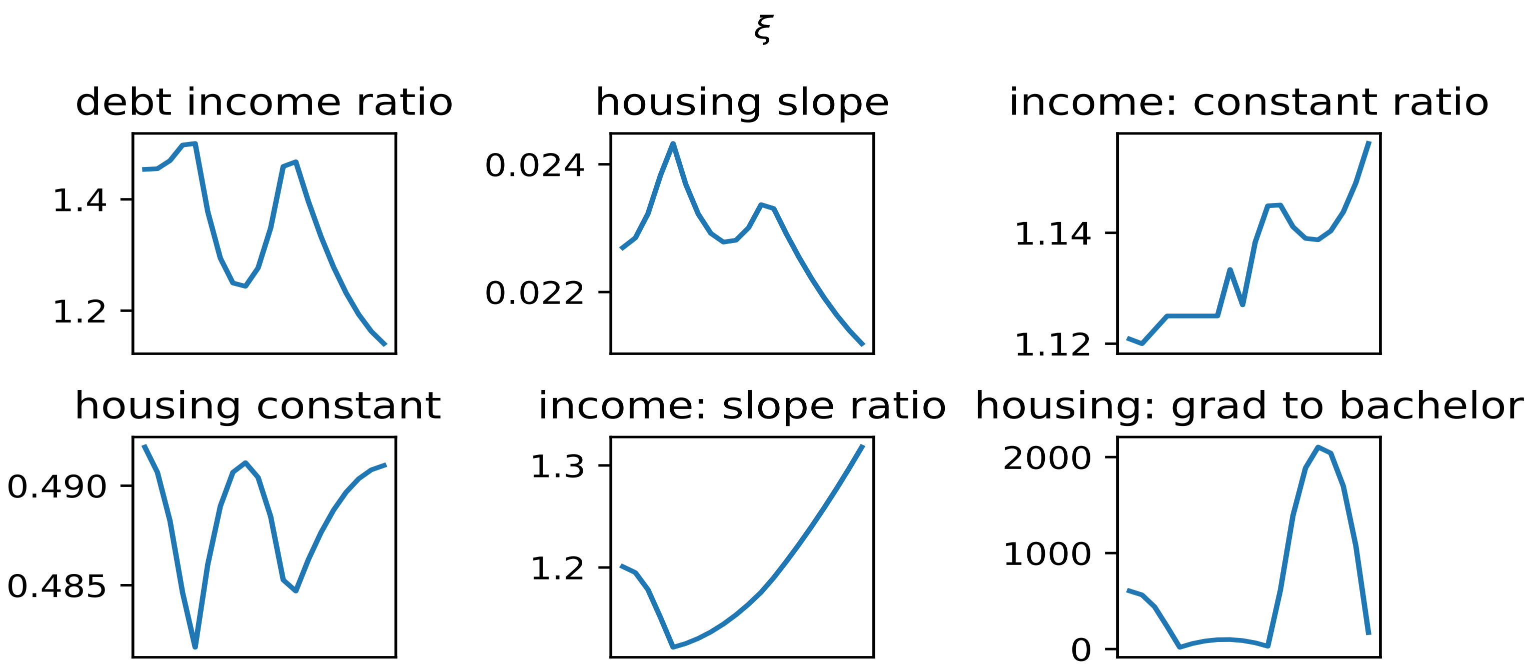
## Identification

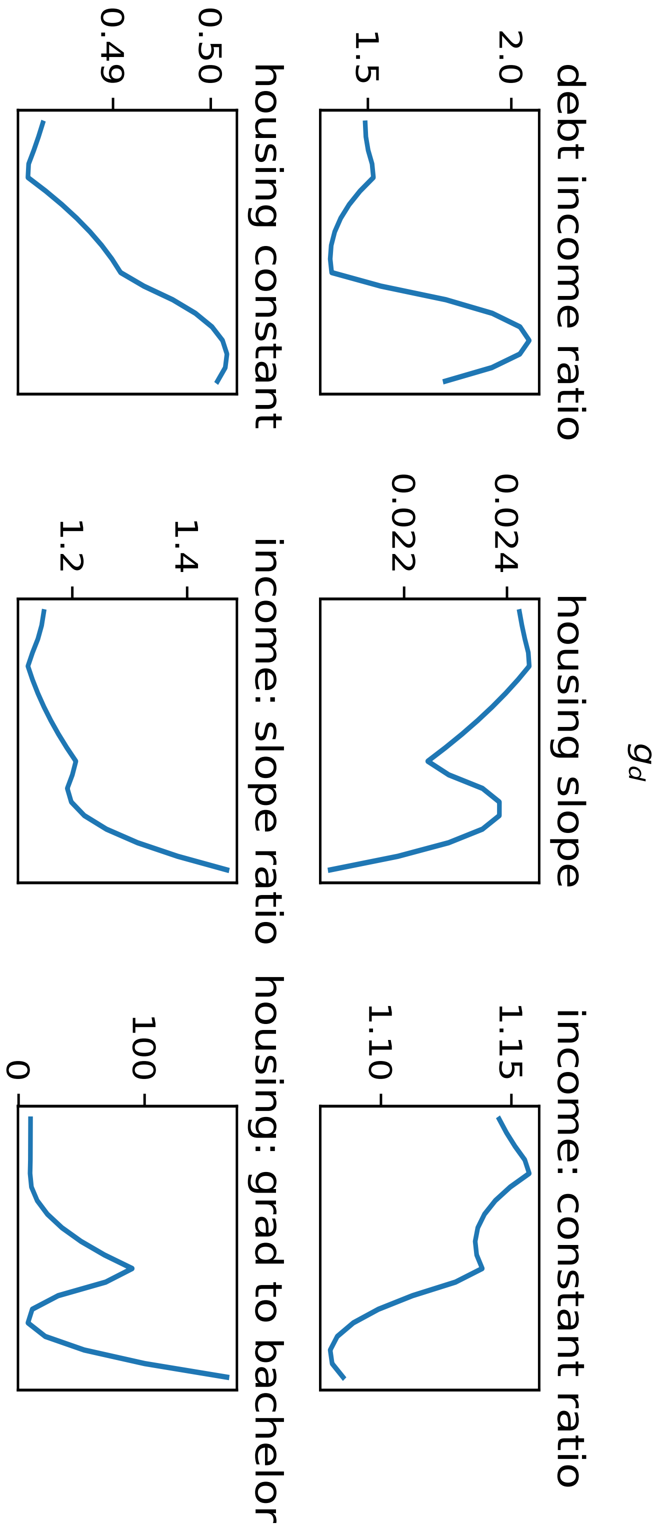


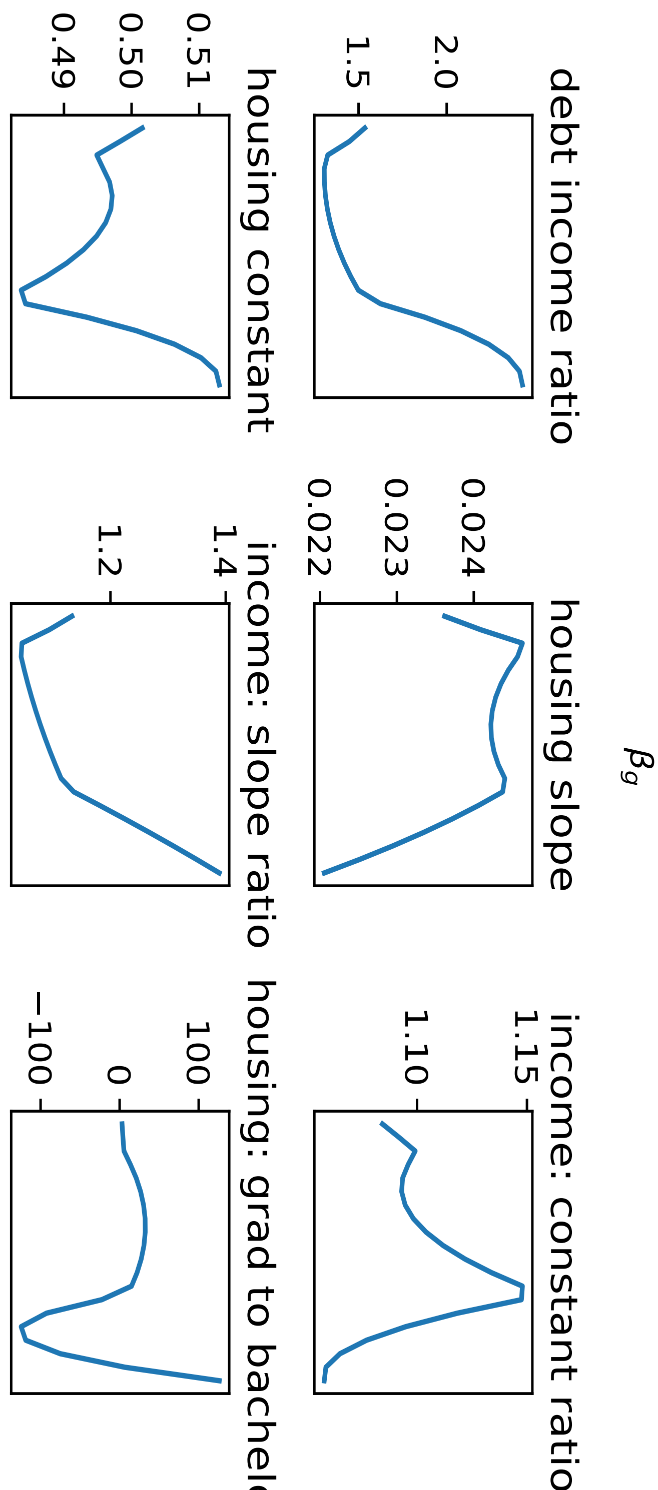


$\xi_1$
debt income ratio
1.5
1.0
0.5
housing slope
0.0250
0.0248
0.0246
income: constant ratio
1.8
1.6
1.4
1.2
housing constant
0.480
0.475
0.470
0.465
income: slope ratio
1.1
1.0
0.9
housing: grad to bachelor
200
100
0


$\zeta_2$
debt income ratio
2.5
2.0
1.5
housing slope
0.024
0.022
income: constant ratio
1.15
1.10
1.05
housing constant
0.52
0.50
income: slope ratio
1.4
1.2
housing: grad to bachelor
400
200
0

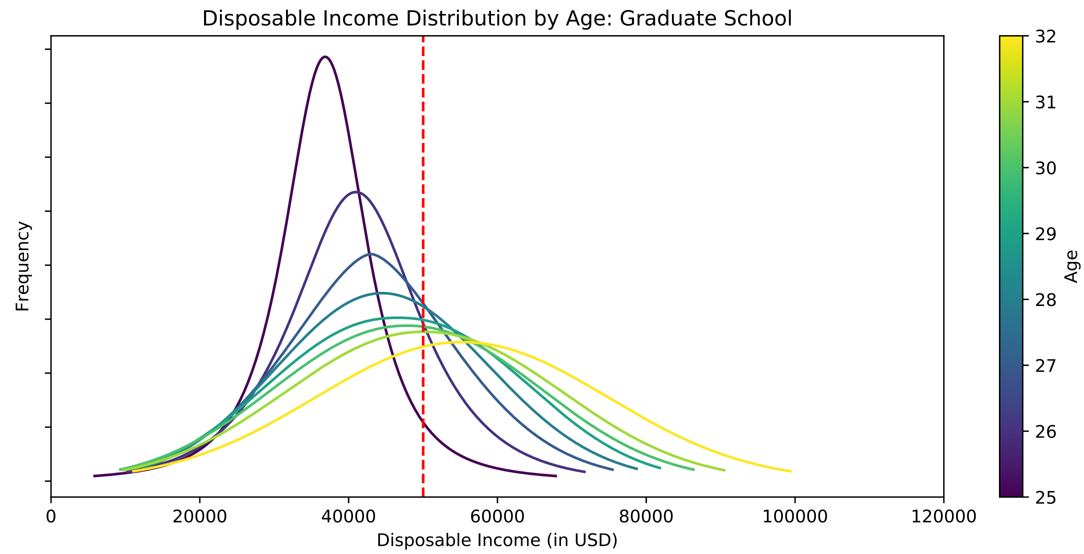

ξ
debt income ratio
1.4
1.2
housing slope
0.024
0.022
income: constant ratio
1.14
1.12
housing constant
0.490
0.485
income: slope ratio
1.3
1.2
housing: grad to bachelor
2000
1000
0

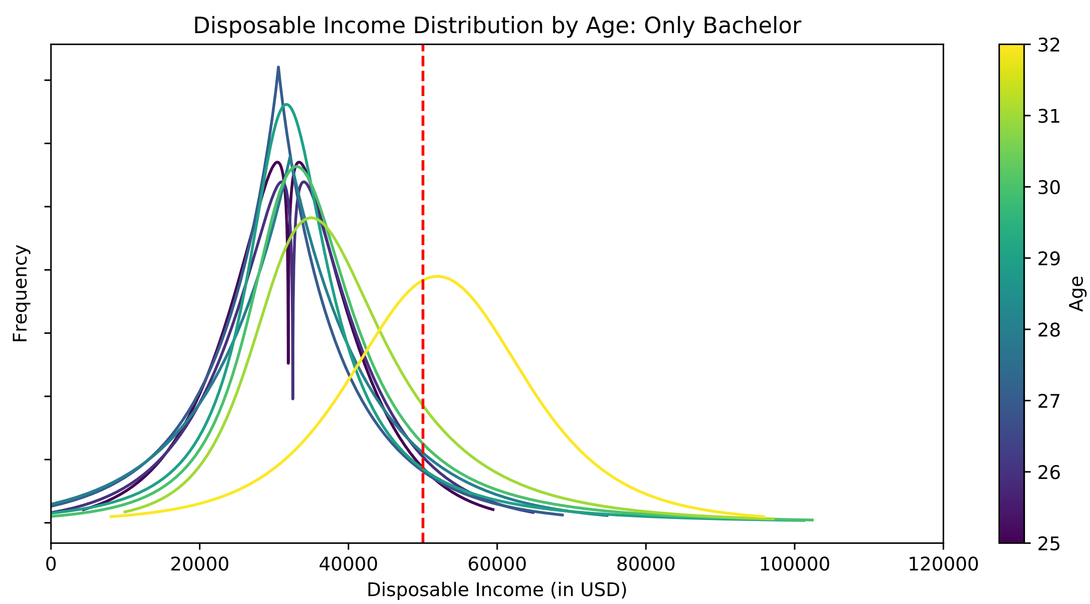

$g_d$
debt income ratio
2.0
1.5
housing slope
0.024
0.022
income: constant ratio
1.15
1.10
housing constant
0.50
0.49
income: slope ratio
1.4
1.2
housing: grad to bachelor
100
0



$\beta_g$
debt income ratio
housing slope
income: constant ratio
housing constant
income: slope ratio
housing: grad to bachelor

## Additional Figures




Distribution of yearly disposable income, i.e. labor wages plus net liquid asset holdings, minus debt payments and housing expenditures for workers with graduate school education. The red line represents the downpayment constraint

Figure 2.22: Graduate School Educated Workers and Downpayment Constraint

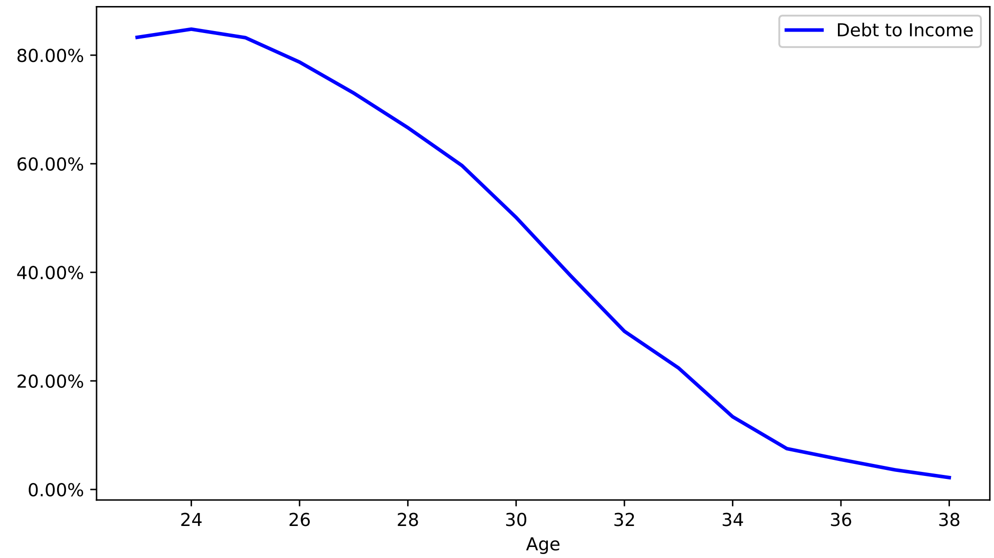
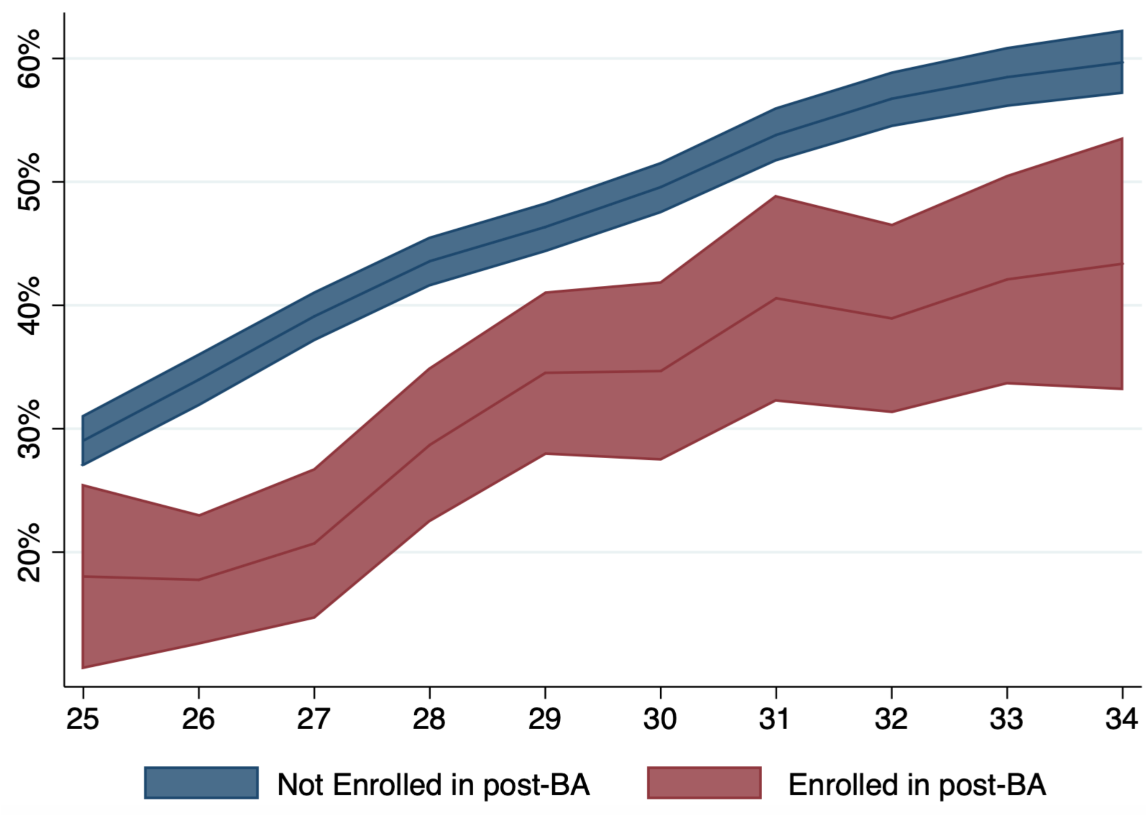


Figure 2.23: Ratio of Student Loan Debt to Income, All Workers



Data:

Current Population Survey for workers aged 25-34, years 2000-2018
Figure 2.24: Entry into Home-Ownership by Education

# Chapter 3

# LEVERAGING ON HUMAN CAPITAL: LABOR RIGIDITIES AND SORTING OVER THE BUSINESS CYCLE

By Edoardo Acabbi [¶], Andrea Alati [†], and Luca Mazzone [‡]

## 3.1 Introduction

Business cycle fluctuations and the accumulation of human capital are strictly intertwined. Recessions have an impact on the matching process between workers and firms, altering the job ladder, and thus accumulation of human capital on the job. At the same time, limits to investment in human capital can produce relevant feedback effects and delay recoveries from recessions. This article proposes a framework to evaluate the persistence of losses in human capital along worker careers and aggregate economic activity caused by temporary fluctuations in the business cycle. In addition, we highlight the role of frictions in shaping the pattern of workers displacement and sorting along the business cycle. Specifically, we show that limited liability on the firm side determines downward rigidity in wages, and induce endogenous separations during recessions. Moreover, search frictions and limits to investments in human capital amplify the impact of recessions on the job ladder of workers.

In order to perform these tasks, we introduce a structural model of the labor market in which we nest a dynamic contract setting between risk-averse workers and risk-neutral firms. Our theoretical framework allows to characterize the interactions between workers human capital accumulation over the life cycle and the matching with heterogeneous firms. The model also characterizes the optimal amount of worker income insurance within the firm and its interaction with workers' human capital accumulation.

¶ Harvard University.
†London School of Economics and Political Science.
‡University of Zürich and Swiss Finance Institute.

In our model economy, heterogeneous workers accumulate on the job experience which augments their skills and helps them climbing the job ladder. Search frictions and the presence of aggregate uncertainty distort sorting of workers with firms and generate inefficiencies in the economy. The characterization of optimal contracts between heterogeneous workers and firms allows us to evaluate the impact of rigidities on the pattern of job destruction at the onset of recessions. Our model predicts that jobs in firms with lower productivity are more vulnerable to output fluctuations. Because of compositional effects, this is true especially for matches between younger workers, and workers that are closer to retirement. This replicates dynamics that are observed in our specific labor market setting, the Italian labor market.

In order to correctly model the cohort effects of aggregate shocks we populate the economy with overlapping generations of workers. Each cohort is exposed to different aggregate conditions at the start of the working career, a feature which determines a time-varying cross-sectional distribution of workers. The OLG structure allows us to identify the different sources of long-run changes in job sorting, wage growth and human capital accumulation affecting each age group.

Human capital can play a persistent role in affecting economic performance, which makes it similar to physical capital. An important difference, however, is that the intensive and extensive margins of investment in (and acquisition of) human capital on the job are likely much more limited given workers limited lifetime. This feature amplifies output fluctuations and keeps workers' productivity below potential for a period longer than the duration of a temporary negative TFP shock. Not only workers who enter the labor market in bad times, but also those who are displaced during recessions, face a worsened job ladder and trade worse employment prospects for a higher likelihood of exiting from unemployment. In this sense even a transitory shock, if intense enough or protracted enough, will generate a permanent loss in the human capital of the labor force that is not going to be completely offset as long as the treated cohorts of workers are part of the workforce.[32]

The economic cycle, by altering the sorting and allocation of workers to firms can also impact overall earnings inequality and potentially alter the shape of the entire income distribution. As long as some workers are less protected by or less attached to firms, the way in which an aggregate shock propagates throughout the economy and towards firms and workers matches will have heterogeneous effects depending on workers' and firms' characteristics and contracts. Therefore, a comprehensive assessment of the impact of

[32]The first study to advance this hypothesis is Ljungqvist and Sargent 1998, in which the *eurosclerosis* of the 1990s' is associated with an analysis of a rigid labor market with slow human capital accumulation by workers.

business cycle shocks on the labor market and their propagation to the entire economy cannot abstract from taking heterogeneity seriously into account.

We capture these rich business cycle effects as in Menzio and Shi 2010; Menzio, Telyukova, and Visschers 2016 by adopting a directed search framework in which we nest human capital accumulation. We also generate endogenous separations through endogenous wage rigidity, resulting in inefficient separations when a high wage induces negative continuation values for the firm. Rigidity results from an optimal insurance contract between firms and workers where firms are subject to commitment but also have limited liability, so they exit when their continuation value is negative. Our framework has the advantage of explicitly modeling the dynamics of aggregate shocks and their effects on the entire earnings-skills distribution and clearly presenting the trade-off between insurance incentives, contractual rigidities and the efficient allocation of workers in the labor market. The model thus allows to assess how these trade-offs vary with the business cycle and what is the effect of having in the economy different cohorts of workers who experienced recessionary periods.

An accurate analysis of the macro-dynamics of sorting in the labor market is particularly relevant for a country like Italy, which is the country we plan to calibrate the model on, using administrative worker-level data. The relevance of our analysis, however, goes beyond the application to the Italian case. Italy is characterized by a dual, rigid labor market in which, in case of layoffs, the re-absorption into employment of workers is limited, overall labor market fluidity is scarce and younger cohorts are especially likely to be less protected.[33] These characteristics, in different ways, extend to most European labor markets. Especially during the latest recessionary periods, throughout the Global Financial Crisis (2007-2009) and the Sovereign Debt Crisis (2010-2012), it became evident that labor rigidities and missing investments in human capital could be potentially very harmful for workers who underwent periods of heightened instability and insecurity on the job, and in many cases were forced to accept under-qualified, precarious employment positions OECD2014. An assessment of aggregate dynamics of the labor market around these kinds of recessionary events is thus instrumental to inform the policy debate regarding the optimal level of flexibility of the market, and the possibility of enacting counter-cyclical policies, such as putting into place targeted unemployment benefits, targeted hiring subsidies, training programs or fiscal devaluations of labor costs to support unemployed cohorts and employment in recessions.

[33]For a cross-country comparison of labor market rigidity, see World Bank 2018 and OECD 2018. The OECD indicator clearly ranks Italy as one of the European countries with the strongest rigidity on labor market regulations, together with Portugal, Czech Republic, Germany and the Netherlands. The World Bank Labor Market tables, instead, offer a more granular representation of employment protection legislation across countries, without an explicit ranking.

## 3.2 Literature Review

Our paper relates to strands of research in labor and macroeconomics analyzing the effects and costs that business cycles can have on workers' careers in the longer term. There are mainly two separate methodologies that are usually employed in order to address research questions about these topics: (i) event studies and other kinds of reduced form empirical approaches, leveraging on an identification (necessarily in partial equilibrium) based on the observation of quasi-exogenous separation shocks or the possibility of matching workers in the data with very similar characteristics but different employment dynamics; and (ii) structural theoretical models, characterized by search frictions in job markets, which attempt to describe the effects of different kinds of shocks, both aggregate and idiosyncratic, on wage dynamics, distributions, and matching of workers to firms. This paper follows the second approach.

A number of recent studies have analyzed the long-run effects of unemployment for workers' earnings, and the importance of the economic cycle on lifetime outcomes. Many empirical studies, following the seminal paper by Jacobson, LaLonde, and Sullivan 1993, have analyzed the impact on earnings and work careers of losing a job during a recession finding large income penalties for workers that lose their job in bad economic times. A related literature, following in particular Kahn 2010, Oreopoulos, Von Wachter, and Heisz 2012, Schwandt and Wachter 2019, focused instead on the long term effects on graduates' and young workers' careers of entering the labor market in a recession, switching the analysis from matches that are already formed to the analysis of how aggregate conditions influence labor-market matches in the first place.[34] These studies use mainly geographical variation in aggregate labor market conditions across cohorts of new graduates and they estimates a significant and persistent effect of entering the labor force under worse economic conditions.

In this line of research and closely related in spirit to our work, Arellano-Bover 2020 estimates how much of the effects of graduating in a recession can be explained by the size of the first employer. He finds that up to 15% of the losses from entering the labor market in periods of high unemployment could be attributed to the fact that during recessions young workers are more likely to match with a smaller first employer. Firm size is taken as a proxy for otherwise difficult to measure firm attributes that can directly impact workers' careers, such as the availability of training programs or simply being exposed to better management practices.[35] There are other recent studies that provide some evidence on

[34]See also, among others, Schmieder, Wachter, and Heining 2018, Lachowska, Mas, and Woodbury 2017, Altonji, Kahn, and Speer 2016, Huckfeldt 2018.

[35]The model presented in this paper does not feature an explicit firm size distribution as we adopt the approach of one-job-one-firm, which is standard in the search and matching literature. As a consequence,

the dynamic role of employment at heterogeneous firms also for job mobility and other labor market outcomes. Abowd, McKinney, and Zhao 2018, for example, show getting the first job in a top-paying firm can lead to up-ward movements in the earnings distribution later in workers careers. Similarly, Bonhomme, Lamadon, and Manresa 2019 show how past firm types can have an impact on future earnings after changing jobs. Therefore, there is a large body of empirical evidence that estimates how the economic cycle can have a substantial impact on cohorts currently in the job market and especially on the ones entering the job market for the first time in a recession. Even if the impact of this condition has already been analyzed as regards the impact on lifetime earnings by means of a reduced form approach, no analysis has been carried out about the potentially persistent effects for the entire economy and for overall labor and firm productivity of having cohorts with a slower accumulation of human capital in the context of general equilibrium model of the labor market.[36]

From a theoretical standpoint, we contribute to the extensive literature analyzing long-term contracting with limited commitment (e.g. Harris and Holmstrom 1982, Thomas and Worral 1988, Krueger and Uhlig 2006, Xiaolan 2014, Lamadon 2016). Contrary to these papers, however, our main focus is to quantify the aggregate effects of the interaction of insurance incentives, provided by the contract and crucial in determining the patterns of job separations along the business cycles, and the long-term term effects on the macroeconomy of having cohorts of workers exposed to worse employment relationship early in their careers.

Models of the labor market focusing on related topics have been recently proposed, among others, by Jarosch 2015 to analyze how workers value job security with respect to the salary, or by Burdett, Carrillo-Tudela, and Coles 2016 to structurally estimate the cost of job loss. We complement these studies by explicitly modelling the contracting problem between the worker and the firm, taking into account the relevance of human capital accumulation for both the workers' careers and the aggregate economy.

In addition, we show that it is possible to describe the recursive structure of our contractual framework using both the formulation based on the inclusion of promised utilities as additional state variables, as in Spear and Srivastava 1987, and with the more computationally feasible formulation of recursive contracts developed by Marcet and Marimon 2019. The latter approach is known as the recursive Lagrangean method.

we are silent on which are the specific channels through which firm quality influences human capital. The mechanism we have in mind, however, is akin to the effect of being exposed to better management practices and more efficient organizations.

[36]Theoretical analyses of the impact of recessions on workers careers are presented in Guo 2018, Wee 2016. Their models share some of the features of our model, but do not feature optimal dynamic contract and endogenous separations, or heterogeneous firms.

Notable contributions, besides the seminal paper, include Cole and Kubler 2012, Messner, Pavoni, and Sleet 2012, and Mele 2014.

At this time, to the best of our knowledge, few studies have managed to incorporate the influence of the economic cycle together with the dynamics of human capital accumulation and firms-workers matches in a structural model of the labor market. Also, little is known about how the contractual framework influences firms decisions' on workforce mix, training programs and overall hiring strategies, especially in relationship to business cycle fluctuations. The search-and-matching literature has recently been trying to address how the sorting of workers and firms varies cyclically (see for instance Lise and Robin 2017). This study provides an important contribution to this strand of research.

The paper is divided as follows: in **Section 3.3** we provide preliminary evidence regarding the dynamics of sorting, matching and the relevance of human capital accumulation along the business cycle; in **Section 3.4** we present the model; in **Section 3.5** we discuss features of the solution of the model and of the equilibrium conditions (proofs are mostly in Appendix); finally, in **Section 3.6** we conclude.

## 3.3 Preliminary motivating evidence

The model and the contractual environment developed in this paper imply a strong dependence of workers' careers on the history of aggregate shocks they are exposed to throughout their working lives. In addition, the fact that workers can accumulate human capital while working paired with the assumption that working lives are finite, gives a disproportionate importance of aggregate shocks to matches early in workers lives. The intuition of why this is the case relies on noticing that early in workers lives human capital accumulation is more important both because the level of human capital workers are endowed with is lower and because the net present value of higher human capital levels in the future are greater the longer is the period a worker is able to reap the benefits of these higher levels. Both these reasons are stronger for younger workers searching for their first job than for older workers that are possibly trying to reallocate. As a consequence, any shock that induces missing investments or impairs the accumulation of human capital, especially early in workers careers, will generate a persistent loss in workers' earnings.[37]

In our model human capital accumulation is a byproduct of matching between firms and workers and the only exogenous process that drives the matching process is aggregate

[37] Among others, Arellano-Bover 2020 shows that the quality of initial matches between firms and workers in Spanish data has persistent effects on workers careers: he shows that when the first job of a worker is in a firm one-standard deviation higher than the average, this results in a lifetime income (20 years) one-third higher than the average.

productivity. Hence a direct prediction of our model is that aggregate conditions, especially at the beginning of workers' careers have persistent effects on their labor market outcomes.

We discuss here our empirical strategy for the analysis of administrative data and preliminary stylized results. [38]

## Cohort-effects in the Italian labor market

Age-period-cohort models allow to isolate the effect of belonging to a particular cohort from the effects of contemporaneous aggregate conditions and aging, which in the context of labor market outcomes is equivalent to accumulating experience.

Formally a generic age-period-cohort model could be expressed as follows:

$$y_{i,c,y} = \Gamma_y + \Gamma_{y-c} + \Gamma_c + \gamma' \mathbf{X}_i + \varepsilon_{i,c,y}$$

where the dependent variable is our outcome of interest for a worker $i$, belonging to cohort $c$ in year $y$ and where the right hand side is composed by a collection of functions aimed at controlling for year, cohort and age effects respectively and, potentially a set of worker level controls.

Ideally, instead of assigning parametric structures to the functions $\Gamma_j$ we would use fixed effects to allow the data to determine the functional for for each of these effects. However this approach would lead to a well-know identification problem as the three fixed effects age-period-cohort are perfectly colliner between each other. As a consequence identifying the levels of these three factors requires additional normalizations or exclusion assumptions. The approach we follow to construct the preliminary evidences, as discussed in Heckman and Robb 1985, consists of treating age, period and cohort effects as proxy variables for underlying variables that are not themselves linearly dependent. We use cyclical real GDP realizations a the time of first job as proxies for cohort effects.

In our baseline specification we estimate the age-period-cohort model discussed above using the cyclical components of Italian real GDP as a proxy variable for cohort effects.[39] We formally evaluate the correlation between annual wage and aggregate conditions at the time of firs job by estimating the following regression:

$$\log(w_{i,c,y}) = \alpha + \beta \tilde{Y}_c + \phi_e + \phi_y + \gamma' \mathbf{X}_i + \varepsilon_{i,c,y}, \tag{3.1}$$

[38] As the empirical part of the project awaits INPS approval, we are not at freedom at this time to disclose any empirical result beyond the ones briefly presented below.

[39] We use the Hamilton filtered (HF) series of real log-GDP with 1 period lag and 2 periods horizon. As a robustness with check also the Hodrick-Prescott (HP) filtered series (smoothing parameters, 6.25) and the quadratically detreneded real-GDP. The results are qualitatively similar but we choose HF for our baseline specification as it is a filter that does not use any future information, which could be problematic in our regressions, and captures business cycles better than the quadratic detrended measure.

in which the dependent variable is the logarithm of annual real wage for worker $i$, belonging to cohort $c$ in year $y$. As each cohort is identified by the year of entry in the labor market, the function for age is replaced by a set of experience fixed effects $\phi_e$ while the function for the period effects is substituted by a set of year fixed effects $\phi_y$. The matrix of worker level controls $\mathbf{X}$ includes a series of fixed effects aimed at controlling for worker specific factors, such as sex, type of contract (part-time vs full-time), contract maturity (fixed term vs open ended), sector and qualification. Under the standard exogeneity restrictions the coefficient $\beta$ measures the average percentage change in annual wage resulting from a one-percent cohort-specific variation in the business cycle measure $\tilde{Y}$. [40]

**Persistence of cohort effects.** The model in **Equation** (3.1) does not allow the effects of business cycles to fade with time we implement also the following specification to measure the persistence of initial aggregate conditions on workers' annual wages:

$$\log(w_{i,c,y}) = \alpha + \sum_{e=1}^{\bar{e}} \beta_e \tilde{Y}_c \mathbf{1}\{s_{i,t} = e\} + \phi_e + \phi_y + \gamma' \mathbf{X}_i + \varepsilon_{i,c,y}, \tag{3.2}$$

where $\mathbf{1}\{s_{i,t} = e\}$ is an indicator function that takes value one if the worker has $e$ years of experience. In this specification the set of coefficients $\{\beta_e\}$ measures the effect of business cycles at each point of the worker career.

**Preliminary results.** Preliminary analyses on the correlations between aggregate conditions at the moment of first job indicate a large and persistent effect of aggregate conditions at the time of first employment. The initial results indicate that when business cycles are one standard deviation below trend on average workers experience a 23% loss in earnings over the first 10 years.[41] Recent estimates by Arellano-Bover 2020 on Spanish data indicate that the size of first employer can explain up to 15% of the scarring effects of recessions.[42]

[40] An obvious threat to identification would be the ability of workers to withdraw from the labor market in according to changes in aggregate conditions, while at the same time having access to a technology for investing in human capital. Depending on their individual traits, e.g. their learning ability, there would be heterogeneity in responding to downturns. Due to characteristics of Italian labor demand and its the education system, however, it is reasonable to assume that workers in the country do not have access to such a technology.

[41] This results are in the ballpark of (but smaller than) what Schwandt and Wachter 2019 found for the US, who show results for the period from 1976 to 2015.

[42] For the Spanish labor market, that shares many similarities to the Italian context (e.g. high youth unemployment and high degree of segmentation) Fernández-Kranz and Rodríguez-Planas 2017 estimate the loss of an average recession in a range between 6-12% of annual earnings over 10 years, stronger and more persistent for less educated workers (7 years persistence for high-school graduates versus 5 years for college graduates).

## 3.4 Model

In this section we present our model of the labor market. We start by discussing the environment, the timing and the preference structure of the economy. Then we discuss the features of a frictional labor market with directed search, and finally we characterize the workers problem and the optimal recursive contract.

### Environment

Time is discrete, runs forever and is indexed by $t \in \mathbb{Z}$. The economy is populated by two kinds of agents: a unit measure of finitely-lived risk-averse households (workers) and a continuum of risk neutral entrepreneurs who have the ability to invest in enterprises and thus run an endogenously chosen number of operating firms. All agents in the economy share the same discount factor $\beta \in (0,1)$.

Following Menzio, Telyukova, and Visschers 2016 we populate the economy with $T \geq 2$ overlapping generations of households, which face both aggregate and idiosyncratic risk. Each household lives for $T$ periods deterministically, with age $\tau \in \mathcal{T} \equiv \{1,2,3,\ldots,T\}$. Every period workers participate to the labor market and, as in Shi 2009, Menzio and Shi 2010, direct their search towards different submarkets. Workers can only search for work and consume, as we do not model saving decisions.[43] The objective of the household is to maximize its own life-time flow-utility from non durable consumption:

$$\mathbb{E}_{t_0}\left(\sum_{\tau=1}^{T}\beta^{\tau}u(c_{\tau,t_0+\tau})\right)$$

where $t_0$ characterizes the time of entry into the labor market and $\tau$ characterizes the age of the agent. Workers can either be employed or unemployed, and we denote by $e$ and $u$ their employment status. Workers are characterized by heterogeneous human capital levels $h$, with $h \in \mathcal{H} \equiv [h,\overline{h}]$. Workers start off their life with a baseline level of human capital drawn from an initial exogenous continuous distribution with density $l(h)$ and can get training on-the-job over the course of their working career. The way in which they get training is another source of heterogeneity in the model across workers. Workers are matched with firms characterized by different levels of (permanent) firm quality $y \in \mathcal{Y} \equiv [y,\overline{y}]$, which in our model are isomorphic to capital levels. This maps into the dynamics of human capital as explained below.

[43] Modeling wealth accumulation in a model with two-sided heterogeneity, life-cycle, human capital accumulation and directed search (also on-the-job) is undoubtedly interesting and important and left to future research.

The only form of human capital accumulation in the model is on-the-job. Following Lise and Postel-Vinay 2019, we model human capital accumulation in this way: depending on the level of quality of the firm and her own level of ability, the worker might accumulate human capital according to some law of motion $g(h,y): \mathcal{H} \times \mathcal{Y} \rightarrow \mathcal{H}$. As in Lise and Postel-Vinay 2019, training is similar to "catching-up" of the firm quality with respect to the "training" ability of the firm, up to a point (depending on $y$) when the worker will not be able to learn anymore from the match and would possibly like to transit to a higher $y$ match. At the same time, coherently with the concept of "mismatch", workers who lost their job and only manage to re-match with a low quality firm see their ability progressively deteriorating with the same $g(h,y)$ function. The function $g(h,y)$ is a concave function in both arguments. [44] Firms are, as common in labor-search studies, just one worker-one job matches, and we are thus abstracting from firm size.[45]

We denote future values in recursive expressions by adding a $'$ to them, or index elements by $t$ in non-recursive ones. The aggregate state of the economy $\Omega$ is characterized by the level of aggregate productivity $a \in \mathcal{A} \subset \mathbf{R}^+$ and by the distribution of agents across states $\mu \in \mathcal{M} : \{e,u\} \times \mathcal{H} \times \mathcal{Y} \times \mathcal{T} \rightarrow [0,1]$. Let $\Omega = (a,\mu) \in \mathcal{A} \times \mathcal{M}$ represent the aggregate state of the economy and let $\mathcal{M}$ represent the set of distributions $\mu$ over the states of the economy. Let $\mu' = \Phi(\Omega, a')$ be the law of motion of the distribution. Aggregate productivity evolves as a stationary increasing Markov process, namely $a' \sim F(a'|a) : \mathcal{A} \rightarrow \mathcal{A}$, with the Feller property.

The timing of each period is as follows: a productivity shock for the period is drawn; entrepreneurs open vacancies across the submarkets and post their offers; workers search from unemployment or on-the-job, and workers transition to a new job if on-the-job search is successful; production of both surviving and newly created firms takes place; employed workers accumulate human capital; an exogenous share of matches breaks down; at the same time, and before knowing what the next period productivity draw will be, incumbent firms decide whether to shut down, endogenously destroying their matches, or continue producing. State contingent policies prescribe an action for each realization of the story of worker-firm matches. For ease of notation, we denote the sequence of stories as $\{s^\tau\}_{\tau=1}^{T}$. The sequence of actions just described is summarized in **Figure 3.1**.

[44] This assumption is needed to ensure that the firm's optimal profit function is smooth and does not exhibit kinks as workers start to accumulate human capital.

[45] Modeling multi-worker firms in our context is an immensely interesting advancement that we leave for future research.

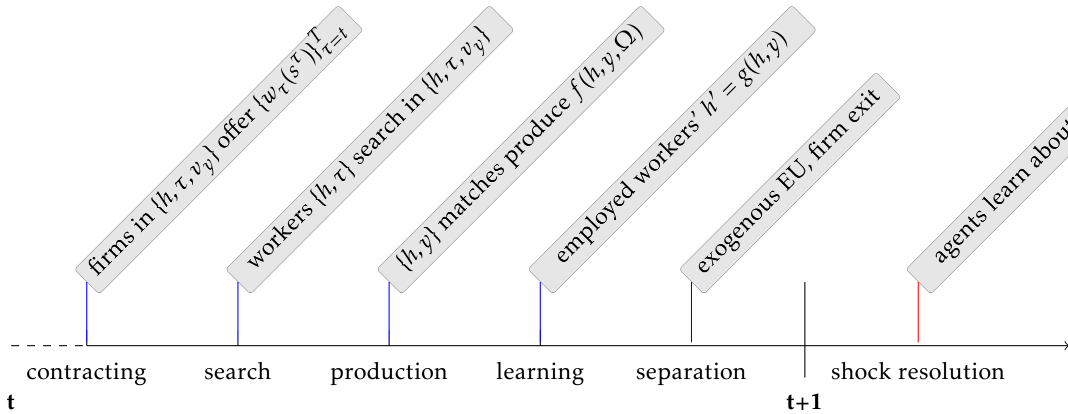


Figure 3.1: Timeline of worker-firm match

## Labor markets

Search is directed. Each labor market is organized as a continuum of submarkets indexed by the expected lifetime utility offered $v_y \in \mathcal{V} \equiv [v, \overline{v}]$. Each worker characterized by $(h,\tau)$ directs search, and entrepreneurs decide which kind of firms $y$ to open and, correspondingly, an offered lifetime value $v_y$.[46] There is free entry for entrepreneurs in submarkets. The process of opening a firm, which amounts to posting a vacancy at a quality-specific cost $\kappa(y)$, will be described in **Section 3.4**. We will also prove that, given a choice of worker $(h,y)$ to whom an offer is made, there is going to be *only* one kind of firm $y$ offering a defined value $v_y$. In other words, given $(h,\tau)$, $v_y$ is an injective function $f_v : \mathcal{Y} \to \mathcal{V}$, and any vacancy in submarket $(h,\tau,v_y)$ is actually offered by the *same* $y$ kind of firm.

The search process is characterized by a constant return to scale twice continuously differentiable matching function $M(u,\nu)$ for each submarket, where the tightness of each submarket is as usual defined as $\theta = \frac{v}{\nu}$. Households job finding rates are defined as $p(\theta(h,\tau,v_y;\Omega)) = \frac{M(u(h,\tau,v_y;\Omega),\nu(h,\tau,v_y;\Omega))}{u(h,\tau,v_y;\Omega)}$, where $p() : \mathbf{R}^+ \to [0,1]$ is twice continuously differentiable, strictly increasing and strictly concave function with $p(0) = 0$, $\lim_{\theta\to+\infty} p(\theta) = 1$ and $p'(0) < \infty$, whereas the vacancy-filling is $q(\theta(h,\tau,v_y;\Omega)) = \frac{M(u(h,\tau,v_y;\Omega),\nu(h,\tau,v_y;\Omega))}{\nu(h,\tau,v_y;\Omega)}$, where $q() : \mathbf{R}^+ \to [0,+\infty]$ is twice continuously differentiable, strictly decreasing and strictly convex, with $q(0) = 1$, $\lim_{\theta\to+\infty} q(\theta) = 0$ and $q'(0) < 0$. We have that $q(\theta) = p(\theta)/\theta$, and $p(q^{-1}(\cdot))$ is concave.

Upon match, workers produce next period according to the twice-continuous increasing and concave production function $f(a,h,y) : \mathcal{A} \times \mathcal{H} \times \mathcal{Y} \to \mathbf{R}^+$. The compensation of the worker depends on workers' and firms' kinds, and is defined by means of dynamic

[46] As in Menzio and Shi 2010, given a menu of offers from any firm, workers able to search in any submarket will separate by type in equilibrium, and any given type $(h,\tau)$ visits a particular market. For this reason submarkets can then be represented directly by $(h,\tau,v)$.

contracts through which firms deliver a promised utility. Contracts are going to be described in **Section 3.4**.

Matches are destroyed at an exogenous rate $\lambda$ each period. Moreover, firms are subject to limited commitment, and matches also separate endogenously either if the worker is poached by another firm (quit) or if the value of the match for the firm become negative (fires).[47] Workers are always allowed to search while unemployed and search while employed with probability $\lambda_e$. Notice that the timing of each period implies that newly separated workers can immediately search in the same period.

## Informational and contractual structure

A contract defines a transfer of utility from the risk neutral firm to the risk averse worker with the match for all future possible histories of shocks. Given a match formed at a generic hiring time $t_0$, the state of the match is defined by $s_{t_0} = (h_{t_0}, \tau_{t_0}, a^{t_0}, \mu^{t_0}) \in \mathcal{S} = \mathcal{H} \times \mathcal{T} \times \Omega^{t_0}$, that is the worker skill, age and the history of aggregate productivity shocks and workers' distributions across employment states and submarkets (specifically, the specific worker's history of employment). We define $s^{t_0+(T-\tau_{t_0})}$ as the history of realizations between $t_0$, the time of hiring of the worker, and $t_0 + (T - \tau_{t_0})$, the time of maximum duration of the match with the worker before retirement ($\tau_{t_0}$ is the age at which the worker is hired and $T$ is the retirement age).

The workers' history and the history of productivity are common knowledge, and histories are fully contractible. In this sense, the contract is fully state-contingent. Nevertheless, the markets are incomplete, given that workers' actions are private knowledge in the search stage, and firms are thus unable to counter outside offers. The contract offered by the firm can thus be defined as:

$$\mathcal{C} := (\mathbf{w}, \zeta) \text{ with } \mathbf{w} := \{w_t(s^{\tau_t - \tau_{t_0} + t_0})\}_{t=t_0}^{t_0+(T-\tau_{t_0})}, \text{ and } \zeta := \{v_t(s^{\tau_t - \tau_{t_0} + t_0})\}_{t=t_0}^{t_0+(T-\tau_{t_0})} \tag{3.3}$$

According to the contract the firm promises a series of state-contingent wages, to which the worker replies by enacting its own state-contingent search strategy, defined by the series of $v_t$ sought at each node of the history.[48] As in Lamadon 2016, $\zeta$ is the action suggested by the contract, which in our analysis is bound to be incentive compatible for the worker. The contract is otherwise fully flexible in the degree to which the firm can determine wage levels and adjustment paths over the match histories.

[47]Notice that separations happen in two waves during the same period. Fires and exogenous separations happen before the worker's search (at the end of previous period), whereas quits happen afterwards.

[48]Similarly to Menzio and Shi 2010; Tsuyuhara 2016; Lamadon 2016, and in order to guarantee, at least for the general proofs, that the problem is well behaved and the firm profit function is concave, the contract will require a randomization, a two-point lottery, which specifies probabilities over the actions prescribed.

### Worker problems

Given the fact that the relationship of workers and firms is going to be characterized by a recursive contract with forward looking constraints, the state space of the problem needs to include the current utility promised to employed workers (or the current utility of unemployed agents), as in Spear and Srivastava 1987. Job seekers, both all unemployed agents and the share $\lambda_e$ of incumbents, face the same kind of search problem. Given a generic current lifetime utility $V$, any job seeker characterized by human capital $h$ and age $\tau$ has to decide in which submarket to direct the search. Submarkets are indexed by the posted offered utility $v_y$. As it will be proved in **Section 3.4**, the choice over $v$ will also indirectly determine which kind of firm $y$ the worker matches with, and thus the human capital accumulation path. For now, assume that this mapping exists, and thus that, given $(h,\tau)$, the function $v_y(y)$ is an injective function $f_v : \mathcal{Y} \to \mathcal{V}$ such that any value $v$ ends up being offered to the *same* kind of worker $(h,\tau)$ by one specific $y$ firm only. This means that, even if workers only care about offered life-time utilities $v$, their choices determine which firm quality $y$ they can match with and the human capital accumulation that concurs to deliver the promised utility $v$ itself.

A worker characterized by $(h,\tau)$ who got the opportunity to search enters the search stage with lifetime utility $V + \max\{0, R(h,\tau,V;\Omega\}$, where the second component of the expression embeds the option value of the search, and $R$ is the search value function. $R$ is defined as:

$$R(h,\tau,V;\Omega) = \sup_{\{v_{y,\Omega}\}} \Big[ p(\theta(h,\tau,v_{y,\Omega};\Omega)) \big[ v_{y,\Omega} - V \big] \Big] \tag{3.4}$$

We denote the solution of the search problem as $v_y^* = v^*(h,\tau,V;\Omega)$, and $p^*(h,\tau,v^*_{y,\Omega};\Omega) = p(\theta(h,\tau,v^*_{y,\Omega};\Omega))$ as the associated optimal job-finding probability. Notice that, given the timing of the choices outlined in **Figure 3.1**, a job seeker can devise search strategies that are *contingent* on the state in which the search actually takes place.

The lifetime utility of an unemployed worker at the beginning of the production stage can be define as

$$\begin{aligned} U(h,\tau;\Omega) = \ & u(b(h,\tau)) + \beta \mathbb{E}_\Omega \Big( U(h,\tau+1;\Omega') \\ & + \max\{0, R(h,\tau+1, U(h,\tau+1;\Omega');\Omega')\} \Big) \end{aligned} \tag{3.5}$$

where $b(h,\tau)$ is a (possibly) skill and age dependent unemployment benefit. Given finite workers' lives, $U(h,\tau;\Omega) = 0 \ \forall (h,\tau;\Omega) \in \mathcal{H} \times \mathcal{T} \times \mathcal{A} \times \mathcal{M}$ where $\tau > T$.

The corresponding lifetime utility of an employed worker with current promised utility $V_{y,\Omega}$ at the beginning of the production stage can be expressed as:

$$V_{y,\Omega} = u(w_y) + \beta \mathbb{E}_\Omega \Bigg( \lambda U(g(h,y), \tau+1; \Omega') + (1-\lambda) \Big[ V_{y,\Omega'} \\ + \lambda_e \max\{0, R(g(h,y), \tau+1, V_{y,\Omega'}; \Omega')\} \Big] \Bigg) \tag{3.6}$$

where $w_y$ is the currently promised wage and $V_{y,\Omega'}$ is next period's state-contingent promised lifetime utility of remaining in the current firm, which becomes the "outside option" in the search problem. Notice that in the worker's problem there is nothing specific to firm quality $y$ *per se*. The worker targets its search towards a desired level of promised utility, which we are going to show is going to be offered by one kind of $y$ firm only in equilibrium. For this reason, we can index wages and utilities by $y$ (and the aggregate state $\Omega$, given the state-contingency of promises). The promised utilities $V$ are an equilibrium object themselves, as they are the outcome of the firm optimization of the dynamic contract.

By means of their search strategy workers indirectly have an impact on their current contract too, as firms internalize workers' strategies in their optimization, and post wages and utility offers to maximize profits and thus retention. In fact, a worker quit drives profits to zero, independently of their previous level. Workers future promised utility incorporates both higher wages and higher option values of search, also through the human capital accumulation dynamics defined by $g(h,y)$.

The policy functions are uniquely defined, and allow to identify $y$ uniquely as long as there exists a injective mapping between the offered utility $v$ and $y$ given $\{h,\tau,\Omega\}$, which we assume for now and prove given the structure of contracts in **Section 3.4**. Proofs for the uniqueness of the policy functions and the optimal policy are provided in **Appendix 3.A**.

The solution of employed workers' on-the-job search problem implicitly defines two "policy" functions, which incorporate workers' incentive compatibility which firms internalize in their optimization.

**Definition 3.4.1 (Optimal retention probability and utility return)** *The solution of the worker's problem defines a retention function* $\widetilde{p} : \mathcal{H} \times \mathcal{T} \times \mathcal{V} \times \Omega \to [(1-\lambda)(1-\lambda_e), 1-\lambda]$ *and a utility return* $\widetilde{r} : \mathcal{H} \times \mathcal{T} \times \mathcal{V} \times \Omega \to \mathcal{V}$*:*

$$\widetilde{p}(h,\tau,V_{y,\Omega};\Omega) \equiv (1-\lambda)(1-\lambda_e p^*(h,\tau,v^*_{y,\Omega};\Omega)) \tag{3.7}$$

$$\widetilde{r}(h,\tau,V_{y,\Omega};\Omega) \equiv \lambda U(h,\tau;\Omega) + (1-\lambda)\Big[V_{y,\Omega} + \lambda_e \max\{0, R(h,\tau,V_{y,\Omega};\Omega)\}\Big] \tag{3.8}$$

The two functions $\widetilde{r}$ and $\widetilde{p}$ incorporate the optimal behavior of the worker, and thus its incentive-compatible best replies when evaluating offers by firms. These functions are what the firms internalize while setting the optimal contract.

## Contract

The contract is structured in such a way that the firm is subject to limited liability but commits to the delivery of a utility value to the worker, who on the other does not have to commit. Specifically, this means the worker is able to search at any period, and the firm is not able to counteract with another offer when its employee matches with another firm. The sequence of stories $s^t$ is common knowledge, and while the firm cannot observe any of the actions of its workers, it has enough information to incorporate the worker's optimal policy decision.

Our choice of timing of exit decision is such that exiting firms know from the start of the period whether the productivity level is below the critical one $a*$ for the match $(h, \tau, y, W_y)$, and thus whether they will exit or not. In particular, given the current state we can define the following indicator function

**Definition 3.4.2 (Exit policy)** *Incumbent firms make their exit decisions before the realization of aggregate productivity. The following indicator takes value one if the firm does not decide to exit in the following period:*

$$\eta_{t+1} = \begin{cases} 1 \ \ if \ \ a \geq \max\{0, a^*\} \\ 0 \ \ otherwise \end{cases}$$

*with the productivity threshold defined as*

$$a^*(h, \tau, y, W_{y,\Omega}) : \mathbb{E}_\Omega[J_{t+1}(h', \tau+1, y, W_{y,\Omega'}; a', \mu')|h, \tau, y, W_{y,\Omega}, a, \mu] = 0. \tag{3.9}$$

Given $\eta_{t+1} = 1$, the value function of a continuing incumbent in state $(h, \tau, W_{y,\Omega}; \Omega)$ can be rewritten recursively using the promised utilities to the workers as additional state variables as:

$$J_t(h, \tau, y, W_{y,\Omega}; \Omega) = \sup_{\pi_i, w_i, \{W_{iy,\Omega'}\}} \Big( f(y, h; \Omega) - w_i$$
$$+ \mathbb{E}_\Omega \left[ \widetilde{p}(g(h,y), \tau+1, W_{iy,\Omega'}; \Omega')(J_{t+1}(g(h,y), \tau+1, y, W_{iy,\Omega'}; \Omega') \right] \Big) \tag{3.10}$$

$$s.t. \ W_{y,\Omega} = \mathbb{E}_\Omega \left( u(w_i) + \mathbb{E}_\Omega \widetilde{r}(g(h,y), \tau+1, W_{iy,\Omega'}; \Omega') \right), \tag{3.11}$$

$$\sum_{i=1,2} \pi_i = 1 \tag{3.12}$$

where **Equation** (3.11) is the promise keeping constraint ensuring that the current value of the contract is indeed based on the current wage and future utility promises with $\widetilde{r}_t()$ implicitly including the incentive constraint of the worker.

In this kind of contracts the firm (principal) optimizes over its possible offers taking into account the utility of the worker (agent) and its incentive compatible best replies. The resulting equilibrium is a subgame perfect Nash equilibrium of the kind identified in leader-follower sequential games, as in Von Stackelberg 1934. The problem also resembles a Ramsey optimal policy problem, in that the principal in our case is akin to a policy-maker who maximizes aggregate utility according to some Pareto weights, taking into account optimization on the part of the agents in the economy (worker-firm match).

## Vacancy opening and free entry

The economy is populated by a continuum of risk-neutral entrepreneurs. Each entrepreneur can invest to reach the desired level of firm quality $y$. The start-up costs of the firm are priced in terms of the consumption good and they consist of the posting a vacancy in the frictional labor market.

The cost of each vacancy is proportional to the quality of the firm being created, hence in order to post a vacancy for the creation of a firm with quality $y$ the entrepreneur is forced to pay $c(y)$ in terms of the consumption good.

Thus, each entrepreneur at a generic time $t$, chooses in which submarket to post the vacancy selecting a lottery over the offered utility $W_y$, which maps into the set of firms' qualities $y \in \mathcal{Y}$, and worker characteristics $(h, \tau) \in \mathcal{H} \times \mathcal{T}$.

As the entrepreneur chooses the submarket in which to open a vacany, he faces the following problem internalizing the optimal contract dynamics:

$$\Pi_t(h, \tau, y, W_{y,\Omega}; \Omega) = \sup_{y,h,\tau,W_{y,\Omega}} -c(y) + q(\theta(h, \tau, W_y; \Omega))\beta[J_t(h, \tau, y, W_{y,\Omega}; \Omega)] \tag{3.13}$$

and, given perfect competition, free entry and the possibility for all entrepreneur to choose *any* possible firm kind $y$ the profits from opening a vacancy should be driven down to 0 in submarkets which actually open:

$$\Pi_t(h, \tau, y, W_{y,\Omega}; \Omega) \leq 0 \text{ for } \forall \{h, \tau, y, W_{y,\Omega}; \Omega\} \in \{\mathcal{Y} \times \mathcal{V} \times \mathcal{S}\} \tag{3.14}$$

Assuming that $q(\cdot)$ is invertible, it delivers the equilibrium definition of marker tightness in each submarket:

$$\theta_t(h, \tau, W_{y,\Omega}; \Omega) = q^{-1}\left(\frac{c(y)}{\beta J(h, \tau, y, W_{y,\Omega}; \Omega)}\right). \tag{3.15}$$

### Equilibrium definition

**Recursive Equilibrium.** Let $\Theta = \mathcal{A}\times\mathcal{M}\times\mathcal{H}\times\mathcal{T}$. A recursive equilibrium in this economy consists of a market tightness $\theta : \Theta\times\mathcal{V}\to\mathbb{R}_+$, a search value function $R:\Theta\times\mathcal{V}\to\mathbb{R}$, a search policy function $v^*:\Theta\times\mathcal{V}\to\mathcal{V}$, an unemployment value functions $U:\Theta\to\mathbb{R}$, a series of firm value functions, $\{J_t\}_{t=1}^T : \mathcal{S}\times\mathcal{V}\times\mathcal{Y}\to\mathbb{R}$, a series of contract policy functions $\{c_t\}_{t=1}^T : \mathcal{S}\times\mathcal{V}\times\mathcal{Y}\to\mathcal{C}$, a mapping between firm qualities and promised utilities $f_v:\mathcal{Y}\to\mathcal{V}$, an exit threshold for aggregate productivity $a^*:\mathcal{S}\times\mathcal{V}\times\mathcal{Y}\to\mathcal{A}$ and a law of motion for the aggregate state of the economy $\Phi_{\Omega,a'}:\mathcal{A}\times\mathcal{M}\to\mathcal{A}\times\mathcal{M}$ such that:

1. given the mapping $f_v$, market tightness satisfies **Equation** (3.15)
2. the unemployment value functions solves **Equation** (3.5)
3. the search value function solves the search problem in **Equation** (3.4) and $v^*$ is the associated policy function
4. the series of firm value functions and the associated contract policy functions are a solution to **Equation** (3.10) for each $t \le T$
5. the exit threshold satisfies **Equation** (3.9)
6. the law of motion for the aggregate state of the economy respects the search and contract policy functions and the exogenous process of aggregate productivity

**Definition 3.4.3 (Block Recursive Equilibrium)** *A block recursive equilibrium is a recursive equilibrium such that the value and policy functions depend on the aggregate state only through aggregate productivity, $a\in\mathcal{A}$ and not through the distribution of agents across states $\mu\in\mathcal{M}$.*

We provide a proof for the existence of a BRE equilibrium in **Appendix 3.A**.

## 3.5 Discussion

In this section we briefly discuss the properties of the equilibrium of the model economy developed in the previous sections. All propositions and corresponding proofs are reported in **Appendix 3.A** and **3.A**.

### Workers optimal behavior

In the following proposition we summarize the main results regarding the behavior of the workers and their objective functions.

**Proposition 1** *Given the worker search problem, the following properties hold:*

*(i) The returns to search, $p(\theta(h,\tau,v_{y,\Omega};\Omega))\Big[v_{y,\Omega}-V\Big]$, are strictly concave with respect to promised utility, $v_{y,\Omega}$.*

*(ii) The optimal search strategy*

$$v^*(h,\tau,V;\Omega)\in\arg\max_{v_y}\Big\{p(\theta(h,\tau,v_{y,\Omega};\Omega))\Big[v_{y,\Omega}-V\Big]\Big\}$$

*is unique and weakly increasing in $V$.*

*(iii) For all promised utilities, the search gain $R(h,\tau,V;\Omega)$ is positive, weakly decreasing in $V$.*

*(iv) The survival probability of the match, given the optimal choice of the worker, is increasing in the value of promised utilities, so $\widetilde{p}_t(h,\tau,W_{y,\Omega};\Omega)$ is increasing in $W_{y,\Omega}$.*

See **Proposition 6** in **Appendix 3.A**.

The first statement implies that the marginal returns of searching towards better firms are decreasing. The intuition is that as workers search for work at better firms, their job-finding probability decreases as better employment prospects are also subject to higher competition.

As a consequence of the strict concavity established in the first statement, we can say that the optimal search strategy of each worker is increasing in the value of life-time utilities granted by the current contract. The intuition is that, as the search stage happens after the realization of the aggregate state and workers only care about the posted utility offers, given their type $(h,\tau)$, workers have a unique preferred option among utility offers.

The third statement follows from the fact that returns to search are decreasing and the set of utility promises is compact. The intuition is that employees at firms that promise an higher value of future utilities don't have a lot of incentives to search. To clarify this, imagine an individual working under a contract that guarantees her the best possible utility: there is no point in searching elsewhere as no other firms could match her current option. This leads directly to the finding that when firms offer better prospects to their workers, workers also have fewer incentives to leave. This guarantees a longer expected duration of the match, and generates retention probabilities that are increasing in promised utilities.

As human capital accumulation is tightly linked to the quality of the employer, workers that are able to start their working careers in good times have a higher chance of finding themselves on an higher path of human capital growth. As worker careers are limited and human capital accumulation follows a slow-moving process, business cycle effects are hard to fade and the quality of initial matches bears a long-standing effect on workers careers.

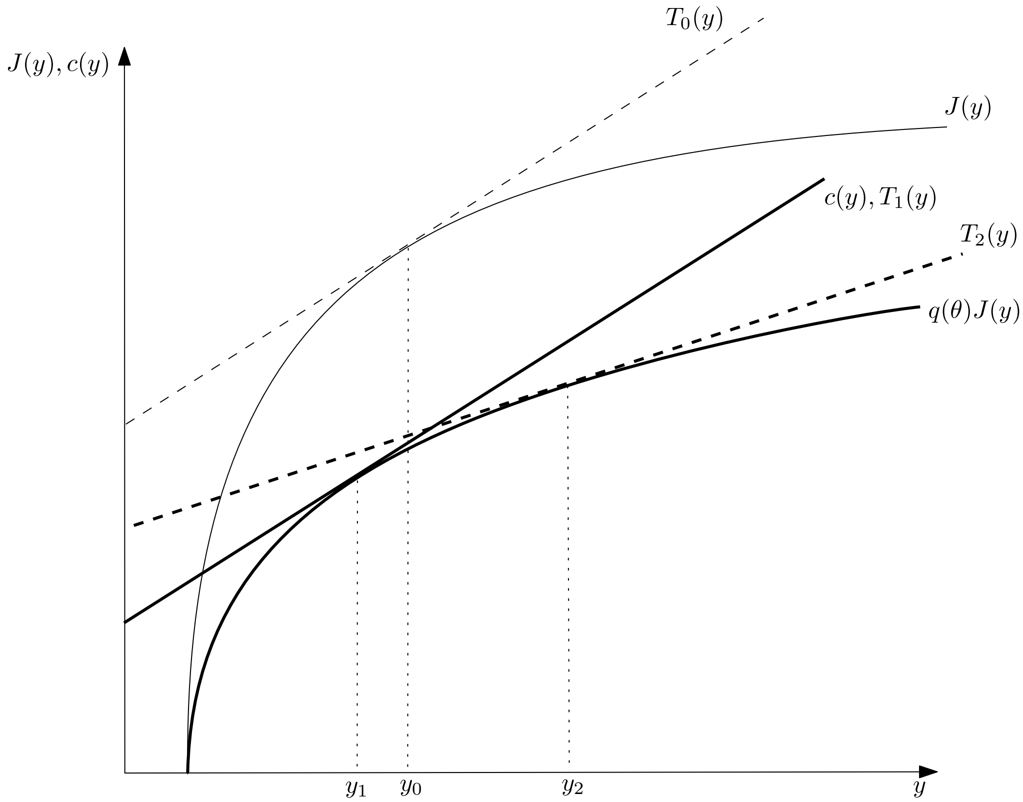



Figure 3.2: Optimal quality choice for entrepreneurs

Note: The figure shows the intuition behind the proof of Proposition 2. Given $(h, \tau, W_y, \Omega)$, an entrepreneur, ex-ante, will ideally select the firm quality that guarantees the maximum possible profits. Free entry in vacancy posting, however, guarantees that, as long as positive profits are available in the submarket, entrepreneurs will continue posting vacancies driving down the vacancy filling probability, $q(\theta)$, and lowering the expected value of a match and progressively pushing firms out. In equilibrium, only entrepreneurs with one firm quality, in the figure's example $y_1$, will find it profitable to remain in the market.

## Characteristics of the optimal contract

The contracting problem between the worker and the firms allows us to analyse directly the trade-offs between the insurance provision and the optimal search behaviour of the workers. The following proposition allows us to characterize the various incentives along the business cycle.

**Proposition 2** *The Pareto frontier $J(h, \tau, y, W_{y,\Omega}; a, \mu)$ is increasing in the aggregate productivity shock $a$, while retention probabilities, $\widetilde{p}(h, \tau, W_{y,\Omega}; a, \mu)$ decrease in aggregate productivity.*

See **Proposition 7** in **Appendix 3.A**.

The intuition behind this proposition relies on the observation that higher productivity realization are associated not only with better outcomes on impact but also to better future prospects due to the monotonically increasing Markov process of productivity. In addition, the model allows us to characterize the optimal behaviour of the workers along the business cycle. The following proposition summarizes how the search strategy changes depending on the realization of aggregate productivity.

**Corollary 3.5.1** *The optimal search strategy of the workers is increasing in aggregate productivity.*

The claim follows directly from the fact that retention probabilities at the Pareto frontier, $\widetilde{p}$, are decreasing in $a$ as discussed in **Proposition** 2.

As discussed more in detail in **Proposition** 2, positive shocks to aggregate productivity will induce workers to search in submarkets that offer higher life-time utilities as firms optimally react to the positive shock increasing the number of vacancies posted. As firm promises and firm-qualities are linked by a one-to-one mapping, as discussed in **Proposition** 4, in booms workers manage to get matched with better firms.[49]

The combination of these two effects imply that as firms became more productive, they are willing to post more vacancies in each submarket to increase the prospect of hiring in good times incorporating the higher benefits due to to the improves future prospects of aggregate productivity. The resulting higher tightness impacts workers' optimal search behaviour as the job finding probability increases in all submarkets. As a consequence, workers respond optimally to the productivity increase searching in submarkets that guarantee higher life-time utility promises.

Firms are willing to commit to higher utility promises as the optimal contract guarantees that the provision of insurance ensures wage paths that are rigid downward and can only partially follow productivity realizations. The following propositions provide a clear picture of the growth path prescribed by the optimal contract for a continuing firm. First, let us define the productivity threshold that determines whether a worker-firm match does not survive.

**Corollary 3.5.2** *There exists a productivity threshold $a^*(h, \tau, y, W_{y,\Omega})$ below which firms will not continue the contract.*

The intuition of why this has to be the case is linked to the fact that the Pareto frontier is strictly increasing in $a$ and decreasing in the level of promised utilities to the worker. Hence once the aggregate state realizes a firm is able to perfectly predict whether next period it will exit the market or stay in (given the timing, the decision is based on expected profits, and is thus *not* state-contingent to next period's productivity). The choice is taken *before* new realizations of productivity, so the firm might for one period at most end up staying but make negative profits.

[49]In our model a better firm is an higher quality firm. We do not specifically model the determinant of quality heterogeneity but we take the existence of profound differences in firm quality as a fact.

**Proposition 3** *For each state in which the firm is willing to continue the contract, the optimal contract delivers a wage path that follows firms profits according to the wage Euler equation:*

$$\frac{\partial \widetilde{p}(\Theta)}{\partial W_{iy,\Omega'}} \frac{J_{t+1}(\Theta)}{\widetilde{p}(\Theta)} = \frac{1}{u'(w_{i,\Omega'})} - \frac{1}{u'(w_i)} \tag{3.16}$$

*with* $\Theta \equiv (g(h,y), \tau+1, W_{iy,\Omega'}; \Omega')$ *being the definition of the relevant state and* $w_{i,\Omega'}$ *is the wage paid in the future state.*

See **Proposition 10** in **Appendix 3.A**.

The optimal contract links the wage growth to the realization of firms profits. The right hand side of **Equation 3.16** shows that, in providing insurance to the worker, the firm links wage growth to profits and to the incentive to maximize retention, incorporated in $\frac{\partial \log \widetilde{p}}{\partial W_y}$ As the production stage happens *after* exit choices are taken by the incumbent firms, the wage growth related to the continuation value of the contract is bound to be (weakly) positive, hence workers enjoy an non-decreasing wage profile under the optimal contract.[50]

A feature that the optimal contract derived in our model shares with the literature on long-term contracts with lack of commitment on the worker side is the backloading of wages. Workers in our model make search decisions that are going to affect the survival probability of the match. However, in doing so they do not appropriate the full future value of the match when making these decisions (unless the firm makes zero profits). This makes it optimal for the firm to front-load profits and back-load wages, as already among others noted by Tsuyuhara 2016, and Lamadon 2016. The reason is that the firm provides insurance and income smoothing to the worker, but given its risk neutrality it prefers to front-load its profits and provide an increasing compensation path to maximize retention. The contract thus optimally balances the consumption smoothing motives (i.e. the insurance provision of the contract) with the commitment problem of the worker.

The next proposition, instead, confirms our initial conjecture that in equilibrium firm qualities and utility promises are related to a one-to-one mapping.

**Proposition 4** *The mapping defined by the function* $f_v : \mathcal{Y} \to \mathcal{V}$ *is an injective function for each worker characteristic* $(h, \tau)$.

See **Proposition 8** in **Appendix 3.A**.

[50] As the exit decision takes place by considering *expected* profits next period, a firm might continue operating low but positive expected profits and end up, at most for a period, to have a negative continuation value. This would imply that wage growth *can* be negative before a firm's closure, which is actually a common finding in empirical studies (firstly observed in Ashenfelter 1978).

The intuition for this result is better expressed graphically as in **Figure** 3.2. The figure shows the optimal choice of an entrepreneur that ex-ante has to choose in a submarket where the utility promise a generic $W$. Out of the equilibrium path, for a given $\theta$, the optimal choice for the entrepreneur would be to pick the firm quality that guarantees the highest possible return to open a vacancy. The possibility of making positive profits however, attracts additional potential entrants. The increase in vacancy posting in the submarket drives the tightness up, lowering the ex-ante return to opening a vacancy to 0. As the firm value function is concave in $y$, in equilibrium, only one type of firm would be able to fend off the competition and remain active in the submarket.

Finally, we provide the alternative recursive formulation for the contracting problem described in the paper. The saddle-point functional equation that can be, alternatively used, to define the recursive contract in **Equation** (3.10) is expressed in the following proposition.

**Proposition 5** *The solution to the contracting problem in **Equation** (3.10) is the same as the solution to the following saddle-point functional equation:*

$$\begin{aligned}\mathcal{P}_t(h_t, \tau_t, y_t, a_t, \gamma_t) =& \underset{\gamma_t}{inf}\, \underset{w_t}{sup}(f(a_t, y_t, h_t) - w_t) + \mu_t^1 W_{y,t} - \gamma_t^1 (W_{y,t} - u(w_t)) + \\ & \beta \mathbb{E}_t(\lambda U_{t+1} + (1-\lambda)\lambda_e p_{t+1} v^*_{t+1}) + \beta \mathbb{E}_t \widetilde{p}_{t+1} \mathcal{P}_{t+1}(h_{t+1}, \tau_{t+1}, y_{t+1}, a_{t+1}, \gamma_{t+1})\end{aligned}$$

*with $\mu_t = \gamma_{t_1}$ for some starting $\gamma_0$.*

See **Appendix 3.A** for the details of the derivation of the SPFE following Marcet and Marimon 2019.

## 3.6 Conclusion

In this paper we develop a rich model of on-the-job search and human capital accumulation that features heterogeneity both on the worker and on the firm side. In the model ex-ante heterogeneous workers accumulate on-the-job experience which augments their skills and helps them climb the job ladder. Search frictions in the labor market and the presence of aggregate uncertainty prevent an efficient allocation of workers to firms and expose different cohort of workers to different human capital accumulation paths depending on the aggregate state at the time of entry in the labor force.

We construct a contractual framework that endogenously accounts for the difference incentives between risk-averse workers and risk-neutral entrepreneurs. We characterize how insurance incentives are of paramount importance in shaping the response of the labor market, the efficiency of workers-firms matches and the overall dynamic of human capital accumulation. More relevantly, we show how even in absence of institutional frictions optimal contract can endogenously generate rigidities in compensation. Employment relationships are subject to a one-sided limited commitment problem, where the firm can commit to a contingent wage path but has limited liability, and are regulated by a dynamic contract that endogenously determines the optimal provision of insurance to workers. Within this framework we show that limited liability on the firm side generates downward wage rigidity as the optimal contract prescribes the firm to pay the worker a (almost) never decreasing compensation path, and that aggregate fluctuations have the ability to affects workers search decisions. We establish that workers that look for employment in bad economic times direct their search towards less productive firms. This limits their ability to accumulate human capital and imposes a drag on the overall labor productivity of the economy that persists as long as these cohorts of workers are active in the labor force. We are currently working on providing a numerical solution of the model, in order to bring it to the data in order to quantify the effects of the mechanisms discussed for workers careers and the overall economic activity.

## References


Abowd, John M, Kevin L McKinney, and Nellie L Zhao (2018). "Earnings inequality and mobility trends in the United States: Nationally representative estimates from longitudinally linked employer-employee data". In: *Journal of Labor Economics* 36.S1, S183–S300.

Altonji, Joseph G, Lisa B Kahn, and Jamin D Speer (2016). "Cashier or consultant? Entry labor market conditions, field of study, and career success". In: *Journal of Labor Economics* 34.S1, S361–S401.

Arellano-Bover, Jaime (2020). *Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size*. Working Paper.

Ashenfelter, Orley (1978). "Estimating the effect of training programs on earnings". In: *The Review of Economics and Statistics*, pp. 47–57.

Benveniste, L.M. and J.A. Scheinkman (1979). "On the Differentiability of the Value Function in Dynamic Models of Economics". In: *Econometrica* 47.3, pp. 727–732.

Bonhomme, Stephane, Thibaut Lamadon, and Elena Manresa (2019). "A Distributional Framework for Matched Employer Employee Data". In: *Econometrica* 87 (3), pp. 1–71.

Burdett, Kenneth, Carlos Carrillo-Tudela, and Melvyn Coles (2016). "The Cost of Job Loss". In:

Cole, Harold and Felix Kubler (2012). "Recursive contracts, lotteries and weakly concave pareto sets". In: *Review of Economic Dynamics* 15.4, pp. 479–500.

Fernández-Kranz, Daniel and Núria Rodríguez-Planas (2017). *The Perfect Storm: Graduating in a Recession in a Segmented Labor Market*. IZA Discussion Paper.

Guo, Naijia (2018). "The effect of an early career recession on schooling and lifetime welfare". In: *International Economic Review* 59.3, pp. 1511–1545. ISSN: 14682354. DOI: 10.1111/iere.12312.

Harris, Milton and Bengt Holmstrom (1982). "A Theory of Wage Dynamics". In: *The Review of Economic Studies* 49.3, p. 315. ISSN: 00346527. DOI: 10.2307/2297359.

Heckman, James and Richard Robb (1985). "Using longitudinal data to estimate age, period and cohort effects in earnings equations". In: *Cohort analysis in social research*. Springer, pp. 137–150.

Herkenhoff, Kyle F., Gordon Phillips, and Ethan Cohen-Cole (2019). "How Credit Constraints Impact Job Finding Rates, Sorting & Aggregate Output". In: pp. 1–78. ISSN: 0717-6163. DOI: 10.3386/w22274. arXiv: arXiv:1011.1669v3.

Huckfeldt, Christopher (2018). *Understanding the Scarring Effect of Recessions*. Tech. rep. Society for Economic Dynamics.

Jacobson, Louis J., Robert J. LaLonde, and Daniel G. Sullivan (1993). “Earnings Losses of Displaced Workers”. In: *The American Economic Review* 83.4, pp. 685–709.

Jarosch, Gregor (2015). “Searching for Job Security and the Consequences of Job Loss”. In:

Kahn, Lisa B. (2010). “The long-term labor market consequences of graduating from college in a bad economy”. In: *Labour Economics* 17.2, pp. 303–316. ISSN: 09275371. DOI: 10.1016/j.labeco.2009.09.002.

Kocherlakota, Narayana R (1996). “Implications of efficient risk sharing without commitment”. In: *The Review of Economic Studies* 63.4, pp. 595–609.

Krueger, Dirk and Harald Uhlig (2006). “Competitive risk sharing contracts with one-sided commitment”. In: *Journal of Monetary Economics* 53.7, pp. 1661–1691. ISSN: 03043932. DOI: 10.1016/j.jmoneco.2005.07.019.

Lachowska, Marta, Alexandre Mas, and Stephen A. Woodbury (2017). “Sources of Displaced Workers’ Long-Term Earnings Losses”. In:

Lamadon, Thibaut (2016). “Productivity Shocks , Long-Term Contracts and Earnings Dynamics”. In: pp. 1–86.

Lise, Jeremy and Fabien Postel-Vinay (2019). “Multidimensional Skills , Sorting , and Human Capital Accumulation”.

Lise, Jeremy and Jean-Marc Robin (2017). “The macrodynamics of sorting between workers and firms”. In: *American Economic Review*. ISSN: 00028282. DOI: 10.1257/aer.20131118.

Ljungqvist, Lars and Thomas J Sargent (1998). “The European Unemployment Dilemma”. In: *Journal of Political Economy* 106.3, pp. 514–550.

Mele, Antonio (2014). “Repeated moral hazard and recursive Lagrangeans”. In: *Journal of Economic Dynamics and Control* 42, pp. 69–85.

Menzio, Guido and Shouyong Shi (2010). “Block recursive equilibria for stochastic models of search on the job”. In: *Journal of Economic Theory*. ISSN: 00220531. DOI: 10.1016/j.jet.2009.10.016.

Menzio, Guido, Irina A Telyukova, and Ludo Visschers (2016). “Directed Search over the Life Cycle ”. In: *Review of Economic Dynamics* 124.3, pp. 771–825.

Messner, Matthias, Nicola Pavoni, and Christopher Sleet (2012). “Recursive methods for incentive problems”. In: *Review of Economic Dynamics* 15.4, pp. 501–525.

OECD (2018). *Indicators of Employment Protection*. http://www.oecd.org/employment/emp/oecdindicatorsofemploymentprotection.htm.

Oreopoulos, Philip, Till Von Wachter, and Andrew Heisz (2012). “The short-and long-term career effects of graduating in a recession”. In: *American Economic Journal: Applied Economics* 4.1, pp. 1–29.

Rudin, Walter (1976). *Principles of Mathematical Analysis*. McGraw-Hill Book Company.

Schmieder, Johannes F., Till von Wachter, and Joerg Heining (2018). "The Costs of Job Displacement over the Business Cycle and Its Sources : evidence from Germany". In: *Working Paper*.

Schwandt, Hannes and Till von Wachter (2019). "Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets". In: *Journal of Labor Economics* 37.S1, S161–S198.

Shi, Shouyong (2009). "Directed search for equilibrium wage-tenure contracts". In: *Econometrica* 77.2, pp. 561–584.

Spear, Stephen E. and Sanjay Srivastava (1987). "On Repeated Moral Hazard with Discounting". In: *The Review of Economic Studies* 54.4, p. 599. ISSN: 00346527. DOI: 10.2307/2297484.

Thomas, Jonathan and Tim Worral (1988). "Self-enforcing wage contracts ". In: *The Review of Economic Studies* 55.4, pp. 541–553.

Tsuyuhara, Kunio (2016). "Dynamic Contracts With Worker Mobility Via Directed on-the-Job Search". In: *International Economic Review* 57.4, pp. 1405–1424. ISSN: 14682354. DOI: 10.1111/iere.12202.

Von Stackelberg, Heinrich (1934). *Marktform und gleichgewicht*. J. springer.

Wee, Shu Lin (2016). "Delayed Learning and Human Capital Accumulation : The Cost of Entering the Job Market During a Recession".

World Bank (2018). *Doing Business*. Labor Market Regulation, http://www.doingbusiness.org/~/media/WBG/DoingBusiness/Documents/Annual-Reports/English/DB2018-Full-Report.pdf.

Xiaolan, Mindy Z. (2014). *Who Bears Firm-Level Risk? Implications for cash-flow volatility*. Working Paper.

## 3.A Appendix

### Properties of worker optimal behavior

The following propositions characterize the properties of workers' optimal search strategies that solve the search problem in (3.4), restated here for convenience:

$$R(h,\tau,V;\Omega)=\sup_{\{v_{y,\Omega}\}}\Big[p(\theta(h,\tau,v_{y,\Omega};\Omega))\big[v_{y,\Omega}-V\big]\Big]. \tag{3.17}$$

**Lemma 3.A.1** *The composite function $p(\theta(h,\tau,v;\Omega))$ is strictly decreasing and strictly concave in $v$.*

For this proof we follow closely Menzio and Shi 2010, Lemma 4.1 (ii). From the properties of the matching function we know that $p(\theta)$ is increasing and concave in $\theta$, while $q(\theta)$ is decreasing and convex. Consider that the equilibrium definition of $\theta(\cdot)$ is

$$\theta(h,\tau,v;\Omega)=q^{-1}\left(\frac{c(y)}{\beta J(h,\tau,y,v;\Omega)}\right),$$

and that the first order condition for the wage and the envelope condition on $V$ of the optimal contract problem in (3.10) implies

$$\frac{\partial J(h,\tau,y,v;\Omega)}{\partial v}=-\frac{1}{u'(w)}.$$

so that as $u'(\cdot)>0$, $J(\cdot)$ is decreasing in $v$.

From the equilibrium definition of $\theta(\cdot)$ and noting that $q^{-1}(\cdot)$ is also decreasing due to the properties of the matching function we have that

$$\frac{\partial\theta(h,\tau,v;\Omega)}{\partial v}=\left.\frac{\partial q^{-1}(\xi)}{\partial\xi}\right|_{\xi=\frac{c(y)}{\beta J(h,\tau,y,v;\Omega)}}\cdot\left(-\frac{\partial J(h,\tau,y,v;\Omega)}{\partial v}\right)\cdot\frac{c(y)}{\beta(J(h,\tau,y,v;\Omega))^2}<0,$$

which, in turn, implies that

$$\frac{\partial p(\theta(h,\tau,v;\Omega))}{\partial v}=\left.\frac{\partial p(\theta)}{\partial\theta}\right|_{\theta=\theta(h,\tau,v;\Omega)}\cdot\frac{\partial\theta(h,\tau,v;\Omega)}{\partial v}<0.$$

Suppressing dependence on the states $(h,\tau,y,\Omega)$ for readability, to prove that $p(\theta(v))$ is concave, consider that $J(v)$ is concave[51] and a generic function $\frac{c}{v}$ is strictly convex in $v$. This implies that with $\alpha\in[0,1]$ and $v_1,v_2\in\mathcal{V}$:

$$\frac{c}{J(\alpha v_1+(1-\alpha)v_2)}\le\frac{c}{\alpha J(v_1)+(1-\alpha)J(v_2)}<\alpha\frac{c}{J(v_1)}+(1-\alpha)\frac{c}{J(v_2)}.$$

[51] $J()$ concave give the two-point lottery in the structure of the contract. See Menzio and Shi 2010 Lemma F.1.

As $p(q^{-1}(\cdot))$ is strictly decreasing the inequality implies that

$$\begin{aligned} p\left(q^{-1}\left(\frac{c}{J(\alpha v_1+(1-\alpha)v_2)}\right)\right) &\geq p\left(q^{-1}\left(\frac{c}{\alpha J(v_1)+(1-\alpha)J(v_2)}\right)\right) \\ &> \alpha p\left(q^{-1}\left(\frac{c}{J(v_1)}\right)\right)+(1-\alpha)p\left(q^{-1}\left(\frac{c}{J(v_2)}\right)\right), \end{aligned}$$

and as $\theta(v) = q^{-1}(\frac{c}{J(v)})$:

$$p(\theta(\alpha v_1 + (1-\alpha)v_2)) > \alpha p(\theta(v_1)) + (1-\alpha)p(\theta(v_2))$$

so that $p(\theta())$ is strictly concave in $v$.

**Proposition 6** *Given the worker search problem, the following properties hold:*

*(i) The returns to search,* $p(\theta(h,\tau,v_{y,\Omega};\Omega))\Big[v_{y,\Omega} - V\Big]$, *are strictly concave with respect to promised utility,* $v_{y,\Omega}$.

*(ii) The optimal search strategy*

$$v^*(h,\tau,V;\Omega) \in \arg\max_{v_y} \Big\{p(\theta(h,\tau,v_{y,\Omega};\Omega))\Big[v_{y,\Omega} - V\Big]\Big\}$$

*is unique and weakly increasing (and Lipschitz continuous) in* $V$.

*(iii) For all promised utilities, the search gain* $R(h,\tau,V;\Omega)$ *is positive, weakly decreasing in* $V$.

*(iv) The survival probability of the match, given the optimal choice of the worker, is increasing in the value of promised utilities, so* $\widetilde{p}_t(h,\tau,W_{y,\Omega};\Omega)$ *is increasing (and Lipschitz continuous) in* $W_{y,\Omega}$.

The proofs follow closely Shi 2009, Lemma 3.1 and Menzio and Shi 2010, Lemma 4.4. More formally, for each triplet $(h,\tau,\Omega)$ given at each search stage, we can re-define the search objective function as $K(v,V) = p(\theta(v))(v-V)$ and $v^*(V) \in \arg\max_v K(v,V)$ as the function that maximises the search returns (i.e. the optimal search strategy of the worker) and prove the following

(i) To show that $K(v,V)$ is strictly concave in $v$ consider two values for $v$, $(v_1,v_2)$ such that $v_2 > v_1$ and define $v_\alpha = \alpha v_1 + (1-\alpha)v_2$ for $\alpha \in [0,1]$.

Then by definition:

$$\begin{aligned} K(v_\alpha,V) &= p(\theta(v_\alpha))(v_\alpha - V) \\ &\geq [\alpha p(\theta(v_1)) + (1-\alpha)p(\theta(v_2))][\alpha(v_1 - V) + (1-\alpha)(v_2 - V)] \\ &= \alpha K(v_1,V) + (1-\alpha)K(v_2,V) + \alpha(1-\alpha)[(p(\theta(v_1)) - p(\theta(v_2))](v_2 - v_1) \\ &> \alpha K(v_1,V) + (1-\alpha)K(v_2,V) \end{aligned}$$

where the first inequality follows from the concavity of $p(\theta(\cdot))$ (this is true if $J()$ concave with respect to $V$) and the second inequality stems from the fact that $p(\theta(\cdot))$ is strictly decreasing hence $\alpha(1-\alpha)[(p(\theta(v_1))-p(\theta(v_2))](v_2-v_1)>0$.

(ii) Given that $v \in [v, \overline{v}]$, and submarkets are going to open depending on realizations of the aggregate productivity, $a$, there is only one region in the set of promised utilities where the search gain is positive, conditional on being in a job that pays lifetime utility $V$. That is $[V, v(a)]$ with $v(a)$ being the highest possible offer that a firm makes in the submarket for the worker $(h, \tau)$. As any submarket that promises higher than $v(a)$ is going to have zero tightness, the optimal search strategy for $V \geq v(a)$ is $v^*(V) = V$. For $V \in [V, v(a)]$, instead, as $K(v, V)$ is bounded and continuous, the solution $v^*(V)$ has to be internal and therefore respect the following first order condition

$$V = v^*(V) + \frac{p(\theta(v^*(V)))}{p'(\theta(v^*(V)) \cdot \theta'(v^*(V))}. \tag{3.18}$$

Now consider two arbitrary values $V_1$ and $V_2$, $V_1 < V_2 < \overline{v}$ and their associated solutions $W_i = v^*(V_i)$ for $i = 1, 2$.

Then, $V_1$ and $V_2$ have to generate two different values for the right-hand side of (3.18). Hence, $v^*(V_1) \cap v^*(V_2) = \emptyset$ when $V_1 \neq V_2$. This also implies that the search gain evaluated at the optimal search strategy is higher than the gain at any other arbitrary strategy so that $K(W_i, V_i) > K(W_j, V_i)$ for $i \neq j$. This implies that

$$\begin{aligned} 0 > &[K(W_2, V_1) - K(W_1, V_1)] + [K(W_1, V_2) - K(W_2, V_2)] \\ &= (p(\theta(W_2)) - p(\theta(W_1)))(V_2 - V_1). \end{aligned}$$

Thus, $p(\theta(W_2)) < p(\theta(W_1)))$ and as $p(\theta(\cdot))$ is strictly decreasing (Corollary 3.A.1) $v^*(V_1) < v^*(V_2)$. Uniqueness follows directly from strict concavity shown in (i). Lipschitz continuity still to show but coming from assumption of $J()$ being bi-Lipschitz continuous and $\theta(), p()$ being bounded functions.

(iii) The Bellman equation for the search problem is:

$$R(h, \tau, V; \Omega) = \sup_{\{v_{y,\Omega}\}} \Big[ p(\theta(h, \tau, v_{y,\Omega}; \Omega)) \big[ v_{y,\Omega} - V \big] \Big]$$

hence a simple envelope argument shows that

$$\frac{\partial R(h, \tau, V; \Omega)}{\partial V} = -p(\theta(h, \tau, v_y; \Omega)) \leq 0,$$

as the job finding probability is weakly positive for all utility promises.

As $p(\theta(\cdot)) \geq 0$, $v^*(\cdot) \in [v, \overline{v}]$ then $R(\cdot) \geq 0$.

(iv) Given the optimal search strategy, $v^*(h,\tau,V;\Omega)$, we can define the survival probability of the match as in (3.7):

$$\widetilde{p}(h,\tau,V_{y,\Omega};\Omega) \equiv (1-\lambda)(1-\lambda_e p(\theta(h,\tau,v^*_{y,\Omega};\Omega))).$$

Then, given $(h,\tau,\Omega)$

$$\frac{\partial \widetilde{p}(V)}{\partial V} = -\beta(1-\lambda)\lambda_e \left.\frac{\partial p(\theta)}{\partial \theta}\right|_{\theta=\theta(v^*)} \left.\frac{\partial \theta(v)}{\partial v}\right|_{v=v^*(V)} \frac{\partial v^*(V)}{\partial V} > 0,$$

because $p(\cdot)$ and $v^*(\cdot)$ are both increasing functions while $\theta(\cdot)$ is a decreasing function in promised utilities.

## Properties of the optimal contract

**Lemma 3.A.2** *The Pareto frontier $J(h,\tau,y,W_{y,\Omega};\Omega)$ is concave in $W_{y,\Omega}$.*

This is a direct consequence of using a two-point lottery for $\{w_i, W_{iy,\Omega'}\}$ as shown by Menzio and Shi 2010, Lemma F.1.

**Lemma 3.A.3** *The Pareto frontier $J(h,\tau,y,W_{y,\Omega};\Omega)$ is increasing in $y$.*

The intuition for this proof follows the fact that a higher $y$ firm, once the match exists, can always deliver a certain promise $V$ and have resources left over. Within a dynamic contract, future retention is already optimized as the match is formed. This means that the promise $V$ can be delivered by the greater capacity on the part of producing with respect to a close $y$ firm. In presence of human capital accumulation, the worker is compensated through greater option values in the future, which again means that, even with lower retention, the firm cashes in more profits while decreasing wages (and respecting the $V$ promise). The reason why one does not have to worry about, for instance, variation in retention is that we are evaluating changes in $y$ given the optimal contract, and given that by definition $J$ is maximized, any indirect derivative of controls over $y$ will get to their respective first order conditions and thus have no direct impact on the comparative static.

One can get to the same conclusion by starting from time $T$, noticing that the function $J$ is trivially increasing in $y$ in the last period, and the stepping back. At $T-1$, given $V$, any higher $y$ function can make greater profits with the same delivery of value $V$, given the contract's optimal promise, which is a fortiori true with human capital accumulation (the option value is greater, so the firm can decrease $w$ as a response).

**Proposition 7** *The Pareto frontier $J(h,\tau,y,W_{y,\Omega};a,\mu)$ is strictly increasing in the aggregate productivity shock $a$, while retention probabilities, $\widetilde{p}(h,\tau,y,W_{y,\Omega};\Omega)$ decrease in aggregate productivity.*

For a generic period $t$, a firm matched to a worker in submarket $\{h, T-1, y, W_{y,\Omega}\}$ will face the following Pareto frontier

$$J_t(h, T-1, y, W_{y,\Omega}; a, \mu) = \sup_{w_i, \{W_{iy,\Omega'}\}} \Big( f(y, h; \Omega) - w \\ + \mathbb{E}_\Omega \Big[ \widetilde{p}(h', \tau+1, W_{y,\Omega'}; a', \mu')(f(y, h; a') - w') \Big] \Big)$$

The fact that period flows are increasing in $a$ is immediate and follows from the properties of contracts with one-sided lack of commitment, as in Thomas and Worral 1988, Kocherlakota 1996 or Krueger and Uhlig 2006. At the same time, following the logic of **Lemma 3.A.3**, the envelope condition on controls guarantees that one does not have to worry about the variation in optimal retention. This proves that $J$ is increasing in $a$.

For the second part of the statemnt, notice that, in equilibrium,

$$\frac{\partial p(\theta)}{\partial a} = \underbrace{\frac{\partial p(\theta)}{\partial \theta}}_{>0} \cdot \underbrace{\frac{\partial \theta}{\partial J(\cdot)}}_{>0} \cdot \frac{\partial J(\cdot)}{\partial a}$$

where the sign of the second derivative on the right hand side comes from the free entry condition and the properties of vacancy filling probability function $q(\cdot)$. Given this, it has to be that $\frac{\partial p(\theta)}{\partial a}$ and $\frac{\partial J(\cdot)}{\partial a}$ have the same sign in equilibrium. This immediately implies that $\frac{\partial \widetilde{p}}{\partial a} < 0$ according to the optimal contract.

**Corollary 3.A.1** *There exists a productivity threshold $a^*(h, \tau, y, W_{y,\Omega})$ below which firms will not continue the contract.*

The proof follows immediately from **Proposition 7** and the timing of the shock. Given the timing of the shock, exit is fully determined by the current productivity shock and incumbent firms know in advance whether they are willing to produce in the next period.

Therefore, as the Pareto frontier is strictly increasing in $a$, firms are willing to continue the contract if $\mathbb{E}_\Omega[J_{t+1}(h', \tau+1, y, W_{y,\Omega'}; a', \mu') | h, \tau, y, W_{y,\Omega}, a, \mu] \geq 0$, so that the threshold that determines exit is

$$a^*(h, \tau, y, W_{y,\Omega}) : \mathbb{E}_\Omega[J_{t+1}(h', \tau+1, y, W_{y,\Omega'}; a', \mu') | h, \tau, y, W_{y,\Omega}, a, \mu] = 0.$$

**Corollary 3.A.2** *The productivity threshold $a^*(h, \tau, y, W_{y,\Omega})$ below which firm $y$ in match with worker $(h, \tau)$ and given promised utility $W_{y,\Omega}$ in the aggregate state $\Omega$ is decreasing in $y$.*

Consider two firms characterized by $y_1, y_2$ with $y_1 < y_2$. Consider the threshold for firm $y_1$, $a_1^* = a^*(h, \tau, y_1, W_{y,\Omega})$. Firm $y_1$ makes 0 profits if state $a_1^*$ materializes next period. Consider firm $y_2$ trying to mimic the current contract offered by $y_1$ to $(h, \tau)$. We know that $J$ is increasing in $y$ from **Lemma** 3.A.3, which implies that the firm is making a profit at $a_1^*$. This completes the proof.

**Lemma 3.A.4** *The Pareto frontier $J(h, \tau, y, W_{y,\Omega}; \Omega)$ is strictly concave in $y$.*

The proof follows from the fact that the flow component of the profit function is always a concave function in $y$.

More formally, start from the last period $T$. The concavity is trivially given by the concavity of $f$. Now moving backwards to the problem at $\tau = T - 1$, one can still consider the behavior of $J$ given a promise $W_{y,\Omega}$. Again, given the option to search, the flow value is concave in $y$, retention probability is constant in $W_{y,\Omega}$, and the continuation value is a concave function. By induction, the statement holds for $J$ at all $\tau \in [0, T]$.

**Corollary 3.A.3** *As $J_t(h, \tau, y, W_{y,\Omega}; \Omega)$ is concave, the tangent line at a generic $y_0 \in \mathcal{Y}$ is above the graph of $J_t(h, \tau, y, W_{y,\Omega}; \Omega)$ so that*

$$J_t(h, \tau, y_0, W_{y,\Omega}; \Omega) + \left.\frac{\partial J_t(h, \tau, y, W_{y,\Omega}; \Omega)}{\partial y}\right|_{y=y_0} (y - y_0) \geq J_t(h, \tau, y, W_{y,\Omega}; \Omega).$$

Dropping dependence on $(h, \tau, W_{y,\Omega}; \Omega)$, consider two values for firm quality $y_0 < y_1$ both in $\mathcal{Y}$. Then, as $J_t(\cdot)$ is concave in $y$, taking $\alpha \in [0, 1]$ the following inequalities are true:

$$\begin{aligned}
&J(\alpha y_0 + (1-\alpha) y_1) \geq \alpha J(y_0) + (1-\alpha) J(y_1) \\
\Rightarrow &J(\alpha y_0 + (1-\alpha) y_1) - J(y_0) \geq (1-\alpha)(J(y_1) - J(y_0)) \\
\Rightarrow &\frac{J(\alpha y_0 + (1-\alpha) y_1) - J(y_0)}{\alpha y_0 + (1-\alpha) y_1 - y_0} \geq \frac{J(y_1) - J(y_0)}{y_1 - y_0}.
\end{aligned}$$

where the third inequality comes from noting that $y_1 > y_0$ and $\alpha y_0 + (1-\alpha) y_1 - y_0 = (1-\alpha)(y_1 - y_0)$.

Taking the limit for $\alpha \to 1$, we have that the left hand side tends to $\left.\frac{\partial J_t(y)}{\partial y}\right|_{y=y_0}$ and hence

$$J(y_0) + \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_0} (y_1 - y_0) \geq J(y_1). \tag{3.19}$$

Note that if $y_0 > y_1$ then $\frac{J(\alpha y_0 + (1-\alpha) y_1) - J(y_0)}{\alpha y_0 + (1-\alpha) y_1 - y_0} \leq \frac{J(y_1) - J(y_0)}{y_1 - y_0}$ but multiplying again the left hand side and the right hand side for $(y_1 - y_0) < 0$ still delivers (3.19).

**Proposition 8** *Define the mapping between promised values and firm installed capital by the function $f_v : \mathcal{Y} \rightarrow \mathcal{V}$. Then $f_v$ is an injective function for each couple of worker characteristics $(h, \tau)$.*

Note: throughout the proof we drop the dependence of the functions to the state $(h, \tau, \Omega)$ to ease readability.

If the function $f_v$ is an injective function then it defines a one-to-one mapping between $\mathcal{Y}$ and $\mathcal{V}$ so that for $(y_1, y_2) \in \mathcal{Y}$, and $f_v(y_1) = W_1$ and $f_v(y_2) = W_2$, $(W_1, W_2) \in \mathcal{V}$, $f_v(y_1) = f_v(y_2) \Rightarrow y_1 = y_2$.[52] We proceed by contradiction. To begin, assume that $f_v(y_1) = f_v(y_2)$ and $y_1 \neq y_2$.

As the optimal contract is a concave function in firm quality, we know that the tangents at each point are above the graph of the function. Thus, we can define the tangents at the two points $y_1, y_2$ as

$$T_1(y) \equiv J(y_1) + \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_1} (y - y_1) \quad \text{and} \quad T_2(y) \equiv J(y_2) + \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_2} (y - y_2).$$

Without loss of generality, consider the case in which $y_2 > y_1$. Knowing that $T_i(y) \geq J(y)$ for $i = 1, 2$ due to the concavity of $J(\cdot)$, we can define the following inequalities:

$$T_1(y_2) - J(y_2) \geq 0 \quad \text{and} \quad T_2(y_1) - J(y_1) \geq 0.$$

Using the definitions for the tangents at $y_1$ and $y_2$ they imply that

$$\frac{J(y_2) - J(y_1)}{y_2 - y_1} \leq \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_1} \quad \text{and} \quad \frac{J(y_2) - J(y_1)}{y_2 - y_1} \geq \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_2},$$

hence combining the inequalities we get that

$$\left.\frac{\partial J(y)}{\partial y}\right|_{y=y_2} \leq \frac{J(y_2) - J(y_1)}{y_2 - y_1} \leq \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_1}. \tag{3.20}$$

However, the free-entry condition in vacancy posting implies that in the submarket $(h, \tau, W)$ both firms must be respecting $c(y_i) = q(\theta)\beta J(y_i)$ for $i = 1, 2$. As $c(y_i)$ is a linear function of firm quality $\frac{\partial c(y_i)}{\partial y_i} = c$ for $i = 1, 2$ and therefore from the free-entry condition:

$$c = q(\theta)\beta \left.\frac{\partial J(y)}{\partial y}\right|_{y=y_i}$$

which is a contradiction of the slopes of the two tangents being decreasing as shown in **Equation** (3.20). Note that if $c(y)$ is convex and twice differentiable, then the derivatives

[52] As the contrapositive of Definition 2.2 in Rudin 1976, that defines a one-to-one mapping for $(x_1, x_2) \in A$ as $x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$.

of $c(y)$ are increasing in $y$ while the derivatives of $J(\cdot)$ are decreasing leading again to a contradiction. The proof for the case in which $y_1 > y_2$ follows the same arguments and leads to a similar contradiction on the implied slopes of the optimal contract and those implied by the free entry condition.

**Lemma 3.A.5** *Given a state $(y, h, \tau, W_{y,\Omega})$ the optimal contract implies that*

$$-\frac{\partial J_t(h, \tau, y, W_{y,\Omega}; \Omega)}{\partial W_{y,\Omega}} = \frac{1}{u'(w)}$$

*so that promised utilities and wages move in the same direction.*

The proof follows directly from the envelope theorem and the concavity of the utility function $u(\cdot)$, as discussed in the proof of Proposition 10.

**Corollary 3.A.4** *The Pareto frontier $J(h, \tau, y, W_{y,\Omega}; \Omega)$ is decreasing in promised utilities $W_{y,\Omega}$.*

The envelope condition in **Lemma 3.A.5** and note that $u'() \geq 0$.

**Proposition 9** *Assume $q(\theta(h, \tau, W_{y,\Omega}; \Omega))$ is not too convex. Then utility promises are unique and increasing in $y$, $\frac{\partial W}{\partial y} > 0$.*

Uniqueness follows directly from the concavity of the composite function.

The increasing property follows from the maximization of the entrepreneur in the free entry condition **Equation 3.14**.

Assuming the same $(h, \tau, y)$, the entrepreneur has to choose which is the optimal value $W_{y,\Omega}$ to deliver in the contract. We know it is unique by assuming concavity of the composite function (which eventually amounts to assuming that the functional form of $q(\theta(W)))$ is not too convex in $W$.

For the rest of the proof we consider as given the dependence of the functions on $(h, \tau)$ and consider directly the composite function $q(\theta(W))$ as $q(W)$. The optimization involves a trade-off which respects the following first order condition:

$$q_W J(y, W) + q(W) J_W = 0 \tag{3.21}$$

For this to be a unique sup, the second order condition must be negative:

$$q_{WW} J + 2 q_W J_W + q J_{WW} < 0 \tag{3.22}$$

where, as mentioned above, the only element which might lead to a violation is $q_{WW}$ in case it is too convex ($J_{WW} < 0$) by 3.A.2. Notice this hypothesis amounts to assuming that $q(\theta(h,\tau,W_{y,\Omega};\Omega))J(h,\tau,y,W_{y,\Omega};\Omega)$ is concave.

By the implicit function theorem, the derivative of **Equation** 3.21 is:

$$(q_{WW}J + 2q_W J_W + qJ_{WW})W_y + q_W J_y + qJ_{Wy} = 0 \tag{3.23}$$

The first term in parenthesis is negative, as second order condition. The second term is positive, given **Lemma** 3.A.3 and the fact that $q_W$ is positive. The third term is 0, as the partial derivative of $J$ in $y$ does not contain $V$ (which is the reason why **Lemma** 3.A.3 trivially holds). This means that, in order for the equality to be respected, $W_y > 0$.

**Proposition 10** *For each state in which the firm is willing to continue the contract, the optimal contract delivers a wage path that follows firms profits according to the wage Euler equation:*

$$\frac{\partial \widetilde{p}(\Theta)}{\partial W_{iy,\Omega'}}\frac{J_{t+1}(\Theta)}{\widetilde{p}(\Theta)} = \frac{1}{u'(w_{i,\Omega'})} - \frac{1}{u'(w_i)},$$

*with* $\Theta \equiv (g(h,y), \tau+1, W_{iy,\Omega'};\Omega')$ *being the definition of the relevant state.*

Consider the firm problem in **Equation** (3.10), restated here for convenience

$$\begin{aligned} J_t(h,\tau,y,W_{y,\Omega};\Omega) = \sup_{\pi_i, w_i, \{W_{i,\Omega'}\}} \sum_{i=1,2} \pi_i \Big( & f(y,h;\Omega) - w_i \\ & + \mathbb{E}_\Omega \left[ \widetilde{p}(h',\tau+1,W_{iy,\Omega'};\Omega') J_{t+1}(h',\tau+1,y,W_{i,\Omega'};\Omega') \right] \Big) \end{aligned}$$

$$\begin{aligned} s.t.\ [\lambda]\ W_{y,\Omega} &= \sum_{i=1,2} \pi_i \left( u(w_i) + \mathbb{E}_\Omega \widetilde{r}(h',\tau+1,W_{iy,\Omega'};\Omega') \right), \\ \sum_{i=1,2} \pi_i &= 1, \quad h' = g(h,y). \end{aligned}$$

For $i = 1, 2$, the first order conditions with respect to the wage and the promised utilities are:

$$[w_i]:\ \lambda = \frac{1}{u'(w_i)} \tag{3.24}$$

$$[W_{iy,\Omega'}]:\ \pi_i \frac{\partial \widetilde{p}()}{\partial W_{iy,\Omega'}} J_{t+1}() + \widetilde{p}() \frac{\partial J_{t+1}()}{\partial W_{iy,\Omega'}} + \lambda \frac{\partial \widetilde{r}()}{\partial W_{iy,\Omega'}} = 0. \tag{3.25}$$

Note that by definition,

$$\widetilde{r}(h,\tau,V_{y,\Omega};\Omega) \equiv \lambda U(h,\tau;\Omega) + (1-\lambda)\Big[ W_{y,\Omega} + \lambda_e \max\{0, R(h,\tau,V_{y,\Omega};\Omega)\} \Big]$$

therefore we can use the envelope theorem as in Benveniste and Scheinkman 1979, Theorem 1 and the definition in **Equation** (3.7) to derive an expression for the derivative of the employment value in $t+1$ as the period ahead of the following:

$$\frac{\partial \widetilde{r}(h,\tau,W_{y,\Omega};\Omega)}{\partial W_{y,\Omega}} = \widetilde{p}(h,\tau,W_{y,\Omega};\Omega).$$

Similarly, using the envelope condition on the firm problem and the first order condition for the wage, we can establish that

$$\frac{\partial J_t(h,\tau,y,W_{y,\Omega};\Omega))}{\partial W_{y,\Omega}} = -\lambda \;\therefore\; \frac{\partial J_t(h,\tau,y,W_{y,\Omega};\Omega))}{\partial W_{y,\Omega}} = -\frac{1}{u'(w_i)}. \tag{3.26}$$

Moving these two expressions one period ahead, substituting them in (3.25), taking $\pi_i > 0$ and rearranging we have that:

$$\frac{\partial \widetilde{p}(\Theta)}{\partial W_{y,\Omega'}} \frac{J_{t+1}(\Theta)}{\widetilde{p}(\Theta)} = \frac{1}{u'(w_{\Omega'})} - \frac{1}{u'(w)},$$

with $\Theta \equiv (g(h,y),\tau+1,W_{y,\Omega'};\Omega')$ and where $w_{\Omega'}$ is the wage next period in state $\Omega'$.

## Derivation of recursive contract SPFE

Solving the optimal contract and the overall model given the recursive structure obtained by following the promised utility method of Spear and Srivastava 1987 is computationally infeasible. This is due to the fact that the optimal contract requires to define a valid recursive domain and codomain of promised values that respects all the future forward looking constraints. Known solution methods for these kinds of models Abreu1990, although robust, easily become computationally unmanageable as the number of states of the model increases. We thus follow Marcet and Marimon 2019 in deriving a recursive expression for the optimal contract in which the Lagrange multiplier for the promise keeping constraint **Equation 3.26** is added as a co-state of the model, and allows us to circumvent the problem of searching for valid promised values domains altogether.

The reason why the recursive contracts method in Marcet and Marimon 2019 simplifies our problem is simple. As shown in **Equation 3.26**, wage growth and levels in any next period and at every node are determined by the state-contingent multiplier on tomorrow's promise keeping constraints. This considerably reduces the complexity of the problem, as by definition Lagrange multipliers are defined over $\mathbb{R}^+$.

We follow Marcet and Marimon 2019 (hereby MM) and their terminology to define how a recursive saddle point functional equation (SPFE) can be obtained from the sequential formulation of the problem. For the present exposition of the constructive method to obtain the SPFE, for simplicity and without loos of generality, we ignore the randomization of the contract over the lotteries and the limited liability constraint. The latter choice, in particular, does not create any problem in terms of thinking about of developing the sequential problem over time: our choice of timing of exit decision is such as that exiting firms know form the start of their period whether the productivity level is below the critical one $a^*_{h,\tau,y,W}$ for the match $(h,\tau,y,W_y)$, and thus whether they will exit or not. The lack of uncertainty and optimization over the next periods makes the problem of these firms, at some low states, equivalent to the problem of a firm with a lower maximum length (which is $T$, the retirement age, in general). At an exiting state $t$ the firm knows *with certainty* that any $J_j = 0$ for $j > t$, match with a worker of age $T$.

Consider the problem

$$
\begin{aligned}
J_t(h_t,\tau_t,y_t,W_{y_t},a_t) = \sup_{w_t,\{W_{y,s^{t+1}}\}} \Big( & f(a_t,y_t,h_t) - w_t \\
& + \mathbb{E}_{s^t}\Big[\widetilde{p}(h_{t+1},\tau_{t+1},W_{y,s^{t+1}},a_{s^{t+1}})(J_{t+1}(h_{t+1},\tau_{t+1},y_t+1,W_{y,s^{t+1}},a_{s^{t+1}})\Big]\Big) \qquad (3.27)\\
s.t.\ W_t = u(w_t) + \beta\mathbb{E}_{s^t}\Big( & \lambda U_t(h_{t+1},\tau_{t+1},a_{t+1}) + \\
& (1-\lambda)(\lambda_e p_{t+1}(h_{t+1},\tau_{t+1},W_{y,s^{t+1}},a_{s^{t+1}})v^*(h_{t+1},\tau_{t+1},W_{y,s^{t+1}},a_{s^{t+1}}) \\
& + (1-\lambda_e p_{t+1}(h_{t+1},\tau_{t+1},W_{y,s^{t+1}},a_{s^{t+1}}))W_{y,s^{t+1}})\Big) \qquad (3.28)
\end{aligned}
$$

We define as endogenous states $\mathbf{x}_t = [h_t,\tau_t,y_t,W_{y,t}]$, controls $\mathbf{c}_t = [w_t,W_{y,s^{t+1}}]\ \forall t,s^{t+1}$, whereas the only exogenous state is $a_t$. The endogenous states follow the law of motion

$$
\mathbf{x}_{t+1} = \begin{bmatrix} h_{t+1} \\ \tau_{t+1} \\ y_{t+1} \\ W_{y,s^{t+1}} \end{bmatrix} = l(\mathbf{x}_t,\mathbf{c}_t,a_{s^{t+1}}) = \begin{bmatrix} g(h_t,y_t) \\ \tau_t+1 \\ y_t \\ W_{y,s^{t+1}}) \end{bmatrix} \qquad (3.29)
$$

In the subsequent notation, where appropriate, we omit listing all states on which elements in the equation, and subsume their dependence under just listing the time $t$. $J$ can be rewritten, by developing forward the recursion until time $T$, at which the match surely dissolves, as

$$
J_t(\{h_t,\tau_t,y_t,W_{y,t},a_t\}_{t=t_0}^{T-t_0}) = \mathbb{E}_{t_0}\sum_{t=t_0}^{T}\beta^{t-t_0}\prod_{i=0}^{t-t_0}\widetilde{p}_{t_o+i}\Big(f(a_t,y_t,h_t) - w_t\Big) \qquad (3.30)
$$

where $\widetilde{p}_{t_0} = 1$. Notice that the forward-looking constraint in **Equation 3.28** is state contingent and an instance of it applies at *every* node of any possible history $s^t$ $\forall t$ given the prevailing $W_y$ promised at that node. The equilibrium is an instance of subgame perfect Nash equilibrium in which an agent chooses its strategies while anticipating the best response of the following agent, as common in dynamic games with a leader-follower component introduced by Von Stackelberg 1934. The structure of the problem and the solution also shares some commonality with Ramsey optimal policy problems in which a policy maker (in this case the firm) optimizes the utility of all agents according to some weights and taking into account their optimal behavior. [53]

We can redefine the problem:

$$V_{t_0}(\mathbf{x}_t, a_t) = \sup_{\{w_{s^t}, W_{y,s^t}\}} \mathbb{E}_{t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_o+i}\Big(f(a_t, y_t, h_t) - w_t\Big) \tag{3.31}$$

$$s.t. \, [j=0]: \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_o+i}\Big(f(a_t, y_t, h_t) - w_t\Big) - R \geq 0 \tag{3.32}$$

$$\begin{aligned}[][j=1, s^t]: \; & W_{y,s^t} - u(w_{s^t}) - \beta \mathbb{E}_{s^t}\Big(\lambda U_{s^{t+1}} + \\ & (1-\lambda)(\lambda_e p_{s^{t+1}} v^*_{s^{t+1}}) + \widetilde{p}_{s^{t+1}} W_{y,s^{t+1}})\Big) \geq 0 \end{aligned} \tag{3.33}$$

where the constraint 3.39 is a slack participation constraint for a sufficiently small $R$, so that the principal (the firm) is willing to enter the contract in the first place.

In the terminology of MM we can label

$$h_0^0(\mathbf{x}_t, \mathbf{c}_t, a_t) = f(a_t, y_t, h_t) - w_t \tag{3.34}$$

$$h_1^0(\mathbf{x}_t, \mathbf{c}_t, a_t) = f(a_t, y_t, h_t) - w_t - R \tag{3.35}$$

$$h_0^1(\mathbf{x}_t, \mathbf{c}_t, a_t) = W_{y,t} \tag{3.36}$$

$$h_1^1(\mathbf{x}_t, \mathbf{c}_t, a_t) = W_{y,t} - u(w_t) + \beta \mathbb{E}_t(\lambda U_{t+1} + (1-\lambda)\lambda_e p_{t+1} v^*_{y,t+1}) \tag{3.37}$$

and define the Pareto problem ($\mathbf{PP}_\mu$)

[53] In the terminology of MM, we treat constraints coming from **Equation 3.28** as a set of one period ahead forward looking constraint, which makes the analysis of our case akin to their case where one have $j=1$ forward looking constraints, and $N_1 = 0$. The difference with their problems, however, is that our problem features finite time, and thus each one period ahead forward looking constraint technically applies to a *different* function $j_t$ (indexed by $t$).

$$\mathbf{PP}_\mu : V_{\mu,t_0}(\mathbf{x}_t, a_t) = \underset{\{w_{s^t}, W_{y,s^t}\}}{sup} \mathbb{E}_{t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_0+i} \mu^0 \Big( f(a_t, y_t, h_t) - w_t \Big) + \mu^1 W_{y,t_0} \tag{3.38}$$

$$s.t.\ [j=0;\ \gamma^0]:\ \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_o+i} \Big( f(a_t, y_t, h_t) - w_t \Big) - R \geq 0 \tag{3.39}$$

$$[j=1, s^t;\ \gamma^1_{s^t}]:\ W_{y,s^t} - u(w_{s^t}) - \beta \mathbb{E}_{s^t} \Big( \lambda U_{s^{t+1}} + (1-\lambda)(\lambda_e p_{s^{t+1}} v^*_{s^{t+1}}) + \widetilde{p}_{s^{t+1}} W_{y,s^{t+1}}) \Big) \geq 0 \tag{3.40}$$

Still following the notation from Marcet and Marimon 2019, we can define the Saddle Point Problem ($\mathbf{SPP}_\mu$) as:

$$\begin{aligned} \mathbf{SPP}_\mu : SV_{\mu,t_0}(\mathbf{x}_{t_0}, a_{t_0}) = & \underset{\{\gamma \in \mathbb{R}^l_+\}}{inf} \underset{\{w_{s^t}, W_{y,s^{t_0}}\}}{sup} \mu^0 \Big( f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} \Big) + \mu^1 W_{y,t_0} + \\ & + \beta \mathbb{E}_t \left( \phi(\mu, \gamma) \sum_{i=0}^{T-t_0} \left[ \beta^{t_0+i} \prod_{i=0}^{T-t_0-1} \widetilde{p}_{t_0+1+i} \Big( f(a_{t_0+i}, y_{t_0+i}, h_{t_0+i}) - w_{t_0+i} \Big) + W_{y,t_0+i} \right] \right) + \\ & + \gamma^1 \Big( u(w_{t_0} + \beta \mathbb{E}_{t_0} \Big( \lambda U_{t_0+1} + (1-\lambda) \lambda_e p_{t_0+1} v^*_{y,t_0+1} \Big) \Big) + \\ & + \gamma^0 \Big( f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} - R \Big) \end{aligned} \tag{3.41}$$

The problem can be restated as a saddle-point problem over a Lagrangian equation

$$\begin{aligned} \underset{\gamma_t}{inf} \underset{\{w_{s^t}, W_{y,s^t}\}}{sup} & \mu^0 \Big( f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} \Big) + \mu^1 W_{y,t_0} + \\ & \gamma^0 \Big( (f(a_{t_0}, y_{t_0}, h_{t_0}) - w_{t_0} \Big) - R) + \\ & \gamma^1_{t_0} \Big( -W_{y,t_0} + u(w_{t_0}) + \beta \mathbb{E}_{t_0} (\lambda U_{t_0+1} + (1-\lambda)(\lambda_e p_{t_0+1} v^*_{t_0+1} + \widetilde{p}_{t_0+1} W_{y,t_0+1}) \Big) + \\ & \beta \mathbb{E}_{t_0} \Bigg[ (\mu_0 + \gamma_0) \sum_{t=t_0+1}^{T} \beta^{t-t_0-1} \prod_{i=0}^{T-t_0-1} \widetilde{p}_{t_0+1+i} \Big( f(a_t, y_t, h_t) - w_t \Big) + \\ & \sum_{t=t_0+1}^{T} \mathbb{E}_t \beta^{t-t_0-1} \prod_{i=0}^{t-t_0-1} \widetilde{p}_{t_0+1+i} \gamma^1_t \Big( -W_{y,t} + u(w_t) + \\ & \beta(\lambda U_{t+1} + (1-\lambda)(\lambda_e p_{t+1} v^*_{t+1} + \widetilde{p}_{t+1} W_{y,t+1}) \Big) \Bigg] \end{aligned} \tag{3.42}$$

which, thanks to some algebra and the law of iterated expectations becomes

$$\underset{\gamma_t}{inf} \underset{\{w_{s^t}, W_{y,s^t}\}}{sup} -\gamma^0 R + \mathbb{E}_{t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} \prod_{i=0}^{t-t_0} \widetilde{p}_{t_0+i} \Bigg[ \Big(\mu_t^0 + \gamma_t^0\Big)\Big(f(a_t, y_t, h_t) - w_t\Big) + \mu_t^1 W_{y,t} -$$
$$\gamma_t^1 \Big( W_{y,t} - u(w_t) - \beta(\lambda U_{t+1} - (1-\lambda)\lambda_e p_{t+1} v_{t+1}^*\Big) \Bigg] \tag{3.43}$$

where $\mu_t^0 = \mu^0 = 1$, $\gamma_t^0 = \gamma_0 = 0$, $\mu_t^1 = \gamma_{t-1}^1$ for some starting $\gamma_{t_0-1}^1$.
The problem can now be written in recursive form. Define

$$\mathcal{P}_t(h_t, \tau_t, y_t, a_t, \gamma_t) = \underset{W_{y,t}}{sup}\, J_t(h_t, \tau_t, y_t, W_{y,t}, a_t) + \mu_t^1 W_{y,t} \tag{3.44}$$

Given **Equation 3.43** the SPFE of the problem can be written as

$$\mathcal{P}_t(h_t, \tau_t, y_t, a_t, \gamma_t) = \underset{\gamma_t}{inf}\, \underset{w_t}{sup} (f(a_t, y_t, h_t) - w_t) + \mu_t^1 W_{y,t} - \gamma_t (W_{y,t} - u(w_t)) +$$
$$\beta \mathbb{E}_t (\lambda U_{t+1} + (1-\lambda)\lambda_e p_{t+1} v_{t+1}^*) + \beta \mathbb{E}_t \widetilde{p}_{t+1} \mathcal{P}_{t+1}(h_{t+1}, \tau_{t+1}, y_{t+1}, a_{t+1}, \gamma_{t+1}) \tag{3.45}$$

One can easily verify that the solution of this equation is the same we found in the maximization of **Equation 3.10** in the main text. Take the first order conditions and compute the envelope condition:

$$[FOC\ w_t]: \ -1 + \gamma_t u'(w_t) = 0 \tag{3.46}$$
$$[ENV\ W_{y,t}]: \ \frac{\partial \mathcal{P}_t}{\partial W_{y,t}} = \mu_t^1 - \gamma_t \tag{3.47}$$
$$[FOC\ W_{y,t+1}]: \ -\widetilde{p}_{t+1} W_{y,t+1} \gamma_t + \frac{\partial \widetilde{p}_{t+1}}{\partial W_{y,t+1}} \mathcal{P}_{t+1} + \widetilde{p}_{t+1} \frac{\partial \mathcal{P}_{t+1}}{\partial W_{y,t+1}} = 0 \tag{3.48}$$

where **Equation**3.48 is obtained by adding and subtracting from **Equation 3.45** $\beta \gamma_t \widetilde{p}_{t+1} W_{y,t+1}$. The reader should also keep in mind that the condition in **Equation 3.48** is actually state contingent and applied to *all* future states next period, with a different set of co-states $\gamma_{s^{t+1}}$ for each realization of $a_{t+1}$.

Some rearranging of the **Equation 3.48** leads to the following result

$$\frac{\partial \log \widetilde{p}_{t+1}}{\partial W_{y,t+1}} \Big( \mathcal{P}_{t+1} - \gamma_t W_{y,t+1} \Big) = \gamma_{t+1} - \mu_{t+1}^1 \tag{3.49}$$

which, given the law of motion of the co-states and the definition in **Equation 3.44** can be re-written as:

$$\frac{\partial \log \widetilde{p}_{t+1}}{\partial W_{y,t+1}} J_{t+1} = \frac{1}{u'(w_{t+1})} - \frac{1}{u'(w_t)} \tag{3.50}$$

which is exactly **Equation 3.16**, namely the Euler equation that governs the behavior of wage setting and disciplines the provision of insurance within the contract.

## Existence of a Block Recursive Equilibrium

In order to show that a Block Recursive Equilibrium (BRE) exists in our model we need to show that the equilibrium contracts, the workers' and the entrepreneurs value and policy functions do not depend on the distribution of employed and unemployed workers.

Most of the results are tightly linked to our search protocol, directed versus random search, and our contracting structure whereby workers have finite lives and therefore contracts end in finite time. The intuition for why directed search is paramount for the existence of a BRE is linked to the fact that with directed search, workers that are matched with a particular job accept that job with certainty as they are actively looking for it in the labor market. This certainty of acceptance makes the probability of filling a vacancy, and consequently the return of opening it in a particular submarket, independent from the type of worker a firm meets. This implies that the only element of the aggregate state that matters for a firm when making an hiring decision is the state of aggregate productivity but not the distribution of worker types (e.g. employed vs unemployed).

**Proposition 11** *A block recursive equilibrium as defined in Definition 3.4.3 exists.*

We follow the approach in Menzio, Telyukova, and Visschers 2016; Herkenhoff, Phillips, and Cohen-Cole 2019 and prove the existence of a BRE using backward induction.

Consider the lifetime values of an unemployed and an employed worker before the production stage in the last period of households lives with $\tau = T$ :

$$U(h, T; \Omega) = u(b(h, T)) \tag{3.51}$$

$$V(h, T, W; \Omega) = u(w(a)), \tag{3.52}$$

their values trivially do not depend on the distribution of types as both valuations are 0 from $T+1$ onward. Hence, $U(h, T; \Omega) = U(h, T; a)$ and $V(h, T, W; \Omega) = V(h, T, W; a)$.

The optimal contract for agents aged $\tau = T$, instead, solves the following problem

$$J_t(h, T, y, W; \Omega) = \sup_w [f(y, h; a) - w] \quad s.t.\ W = u(w),$$

that clearly does not depend on the distribution of worker types due to the directed search protocol and where the aggregate state only affects the promised utility and the optimal wage through realization of the aggregate productivity processes. Therefore, $J_t(h,T,y,W;\Omega) = J_t(h,T,y,W;a)$.

This also implies that the equilibrium market tightness

$$\theta(h,T,W;\Omega) = q^{-1}\left(\frac{c(y)}{J_t(h,T,y,W;a)}\right)$$

is independent from the distribution of worker types and it is only affected by realization of aggregate productivity, so $\theta(h,T,W;a)$.

This in turn implies that the search problem workers face at the beginning of the last period of their lives depends on the aggregate state only through aggregate productivity $a$:

$$R(h,T,V;a) = \sup_{\{v_{y,\Omega}\}}\Big[p(\theta(h,T,v_{y,\Omega};a))\big[v_{y,\Omega} - V\big]\Big],$$

does not depend on the distribution of worker types.

Stepping back at $\tau = T-1$, the value functions for the unemployed and the employed agents are solutions to the following dynamic programs

$$\sup_{\{v_{y,\Omega'}\}} u(b(h,T-1)) + \beta\mathbb{E}_\Omega\Big(U_{t+1}(h,T;a') + p(\theta(h,T,v_{y,\Omega'};a'))\big[v_{y,\Omega'} - U_{t+1}(h,T;a')\big]\Big)$$

$$u(w) + \beta\mathbb{E}_\Omega\left(\begin{array}{c}\lambda U_{t+1}(g(h,y),T;a') + \beta(1-\lambda)W_{\Omega'} + \\ +\beta(1-\lambda)\lambda_e \max(0, R(g(h,y),T,W_{\Omega'});a')]\big]\end{array}\right),$$

where both do not depend on the distribution of worker types.

The optimal contract at this step is a solution to

$$J_t(h,T-1,y,V;a) = \sup_{w_i,\{W_{i,\Omega'}\}} \sum_{i=1,2} \pi_i\Big(f(y,h;a) - w_i$$

$$+\mathbb{E}_\Omega\left[\widetilde{p}_{t+1}(h',T,W_{i,\Omega'};a')(J_{t+1}(h',T,y,W_{i,\Omega'};a')\right]\Big)$$

$$s.t.\ V = \sum_{i=1,2} \pi_i\left(u(w_i) + \mathbb{E}_\Omega \widetilde{r}_{t+1}(h',T,W_{i,\Omega'};a')\right),\ h' = g(h,y)$$

$$\mathbb{E}_\Omega \sum_{i=1,2} \pi_i\left(\mathbb{E}_\Omega J_{t+1}(h',T,y,W_{i,\Omega'};a')\right) \geq 0 \text{ and } t \leq T$$

which does not depend on types distribution.

Therefore, also the equilibrium tightness and the search gain at $T-1$ are independent from types' distributions, as

$$\theta(h,T-1,W;a) = q^{-1}\left(\frac{c(y)}{J_t(h,T-1,y,W;a)}\right)$$

$$R(h,T-1,V;a) = \sup_{\{W_{y,\Omega}\}}\Big[p(\theta(h,T-1,v_{y,\Omega};a))\big[v_{y,\Omega} - V\big]\Big].$$

Stepping back from $\tau = T-1, \ldots, 1$ and repeating the arguments above completes the proof.

# LUCA MAZZONE

born: 31 / 08 / 1987

## PROFESSIONAL WORK EXPERIENCE

**International Monetary Fund** *2020 – present*
Economist - EP Program

**University of Pennsylvania** *2018 – 2020*
Visiting Scholar, Economics Department

**Inter American Development Bank** *2012 – 2013*
Research Fellow

## EDUCATION

**University of Zurich and Swiss Finance Institute** *2014 – 2020*
PhD Program in Finance

**Universitat Pompeu Fabra and Barcelona Graduate School of Economics** *2013 – 2014*
Master in Economics and Finance

**Università Bocconi** *2009 – 2012*
Master of Science in Economics and Social Sciences

**Università Bocconi** *2006 – 2009*
Bachelor of Science in Economics and Social Sciences

## HONORS AND DISTINCTIONS

Swiss National Science Foundation - Doc-Mobility Scholarship (2019), Leading House for the Latin American Region - Mobility Grant (2019), Swiss Finance Institute Fellowship (2014), Fundacion Ramon Areces Scholarship (2013), "Bonaldo Stringher" Scholarship Honorable Mention, Bank of Italy (2012)

## INVITED TALKS AND PRESENTATIONS

2021: WEAI Conference*, IMF Research Department*, INPS Seminar*, Society for Economic Dynamics (Minneapolis) 2020: European Economic Association Congress*, 2nd Bank of Italy Human Capital Workshop*, the Young Economist Symposium*, University of Copenhagen Economics Department, ICEF at the Higher School of Economics (Moscow), New Economic School (Moscow)
2019: Swiss Finance Institute Workshop (Lausanne), Swiss Finance Institute Research Days (Gerzensee), University of Pennsylvania Economics Department, Wharton Finance Department, University of Zurich
2018: PASC 2018 (Basel), Society for Computational Economics Conference 2018 (Milan), Sparse Grids and Applications 2018 (Münich)

## REFEREEING ACTIVITY

American Economic Journal: Macroeconomics, Small Business Economics Journal