# Identification and estimation of heterogeneous agent models: A likelihood approach*

Juan Carlos Parra-Alvarez[†1,3], Olaf Posch[2,3], and Mu-Chun Wang[2]

[1] *Aarhus University*
[2] *Hamburg University*
[3] *CREATES*

July 29, 2015

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE

**Abstract**

In this paper, we study the statistical properties of heterogeneous agent models with incomplete markets. Using a continuous-time version of the Bewley-Hugget-Aiyagari model we obtain the equilibrium density function of wealth and show how can it be used for likelihood inference. We investigate the identifiability of the model parameters both at the population and at the sample level when the only information available to the econometrician is a large cross-section of individual wealth. We examine the finite sample properties of the maximum likelihood estimator and study the consequences for parameter identification of combining calibration and estimation. We find that the cross-sectional data on wealth alone identifies all the structural parameters of the model. We also find that small samples generate substantial upward biases in the capital share of production and the capital depreciation rate.

***Keywords***: Heterogeneous agent models, Continuous-time, Fokker-Planck equations, Identification, Maximum likelihood.

***JEL classification***: C10, C13, C63, E21, E24.

*We would like to thank Klaus Wälde, George-Marios Angeletos, Kjetil Storesletten and participants at the 10th NNH-UiO Workshop on Economic Dynamics (Oslo, 2015) for their helpful comments and discussions. The first two authors acknowledge financial support from the Center for Research in Econometric Analysis of Time Series (DNRF78), CREATES, funded by The Danish National Research Foundation.

†Corresponding author: Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark. Email address: `jparra@econ.au.dk`

# 1 Introduction

Heterogeneous agent models have become an extensively used tool in macroeconomics for the study and evaluation of the welfare implications and desirability of business cycle stabilization policies. They have also been used to address questions related to social security reforms, the precautionary savings behavior of agents, employment mobility and wealth inequality. A comprehensive review on the developments made in the field during the last two decades can be found in Ríos-Rull (1995, 2001) and Heathcote et al. (2009). These models extend the standard representative agent framework, introduced in Kydland and Prescott (1982), by bringing forward different types of heterogeneity across households and firms that has allowed for a change of focus in macroeconomics from the analysis of average values to the study of the entire distribution of economic aggregates.

Currently, the main workhorse in the heterogeneous agent literature is based on the contributions of Bewley (Undated), Huggett (1993) and Aiyagari (1994). Their theories are motivated by the empirical observation that individual earnings, savings, wealth and labor exhibit much larger fluctuations over time than per-capita averages, and accordingly significant individual mobility is hidden within the cross-sectional distributions. These ideas have been formalized with the use of general equilibrium, dynamic and stochastic models of a large number of rational consumers that are subject to idiosyncratic income fluctuations against which they cannot fully insure due to market incompleteness. The individual choices of agents determine the aggregate amount of capital stock and effective labor supplied for production which combined with their optimal demands determine the general equilibrium prices that clear the markets in the whole economy.

To date, calibration is the standard methodology used to examine the quantitative properties of these models. This procedure fixes the value of the model parameters to those encountered in external sources, e.g. from studies that rely on micro data or cross-sectional observations on individual allocations, or from long-run averages of macroeconomic aggregates. In general, calibrated models, as opposed to statistically estimated models, can not make statements regarding the uncertainty surrounding the values used, their statistical significance and how well the models fit the data.

One possible explanation of why these models have not yet been statistically estimated is that their solution imposes a computational burden that makes any econometric procedure infeasible. It is a well known fact that the numerical approximation of the density function of the state variables of the model increases considerably the computing time of the model's solution. However, recent advances in continuous-time heterogeneous agent modeling have proven successful in reducing this computational complexities making possible the implementation of standard econometric methods to extract information from observed data.

Despite calibration being very illustrative for the study of any model's dynamics, the use of econometric methods provide some important advantages by allowing: (i) to impose on the data the restrictions arising from the economic theory associated with a particular model; (ii) to assess the uncertainty surrounding the parameter values which ultimately provides a framework for hypothesis testing, (iii) the use of standard tools of models selection and evaluation. However, little has been done in this regard within the heterogeneous agents framework with the exception of the recent work of Kirkby (2014) for the type of household heterogeneity considered in this chapter, and the structural estimation of search and matching models applied to labor economics carried out in Postel-Vinay and Robin (2002), Flinn (2006), Cahuc et al. (2006) and Launov and Wälde (2013). A more exhaustive list of examples can be found in the recent book by Wolpin (2013).

The first contribution of this paper is to take a step in this direction by introducing a simple framework to estimate the structural parameters of heterogeneous agent models by using the information content in the cross-sectional distribution of wealth. The approach is based on the economy-wide stationary probability density function which can be later used to derive the likelihood function of the model. Our framework belongs to the class of full information estimators since it uses all the restrictions imposed by the economic model through the density function.

The computation of the probability density function of wealth in heterogeneous agent models is not straightforward as it turns out to be a complicated endogenous and non-linear object that usually has to be numerically approximated. For continuous-time setups, Bayer and Wälde (2010b,a, 2011, 2013), Achdou et al. (2014a) and Gabaix et al. (2015) have recently suggested the use of Fokker-Planck equations for the derivation and analysis of endogenous distributions in macroeconomics[1]. These partial differential equations describe the entire dynamics of any probability density function in a very general manner without the need to impose particular functional forms and for which analytical solutions can be found. When combined with the standard Hamilton-Jacobi-Bellman equation that describes the optimal behavior of economic agents, they form a system of coupled partial differential equations, termed Mean-Field game following the work of Lasry and Lions (2007), that can be numerically solved with high degree of accuracy and efficiency on the entire state-space of the model using the finite difference methods described in Candler (1999) and Achdou et al. (2014b).

To illustrate our approach we make use of a continuous-time version of a Bewley-Hugget-Aiyagari model in which a large number of households face idiosyncratic and uninsurable income risk in the form of exogenous shocks to their productivity modeled through a Poisson process with two states. In our economy, households make rational and independent consumption and saving decisions taken the aggregate prices as given. When aggregated, their individual choices define the aggregate labor and capital stock supplied to a representative firm that produces the only good in the economy. The interaction between demand and supply determine the prices of the production

[1]The Fokker-Planck equations are also often called Kolmogorov Forward equations and both terms are equally used in the economic literature.

factors that clear the markets. At this stage, we do not consider the effects of aggregate uncertainty, but we hope to study them in future research.

The prototype economy is then solved for the stationary competitive equilibrium which is characterized by a time-invariant distribution of wealth and aggregate variables that do not grow over time. For a given set of parameter values this solution defines the true distribution of wealth from which we can simulate multiple samples of i.i.d. observations by randomly drawing from the approximated probability density function of wealth. These samples are later used for the construction of the likelihood function that define our econometric framework.

A well established condition for the maximum likelihood estimator to deliver consistent estimates of the model parameters and a valid asymptotic inference is that of identification (See Newey and McFadden, 1986). Roughly speaking, identification refers to the fact that the estimator's objective function must have a unique maximum at the true parameter vector and at the same time display enough curvature in all its dimensions. Lack of identification leads to misleading statistical inference, suggesting the existence of some features in the data that are actually absent. Therefore, it is important to verify the identification condition prior to the application of any estimation strategy. The recent contributions of Canova and Sala (2009), Iskrev (2010b), Komunjer and Ng (2011) and Ríos-Rull et al. (2012) point out in that direction by providing tools that help to check the identifiability of structural parameters in the context of linearized representative agent DSGE models.

The second contribution of this paper is to investigate whether it is possible, and to what extent, to (locally) identify the structural parameters of heterogeneous agent models in a likelihood-based framework when the only available information to the econometrician is a cross-sectional sample of individual wealth. Given that the mapping between the deep parameters of the model and the estimator's objective function is highly nonlinear and not available in closed form, we assess the identification power of the likelihood function in an indirect way by using some of the simulation and graphical diagnostics proposed in Canova and Sala (2009). The identification analysis is carried out on both the population and the sample objective function. We finally investigate the small sample properties of the maximum likelihood estimator and their relation to the identification condition.

The remainder of the paper is organized as follows. Section 2 introduces a prototype heterogeneous agent model, defines the stationary competitive equilibrium and describes the numerical algorithms used to: (i) compute the steady state equilibrium at the macro level, and (ii) globally approximate both the solution to the households saving-consumption problem, and the distribution of wealth. Section 3 shows how to use the solution of the economic model to derive a likelihood function that can be used to estimate the parameters of the model by exploiting the information content in the cross-sectional distribution of individual wealth. It also defines some of the main identification problems that arise in heterogeneous agent models, and shows the close link that exists between an estimator's objective function and its identification power. Section 4 investigates the identifiability of the model parameters in the population while Section 5 sets up a Monte Carlo experiment to study the finite sample properties of the maximum likelihood estimator and the implications of the resulting parameter estimates for macroeconomic analysis. Section 6 discusses the consequences, in small samples, of following a mixed calibration-estimation strategy when it is believed *a priori* that some of the model parameters are poorly identified. Finally, Section 7 concludes.

# 2 A prototypical heterogeneous agent model

For our study we consider a prototypical heterogeneous agent models *á la* Bewley-Hugget-Aiyagari set up in continuous-time following Bayer and Wälde (2011), Bayer and Wälde (2013) and Achdou et al. (2014b). In our economy there is no aggregate uncertainty and we assume that all aggregate variables are in their steady state while at the individual level, households face idiosyncratic uninsurable risk and variables change over time in a stochastic way.

## 2.1 Household's problem

Consider an economy with a continuum of unit mass of infinitively lived households that are heterogeneous in their wealth and income and where decisions, are made continuously in time. Each household consists of one agent, and we will speak of households and agents interchangeably. Household $i$ with $i \in (0,1)$ has standard preferences over streams of consumption, $c_t$, defined by

$$U_0 \equiv \mathbb{E}_0 \int_0^\infty e^{-\rho t} u(c_t)\mathrm{d}t, \quad u' > 0,\ u'' < 0, \tag{1}$$

where $\rho > 0$ is the discount rate and the utility function is given by:

$$u\left(c_t\right) = \begin{cases} \frac{c_t^{1-\gamma}}{1-\gamma} & \text{for } \gamma \neq 1 \\ \log\left(c_t\right) & \text{for } \gamma = 1 \end{cases} s$$

where $\gamma > 0$ denotes the coefficient of relative risk aversion. At time $t = 0$, the agent knows his initial wealth and income levels and chooses the optimal path of consumption $\{c_t\}_{t=0}^{\infty}$ subject to

$$\mathrm{d}a_t = (ra_t + we_t - c_t)\mathrm{d}t, \quad (a_0, e_0) \in [\underline{a}, \infty) \times \mathcal{E} \tag{2}$$

where $a_t$ denotes the household's financial wealth per unit of time and $r$ the interest rate. Wealth increases if capital income $ra_t$ plus labor income $we_t$ exceeds consumption $c_t$. At every instant of time, households face uninsurable idiosyncratic and exogenous shocks to their endowment of efficiency labor units, $e_t$, as in Castañeda et al. (2003) making their labor income stochastic; $w$ denotes the wage rate per efficiency unit which is the same across households and determined in general equilibrium together with the interest rate[2]. The fact that there are no private insurance markets for the household specific endowment shock can be explained, for example, by the existence of private

[2] Alternatively, the efficiency levels can be understood as productivity shocks following Heer and Trede (2003).

information on the employee side, like his real ability, that could give rise to adverse selection and moral hazard problems. This would prevent private firms to provide insurance against income fluctuations. However, the wealth accumulation process in Equation (2) creates a mechanism used by agents to self-insure themselves against labor market shocks and allows for consumption smoothing.

Following Huggett (1993), the endowment of efficiency units can be either high, $e_h$, or low, $e_l$. The endowment process follows a continuous-time Markov Chain with state space $\mathcal{E} = \{e_h, e_l\}$ described by:

$$\mathrm{d}e_t = -\Xi \mathrm{d}q_{1,t} + \Xi \mathrm{d}q_{2,t}, \quad \Xi \equiv e_h - e_l \quad \text{and} \quad e_0 \in \mathcal{E}. \tag{3}$$

The Poisson process $q_{1,t}$ counts the frequency with which an agent moves from a high to a low efficiency level, while the Poisson process $q_{2,t}$ counts how often it moves from a low to a high level. As an individual cannot move to a particular efficiency level while being in that same level, the arrival rates of both stochastic processes are state dependent. Let $\phi_1(e_t) \geq 0$ and $\phi_2(e_t) \geq 0$ denote the demotion and promotion rates respectively, with[3]:

$$\phi_1(e_t) = \begin{cases} \phi_{lh} & e_t = e_h \\ 0 & e_t = e_l \end{cases}$$

and

$$\phi_2(e_t) = \begin{cases} 0 & e_t = e_h \\ \phi_{hl} & e_t = e_l. \end{cases}$$

Households in this economy can not run their wealth below $\underline{a}$, where $a_n \leq \underline{a} \leq 0$, and $a_n = -{}^{w_t e_l}/_r$ defines the natural borrowing constraint implied by the non-negativity on consumption. The effects of different values for $\underline{a}$ on the model implications are studied in Aiyagari (1994).

## 2.2 Production possibilities and macroeconomic identity

Aggregate output in this economy, $Y$, is produced by firms owned by the households. They combine aggregate capital, $K$, and aggregate labor, $L$, through a constant return to scale production function:

$$F(K, L) = K^{\alpha} L^{1-\alpha}, \quad \alpha \in (0, 1).$$

in order to maximize their profits.

We further assume that the aggregate capital stock in the economy depreciates at a constant rate, $\delta \geq 0$. Since our focus is on the steady state, all the investment decisions in the economy are exclusively directed towards replacing any depreciated capital. Therefore the macroecomomic identity:

$$Y = C + \delta K \tag{4}$$

[3]The arrival rates of the Poisson processes allows us to compute the implicit transition probabilities between states. Appendix C shows how to go from arrival rates to transition probabilities.

holds at every instant of time, where $C$ denotes aggregate consumption and $\delta K$ is the aggregate investment, and where we have removed the temporal subscript $t$ from all aggregate variables to indicate that the economy is in a stationary equilibrium.

### 2.3 Equilibrium

In this economy, households face uncertainty regarding their future efficiency endowment. This makes their labor income and wealth also uncertain. Hence, the state of the economy at instant $t$ is characterized by the wealth-efficiency process $(a_t, e_t) \in [\underline{a}, \infty) \times \mathcal{E}$ defined on a probability space $(\Omega, \mathcal{F}, G)$ with associated joint density function $g(a_t, e_t, t)$.

As shown in Appendix A, the optimal behavior of each of the households in the economy can be represented by the Hamilton-Jacobi-Bellman equation (HJB):

$$\begin{aligned} \rho V(a_t, e_t) &= \max_{c_t \in \mathbb{R}^+} \big\{ u(c_t) + V_a(a_t, e_t)(r a_t + w e_t - c_t) \\ &\quad + (V(a_t, e_l) - V(a_t, e_h))\phi_1(e_t) + (V(a_t, e_h) - V(a_t, e_l))\phi_2(e_t) \big\} \end{aligned} \tag{5}$$

for given values of $r$ and $w$ and where $V(a_t, e_t)$ denotes the value function of the agent. The first-order condition for an interior solution reads:

$$u'(c_t) = V_a(a_t, e_t), \tag{6}$$

for any $t \in [0, \infty)$ making optimal consumption $c_t^\star = c(a_t, e_t)$ a function only of the states and independent of calendar time, $t$. Equation (6) implies that in the optimum, the instantaneous increase in utility due to marginally consuming more must be exactly equal to the increase in overall utility due to an additional unit of wealth.

Due to the state dependence of the arrival rates in the endowments of efficiency, only one Poisson process will be active for each of the values in $e_t$. This leads to a bivariate system of maximized HJB equations:

$$\rho V(a_t, e_l) = u(c_t^\star) + V_a(a_t, e_l)(r a_t + w e_l - c_t^\star) + (V(a_t, e_h) - V(a_t, e_l))\phi_{hl}, \tag{7}$$

$$\rho V(a_t, e_h) = u(c_t^\star) + V_a(a_t, e_h)(r a_t + w e_h - c_t^\star) + (V(a_t, e_l) - V(a_t, e_h))\phi_{lh}. \tag{8}$$

An interesting feature of our continuous-time setup as opposed to the discrete-time case, is that Equation (6) holds for all $a_t > \underline{a}$ since the borrowing constraint never binds in the interior of the state space. Therefore, the system of equations formed by (7) and (8) does not get affected by the existence of the inequality constraint $a_t \geq \underline{a}$ and instead gives rise to the following state-constraint boundary condition (See Achdou et al., 2014b):

$$u'(r\underline{a} + w e_t) > V_a(\underline{a}, e_t). \tag{9}$$

It can be shown that Equation (9) implies that $r\underline{a} + we_t - c(\underline{a}, e_t) \geq 0$ and therefore the borrowing constraint is never violated.

On the other hand, a representative firm rents capital and labor in competitive markets and therefore production factors are paid their marginal product:

$$r = \alpha K^{\alpha-1} L^{1-\alpha} - \delta \quad \text{and} \quad w = (1-\alpha) K^{\alpha} L^{-\alpha} \tag{10}$$

where the aggregate capital is obtained by aggregating the wealth held by every type of household, and similarly, aggregate labor is obtained by aggregating the efficiency labor units:

$$K = \sum_{e_t \in \{e_l, e_h\}} \int_{\underline{a}}^{\infty} a_t g\left(a_t, e_t\right) \mathrm{d}a_t, \tag{11}$$

$$L = \sum_{e_t \in \{e_l, e_h\}} \int_{\underline{a}}^{\infty} e_t g\left(a_t, e_t\right) \mathrm{d}a_t. \tag{12}$$

Equations (11) and (12) provide the link between the dynamics and randomness that occurs at the micro level and the deterministic behavior at the macro level.

We consider a stationary equilibrium where aggregate variables and prices are constant, the joint distribution of wealth and efficiency units is time-invariant and markets clear. More specifically, the distribution of wealth is constant for both the low and highly efficient workers, and the number of low and highly efficient workers is constant, too. Nonetheless, the individual agents are not characterized by constant wealth and efficiency status over time. In particular:

**Definition 2.1 (Competitive stationary equilibrium)** *A competitive stationary equilibrium is a pair of value functions $V\left(a_t, e_l\right)$ and $V\left(a_t, e_h\right)$, individual policy functions for consumption $c\left(a_t, e_l\right)$ and $c\left(a_t, e_h\right)$, a time-invariant density of the state variables $g\left(a_t, e_l\right)$ and $g\left(a_t, e_h\right)$, time-invariant prices of labor and capital $\{w, r\}$, and a vector of aggregates $\{K, L, C\}$ such that:*

1. *production factors satisfy Equations (11) and (12),*
2. *the consumption functions $c\left(a_t, e_h\right)$ and $c\left(a_t, e_l\right)$ satisfy Equations (8) and (7),*
3. *factor prices satisfy the first order condition in Equation (10),*
4. *the goods market clears, i.e., Equation (4) holds, where $C = \sum_{e_t} \int_{\underline{a}}^{\infty} c\left(a_t, e_t\right) g\left(a_t, e_t\right) \mathrm{d}a_t$,*
5. *the distribution of the state variables is stationary for all $\left(a_t, e_t\right) \in [\underline{a}, \infty) \times \mathcal{E}$ , i.e. $\frac{\partial g\left(a_t, e_t\right)}{\partial t} = 0$.*

## 2.4 Distribution of endowments and wealth

Recall that the state of the economy at instant $t$ can be described by the joint density function $g(a_t, e_t, t)$. In the stationary equilibrium this density is independent of calendar time $t$ and will be denoted by $g(a_t, e_t)$. Given its dependence on one continuous random variable and one discrete random variable, the joint density can be split into two subdensities, one for each element in $e_t$. Following Bayer and Wälde (2013), the (unconditional) density function of wealth is defined as:

$$g(a_t) = g(a_t, e_h) + g(a_t, e_l) \tag{13}$$

where $g(a_t, e_t) \equiv g(a_t \mid e_t)\, p(e_t)$ are not conditional densities but can be interpreted as the product between a conditional probability and the probability of having a given endowment of efficiency. This implies that for an individual currently in state $e_t$:

$$\begin{aligned}\int g(a_t, e_t)\, \mathrm{d}a_t &= \int g(a_t \mid e_t)\, p(e_t)\, \mathrm{d}a_t \\ &= p(e_t) \int g(a_t \mid e_t)\, \mathrm{d}a_t \\ &= p(e_t)\,. \end{aligned} \tag{14}$$

Given our two state Markov process for the endowment of efficiency units it is possible to show that its stationary distribution is given by (See Appendix C):

$$\lim_{t\to\infty} p(e_h, t) \equiv p(e_h) = \frac{\phi_{hl}}{\phi_{hl} + \phi_{lh}} \tag{15}$$

$$\lim_{t\to\infty} p(e_l, t) \equiv p(e_l) = \frac{\phi_{lh}}{\phi_{hl} + \phi_{lh}}. \tag{16}$$

Let $s(a_t, e_t) = ra_t + we_t - c(a_t, e_t)$ denote the optimal savings function for an individual with an endowment of efficiency $e_t$. As shown in Appendix B, the subdensities $g(a_t, e_t)$ in Equation (13) correspond to the solution of the following non-autonomous quasi-linear system of differential equations known as Fokker-Planck equations:

$$s(a_t, e_l) \frac{\partial}{\partial a_t} g(a_t, e_l) = -\left(r - \frac{\partial}{\partial a_t} c(a_t, e_l) + \phi_{hl}\right) g(a_t, e_l) + \phi_{lh} g(a_t, e_h)\,, \tag{17}$$

$$s(a_t, e_h) \frac{\partial}{\partial a_t} g(a_t, e_h) = -\left(r - \frac{\partial}{\partial a_t} c(a_t, e_h) + \phi_{lh}\right) g(a_t, e_h) + \phi_{hl} g(a_t, e_l) \tag{18}$$

where the derivatives with respect to $a_t$ describe the cross-sectional dimension of the density function. The system of equation (17)-(18) takes as given the optimal policy functions for consumption of individuals. This feature creates a recursive structure within the model, called a Mean-Field game by Lasry and Lions (2007), that facilitates its solution: households and firms meet at the market place and make their choices taking prices as given. Prices in turn are determined in general equilibrium and hence depend on the entire distribution of individuals in the economy. Such distribution is determined by the optimal choices of households and the stochastic properties of the exogenous shocks.

Once the subdensities are computed, it is possible to transform them into probability subdistributions according to:

$$G(a_t, e_t) = \int_{\underline{a}}^{a} g(x_t, e_t)\,\mathrm{d}x_t \tag{19}$$

where $G(a_t, e_t)$ denotes the probability that an individual with endowment of efficiency equal to $e_t$ has a wealth level of at most $a$. When $a \to \infty$, Equation (14) implies that $\lim_{a_t \to \infty} G(a_t, e_t) = p(e_t)$. Similar to Equation (13), the probability distribution of wealth is given by:

$$G(a_t) = G(a_t, e_h) + G(a_t, e_l) \tag{20}$$

which can be used to compute the Gini coefficient in the economy:

$$\mathcal{G} = \frac{1}{\mu} \int_{\underline{a}}^{\infty} G(a_t)\,(1 - G(a_t))\,\mathrm{d}a_t \tag{21}$$

with $\mu = \mathbb{E}(a_t)$.

### 2.5 Computation of the equilibrium

A closed form solution for our prototype economy is not available. Therefore, given a set of values for the structural parameters, the stationary competitive equilibrium in Definition 2.1 is numerically approximated. The algorithm we use builds on earlier work by Achdou et al. (2014b) and exploits the recursive nature of the model. It consists of two main blocks: an outer block that computes recursively the stationary equilibrium at the macro level using a relaxation algorithm on the aggregate capital stock; and an inner block that uses an implicit finite difference method to provide an approximate solution to the household's problem at the micro level that is valid on the entire state space. The inner block also uses a finite difference method to approximate the stationary subdensities that solve the system of ordinary differential equations in Equations (17)-(18) taking as given the approximated optimal policy functions for consumption[4]. A detailed description of the algorithms and their implementation can be found in Appendix D.

## 3 Structural Estimation and identification issues

While there is a broad consensus on the importance of heterogeneity in macroeconomics not just for the study of inequality and the distribution of wealth but also for the understanding of aggregates like GDP and the employment rate, there is less aggreement on how these models should

[4]The finite difference method is an efficient algorithm to numerically approximate the solution of differential equations. In our particular case, it is used to solve the partial differential equation that defines the HJB equation and the ordinary differential equation that solve each of the subdensities.

be taken to the data. To date, calibration is the standard approach used by researchers to map observations into parameter values of a structural model. Under this methodology, parameters are determined by minimizing the distance between a set of empirical moments and the same set of moments implied by the model, or by fixing the values of the parameters to those estimated in previous microeconomic studies, or to long-run averages of macroeconomic aggregates.

An alternative way to take structural models to the data is through formal econometric methods. With the exception of Kirkby (2014), little has been done in this regard due to the computational burden imposed by the solution of heterogeneous agent models. The purpose of this section to show how the solution of an economic model like that of Section 2 can be used to derive a likelihood function that can be used to estimate the parameters of the model by extracting information from the cross-sectional distribution of individual wealth.

A crucial but hardly ever verified assumption for the maximum likelihood estimator to deliver consistent estimates and valid asymptotic inference is that of identification. Therefore, this section also defines some of the main identification problems that might arise in heterogeneous agent models, and with a simple example shows the connection that exists between an estimator's objective function and its identification power.

### 3.1 Full information approach: The likelihood function

Let $\mathbf{a} = [a_1, \dots, a_N]$ be a sample of $N$ i.i.d observations on individual wealth and $\boldsymbol{\theta}$ a $\mathcal{K} \times 1$ vector of structural parameters to be estimated. In what follows, we assume that $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{\mathcal{K}}$, where $\boldsymbol{\Theta}$ is the parameter space which is assumed to be compact. The likelihood function of the data can be derived using the approximated subdensity functions in Equation (48). According to the identity in Equation (13) the (unconditional) probability density function of wealth can be computed as:

$$g\left(a_n \mid \boldsymbol{\theta}\right) = g\left(a_n, e_l \mid \boldsymbol{\theta}\right) + g\left(a_n, e_h \mid \boldsymbol{\theta}\right) \tag{22}$$

for each $n = 1, \dots, N$, where we have made explicit the dependence on the vector of parameter values, $\boldsymbol{\theta}$. This implies that the log-likelihood function of the sample is given by:

$$\mathcal{L}_N\left(\boldsymbol{\theta} \mid \mathbf{a}\right) = \sum_{n=1}^{N} \log g\left(a_n \mid \boldsymbol{\theta}\right), \tag{23}$$

whereas the maximum likelihood (ML) estimator, $\hat{\boldsymbol{\theta}}_N$ is defined as:

$$\hat{\boldsymbol{\theta}}_N = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\max}\; \mathcal{L}_N\left(\boldsymbol{\theta} \mid a_1, \dots, a_N\right). \tag{24}$$

Since the density function of wealth, and hence the log-likelihood function, summarizes all the restrictions imposed by the economy model, our maximum-likelihood estimator belongs to the class of full information estimators.

For the implementation of the ML estimation we use a numerical (constrained) nonlinear optimizer with initial value $\boldsymbol{\theta}^0$. For each iteration of the optimization routine the economic model is solved over the state-space $\mathcal{A} \times \mathcal{E}$ using the algorithms introduced in Section 2. In particular, the wealth lattice is discretized using $I \leq N$ grid points on the partially ordered set $\mathcal{A} = [\min(\mathbf{a}), \max(\mathbf{a})]$. Once the density function of wealth has been approximated, the log-likelihood function is constructed in two steps:

1. For each $a_n \in \mathbf{a}$, let $\mathbb{A}_n = \{a \in \mathcal{A} : a \leq a_n\}$. Then, the density function evaluated at sample point $a_n$ is given by $g(a_n \mid \boldsymbol{\theta}) = g(a^\star \mid \boldsymbol{\theta})$ where $a^\star = \max \mathbb{A}_n$.

2. Once $g(a_n \mid \boldsymbol{\theta})$ has been computed for all $a_n \in \mathbf{a}$, the log-likelihood function is built using Equation (23).

Intuitively, step 1 approximates the sample density function of wealth with a histogram by assigning to each observation in the sample a density value equal to that of the grid point in the approximation space for wealth, $\mathcal{A}$, that is closest to its left.

### 3.2 The identification problem

In general, a vector of parameters, $\boldsymbol{\theta}$, is said to be identified if the objective function, $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{a})$, has a unique maximum at its true value, $\boldsymbol{\theta}_0$. Formally, the identification condition establishes that if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ then $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{a}) \neq \mathcal{L}(\boldsymbol{\theta}_0 \mid \mathbf{a})$, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Checking this condition in practice is difficult because the mapping from the structural parameters of the model to the objective function is highly nonlinear and usually not known in closed form. Therefore, the standard rank and order conditions used in linear models can not be applied.

Recently, Canova and Sala (2009) documented the existence of identification issues in the context of linearized DSGE models. These identification problems, which could also emerge in heterogeneous agent models, are related to the shape and curvature of the objective function and have been classified by the authors as follows[5]:

1. *Observational equivalence*: if two vector of parameters, $\hat{\boldsymbol{\theta}}_1 \in \boldsymbol{\Theta}^\dagger \subset \boldsymbol{\Theta}$ and $\hat{\boldsymbol{\theta}}_2 \in \boldsymbol{\Theta}^\dagger \subset \boldsymbol{\Theta}$ deliver the same maximized objective function, they are said to be observational equivalent. In the maximum likelihood case, this occurs whenever $\mathcal{L}\left(\hat{\boldsymbol{\theta}}_1 \mid \mathbf{a}\right) = \mathcal{L}\left(\hat{\boldsymbol{\theta}}_2 \mid \mathbf{a}\right)$ and for any other $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\mathcal{L}\left(\hat{\boldsymbol{\theta}}_j \mid \mathbf{a}\right) > \mathcal{L}\Big(\boldsymbol{\theta} \mid \mathbf{a}\Big)$, for $j = 1, 2$.

2. *Partial-identification*: if for some partition $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2] \in \boldsymbol{\Theta}_1 \times \boldsymbol{\Theta}_2 = \boldsymbol{\Theta}^\dagger \subset \boldsymbol{\Theta}$, $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{a}) = \mathcal{L}(f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \mid \mathbf{a})$, for all $\mathbf{a}$ and for all $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}_1$ and $\boldsymbol{\theta}_2 \in \boldsymbol{\Theta}_2$ where $f$ is a continuous function, then $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are said to be partially identified.

[5] A fifth type of identification problem known as *under-identification* emerges in models where the solution is only locally valid, i.e. approximated using perturbation or linear quadratic methods. In that case, some of the model parameters disappear from the estimator's objective function because they are not present in the rational expectation solution of the model.

3. *Weak identification*: a subset of parameters in $\boldsymbol{\theta}$ is said to be weakly identified if the objective function, even though has a unique maximum, does not show enough curvature. In other words if there exists a $\hat{\boldsymbol{\theta}}$ such that $\mathcal{L}\left(\hat{\boldsymbol{\theta}} \mid \mathbf{a}\right) > \mathcal{L}\left(\boldsymbol{\theta} \mid \mathbf{a}\right)$ for all $\mathbf{a}$ and for all $\boldsymbol{\theta} \neq \hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}^{\dagger} \subset \boldsymbol{\Theta}$. However, $\left\|\mathcal{L}\left(\hat{\theta}_i \mid \mathbf{a}\right) - \mathcal{L}\left(\theta_i \mid \mathbf{a}\right)\right\| < \epsilon$ for some $\theta_i \neq \hat{\theta}_i \in \boldsymbol{\Theta}^{\dagger} \subset \boldsymbol{\Theta}, i = 1, \ldots, \mathcal{K}$.

4. *Asymmetric weak identification*: a group of parameters in $\boldsymbol{\theta}$ is said to exhibit asymmetric weak identification if the objective function is asymmetric in the neighborhood of the maximum, and its curvature is insufficient only in a portion of the parameter space. In other words if there exists a $\hat{\boldsymbol{\theta}}$ such that $\mathcal{L}\left(\hat{\boldsymbol{\theta}} \mid \mathbf{a}\right) > \mathcal{L}\left(\boldsymbol{\theta} \mid \mathbf{a}\right)$ for all $\mathbf{a}$ and for all $\boldsymbol{\theta} \neq \hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}^{\dagger} \subset \boldsymbol{\Theta}$. However, $\left\|\mathcal{L}\left(\hat{\theta}_i \mid \mathbf{a}\right) - \mathcal{L}\left(\theta_i \mid \mathbf{a}\right)\right\| < \epsilon$ for some $\theta_i > \hat{\theta}_i \in \boldsymbol{\Theta}^{\dagger} \subset \boldsymbol{\Theta}$ or for some $\theta_i < \hat{\theta}_i \in \boldsymbol{\Theta}^{\dagger} \subset \boldsymbol{\Theta}, i = 1, \ldots, \mathcal{K}$.

The previous definitions refer to local properties of the objective function and are valid for any extremum estimator. The global identification properties are equivalent and can be obtained by simply letting $\boldsymbol{\Theta}^{\dagger} = \boldsymbol{\Theta}$. For the case of the maximum likelihood estimator, the identifiability of the parameters can be alternatively assessed with the Fisher information matrix:

$$\mathcal{I}_N\left(\boldsymbol{\theta}\right) := \mathbb{E}\left[\left\{\frac{\partial \mathcal{L}_N\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}'}\right\}'\left\{\frac{\partial \mathcal{L}_N\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}'}\right\}\right].$$

The information matrix measures the amount of information that a particular sample provides about the model parameters. As shown in Rothenberg (1971), the vector of parameters $\boldsymbol{\theta}$ is locally identified if and only if the rank of $\mathcal{I}_N\left(\boldsymbol{\theta}\right)$ is constant and equal to $\mathcal{K}$ around $\boldsymbol{\theta}_0$. However, to evaluate $\mathcal{I}_N\left(\boldsymbol{\theta}\right)$ we need the derivatives of the log-likelihood function with respect to $\boldsymbol{\theta}$ which are not available in closed form for the type of models considered here. Therefore, they can only be numerically approximated which is out of the scope of this paper and it is left for future research. An application of the Fisher information matrix for the study of identification issues in linearized DSGE models can be found in Iskrev (2010a).

### 3.3 Full information vs. Limited information

From the different types of identification problems described above, it is clear that the choice of objective function is critical for the ability of any quantitative method to produce identifiable parameters. The objective function gathers the restrictions that arise from any economic model and imposes them to the information available to the econometrician. A limited information estimator uses a portion of the restrictions to build the objective function, while a full information estimator uses all of the restrictions.

To motivate our full information approach and choice of objective function we first illustrate the weaknesses and challenges of the limited information framework by investigating the identification condition in the generalized method of moments (GMM) methodology studied recently in Kirkby

**Table 1. Population parameters, $\boldsymbol{\theta}_0$.**

In the model, time is measured in years and parameter values should be interpreted accordingly. The endowment of efficiency units is given by:

$$\mathrm{d}e_t = -\Xi \mathrm{d}q_{1,t} + \Xi \mathrm{d}q_{2,t}, \quad \Xi \equiv e_h - e_l \quad \text{and} \quad e_0 \in \{e_h, e_l\},$$

where $q_{1,t}$ and $q_{2,t}$ are Poisson process with intensity rates $\phi_{lh}$ and $\phi_{hl}$ respectively. The representative household has standard preferences defined by $U_t = \mathbb{E}_t \left[\int_t^\infty e^{\rho t} u(c_s)\, \mathrm{d}s\right]$ where $u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}$. The macroeconomic identity in the stationary competitive equilibrium is given by:

$$Y = C - \delta K, \quad \text{where} \quad Y = K^\alpha L^{1-\alpha}.$$

| | |
|---|---|
| Relative risk aversion, $\gamma$ | 2.0000 |
| Rate of time preference, $\rho$ | 0.0410 |
| Capital share in production, $\alpha$ | 0.3600 |
| Depreciation rate of capital, $\delta$ | 0.0800 |
| Endowment of high efficiency, $e_h$ | 1.0000 |
| Endowment of low efficiency, $e_l$ | 0.1000 |
| Demotion rate, $\phi_{lh}$ | 0.6697 |
| Promotion rate , $\phi_{hl}$ | 4.4644 |

(2014). The objective function of the GMM measures the (weighted) distance between two sets of distributional moments and constitutes the method behind most of the calibration procedures used in the heterogeneous agent literature (See Castañeda et al. (2003) and Díaz-Giménez et al. (2014)).

For simplicity, suppose we are interested in the power of GMM to identify only one parameter $\theta \in \boldsymbol{\theta}$ at a time while keeping the remaining parameters at their population value. For the model in Section 2, let $\boldsymbol{\theta} = \{\gamma, \rho, \alpha, \delta, e_l, e_h, \phi_{hl}, \phi_{lh}\}$. In order to avoid over-identification issues we consequently choose only one moment condition for the construction of the objective function. In particular, let us choose the Gini coefficient in Equation (21) as our target moment. We then compute the distance between the Gini coefficient in the population, $\mathcal{G}(\theta_0)$, and the Gini coefficient obtained varying one parameter at a time in a reasonable neighborhood of the true value, $\mathcal{G}(\theta)$.

The population values for the structural parameters, $\boldsymbol{\theta}_0$, are given by the calibration in Table 1 which combines that in Huggett (1993) for the endowments of efficiency units and transition probabilities, with that in Aiyagari (1994) for the remaining parameters. In the model, time is measured in years and parameter values should be interpreted accordingly[6]. The transition rates for the Poisson processes are obtained from Equations (15)-(16).

For each of the elements in $\boldsymbol{\theta}$, Figure 1 plots the percentage deviation of the GMM objective function from the Gini coefficient in the population as a function of the parameter space. The results suggest that the objective function attains a unique minimum in the case of parameters

[6] Huggett (1993) reports transition probabilities of $\mathbb{P}(e_h \mid e_l) = 0.5$ and $\mathbb{P}(e_h \mid e_h) = 0.925$ in an model economy with six periods per year.

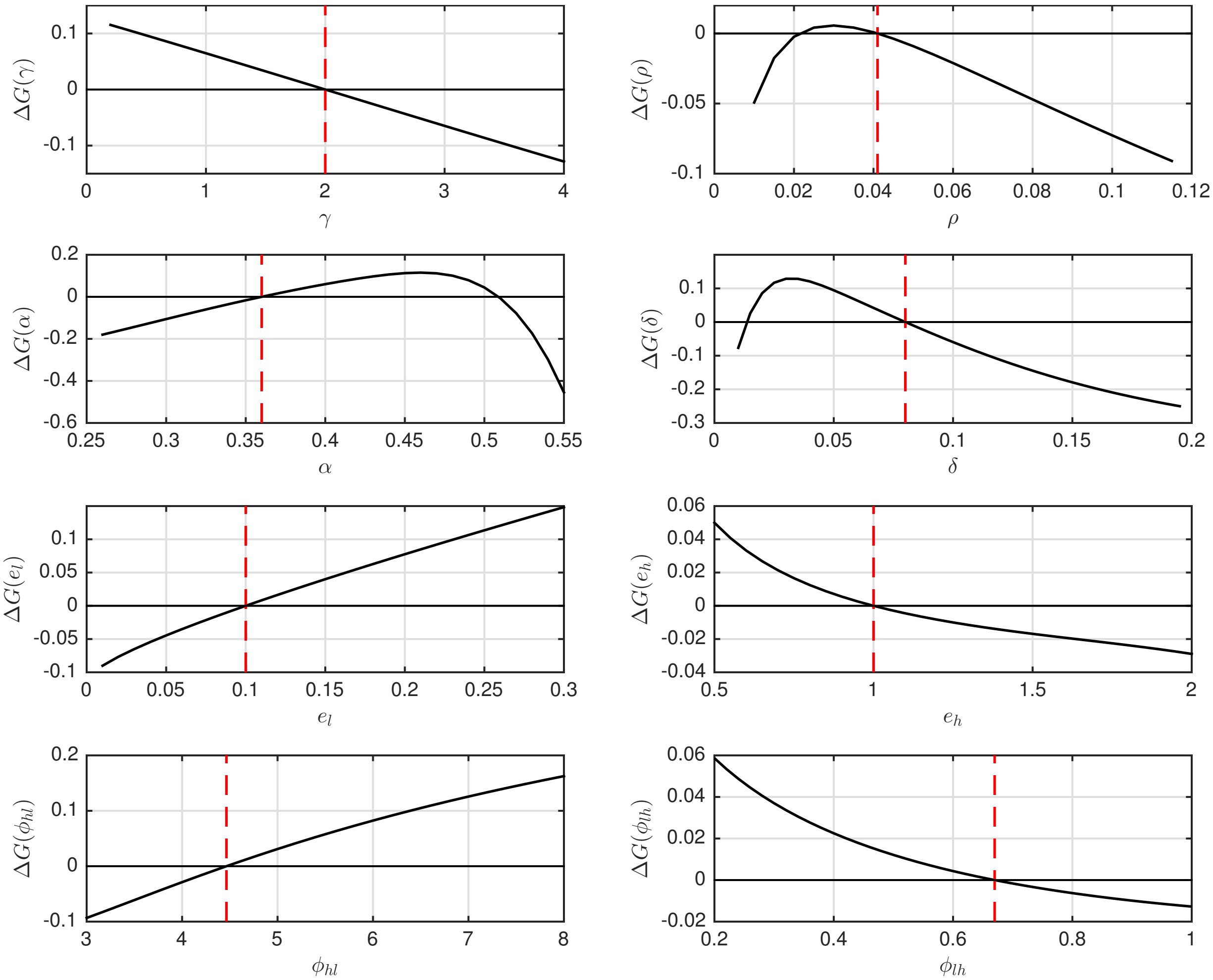


**Figure 1. Relative distance between $\mathcal{G}(\theta)$ and $\mathcal{G}(\theta_0)$.** The graph shows the percentage deviation of the GMM distance criterion as a function of the parameter space. The population values for the structural parameters, $\boldsymbol{\theta}_0$, are given in Table 1 and are represented by the dotted vertical line.

related to the exogenous process in the economy, $\{e_l, e_h, \phi_{hl}, \phi_{lh}\}$ and the coefficient of relative risk aversion, $\gamma$. The minimum corresponds indeed to the true parameter value as indicated by the dotted vertical line. However, for the discount rate, $\rho$, the depreciation rate of capital, $\delta$, and the share of capital in the production function, $\alpha$, the GMM distance function suffers from observational equivalence, i.e. exhibits multiple local minima. Therefore, the Gini coefficient does not contain enough information about certain parameters of the model and if used for estimation it can lead to ill-behaved estimates and invalid asymptotic inference.

In general, minimum distance estimators, like the GMM, exhibit identification deficiencies relative to the maximum likelihood estimator as it will be clear later. While the former uses only partial or limited information through the set of moments in the objective function, the latter uses the whole distribution of the variables in the model to identify the parameters. Canova and Sala (2009) have shown the pervasive consequences of using a limited information approach to conduct inference when identification problems are present in linearized DSGE models. Therefore, we will focus on the statistical properties of a full information likelihood approach. In particular, we will investigate identification issues by exploiting the direct link that exists between the density function of the state variables in the model and the likelihood of the data as shown in Subsection 3.1. The feasibility of our procedure is dictated, in general, by the use of continuous-time methods, and in particular by the Fokker-Planck equations that allow us to approximate the evolution of the distribution of wealth.

# 4 Population identification analysis

We begin our analysis of identification by studying the behavior of the density function of wealth when the sampling process is known. A basic prerequisite for carrying out valid inference about the parameter vector $\boldsymbol{\theta}$, is that distinct values of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ imply distinct density functions. Therefore, this section investigates whether it is possible (or not) to distinguish the model's density function of wealth approximated using the true parameter values, $g(a \mid \boldsymbol{\theta}_0)$, from the density function approximated using a range of parameter values that differ from those in the population, $g(a \mid \boldsymbol{\theta})$, with $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. We refer to this approach as *population identification analysis* since it is independent of the data, and its conclusions remain valid even with samples of infinite size. Formally, we say that the parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is identified if $g(a \mid \boldsymbol{\theta}) = g(a \mid \boldsymbol{\theta}_0)$. As discussed in Section 2, there is no analytical expression for the density function and therefore it must be approximated it using $I$ grid points on the wealth lattice. In what follows, we set $\mathcal{A} = [0, 40]$ and $I = 1000$. In order to make the identifiability condition operational we use the $L_1$ norm to measure the distance between two densities:

$$d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \equiv d(g(a \mid \boldsymbol{\theta}), g(a \mid \boldsymbol{\theta}_0)) = \sum_{i=1}^{I} |g(a_i \mid \boldsymbol{\theta}) - g(a_i \mid \boldsymbol{\theta}_0)|. \tag{25}$$

From a statistical point of view, the probability density function should contain all the relevant information about the value of the parameter vector $\boldsymbol{\theta}$. Therefore, if the distance function in Equation (25) features identification problems, we cannot hope to achieve identification of the model parameters using the likelihood of the data.

Similar to the case of the GMM distance function discussed in Section 3, Figure 2 plots the shape of the distance function $d(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ in each of the elements of $\boldsymbol{\theta}$. In each case, we vary one parameter at the time within an economically reasonable range while keeping all the remaining parameters at their population value. The population value of the parameter vector is indicated by the dotted vertical line.

The figure displays two important features. First, the use of the entire density function provides more information for parameter identification relative to limited information methods. In fact, the distance function is uniquely minimized at $\boldsymbol{\theta}_0$ which rules out identification problems related to observational equivalence. Alternative metrics like the Chi-squared distance and the Earth's Moving

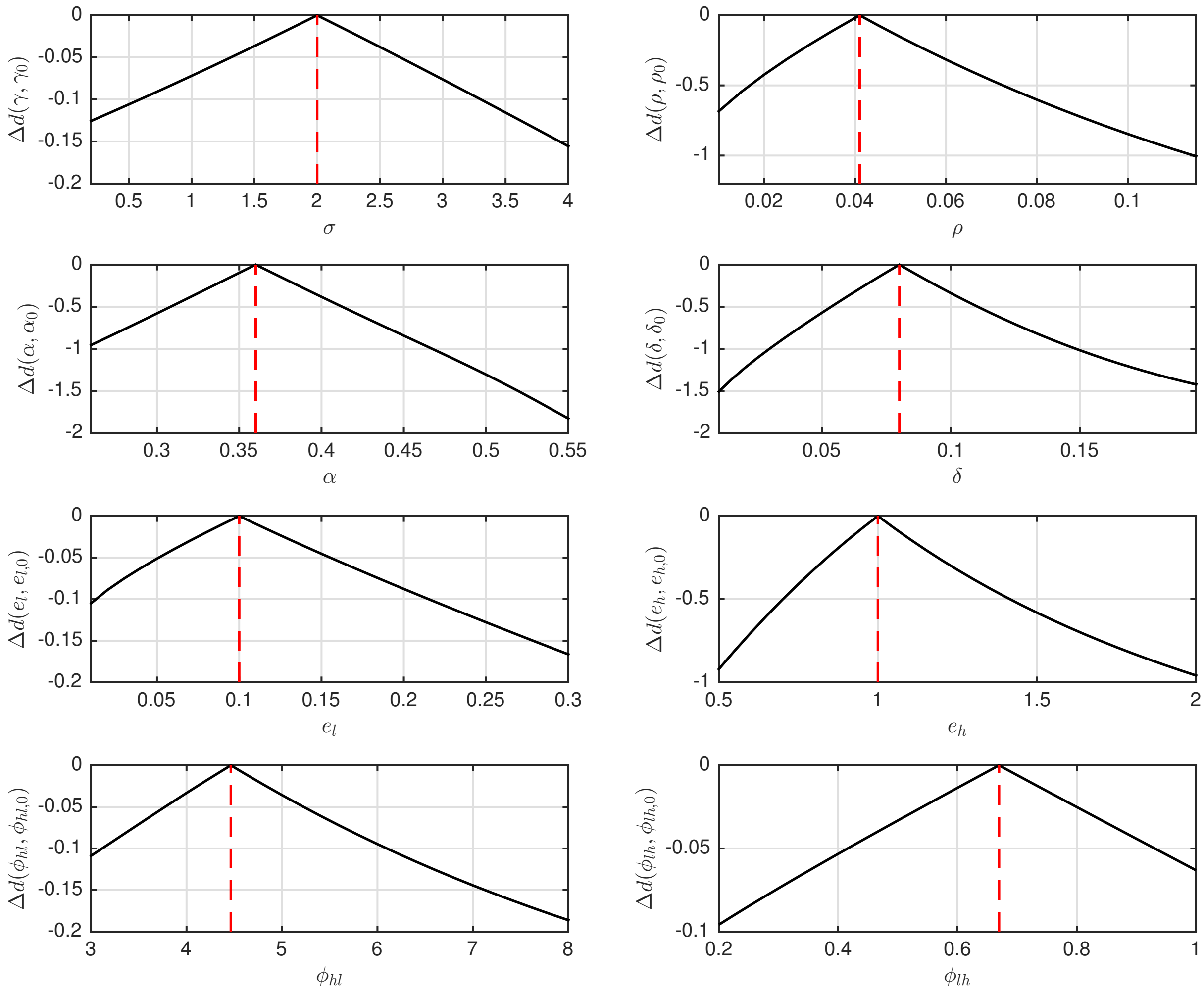


**Figure 2. Distance function** $d\left(g\left(a \mid \theta\right), g\left(a \mid \theta_0\right)\right)$**.** The graph shows the percentage deviation of the $L_1$ distance criterion as a function of the parameter space. The population values for the structural parameters, $\boldsymbol{\theta}_0$, are given in Table 1 and are represented by the dotted vertical line.

Distance function give similar results. They are not displayed here for space considerations but are available upon request. Second, although displaying an unique minimum, the distance functions feature some asymmetries in all dimensions of the parameter space except for $\gamma$.

The figure also suggests that the distance function exhibits enough curvature in the neighborhood of $\boldsymbol{\theta}_0$. To corroborate this conclusion and rule out weak identification issues we follow Canova and Sala (2009) and try to determine whether there exists a region of the parameter space with parameter vectors that produce small deviations from the population model. For each of the structural parameters in $\boldsymbol{\theta}$ we uniformly draw 1000 values from the intervals used to build Figure 2, solve the model, approximate the density function of wealth and measure its distance from the true density function using Equation (25). We then construct the distribution of the distance function and select only those parameter draws for which the distance function is within the 0.1 percentile of the distribution around zero.

**Table 2. Population intervals generating small distance functions.**

The table reports the median, minimum and maximum values of the parameters generating distance functions $d(\theta, \theta_0)$ which are in the 0.1 percentile of the distribution of distance functions around zero. In each case, the percentiles are obtained by uniformly randomizing the parameters within the ranges used in Figure 2.

| $\boldsymbol{\theta}$ | $\boldsymbol{\theta}_0$ | Min. | Mean | Median | Max. |
|---|---|---|---|---|---|
| $\gamma$ | 2.0000 | 1.9290 | 1.9899 | 1.9742 | 2.0732 |
| $\rho$ | 0.0410 | 0.0399 | 0.0408 | 0.0404 | 0.0418 |
| $\alpha$ | 0.3600 | 0.3595 | 0.3601 | 0.3601 | 0.3608 |
| $\delta$ | 0.0800 | 0.0741 | 0.0804 | 0.0806 | 0.0863 |
| $e_l$ | 0.1000 | 0.1000 | 0.1001 | 0.1000 | 0.1004 |
| $e_h$ | 1.0000 | 0.9644 | 1.0010 | 0.9989 | 1.0365 |
| $\phi_{hl}$ | 4.4644 | 4.3543 | 4.4706 | 4.4740 | 4.5745 |
| $\phi_{lh}$ | 0.6697 | 0.6596 | 0.6670 | 0.6656 | 0.6796 |

Table 2 summarizes the results of the experiment. For each of the parameters of the model, it reports its value in the population as well as the minimum, median and maximum value of the generated intervals. In general, the intervals are small and concentrated around the true value, i.e., only parameter values close to the true are compatible with small distances between densities. The composition of these interval suggest that there is no observational equivalence, neither weak identification problems.

Figure 2 and Table 2 only consider one dimension of the parameter space at the time. This prevents from detecting ridges in the objective function that might suggest partial identification problems. Therefore, Figure 3 plots the distance surfaces and their respective contours for selected combinations of parameters while keeping the remaining parameters fixed at their true value.

A thoughtless look to the figure might erroneously lead us to conclude that for some combinations of parameters, the distance function indeed exhibits ridges, e.g. $\alpha$ and $\delta$ and/or $\delta$ and $\rho$. We will be tempted then to conclude that the distribution of wealth alone carries little information about the correct combination of those parameters. However, a careful analysis of the results in Figure 3 indicates that all the parameter combinations selected are correctly identified. This can be seen by looking at the contour plots in the right column, where we have also marked the location of the true parameter values, $\boldsymbol{\theta}_0$, and the location of the parameters that deliver the maximum of the distance function, $\tilde{\boldsymbol{\theta}} = \arg\max d(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$. We can conclude that the model is well identified since $\tilde{\boldsymbol{\theta}} \approx \boldsymbol{\theta}_0$. Small discrepancies between the locations are due to the approximation errors in the solution of the heterogeneous agent model and to the design of the intervals used for the computation of the distance function.

In sum, the graphical evidence provided in this section indicates that the prototype model of Section 2 does not exhibit (local) identification problems in any of the dimensions of the parameter space at the population level.

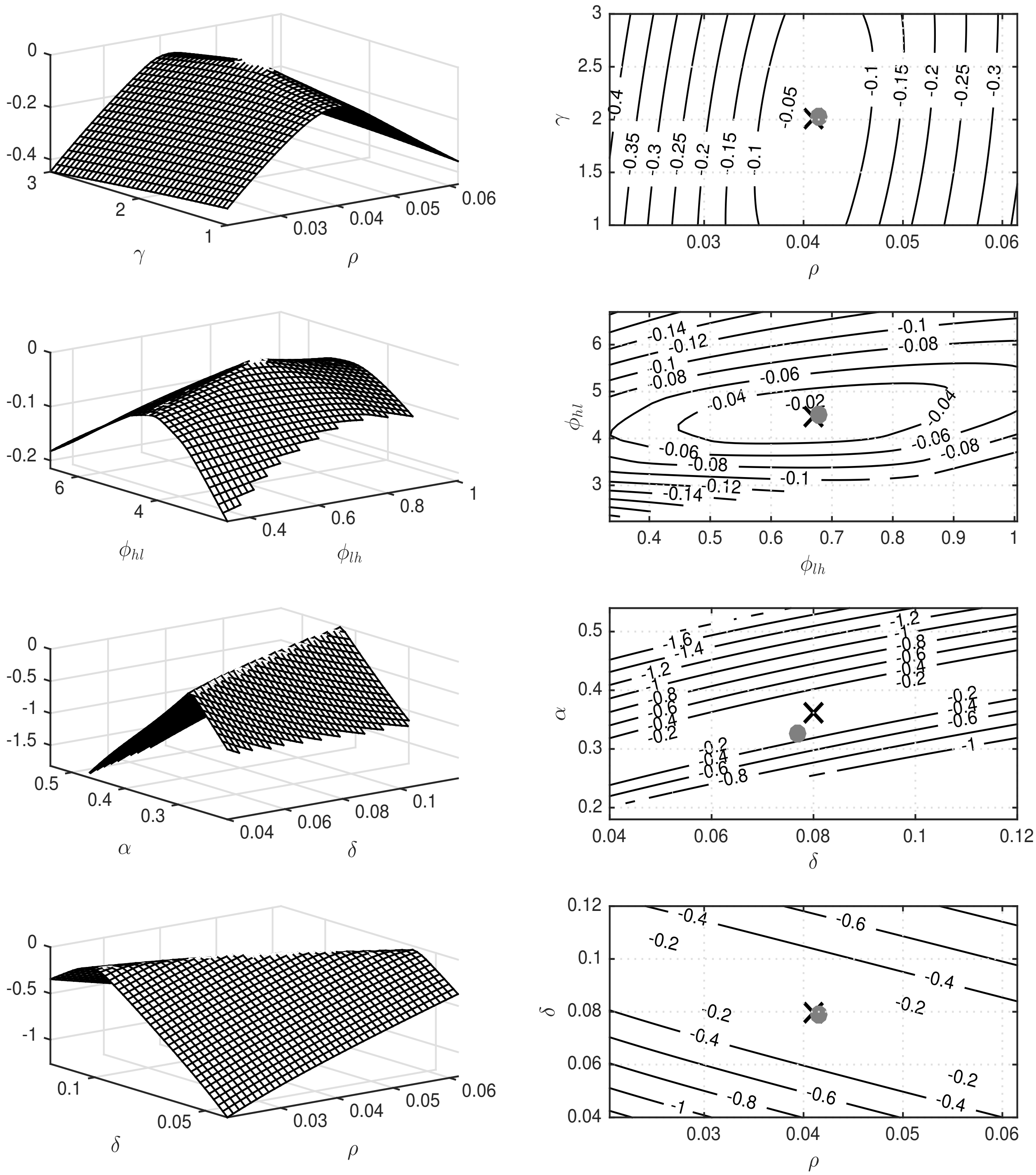


**Figure 3. Distance surface for selected parameters.** The graph shows the percentage deviation of the $L_1$ distance function for selected parameters as a function of the parameter space on the left column and its respective contour plot on the right column. The population values for the structural parameters, $\boldsymbol{\theta}_0$, are given in Table 1. "$\times$" locates the true parameter values, while ● the combination of parameters that deliver the maximum of the distance surface.

# 5 Finite sample properties and identification issues

This section uses Monte Carlo simulations to investigate the properties of the ML estimator in small samples by estimating the model of Section 2 on simulated data sets. The experiment is carried out by simulating $M = 100$ samples of size $N = 5000, 10000, 20000, 50000$ drawn from the model's population probability density function $g(a \mid \boldsymbol{\theta}_0)$. For each sample we estimate the vector of model parameters using the maximum likelihood estimator defined in Equation (24)[7]. The distribution of wealth is numerically approximated using $I \leq N$ grid points in the wealth-lattice spread over the interval $\mathcal{A} = [0, 40]$. The simulated data is generated using the Acceptance-Rejection sampling algorithm described in Appendix E.

The results of the Monte Carlo experiment are summarized in Table 3. For each $N$, it reports the true parameter vector, the mean of the estimates across the $M$ simulated data sets, the 5%, 50% and 95% percentiles of the estimated distribution, the vector of mean percentage biases and the Monte Carlo standard errors (in parenthesis). It is important to mention that the reported standard errors do not capture sampling uncertainty as we are using data from the true data generating process. We have excluded $e_h$ from the estimation exercise since from the model's point of view the endowments of efficiency units are normalized by this level. Therefore, any value for $e_l$ should be interpreted as relative to $e_h$ which in the population we have set to be one.

The Monte Carlo experiment reveals some important features that should be addressed. First, the mean estimate of the risk aversion coefficient, the discount rate and the parameters describing the endowments of efficient units are within a reasonable distance from their population counterparts. However, the mean estimate of the capital share in production and the depreciation rate of capital are positively biased. The bias is far from negligible in small samples. Their mean estimates are both economically and statistically different from those in the population. Second, the biases and the Monte Carlo standard errors decrease significantly as the sample size increases. For the case of $N = 50000$, the maximum bias does not exceed 5%. Third, the standard errors for $\gamma$ and $\phi_{lh}$ although decreasing with the sample size, are high relative to that of the rest of parameters. Fourth, the distribution of estimates seems far from normal making asymptotic inference unreliable. For $N = 20000$, which is similar to the sample size available in real life situations, Figure 4 plots the distribution of parameter estimates. Superimposed with a vertical bar in each box is the true parameter value. The figure gives a clearer idea of the meaningful biases already reported in $\alpha$ and $\delta$ and shows some deficiencies in $\gamma$ and $e_l$ that are not evident from looking at Table 3.

[7] The initial value used in the estimation procedure corresponds to the true parameter vector, $\boldsymbol{\theta}^0 = \boldsymbol{\theta}_0$.

**Table 3. Small sample estimates.**

The table reports finite sample estimates of the structural parameters of the model. The mean, the 5th, 50th and 95th percentile, % bias and Monte Carlo standard errors (in parenthesis) are obtained using $M = 100$ replications of the experiment. The sample size in each replication is given by $N$.

| | $N = 5000$ | | | | | $N = 10000$ | | | | | $N = 20000$ | | | | | $N = 50000$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{\theta}_0$ | Mean | 5% | Median | 95% | Bias | Mean | 5% | Median | 95% | Bias | Mean | 5% | Median | 95% | Bias | Mean | 5% | Median | 95% | Bias |
| $\gamma = 2.00$ | 2.18 | 1.01 | 1.78 | 4.65 | 9.20 (1.15) | 2.48 | 1.00 | 2.40 | 4.78 | 24.22 (1.27) | 2.27 | 1.05 | 2.17 | 3.95 | 13.45 (0.97) | 2.06 | 1.37 | 1.98 | 2.89 | 3.03 (0.45) |
| $\rho = 0.041$ | 0.04 | 0.02 | 0.04 | 0.08 | 7.58 (0.02) | 0.04 | 0.02 | 0.04 | 0.06 | 1.05 (0.01) | 0.04 | 0.03 | 0.04 | 0.06 | 1.19 (0.01) | 0.04 | 0.03 | 0.04 | 0.05 | 0.66 (0.01) |
| $\alpha = 0.36$ | 0.45 | 0.38 | 0.45 | 0.51 | 25.12 (0.05) | 0.43 | 0.34 | 0.43 | 0.50 | 19.40 (0.05) | 0.42 | 0.35 | 0.41 | 0.53 | 17.31 (0.05) | 0.37 | 0.35 | 0.37 | 0.40 | 2.01 (0.02) |
| $\delta = 0.08$ | 0.14 | 0.09 | 0.13 | 0.17 | 69.32 (0.04) | 0.12 | 0.08 | 0.13 | 0.16 | 54.62 (0.03) | 0.12 | 0.08 | 0.12 | 0.18 | 48.44 (0.03) | 0.08 | 0.07 | 0.08 | 0.10 | 4.79 (0.01) |
| $e_l = 0.10$ | 0.12 | 0.01 | 0.06 | 0.35 | 16.81 (0.12) | 0.15 | 0.01 | 0.14 | 0.37 | 46.25 (0.12) | 0.13 | 0.01 | 0.12 | 0.30 | 29.15 (0.10) | 0.11 | 0.01 | 0.10 | 0.22 | 9.06 (0.06) |
| $\phi_{hl} = 4.4644$ | 4.56 | 2.40 | 4.04 | 7.43 | 2.24 (1.57) | 4.38 | 2.69 | 4.09 | 6.81 | -1.98 (1.24) | 4.34 | 2.30 | 4.25 | 6.17 | -2.63 (0.91) | 4.44 | 3.45 | 4.41 | 5.58 | -0.57 (0.62) |
| $\phi_{lh} = 0.6697$ | 0.76 | 0.18 | 0.56 | 1.98 | 13.53 (0.57) | 0.72 | 0.18 | 0.61 | 1.69 | 7.11 (0.44) | 0.70 | 0.28 | 0.60 | 1.46 | 4.75 (0.36) | 0.69 | 0.40 | 0.64 | 1.11 | 2.95 (0.22) |

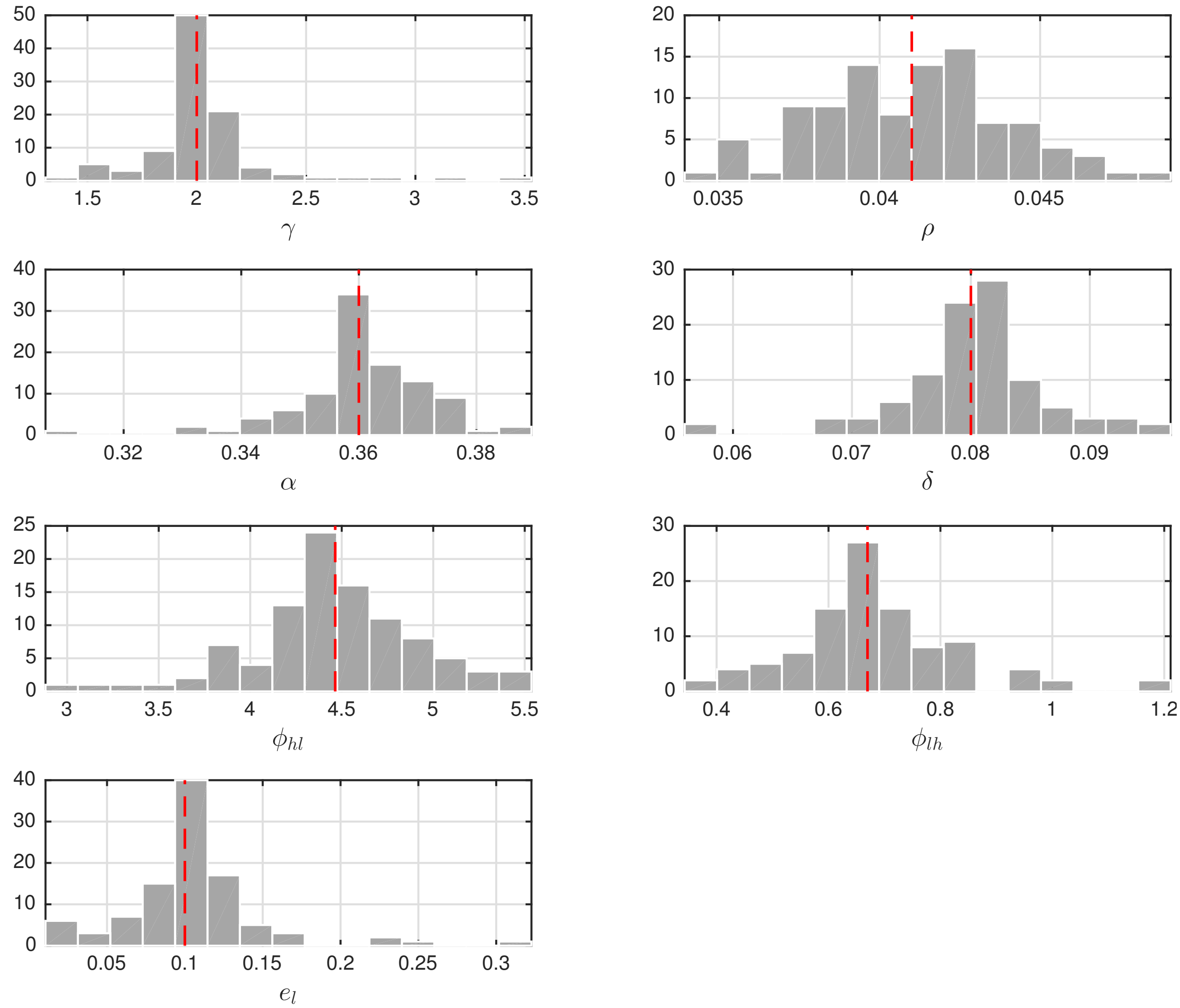


**Figure 4. Finite sample distribution of parameter estimates**. The graph plots the histogram of estimated parameters across $M = 100$ random samples of size $N = 20000$ generated from the true data generating process. The vertical line denotes the true parameter value.

While the results are encouraging for large samples, as they approach the true values in the population, they suggest that the strength of identification power of the maximum likelihood estimator in small samples is reduced in some dimensions of the parameter space. In the case of the prototype economy of Section 2, the data deficiencies induced by the use of samples of reduced size are reflected in a poor estimation of parameters related to the supply side of the economy. In particular, the results imply a higher participation of capital in the production function, and a higher fraction of the depreciated capital stock.

What are the consequences of having biased estimates in some of the model parameters for the macroeconomic aggregates and the theoretical implications of heterogeneous agent models? Figure 5 plots the distribution of estimates for the steady state interest rate, capital-output ratio and

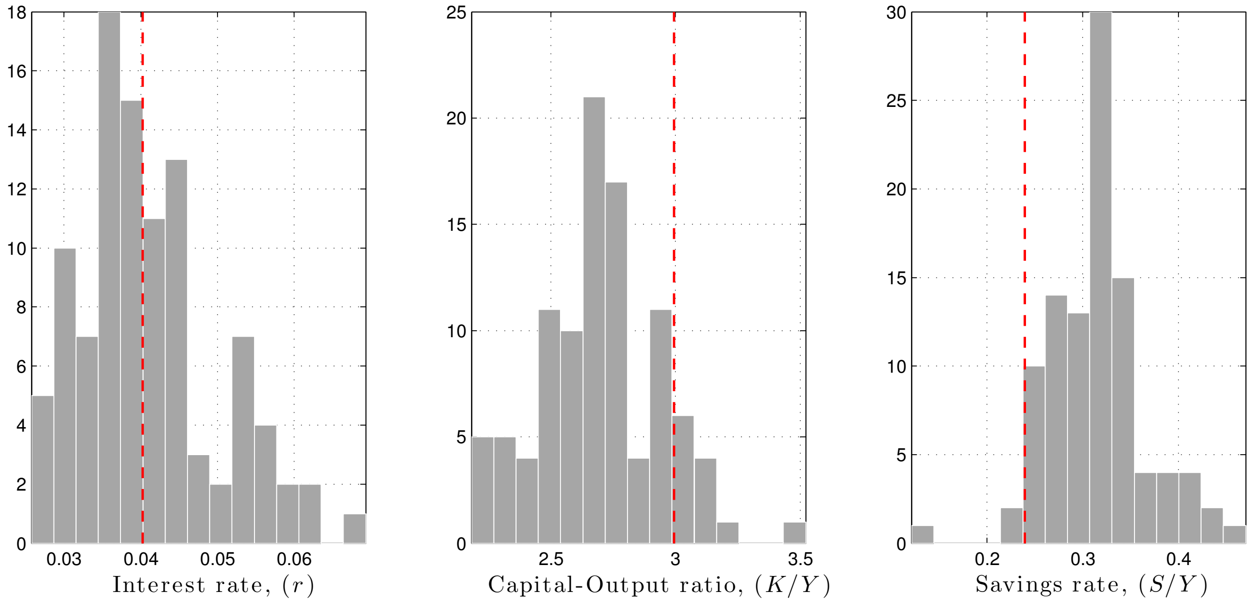


**Figure 5. Steady state macroeconomic aggregates estimates**. The graph shows the implications for the steady state interest rate, capital-output ratio and savings rate of the ML estimation using a sample size of $N = 20000$. The aggregate savings rate in the economy is computed as $^{(Y-C)}/_{Y}$. The vertical line denotes their value in the population.

aggregate savings rate implied by the ML estimation when $N = 20000$. While the small samples deficiencies documented above do not affect the steady state interest rate, they imply a downward bias in the capital-output ratio and an upward bias in the economy-wide savings rate that are far from trivial and could lead to misleading economic policy interpretations and recommendations. In particular, the estimation results overstate the contribution of uninsurable idiosyncratic risks to aggregate savings and therefore might erroneously suggest a degree of precautionary savings that is absent in the economy.

The evidence so far indicates that the accuracy of the ML estimates is poor in small samples. As pointed out in Canova and Sala (2006), it is possible to have situations where identification issues emerge just because of small samples even when the objective function is well behaved. As an example, consider the simple case of OLS with $p$ linear independent regressor, which ensures that the model parameters are identified in the population (See Newey and McFadden, 1986). In finite samples, however, the parameters of the model might not be identified due to data deficiencies. In fact, if $N < p$ the matrix of linear regressors is singular preventing us to recover the parameters of the model. Nonetheless, as $N \to \infty$ the estimated parameters will converge to their values in the population as the OLS sample objective function uniformly converges to a well behaved population objective function.

To assess the previous statement in our likelihood framework we build the log-likelihood profile of a particular sample with $N = 5000$ in each of the elements of $\boldsymbol{\theta}$ by varying one parameter at a time

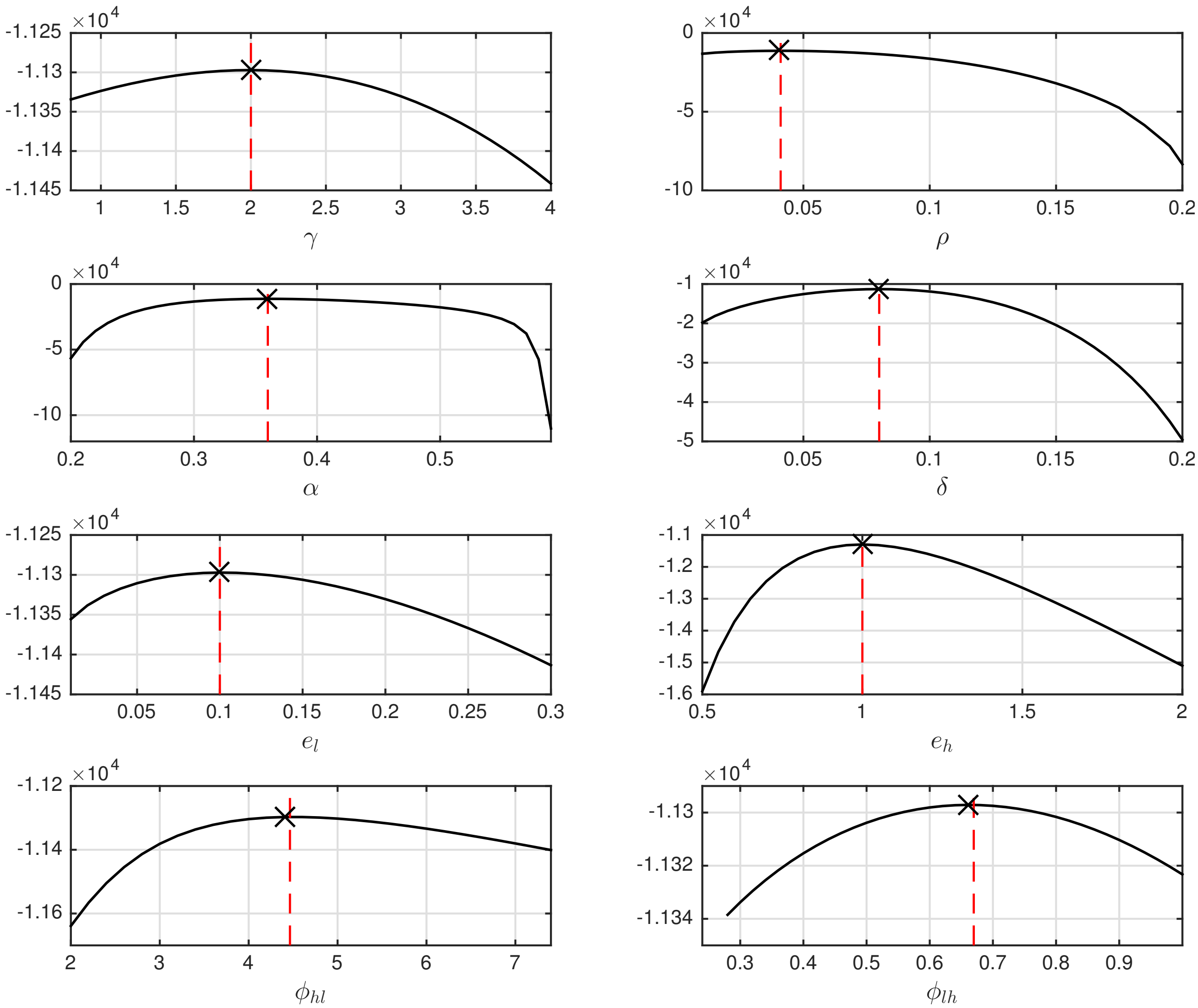


**Figure 6. Individual log-likelihood profile**. The graph shows the log-likelihood function $\mathcal{L}(\theta \mid \mathbf{a})$ for each $\theta \in \boldsymbol{\theta}$ along a neighborhood of the its true value, while keeping the remaining parameters at their population levels. The vertical line denotes the true parameter value and 'X' marks the location of parameter value associated with the maximum value of the log-likelihood profile.

in an economically reasonable neighborhood of its true value while keeping the remaining parameters at their population value. Figure 6 plots the results together with the true parameter values and the parameter value that generates the maximum of the log-likelihood function. Analogous to the population identification analysis of Section 4, the results suggest that the parameters of the model can be fully recovered by using only data on individual wealth. Similar conclusions can be reached by looking at Figure 8 in Appendix F where we plot the bivariate log-likelihood profile. Therefore, our experiment suggest that the biases found in small samples arise from data deficiencies and not from identification issues.

In conclusion, our Monte Carlo evidence suggest that although the model is well identified in

**Table 4. Small sample (conditional) estimates.**

The table reports finite sample estimates for a subset of the structural parameters of the model conditional on the calibrated values reported in the first row. The mean, the 5th, and 95th percentile, and % bias are obtained using $M = 100$ samples each of them of size $N = 5000$.

| | $\alpha = \alpha_0$ | | | | $\delta = \delta_0$ | | | | $\alpha = \alpha_0$ and $\delta = \delta_0$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\boldsymbol{\theta}_0$ | Mean | 5% | 95% | Bias | Mean | 5% | 95% | Bias | Mean | 5% | 95% | Bias |
| $\gamma = 2.00$ | 1.99 | 0.89 | 4.85 | -0.64 | 2.04 | 0.99 | 4.53 | 1.91 | 2.67 | 0.82 | 4.50 | 33.63 |
| $\rho = 0.041$ | 0.05 | 0.02 | 0.08 | 10.58 | 0.05 | 0.02 | 0.08 | 9.54 | 0.04 | 0.03 | 0.05 | 2.04 |
| $\alpha = 0.36$ | - | - | - | - | 0.37 | 0.32 | 0.42 | 1.54 | - | - | - | - |
| $\delta = 0.08$ | 0.08 | 0.05 | 0.10 | -5.08 | - | - | - | - | - | - | - | - |
| $e_l = 0.10$ | 0.09 | 0.01 | 0.37 | -9.31 | 0.10 | 0.01 | 0.36 | 1.03 | 0.17 | 0.01 | 0.44 | 64.78 |
| $\phi_{hl} = 4.4644$ | 4.65 | 2.53 | 7.68 | 4.09 | 4.63 | 2.34 | 7.62 | 3.75 | 4.27 | 2.09 | 7.37 | -4.33 |
| $\phi_{lh} = 0.6697$ | 0.76 | 0.16 | 2.00 | 12.94 | 0.76 | 0.16 | 2.00 | 13.51 | 0.73 | 0.17 | 2.00 | 8.98 |

the population, the use of small samples generates considerable inferential problems. However, as the sample size grows the use of cross-sectional data on individual wealth provides sufficient information to recover all the structural parameters of an otherwise standard heterogeneous agent model using a likelihood framework.

# 6 A note on calibration: Implications for empirical research

Our findings indicate that the maximum likelihood estimator produce parameter estimates that are usually biased and poorly identified in finite samples. A common practice among economists to get around this obstacle is to calibrate those parameters that are problematic by fixing their value and estimate the remaining ones. Therefore, this section investigates the potential consequences of following such an strategy. In the context of unidentified representative agent (linearized) DSGE models, Canova and Sala (2009) conclude that combining both approaches can lead to a biased inference and meaningless parameter estimates. To check whether this is the case in the heterogeneous agent model analyzed here, we conduct three Monte Carlo experiments based on the *a priori* assumption that the share of capital in the production function and the depreciation rate of capital can not be identified properly in small samples.

Table 4 summarizes the results when the sample size is $N = 5000$. It reports the true parameter vector, the mean of the estimates across the $M$ simulated data sets, the 5%, and 95% percentiles of the estimated distribution, and the vector of mean percentage biases for the following cases: (i) $\alpha$ is calibrated to its value in the population; (ii) $\delta$ is calibrated to its value in the population; (iii) $\alpha$ and $\delta$ are calibrated to their respective values in the population. Two main conclusions emerge from the experiment. First, while calibrating only one of the parameters reduces the bias in $\gamma$ and $e_l$ relative to the case where no calibration is used, fixing both at the same time extremely increases the bias. The opposite occurs with the mean estimate of the discount rate: its mean bias is considerably

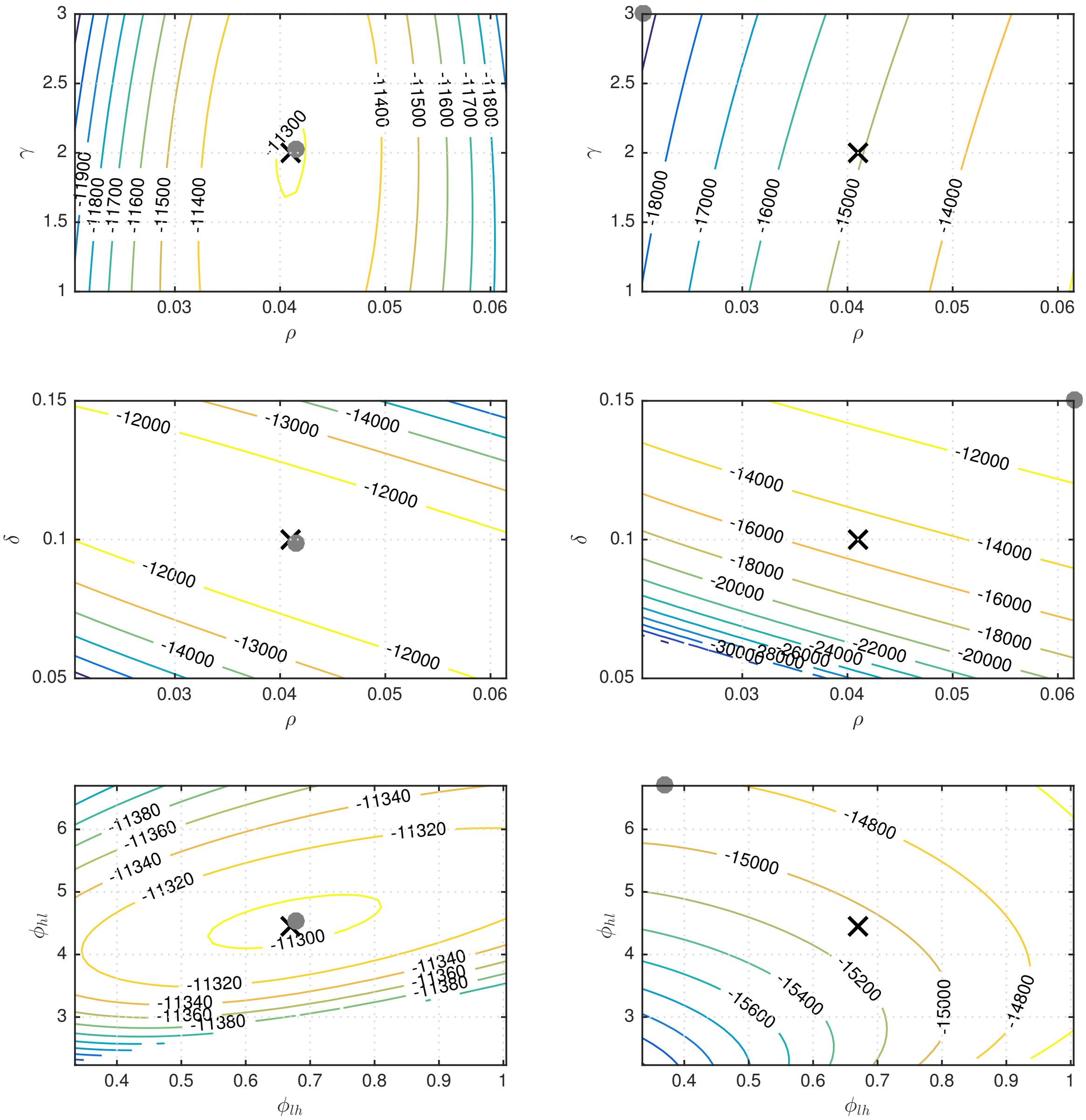


**Figure 7. Log-likelihood profile contours for selected parameters**. The graph shows the contour of the log-likelihood function $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{a})$ for selected combination of parameters using a random sample of size $N = 5000$ generated from the true model. The contours on the left use are identical to those in Figure 8. The contours on the right are built by varying the values of the parameters in the combination set while keeping the remaining parameters at their true population value, except for $\alpha$ and $\delta$ which are miscalibrated to 0.5 and 0.1 respectively. "$\times$" locates the true parameter values, and ● the combination of parameters that deliver the maximum of the log-likelihood function.

reduced by calibrating $\alpha$ and $\delta$ at the same time, while increases when only one of the problematic parameters is fixed. Second, the strategy of mixing calibration and estimation has no considerable effects on the mean estimates of the arrival rates of the Poisson process, $\phi_{lh}$ and $\phi_{hl}$.

The estimation results in Table 4 are mixed about the advantages or disadvantages of following a strategy based on combining calibration and estimation. While some parameters are better identified conditional on some others being calibrated, it could also be the case the some parameters are unaffected or considerably pushed away from their value in the population. However, so far we have assumed that the econometrician knows the population value of the calibrated parameters. In practice this is usually not the case. Therefore, Figure 7 plots the contours of the log-likelihood function for selected combinations of parameters using a random sample of size $N = 5000$ generated from the true model. The contours on the left column are identical to those found in Appendix F while those on the right column correspond to the log-likelihood profile when the problematic parameters $\alpha$ and $\delta$ are miscalibrated to 0.5 and 0.1 respectively. For both cases, we have marked the combination of parameters in the population and the combination of parameters that deliver the maximum of the log-likelihood function. As as result of such miscalibration, the estimator's objective function changes significantly in shape and hight, inducing biases in the estimates of the remaining parameters.

For small samples, the results suggests that estimating a subset of the parameters in the model, conditional on the remaining parameters being calibrated introduce stronger biases than those produced when unconditionally estimating all the structural parameters. In general, fixing some of the parameters does not increase the overall accuracy of the maximum likelihood approach in our heterogeneous agent framework since it may take the optimization routine to search for the maximum of the likelihood function in the wrong portion of the parameter space (See Canova and Sala, 2006). This in turn, will lead to inaccurate and inappropriate economic conclusions.

# 7 Conclusions

Heterogeneous agent models constitute a powerful framework in macroeconomics not just for the study of inequality and the distribution of wealth but also for the understanding of aggregates like GDP and the employment rate. However, there is little aggreement on how these models should be taken to the data. To date, calibration is the standard approach used by researchers to map observations into parameter values. Despite a being very illustrative methodology for the investigate the implications of economic, the use of econometric methods provide some important advantages by allowing: (i) to impose on the data the restrictions arising from the economic theory associated with a particular model; (ii) to assess the uncertainty surrounding the parameter values which ultimately provides a framework for hypothesis testing, (iii) for the use of standard tools of models selection and evaluation.

In this paper we introduce a simple full information likelihood approach to estimate the structural parameters of heterogeneous agent models cast in continuous-time by using the information content in the cross-sectional distribution of wealth. Our approach builds on earlier work by Bayer and Wälde (2011, 2013) and Achdou et al. (2014a) and uses Fokker-Planck equations to compute the stationary probability density function of wealth which can be later used to derive the likelihood function of the model.

We also investigate identification issues in our likelihood-based framework when the only available information to the econometrician is a cross-sectional sample of individual wealth. Given that the mapping between the deep parameters of the model and the estimator's objective function is highly nonlinear and not available in closed form, we assess the identification condition indirectly using simulation and graphical diagnostics as in Canova and Sala (2009). When the sample objective function is available we also investigate the small sample properties of the maximum likelihood estimator and the consequences of mixing calibration with estimation.

Our results indicate that the parameters of our prototype Bewley-Hugget-Aiyagari model are identified in the population, while the data-based likelihood function exhibit some identification issues in small samples. This identification problems lead to parameter estimates that are positively biased in the case of the capital share in the production function and the depreciation rate of the capital stock. However, as the sample size grows, their mean estimates, computed from a Monte Carlo experiment, converge to the value in the population.

We also find that following an empirical strategy that mixes calibration with estimation in an heterogeneous agent framework may deteriorate the properties of the maximum likelihood estimator by inducing changes in the shape and hight of the objective function. The problems are aggravated if the calibrated parameters are fixed to a value considerably different from their value in the population and if the sample size is small.

While our results are encouraging and suggest an important role for likelihood-based inference in heterogeneous agent models there is much more to be done before conclusive and definite evidence is available. Our future research agenda includes the computation of standard errors, investigating the properties of the maximum-likelihood models in more sophisticated models as in Krusell and Smith (1998), Cagetti and Nardi (2006), Angeletos and Calvet (2006), Angeletos (2007) and Benhabib et al. (2011), and testing the models with observed data.

# A Hamilton-Jacobi-Bellman equations

Define the optimal value function

$$V(a_0, e_0; w, r) = \max_{\{c_t\}_{t=0}^{\infty}} U_0 \quad s.t. \quad (2), (3)$$

in which the general equilibrium factor rewards $r$ and $w$ are taken as parametric.

For any $t \in [0, \infty)$, the household's problem can be characterized by the Hamilton-Jacobi-Bellman equation following the principle of optimality:

$$\rho V(a_t, e_t; r, w) = \max_{c_t \in \mathbb{R}^+} \left\{ u(c_t) + \frac{1}{\mathrm{d}t} \mathbb{E}_t \mathrm{d}V(a_t, e_t; r, w) \right\}.$$

Applying the change of variable formula (see Sennewald and Wälde, 2006) the continuation value is given by:

$$\mathrm{d}V(a_t, e_t; r, w) = V_a(a_t, e_t)\,\mathrm{d}a_t + (V(a_t, e_l) - V(a_t, e_h))\mathrm{d}q_1(t) + (V(a_t, e_h) - V(a_t, e_l))\mathrm{d}q_2(t)$$

where $V_a(a_t, e_t)$ denotes the partial derivative of the value function with respect to wealth.

Using Equation (2), and the martingale difference properties of the stochastic integrals under Poisson uncertainty,

$$\begin{aligned}
\mathbb{E}_s\Big[\int_s^t (V(a_t, e_l) - V(a_t, e_h))\mathrm{d}q_1(t) - \int_s^t (V(a_t, e_l) - V(a_t, e_h))\phi_1(e_t)\mathrm{d}t\Big] &= 0 \\
\mathbb{E}_s\Big[\int_s^t (V(a_t, e_h) - V(a_t, e_l))\mathrm{d}q_2(t) - \int_s^t (V(a_t, e_h) - V(a_t, e_l))\phi_2(e_t)\mathrm{d}t\Big] &= 0
\end{aligned}$$

for $s \leq t$, the Hamilton-Jacobi-Bellman equation can be written as:

$$\begin{aligned}
\rho V(a_t, e_t; r, w) = & \max_{c_t \in \mathbb{R}^+} \{u(c_t) + V_a(a_t, e_t; r, w)(ra_t + we_t - c_t) \\
& + (V(a_t, e_l; r, w) - V(a_t, e_h; r, w))\phi_1(e_t) + (V(a_t, e_h; r, w) - V(a_t, e_l; r, w))\phi_2(e_t)\}
\end{aligned}$$

The first-order condition of an interior solution reads:

$$u'(c_t) = V_a(a_t, e_t; r, w), \tag{26}$$

for any $t \in [0, \infty)$ making optimal consumption $c_t^\star = c(a_t, e_t)$ a function only of the states and independent of calendar time, $t$.

Due to the state dependence of the arrival rates in the endowments of efficiency units, only one Poisson process will be active for each of the values of the discrete state variable, $e_t$. Using the first order condition we obtain a bivariate system of maximized HJB equations:

$$\begin{aligned}
\rho V(a_t, e_h; r, w) &= u(c_t^\star) + V_a(a_t, e_h; r, w)(ra_t + we_h - c_t^\star) + (V(a_t, e_l; r, w) - V(a_t, e_h; r, w))\phi_{lh}, \\
\rho V(a_t, e_l; r, w) &= u(c_t^\star) + V_a(a_t, e_l; r, w)(ra_t + we_l - c_t^\star) + (V(a_t, e_h; r, w) - V(a_t, e_l; r, w))\phi_{hl}.
\end{aligned}$$

## B Fokker-Planck equations

Assume there exists a function $f$ whose arguments are the stochastic process $a$ and $e$ and define the agent's optimal savings function as $s(a_t, e_t) = ra_t + we_t - c(a_t, e_t)$. Using the change of variable formula, the evolution of $f$ is given by:

$$\begin{aligned} \mathrm{d}f(a_t, e_t) &= f_a(a_t, e_t)\, s(a_t, e_t)\, \mathrm{d}t \\ &\quad + \left(f(a_t, e_l) - f(a_t, e_h)\right) \mathrm{d}q_1(t) + \left(f(a_t, e_h) - f(a_t, e_l)\right) \mathrm{d}q_2(t) \end{aligned}$$

Due to the state dependence of the arrival rates only one Poisson process will be active. Applying the expectations operator conditional on the information available at instant $t$ and dividing by $dt$ we obtain the infinitesimal generator of $f(a_t, e_t)$, denoted by $\mathcal{A}f(a_t, e_t) \equiv \frac{\mathbb{E}_t \mathrm{d}f(a_t, e_t)}{\mathrm{d}t}$:

$$\begin{aligned} \frac{\mathbb{E}_t \mathrm{d}f(a_t, e_t)}{\mathrm{d}t} &= f_a(a_t, e_t)\, s(a_t, e_t) \\ &\quad + \left(f(a_t, e_l) - f(a_t, e_h)\right) \phi_{lh} + \left(f(a_t, e_h) - f(a_t, e_l)\right) \phi_{hl}. \end{aligned} \tag{27}$$

By means of Dynkin's formula, the expected value of the function $f(\cdot)$ at a point in time $t$ is given by the expected value of the function at instant $s < t$ plus the sum of the expected future changes up to $t$:

$$\mathbb{E}f(a_t, e_t) = \mathbb{E}f(a_s, e_s) + \int_s^t \mathbb{E}\left(\mathcal{A}f(a_\tau, e_\tau)\right) \mathrm{d}\tau \tag{28}$$

Differentiating Equation (28) with respect to time:

$$\begin{aligned} \frac{\partial}{\partial t}\mathbb{E}f(a_t, e_t) &= \frac{\partial}{\partial t}\left\{\mathbb{E}f(a_s, e_s) + \int_s^t \mathbb{E}\left(\mathcal{A}f(a_\tau, e_\tau)\right) \mathrm{d}\tau\right\} \\ &= \frac{\partial}{\partial t}\left\{\mathbb{E}f(a_s, e_s) + \int_s^t \mathbb{E}\left(\frac{\mathbb{E}_\tau df(a_\tau, e_\tau)}{d\tau}\right) \mathrm{d}\tau\right\} \\ &= \frac{\partial}{\partial t}\left\{\mathbb{E}f(a_s, e_s) + \int_s^t \mathbb{E}df(a_\tau, e_\tau)\right\} \\ &= \mathbb{E}\left(\mathcal{A}f(a_t, e_t)\right) \\ &= \sum_{e_t \in \{e_h, e_l\}} \int_{\underline{a}}^{\infty} \mathcal{A}f(a_t, e_t)\, g(a_t, e_t, t)\, \mathrm{d}a_t \end{aligned}$$

that is:

$$\frac{\partial}{\partial t}\mathbb{E}f(a_t, e_t) = \underbrace{\int_{-\infty}^{\infty} \mathcal{A}f(a_t, e_h)\, g(a_t, e_h, t)\, \mathrm{d}a_t}_{\omega_{e_h}} + \underbrace{\int_{-\infty}^{\infty} \mathcal{A}f(a_t, e_l)\, g(a_t, e_l, t)\, \mathrm{d}a_t}_{\omega_{e_l}}$$

where $g(a_t, e_t, t)$ is the joint density function of wealth and endowment of efficiency units at instant $t$.

For illustration consider the case of $e_t = e_h$, i.e., $\omega_{e_h}$. Using the definition of the infinitesimal operator together with Equation (27) we note that:

$$\mathcal{A}f(a_t, e_h) \quad = \quad f_a(a_t, e_h)\, s(a_t, e_h) + (f(a_t, e_l) - f(a_t, e_h))\, \phi_{lh} \tag{29}$$

Hence,

$$\begin{aligned}
\omega_{e_h} &= \int_{\underline{a}}^{\infty} \Big[ f_a(a_t, e_h)\, s(a_t, e_h) + (f(a_t, e_l) - f(a_t, e_h))\, \phi_{lh} \Big] g(a_t, e_h, t)\, \mathrm{d}a_t \\
&= \int_{\underline{a}}^{\infty} f_a(a_t, e_h)\, s(a_t, e_h)\, g(a_t, e_h, t)\, \mathrm{d}a_t + \int_{\underline{a}}^{\infty} (f(a_t, e_l) - f(a_t, e_h))\, \phi_{lh} g(a_t, e_h, t)\, \mathrm{d}a_t
\end{aligned}$$

Using integration by part for the term associated with $f_a$:

$$\int_{\underline{a}}^{\infty} f_a(a_t, e_h)\, s(a_t, e_h)\, g(a_t, e_h, t)\, \mathrm{d}a_t \quad = \quad -\int_{\underline{a}}^{\infty} f(a_t, e_h)\, \frac{\partial}{\partial a_t} [s(a_t, e_h)\, g(a_t, e_h, t)]\, \mathrm{d}a_t$$

where:

$$\frac{\partial}{\partial a_t} [s(a_t, e_h)\, g(a_t, e_h, t)] = \left( r_t - \frac{\partial}{\partial a_t} c(a_t, e_h) \right) g(a_t, e_h, t) + s(a_t, e_h)\, \frac{\partial}{\partial a_t} g(a_t, e_h, t).$$

Hence,

$$\begin{aligned}
\omega_{e_h} = \int_{\underline{a}}^{\infty} f(a_t, e_h) \left[ -\left( r_t - \frac{\partial}{\partial a_t} c(a_t, e_h) \right) g(a_t, e_h, t) - s(a_t, e_h)\, \frac{\partial}{\partial a_t} g(a_t, e_h, t) \right] \mathrm{d}a_t \\
+ \int_{\underline{a}}^{\infty} \Big[ (f(a_t, e_l) - f(a_t, e_h))\, \phi_{lh} \Big] g(a_t, e_h, t)\, \mathrm{d}a_t
\end{aligned}$$

$$\begin{aligned}
\omega_{e_l} = \int_{\underline{a}}^{\infty} f(a_t, e_l) \left[ -\left( r_t - \frac{\partial}{\partial a_t} c(a_t, e_l) \right) g(a_t, e_l, t) - s(a_t, e_l)\, \frac{\partial}{\partial a_t} g(a_t, e_2, t) \right] \mathrm{d}a_t \\
+ \int_{\underline{a}}^{\infty} \Big[ (f(a_t, e_h) - f(a_t, e_l))\, \phi_{hl} \Big] g(a_t, e_l, t)\, \mathrm{d}a_t
\end{aligned}$$

Note that the expected value of $f$ can be written as:

$$\mathbb{E} f(a_t, e_t) \quad = \quad \int_{\underline{a}}^{\infty} f(a_t, e_h)\, g(a_t, e_h, t)\, \mathrm{d}a_t + \int_{\underline{a}}^{\infty} f(a_t, e_l)\, g(a_t, e_l, t)\, \mathrm{d}a_t$$

and therefore:

$$\frac{\partial}{\partial t}\mathbb{E}f\left(a_t,e_t\right) \;=\; \int\limits_{\underline{a}}^{\infty} f\left(a_t,e_h\right)\frac{\partial}{\partial t}g\left(a_t,e_h,t\right)\mathrm{d}a_t + \int\limits_{\underline{a}}^{\infty} f\left(a_t,e_l\right)\frac{\partial}{\partial t}g\left(a_t,e_l,t\right)\mathrm{d}a_t \tag{30}$$

Finally we equate Equations (29) and (30) and collect terms to obtain:

$$\int\limits_{\underline{a}}^{\infty} f\left(a_t,e_h\right)\varphi_{e_h}\mathrm{d}a_t + \int\limits_{\underline{a}}^{\infty} f\left(a_t,e_l\right)\varphi_{e_l}\mathrm{d}a_t = 0 \tag{31}$$

where:

$$\begin{aligned}\varphi_{e_h} = &-\left(r_t - \frac{\partial}{\partial a_t}c\left(a_t,e_h\right)+\phi_{lh}\right)g\left(a_t,e_h,t\right)\\ &\qquad - s\left(a_t,e_h\right)\frac{\partial}{\partial a_t}g\left(a_t,e_h,t\right)+\phi_{hl}g\left(a_t,e_l,t\right)-\frac{\partial}{\partial t}g\left(a_t,e_h,t\right)\end{aligned}$$

$$\begin{aligned}\varphi_{e_l} = &-\left(r_t - \frac{\partial}{\partial a_t}c\left(a_t,e_l\right)+\phi_{hl}\right)g\left(a_t,e_l,t\right)\\ &\qquad - s\left(a_t,e_l\right)\frac{\partial}{\partial a_t}g\left(a_t,e_l,t\right)+\phi_{lh}g\left(a_t,e_h,t\right)-\frac{\partial}{\partial t}g\left(a_t,e_l,t\right)\end{aligned}$$

The Fokker-Planck equations that define these subdensities are obtained by setting:

$$\varphi_{e_l} = \varphi_{e_h} = 0$$

since that is that only way the integral equation (31) can be satisfied for all possible functions $f$. A formal proof can be found in Bayer and Wälde (2013). This results in a system of two non-autonomous quasi-linear partial differential equations in two unknowns $g\left(a_t,e_h,t\right)$, $g\left(a_t,e_l,t\right)$:

$$\begin{aligned}&\frac{\partial}{\partial t}g\left(a_t,e_h,t\right)+s\left(a_t,e_h\right)\frac{\partial}{\partial a_t}g\left(a_t,e_h,t\right)=\\ &\qquad -\left(r_t-\frac{\partial}{\partial a_t}c\left(a_t,e_h\right)+\phi_{lh}\right)g\left(a_t,e_h,t\right)+\phi_{hl}g\left(a_t,e_l,t\right)\end{aligned}$$

$$\begin{aligned}&\frac{\partial}{\partial t}g\left(a_t,e_l,t\right)+s\left(a_t,e_l\right)\frac{\partial}{\partial a_t}g\left(a_t,e_l,t\right)=\\ &\qquad -\left(r_t-\frac{\partial}{\partial a_t}c\left(a_t,e_l\right)+\phi_{hl}\right)g\left(a_t,e_l,t\right)+\phi_{lh}g\left(a_t,e_h,t\right).\end{aligned}$$

The long-run subdensities correspond to the case where the time derivatives ${\partial g(a_t,e_t,t)}/{\partial t}$ are zero for all $e_t \in \mathcal{E}$, which transforms the previous system of equations into one of ordinary differential equations as described by Equations (17) and (18) in the main text.

# C Transition probabilities for the endowment of efficiency units

This appendix shows how to derive the limiting (stationary) probability distribution of the endowment of efficiency units defined in Equations (15) and (16) from the arrival rates of the stochastic process defined by Equation (3).

For illustration purposes consider an individual who is in state $e_t = e_l$ at time $s$. Let $p(e_l, e_h, t) \equiv \mathbb{P}(e_t = e_h \mid e_s = e_l)$ for $s \leq t$ denote the probability that the individual jumps from state $e_l$ at time $s$ from state $e_h$ at time $t$, and $\phi_{lh}$ and $\phi_{hl}$ the instantaneous transition rates at which the stochastic process jumps to state $e_l$ from state $e_h$ and to state $e_h$ from state $e_l$ respectively. Then, the transition probabilities at time $t$ can be computed from solving the following system of Backward Kolmogorov equations (see Ross, 2009):

$$\begin{aligned}\dot{p}(e_h, e_h, t) &= \phi_{lh}\left[p(e_l, e_h, t) - p(e_h, e_h, t)\right],\\ \dot{p}(e_l, e_h, t) &= \phi_{hl}\left[p(e_h, e_h, t) - p(e_l, e_h, t)\right]\end{aligned}$$

where $\dot{p}(e_i, e_j, t) = \lim_{s\to 0}\frac{1}{s}\left[p(e_i, e_j, t+s) - p(e_i, e_j, t)\right]$ for all $i, j \in \mathcal{E}$ and $p(e_h, e_h, s) = 1$ and $p(e_l, e_h, s) = 0$ are initial conditions. The solution to this system of ordinary differential equations is given by:

$$p(e_h, e_h, t) = \frac{\phi_{hl}}{\phi_{hl}+\phi_{lh}} + \frac{\phi_{lh}}{\phi_{hl}+\phi_{lh}} e^{-(\phi_{hl}+\phi_{lh})(t-s)} \tag{32}$$

$$p(e_l, e_h, t) = \frac{\phi_{hl}}{\phi_{hl}+\phi_{lh}} - \frac{\phi_{hl}}{\phi_{hl}+\phi_{lh}} e^{-(\phi_{hl}+\phi_{lh})(t-s)}. \tag{33}$$

Now let $p(e_h, s)$ denote the unconditional probability of being in state $e_h$ at time $s$. The unconditional probability of being in the same state at time $t > s$ can be computed according to:

$$p(e_h, t) = p(e_h, s)\, p(e_h, e_h, t) + (1 - p(e_h, s))\, p(e_l, e_h, t). \tag{34}$$

In the limit as $t \to \infty$ the unconditional probability of having an endowment of high efficiency is given by:

$$\lim_{t\to\infty} p(e_h, t) = p(e_h) = \frac{\phi_{hl}}{\phi_{hl}+\phi_{lh}}. \tag{35}$$

A similar procedure can be used to show that the stationary and unconditional probability of having an endowment of low efficiency is:

$$\lim_{t\to\infty} p(e_l, t) = p(e_l) = \frac{\phi_{lh}}{\phi_{hl}+\phi_{lh}}. \tag{36}$$

The system of equations formed by (32) and (33) together with an appropriate choice of $(t-s)$ can be used to back out the instantaneous transition rates of the Poisson processes, $\phi_{hl}$ and $\phi_{lh}$ from any probability transition matrix. Given the annual frequency used in the calibration of the model of Section 2, we set $(t-s) = 1$ (one year).

## D Computation of the stationary equilibrium

To computation of the stationary density of wealth is done following the method in Achdou et al. (2014b) which consists of two main blocks. The first block computes the stationary general equilibrium at the macro level by using the following fixed point algorithm in the time-invariant aggregate capital stock:

**Algorithm D.1 (Stationary General Equilibrium)** *Make an initial guess for the aggregate capital stock, $K^0$, and then for $j = 0, 1, \ldots$:*

1. *Compute the factor prices $r^j$ and $w^j$ using Equation (10).*

2. *Compute the optimal consumption functions $c^j(a, e_h)$ and $c^j(a, e_l)$ and the subdensities $g^j(a, e_h)$ and $g^j(a, e_l)$.*

3. *Update $K^{j+1}$ according to*

$$K^{j+1} = \omega K^j + (1 - \omega) \sum_{e_t} \int_{\underline{a}}^{\infty} a_t g^j(a_t, e_t) \, \mathrm{d}a_t$$

   *where $\omega \in (0, 1]$ is a relaxation parameter.*

4. *If $\left\| K^{j+1} - K^j \right\| < \epsilon$ stop, otherwise return to step 1.*

Algorithm D.1 does not require to update the aggregate labor supply $L$ at each iteration $j = 0, 1, \ldots$ since in our prototype economy the labor supply is assumed to be exogenous as can be seen from Equation (12).

The second block approximates the solution of the household's problem at the micro level using the finite difference methods suggested in Candler (1999) and Achdou et al. (2014b). This solution, which is required in step 2 of Algorithm D.1 for every iteration $j = 0, 1, \ldots$, consists of two set of functions which can be independently computed. The first set corresponds to the policy functions for consumption associated with the solution to the HJB equations (7) and (8), while the second set corresponds to the subdensity functions associated with the solution to the Fokker-Planck equations (17) and (18).

**Solving the Hamilton-Jacobi-Bellman equations.**

Consider first the solution to the HJB equations. For each $e_t \in \mathcal{E}$, the finite difference method approximates the function $V(a_t, e_t)$ on an equally spaced grid for wealth with $I$ discrete points, $a_i, i = 1, \ldots, I$, where $a_i \in \mathcal{A} = [a_{min}, a_{max}]$ and $a_{min} = \underline{a}$. The distance between points is denoted by $\Delta a$ and we introduce the short-hand notation $V_{e,i} \equiv V(a_i, e)$. The derivative $V_a(a_i, e) \equiv V'_{e,i}$ is computed with either a forward or a backward difference approximation:

$$V'^{F}_{e,i} \approx \frac{V_{e,i+1} - V_{e,i}}{\Delta a} \tag{37}$$

$$V'^{B}_{e,i} \approx \frac{V_{e,i} - V_{e,i-1}}{\Delta a}. \tag{38}$$

Following Candler (1999), the choice of difference operator is based on an upwind differentiation scheme. The correct approximation is based on the direction of the continuous state variable. Thus, if the saving function, $s(a_i, e) \equiv s_{e,i} = ra_i + we - (u')^{-1}\left(V'_{e,i}\right)$, is positive we use a forward operator and if it is negative we use the backward operator. This gives rise to the following upwind operator:

$$V'_{e,i} = V'^{F}_{e,i}\mathbf{1}_{\{s^F_{e,i}>0\}} + V'^{B}_{e,i}\mathbf{1}_{\{s^B_{e,i}<0\}} + \bar{V}'_{e,i}\mathbf{1}_{\{s^F_{e,i}<0<s^B_{e,i}\}} \tag{39}$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function and, $s^F_{e,i}$ and $s^B_{e,i}$ the saving functions computed with the forward and difference operators respectively. Following Achdou et al. (2014b), the concavity of the value function in the wealth dimension motivates the last term in Equation (39) since there could be grid points $a_i \in \mathcal{A}$ for which $s^F_{e,i} < 0 < s^B_{e,i}$. In those cases, they suggest to set savings to be equal to zero which implies that the derivative of the value function is equal to $\bar{V}'_{e,i} = u'(ra_i + we)$.

The finite difference approximation to the HJB equations is then given by:[8]

$$\rho V_{e,i} = u(c_{e,i}) + V'_{e,i}\left[ra_i + ew - c_{e,i}\right] + \phi_{-ee}\left[V_{-e,i} - V_{e,i}\right]$$

for each $e \in \mathcal{E}$, where optimal consumption is given by:

$$c_{e,i} = \left(u'\right)^{-1}\left(V'_{e,i}\right).$$

The upwind representation of the HJB equation reads:

$$\rho V_{e,i} \quad = \quad u(c_{e,i}) + \frac{V_{e,i+1} - V_{e,i}}{\Delta a}(s_{e,i})^{+} + \frac{V_{e,i} - V_{e,i-1}}{\Delta a}(s_{e,i})^{-} + \phi_{-ee}\left[V_{-e,i} - V_{e,i}\right] \tag{40}$$

where:

$$(s_{e,i})^{+} = \max\left\{ra_i + we - \left(u'\right)^{-1}\left(V'^{F}_{e,i}\right), 0\right\} \quad \text{and} \quad (s_{e,i})^{-} = \min\left\{ra_i + we - \left(u'\right)^{-1}\left(V'^{B}_{e,i}\right), 0\right\}$$

denote the positive and negative parts of savings respectively.

Equation (40) defines a highly non linear system of equations in $V_{e,i}$ that can only be solved by iterative methods. We follow Candler (1999) and construct an iterative procedure based on the time-dependent HJB equation, $V^l_{e,i} \equiv V(a_i, e, t)$, and then from an arbitrary initial condition we integrate in time until the solution is no longer a function of the initial condition, i.e. until it converges to

[8] The state-constraint boundary condition in Equation (9) is enforced at the lower bound of the state space, $a_{min}$, by imposing $V'^{B}_{e,1} = u'(ra_1 + we)$.

the time-independent HJB, $V_{e,i}$. The time-updating is carried out by means of an implicit scheme in which the value function at the next time step, $V_{e,i}^{l+1}$, is implicitly defined by the equation:

$$\frac{V_{e,i}^{l+1}-V_{e,i}^{l}}{\Delta}+\rho V_{e,i}^{l+1}=u\left(c_{e,i}^{l}\right)+\frac{V_{e,i+1}^{l+1}-V_{e,i}^{l+1}}{\Delta a}\left(s_{e,i}^{l}\right)^{+} \\ +\frac{V_{e,i}^{l+1}-V_{e,i-1}^{l+1}}{\Delta a}\left(s_{e,i}^{l}\right)^{-}+\phi_{-ee}\left[V_{-e,i}^{l+1}-V_{e,i}^{l+1}\right] \tag{41}$$

where $\Delta$ is the time step size, $c_{e,i}^{l}=(u')^{-1}\left[\left(V_{e,i}^{l}\right)'\right]$ and $\left(V_{e,i}^{l}\right)'$ is given by Equation (39).

Equation (41) constitutes now a system of $2\times I$ linear equations in $V_{e,i}^{l+1}$ with the following matrix representation:

$$\mathbf{A}^{l}\mathbf{V}^{l+1}=\mathbf{b}^{l} \tag{42}$$

where $\mathbf{V}^{l+1}=\left(V_{e_l,1}^{l+1},\dots,V_{e_l,I}^{l+1},V_{e_h,1}^{l+1},\dots,V_{e_h,I}^{l+1}\right)'$, $\mathbf{b}^{l}$ is a vector with elements $b_{e,i}^{l}=u\left(c_{e,i}^{l}\right)+V_{e,i}^{l}/\Delta$ and $\mathbf{A}^{l}$ is the block matrix:

$$\mathbf{A}^{l}=\begin{bmatrix}\mathbf{A}_{e_l} & -\Phi_{hl}\\ -\Phi_{lh} & \mathbf{A}_{e_h}\end{bmatrix}$$

with $\Phi_{-ee}=-\phi_{-ee}\mathbf{I}_I$ and

$$\mathbf{A}_e=\begin{bmatrix} y_{e,1} & z_{e,1} & 0 & \dots & 0 & 0\\ x_{e,2} & y_{e,2} & z_{e,2} & \dots & 0 & 0\\ 0 & x_{e,3} & y_{e,3} & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & 0 & \dots & y_{e,I-1} & z_{e,I-1}\\ 0 & 0 & 0 & \dots & x_{e,I} & y_{e,I}\end{bmatrix}.$$

where

$$\begin{aligned} x_{e,i} &= \frac{\left(s_{e,i}^{l}\right)^{-}}{\Delta a}\\ y_{e,i} &= \frac{1}{\Delta}+\rho+\frac{\left(s_{e,i}^{l}\right)^{+}}{\Delta a}-\frac{\left(s_{e,i}^{l}\right)^{-}}{\Delta a}+\phi_{-ee}\\ z_{e,i} &= -\frac{\left(s_{e,i}^{l}\right)^{+}}{\Delta a}. \end{aligned}$$

and $e\in\mathcal{E}$. The iterative algorithm used to find the solution to the HJB equation can be summarized as follows:

**Algorithm D.2 (Solution of the HJB equation)** *Guess $V_{e,i}^{0}$ for each $e\in\mathcal{E}$ and $i=1,\dots,I$. Then for $l=0,1,2,\dots$:*

1. *Compute $\left(V_{e,i}^{l}\right)'$ using Equation (39).*

2. *Compute $c_{e,i}^{l}=(u')^{-1}\left(V_{e,i}'\right)$.*

*3. Find $V_{e,i}^{l+1}$ by solving the system of equations defined in (42).*

*4. If $\left\|V_{e,i}^{l+1}-V_{e,i}^{l}\right\|<\epsilon$ stop. Otherwise, go to step 1.*

**Solving the Fokker-Planck equations.**

Once the optimal consumption has been computed from Algorithm D.2, we proceed to approximate the solution of the associated Fokker-Planck equations (17) and (18). As before, we use a finite difference method and apply it to:

$$
\begin{aligned}
0 &= -\frac{\partial}{\partial a_t}\left[s\left(a_t,e_l\right)g\left(a_t,e_l\right)\right]-\phi_{hl}g\left(a_t,e_l\right)-\phi_{lh}g\left(a_t,e_h\right), && (43)\\
0 &= -\frac{\partial}{\partial a_t}\left[s\left(a_t,e_h\right)g\left(a_t,e_h\right)\right]-\phi_{lh}g\left(a_t,e_h\right)-\phi_{hl}g\left(a_t,e_l\right) && (44)
\end{aligned}
$$

which corresponds to an alternative representation of Equations (17) and (18) as shown in Appendix B. We further need to restrict the solution to satisfy the integrability condition:

$$
1=\sum_{e_t\in\{e_l,e_h\}}\int_{-\infty}^{\infty}g\left(a_t,e_t\right)\mathrm{d}a. \tag{45}
$$

The system of equations (44) to (45) is discretized as follows:

$$
\begin{aligned}
0 &= -\left[s_{e,i}g_{e,i}\right]'-\phi_{-ee}g_{e,i}-\phi_{e,-e}g_{-e,i} && (46)\\
1 &= \sum_{e_t\in\{e_l,e_h\}}\sum_{i=1}^{I}g_{e,i}\Delta a. && (47)
\end{aligned}
$$

for each $e\in\mathcal{E}$ and where $g_{e,i}\equiv g\left(a_i,e\right)$. To approximate the derivative $\left[s_{e,i}g_{e,i}\right]'$ we use the upwind differentiation scheme:

$$
\left[s_{e,i}g_{e,i}\right]' = \tfrac{(s_{e,i})^{+}g_{e,i}-(s_{e,i-1})^{+}g_{e,i-1}}{\Delta a}+\tfrac{(s_{e,i+1})^{-}g_{e,i+1}-(s_{e,i})^{-}g_{e,i}}{\Delta a},
$$

where $s_{e,i}=ra_i+we-\left(u'\right)^{-1}\left(V_{e,i}'\right)$ is the optimal savings function obtained from the solution to the HJB equation. Equation (46) defines a system of $2\times I$ linear equations in $g_{e,i}$ with matrix representation:

$$
\mathbf{B}\mathbf{g}=\mathbf{0} \tag{48}
$$

where $\mathbf{g}=\left(g_{e_l,1},\ldots,g_{e_l,I},g_{e_h,1},\ldots,g_{e_h,I}\right)'$. The matrix $\mathbf{B}$ is defined as $\mathbf{B}=\tilde{\mathbf{A}}^{\top}$, where $\tilde{\mathbf{A}}=-\mathbf{A}+\left(\rho+\frac{1}{\Delta}\right)\mathbf{I}$. The matrix $\tilde{\mathbf{A}}$ captures the evolution of the continuous-time stochastic processes $\{a_t,e_t\}_{t=0}^{\infty}$. To impose the integrability condition in Equation (45) we follow Achdou et al. (2014b) and fix $g_{e,i}=0.1$ for an arbitrary $i$. Then solve the system of equations in (48) for some $\tilde{\mathbf{g}}$, and proceed to renormalize $g_{e,i}={}^{\tilde{g}_{e,i}}/(\sum_{e,i}\tilde{g}_{e,i}\Delta a)$.

## E Sampling algorithm

To drawn random observations from the model's population probability density function $g(a \mid \boldsymbol{\theta}_0)$ we use the following Acceptance-Rejection algorithm:

**Algorithm E.1 (Acceptance-Rejection algorithm)** *Let $a \in [\underline{a}, \infty)$ be a random variable with target probability density function $g(a \mid \boldsymbol{\theta}_0)$, and let $f$ be an instrumental density defined in the support of $a$ such that $g(a \mid \boldsymbol{\theta}_0) \leq \kappa f(a)$ holds for all $a$ in the support and $\kappa > 1$ is a known constant. Then to draw a random number from the target density function:*

1. *Generate $x$ from $f(x)$.*
2. *Generate $u$ from $U(0, 1)$.*
3. *If $u \leq \frac{g(x|\boldsymbol{\theta})}{\kappa f(x)}$ then set $a = x$ ("accept"); otherwise go back to step 1 ("reject").*

For the implementation of the algorithm we use as instrumental density function that of a gamma-distributed random variable.

# F Log-likelihood profile for combinations of parameters

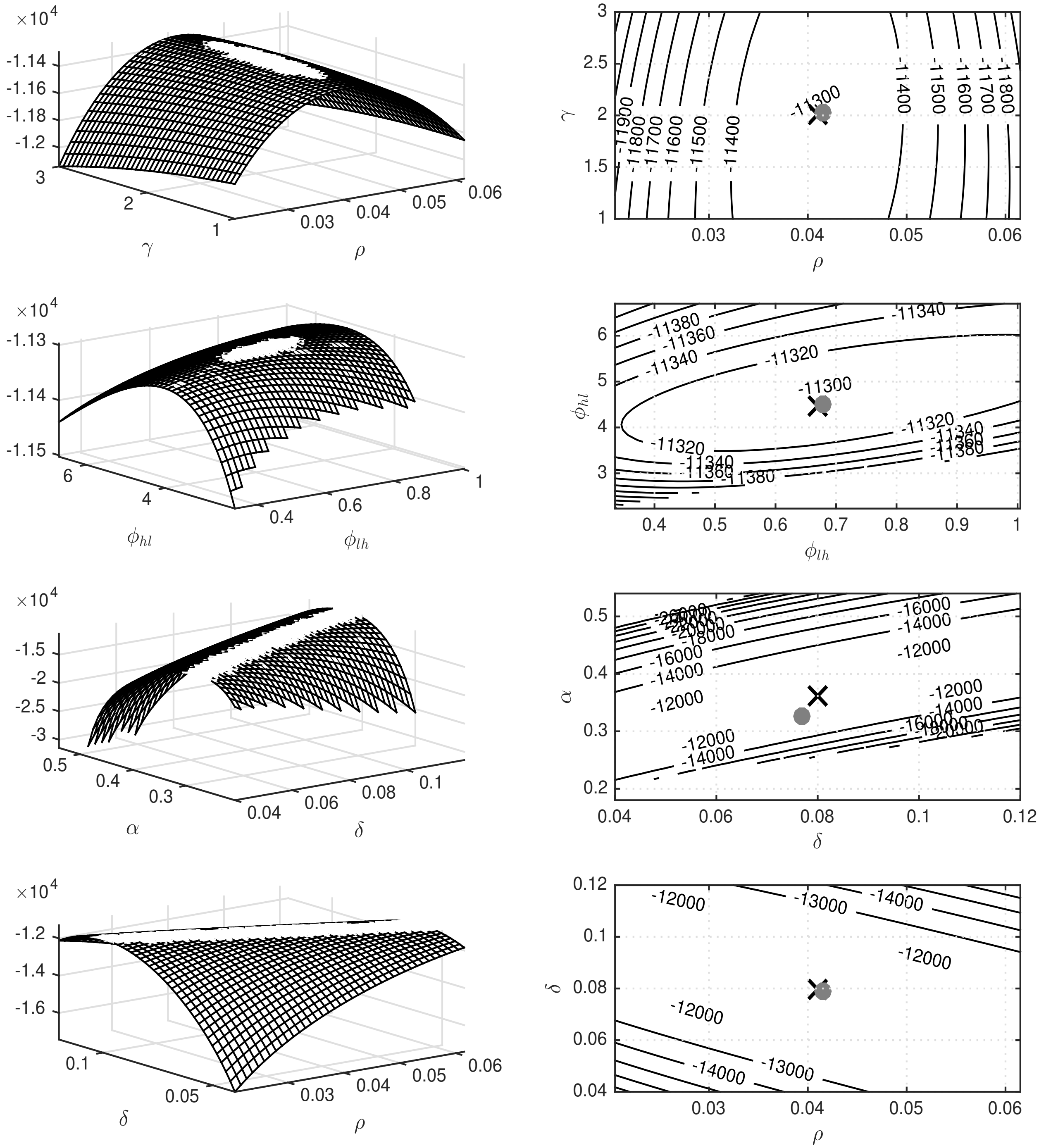


**Figure 8. Log-likelihood profile surfaces for selected parameters.** The graph shows the log-likelihood function $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{a})$ for selected parameters as a function of the parameter space on the left column using a random sample of size $N = 5000$ generated from the true model. The right column plots the associated contour plot. The population values for the structural parameters, $\boldsymbol{\theta}_0$, are given in Table 1. "×" locates the true parameter values, and ● the combination of parameters that deliver the maximum of the log-likelihood surface.

## References

Achdou, Y., F. J. Buera, J.-M. Lasry, P.-L. Lions, and B. Moll (2014a): "PDE Models in Macroeconomics," *Proceedings of Royal Society A*, 1–15.

Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2014b): "Heterogenous Agent Models in Continuous Time," Unpublished.

Aiyagari, S. R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109, 659–684.

Angeletos, G.-M. (2007): "Uninsured Idiosyncratic Investment Risk and Aggregate Saving," *Review of Economic Dynamics*, 10, 1–30.

Angeletos, G.-M. and L.-E. Calvet (2006): "Idiosyncratic production risk, growth and the business cycle," *Journal of Monetary Economics*, 53, 1095–1115.

Bayer, C. and K. Wälde (2010a): "Matching and Saving in Continuous Time: Proofs," CESifo Working Paper Series 3026-A, CESifo Group Munich.

——— (2010b): "Matching and Saving in Continuous Time: Theory," CESifo Working Paper Series 3026, CESifo Group Munich.

——— (2011): "Describing the Dynamics of Distributions in Search and Matching Models by Fokker-Planck Equations," Unpublished.

——— (2013): "The Dynamics of Distributions in Continuous-Time Stochastic Models," Unpublished.

Benhabib, J., A. Bisin, and S. Zhu (2011): "The Distribution of wealth and fiscal policy in economies with finitely lived agents," *Econometrica*, 79, 123–157.

Bewley, T. (Undated): "Interest Bearing Money and the Equilibrium Stock of Capital," Tech. rep., manuscript.

Cagetti, M. and M. D. Nardi (2006): "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, 114, 835–870.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006): "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 74, pp. 323–364.

Candler, G. (1999): *Finite-difference methods for continuous-time dynamic programming,*, in R. Marimon and A. Scott (Eds.): Computational Methods for the Study of Dynamic Economies, Oxford University Press.

Canova, F. and L. Sala (2006): "Back to square one: identification issues in DSGE models," Working Paper Series 0583, European Central Bank.

——— (2009): "Back to square one: Identification issues in DSGE models," *Journal of Monetary Economics*, 56, 431–449.

Castañeda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull (2003): "Accounting for the U.S. Earnings and Wealth Inequality," *Journal of Political Economy*, 111, 818–857.

Díaz-Giménez, J., R. Kirkby, and J. Pijoan-Mas (2014): "Flat Tax Reforms: Investment Expensing and Progressivity," Unpublished.

Flinn, C. J. (2006): "Minimum Wage Effects on Labor Market Outcomes under Search, Matching, and Endogenous Contact Rates," *Econometrica*, 74, 1013–1062.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2015): "The dynamics of inequality," Unpublished.

Heathcote, J., K. Storesletten, and G. L. Violante (2009): "Quantitative Macroeconomics with Heterogeneous Households," *Annual Review of Economics*, 1, 319–354.

Heer, B. and M. Trede (2003): "Efficiency and distribution effects of a revenue-neutral income tax reform," *Journal of Macroeconomics*, 25, 87–107.

Huggett, M. (1993): "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of Economic Dynamics and Control*, 17, 953–969.

Iskrev, N. (2010a): "Evaluating the strength of identification in DSGE models. An a priori approach," Working Papers w201032, Banco de Portugal, Economics and Research Department.

——— (2010b): "Local identification in DSGE models," *Journal of Monetary Economics*, 57, 189–202.

Kirkby, R. (2014): "Estimation of Bewley-Huggett-Aiyagari models: Theory and implementation," Unpublished.

Komunjer, I. and S. Ng (2011): "Dynamic Identification of Dynamic Stochastic General Equilibrium Models," *Econometrica*, 79, 1995–2032.

Krusell, P. and A. A. Smith (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106, 867–896.

Kydland, F. E. and E. C. Prescott (1982): "Time to build and aggregate fluctuations," *Econometrica*, 50, 1345–1370.

Lasry, J.-M. and P.-L. Lions (2007): "Mean field games," *Japanese Journal of Mathematics*, 2, 229–260.

Launov, A. and K. Wälde (2013): "Estimating Incentive And Welfare Effects Of Nonstationary Unemployment Benefits," *International Economic Review*, 54, 1159–1198.

Newey, W. K. and D. McFadden (1986): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. F. Engle and D. McFadden, Elsevier, vol. 4 of *Handbook of Econometrics*, chap. 36, 2111–2245.

Postel-Vinay, F. and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70, pp. 2295–2350.

Ríos-Rull, J. V. (1995): *Models with heterogeneous agents*, in T. Cooley (Ed.): Frontiers of Business Cycle Research, Princeton University Press.

——— (2001): *Computation of equilibria in heterogenous agent models*, in R. Marimon and A. Scott (Eds.): Computational Methods for the Study of Dynamic Economies, Oxford University. Press, 2nd ed.

Ríos-Rull, J.-V., F. Schorfheide, C. Fuentes-Albero, M. Kryshko, and R. Santaeulalia-Llopis (2012): "Methods versus substance: Measuring the effects of technology shocks," *Journal of Monetary Economics*, 59, 826–846.

Ross, S. M. (2009): *Introduction to Probability Models*, Academic Press, 10th ed.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, pp. 577–591.

Sennewald, K. and K. Wälde (2006): "Itô's Lemma and the Bellman equation for Poisson processes: An applied view," *Journal of Economics*, 89, 1–36.

Wolpin, K. I. (2013): *The Limits of Inference without Theory*, Tjalling C. Koopmans Memorial Lectures, The MIT Press.