Bayesian estimation of DSGE models with heterogeneous agents

Jerome Williams*

July 25, 2017

1 Introduction

I present a method for estimating dynamic stochastic general equilibrium (DSGE) models with heterogeneous agents using Bayesian estimation. The method combines the projection and perturbation solution method developed by Reiter (2009) with Bayesian estimation techniques, surveyed in An and Schorfheide (2007). This novel combination allows the estimation procedure to incorporate in the estimation dataset time series of moments of the cross-sectional distribution of agents. I illustrate the model by estimating a simple firm dynamics model on data simulated from the model, including a cross-sectional variable.

Bayesian estimation is frequently used in macroeconomics to estimate the parameters of dynamic stochastic general equilibrium models using aggregate time series data. (See Smets and Wouters (2007), Christiano et al. (2001), Justiniano et al. (2011), for example). A standardized set of tools has been developed (see Fernández-Villaverde et al. (2016), An and Schorfheide (2007) and Adjemian et al. (2011)).

A drawback of this literature is its limited approach to heterogeneity with respect to model agents. Researchers using Bayesian time series methods typically avoid using models in which heterogeneous agents have persistent idiosyncratic state variables, even though such models are now commonplace in macroeconomics (see Algan et al. (2014)). Rather, models of this type (heterogeneous agent DSGE) models are usually taken to data using a mixture of calibration and moment-based estimation (see Khan and Thomas (2008), Shourideh and Zetlin-Jones (2013), for example).

In this note, I combine two strands of this literature to formulate a method for Bayesian estimation of heterogeneous agent DSGE models. I use the projection and perturbation solution method, based on Reiter (2009), to cast the heterogeneous agent DSGE model in a form for which the toolbox of full-information methods DSGE as described by An and Schorfheide (2007) is applicable.

This approach has multiple advantages. Firstly, estimation, relative to calibration or quasi-calibration, allows the researcher to characterize the uncertainty with respect to the parameter estimates, and to perform hypothesis tests (see Hansen and Heckman (1996) for a discussion). Secondly, as Ruge-Murcia (2007) discusses, full-information estimation, relative to moment-based estimation, may result in increased efficiency of the estimators. Finally, and most interestingly in this case, the approach allows the researcher to use time-varying data on the distribution of agents as an empirical target.

*jmw612@nyu.edu
I am extremely grateful to Tim Cogley, Gianluca Violante and Mark Gertler for encouragement and support. I am also grateful to Jaroslav Borovička, Sydney Ludvigson, Katarína Borovičková, Gianluca Clementi, Miguel Faria-e-Castro, Keith O'Hara, James Graham, Daniel Greenwald and Joseba Martinez for comments and suggestions. This note owes a huge debt to Simon Mongey, with whom I worked on many of the ideas herein. All faults and errors are my own. An implementation of the solution and estimation method in MATLAB is available at `https://github.com/jeromematthewcelestine/hadsge`.

I illustrate the method by estimating the parameters of a simple heterogeneous firm dynamics model using simulated data, leveraging its specific advantages by incorporating into the estimation dataset a time series of distributional variable. I show that the method performs well in recovering the true parameter values.

1.1 Outline

I first present a simple model to which the method can be applied in Section 2. I describe in detail the solution method in Section 3, and the estimation procedure in Section 4.

2 A simple model

I illustrate the method for a simple firm dynamics model, a variation of the model presented in Khan and Thomas (2008). For simplicity, I make several changes to the model presented in Khan and Thomas (2008): (i) capital is the sole factor of production, (ii) there is a quadratic adjustment costs on capital, rather than a fixed cost, and (iii) there are three separate aggregate shocks in the economy.

I make these changes in order to engineer the simplest model with which I can demonstrate the usefulness of the solution and estimation method. The primary advantage of the method I present is that it can be used for models where the full distribution of agents is a state variable of the economy. I include the adjustment costs on capital to ensure that this is the case here. In the absence of adjustment costs, firms' capital choice would be independent of their current capital stock, so the current distribution of firms over capital wouldn't be required to solve for next period's distribution of firms over capital and, by extension, for next period's aggregate variables, and therefore wouldn't be a true state variable of the model.

I include three aggregate shocks in anticipation of the estimation step in Section 4. The full-information estimation method I use requires an equal number of shocks and observables. I include three observables, in particular in order to include multiple aggregate variables as well as a distributional variable, so as a result, I include three shocks in the model.

2.1 Environment

Time is discrete and the horizon is infinite. Two types of agents populate the economy: firms and households. There is a single final good, which is produced by the firms, and can be used for consumption and investment. Firms produce the final good using their installed capital and use their output to invest in capital or pay out dividends. Households own shares in the firms and value consumption of the final good, which they receive as dividends from the firms.

2.2 Firms

Production technology There is a fixed unit mass of competitive firms, indexed by $j \in [0, 1]$. Firms hold capital and are heterogeneous in their idiosyncratic productivity $z_{j,t}$. In period t, firms produce a quantity $y_{j,t}$ of the final good using their installed capital stock $k_{j,t}$ with decreasing returns to scale. The quantity produced $y_{j,t}$ is shifted by the aggregate productivity level Z_t and the firm's idiosyncratic productivity level $z_{j,t}$:

$$y_{j,t} = Z_t z_{j,t} k_{j,t}^{\alpha} \tag{1}$$

Aggregate productivity is common to all firms and follows an AR(1) process in logs:

$$\log Z_{t+1} = \rho^Z \log Z_t + \sigma^Z \varepsilon^Z_{t+1}, \qquad \varepsilon^Z_{t+1} \sim N(0,1). \tag{2}$$

Each firm's idiosyncratic productivity is also stochastic and follows an AR(1) process in logs:

$$\log z_{j,t+1} = \rho^z \log z_{j,t} + \sigma^z \varepsilon^z_{j,t+1}, \qquad \varepsilon^z_{j,t+1} \sim N(0,1). \tag{3}$$

Investment After producing in period t, firms choose next period's capital $k_{j,t+1}$. The final good can be converted into capital at a rate of one-to-one. Capital depreciates at rate δ, so in order to arrive in period $t+1$ with capital stock $k_{j,t+1}$, firms must invest an amount $i_{j,t} = k_{j,t+1} - (1-\delta) k_{j,t}$ of the final good. Capital investment is subject to quadratic adjustment costs which scale with installed capital stock: to invest an amount $x_{j,t}$, a firm with capital stock $k_{j,t}$ must pay an additional cost of $X^\phi_t \phi (x_{j,t}/k_{j,t})^2 k_{j,t}$, where ϕ is a fixed parameter and X^ϕ_t is stochastic and common to all firms. The stochastic component of the adjustment cost follows an AR(1) process in logs:

$$\log X^\phi_{t+1} = \rho^\phi \log X^\phi_t + \sigma^\phi \varepsilon^\phi_{t+1}, \qquad \varepsilon^\phi_{t+1} \sim N(0,1). \tag{4}$$

Dividends The firms are owned by the representative household, and firm output net of investment is paid out as dividends $d_{j,t}$. There is no restriction on $d_{j,t}$. If firm's j's $d_{j,t}$ is negative in period t (if investment is large relative to output, say), I interpret this as firm j raising equity finance from the households.

2.3 Households

Preferences I assume a unit measure of identical households. The households own the firms in the economy by holding shares in a mutual fund which in turn holds shares in all firms $j \in [0,1]$. The household is perfectly diversified with respect to the idiosyncratic shocks of firms. Households value consumption and choose their holdings of the mutual fund shares s_t, in order to maximize the present discounted value of utility, given by

$$\mathbb{E}_0 \left\{ \sum_{t=0}^{\infty} \left[\prod_{s=0}^{t} \left(X^\beta_{s-1} \beta \right) \right] u(C_t) \right\}, \tag{5}$$

where u is the households' per-period utility function.

I assume the household inter-period discount factor has two components, a fixed permanent component β, and a stochastic component X^β_t, which follows an AR(1) process in logs:

$$\log X^\beta_{t+1} = \rho^\beta \log X_t + \sigma^\beta \varepsilon^\beta_{t+1}, \qquad \varepsilon^\beta_{t+1} \sim N(0,1). \tag{6}$$

I include the shock to the household discount factor to introduce another source of stochastic variation in the model, in order to demonstrate the effectiveness of the likelihood-based estimation procedure.

Assets The household trades shares in a mutual fund which in turn owns the firms' shares. The quantity of mutual shares held by the household at time t is given by s_t. Shares in the mutual fund are traded in a Walrasian market, so household's take the price at time t, q_t, as given. Firms' net aggregate dividend is paid out to shareholders, and is denoted by d_t. I assume adjustment costs are paid out to households lump-sum, and I denote by Ψ_t aggregate adjustment costs.

2.4 Policies

2.4.1 Household optimization

To solve for the household's optimal policy, I express the household problem recursively. Let $\mathbf{S}$ denote the aggregate state of the economy, which is defined below. Let $W(s;\mathbf{S})$ denote the household's expected present discount value, given share holdings s and aggregate state $\mathbf{S}$. Then $W(s;\mathbf{S})$ satisfies the Bellman equation

$$\begin{aligned} W(s;\mathbf{S}) &= \max_{C,s'} u(C) + X^{\beta}\beta E\left[W(s';\mathbf{S}')\right] && (7) \\ \text{subject to} & \\ q(\mathbf{S})s' + C &= (q(\mathbf{S}) + d(\mathbf{S}))s + \Psi(\mathbf{S}). && (8) \end{aligned}$$

The solution to this problem gives the household's Euler equation

$$\beta\mathbb{E}\left[\frac{u'(C')}{u'(C)}\frac{q(\mathbf{S}') + d(\mathbf{S})}{q(\mathbf{S})}\right] = 0 \tag{9}$$

which delivers the household's discount factor

$$M(\mathbf{S},\mathbf{S}') = X^{\beta}\beta\frac{u'(C'(\mathbf{S}))}{u'(C(\mathbf{S}))}. \tag{10}$$

In equilibrium, the representative household must hold all shares, so $s = 1$. Since the firms are owned by households, firms discount the future according to the household's stochastic discount factor $M(\mathbf{S},\mathbf{S}')$.

2.4.2 Firm optimization

Let $V(k,z;\mathbf{S})$ be the expected present discounted value of the firm, as valued by the households' stochastic discount factor M, given current idiosyncratic state (k,z) and aggregate state $\mathbf{S}$. Then V solves the Bellman equation:

$$\begin{aligned} V(k,z;\mathbf{S}) &= \max_{k'\geq 0}\left\{y(k,z;\mathbf{S}) - i(k',k) - \varphi(k',k;\mathbf{S}) + \mathbb{E}\left[M(\mathbf{S},\mathbf{S}')V(k',z';\mathbf{S}')\right]\right\} && (11) \\ \text{subject to} & \\ y(k,z;\mathbf{S}) &= X^{Z}zk^{\alpha} && (12) \\ i(k',k) &= k' - (1-\delta)k && (13) \\ \varphi(k',k;\mathbf{S}) &= X^{\phi}\phi\left(\frac{k'-(1-\delta)k}{k}\right)^{2}k && (14) \end{aligned}$$

where $M(\mathbf{S},\mathbf{S}')$ is the household's discount factor. The expectation in equation (11) is taken with respect to z' and any stochastic components of $\mathbf{S}'$.

Anticipating the solution method, I define the "price function" $p(\mathbf{S})$ to be the household's marginal utility of consumption given aggregate state $\mathbf{S}$,

$$\mathrm{p}(\mathbf{S}) \equiv u'(\mathbf{S}),$$

and express the household's stochastic discount factor $M(\mathbf{S},\mathbf{S}')$ in terms of $p(\mathbf{S})$ and $p(\mathbf{S}')$:

$$M(\mathbf{S},\mathbf{S}') = X^{\beta}\beta\frac{p(\mathbf{S}')}{p(\mathbf{S})}.$$

The firm's Bellman equation can now be expressed in terms of p rather than M:

$$\begin{aligned} V(k,z;\mathbf{S}) &= \max_{k'\geq 0} y(k,z;\mathbf{S}) - i(k',k) - \varphi(k',k;\mathbf{S}) + X^{\beta}\beta\mathbb{E}\left[\frac{p(\mathbf{S}')}{p(\mathbf{S})}V(k',z';\mathbf{S}')\right] \\ & \quad \text{subject to (12), (13) and (14).} \end{aligned} \tag{15}$$

Let $v(k,z;\mathbf{S}) \equiv p(\mathbf{S})V(k,z;\mathbf{S})$ denote the firm's value function scaled by marginal utility. From (15), it is clear that the function v solves a modified Bellman equation:

$$\begin{aligned} v(k,z;\mathbf{S}) &= \max_{k'\geq 0} p(\mathbf{S})\left[y(k,z;\mathbf{S}) - i(k',k) - \varphi(k',k;\mathbf{S})\right] + \beta X^{\beta}\mathbb{E}\left[v(k',z';\mathbf{S}')\right] \\ & \quad \text{subject to (12), (13) and (14).} \end{aligned} \tag{16}$$

This transformation, based on Khan and Thomas (2008), replaces the dependence of the Bellman equation on the stochastic discount factor $M(\mathbf{S},\mathbf{S}')$ with dependence on the quantity $p(\mathbf{S})$. Since the value function $V(k,z;\mathbf{S})$ is itself of no particular interest, I henceforth work exclusively with the scaled value function $v(k,z;\mathbf{S})$ and its corresponding Bellman equation (16). The transformation is purely for convenience when solving the model. Note that, after solving for $v(k,z;\mathbf{S})$, the value function $V(k,z;\mathbf{S})$ can easily be retrieved if necessary.

2.4.3 First order condition

Let $k^*(k,z;\mathbf{S})$ denote the policy function representing the optimal choice of k', given idiosyncratic state (k,z) and aggregate state $\mathbf{S}$. Given the firm's problem (16), it is straightforward to show that $k^*(k,z;\mathbf{S})$ must satisfy the firm's first-order condition:

$$\begin{aligned} & p(\mathbf{S})\left(1+2\phi\frac{k^*(k,z)}{k}+1-\delta\right) = \beta X^{\beta} \\ & \mathbb{E}\left[p(\mathbf{S}')\int\left\{X^{Z\prime}z'\alpha(k^*)^{\alpha-1}+1-\delta+X^{\phi\prime}\phi\left(\left(\frac{k^*(k^*(k,z),z')}{k^*(k,z)}\right)^2-(1-\delta)^2\right)\right\}d\varepsilon'\right] \end{aligned} \tag{17}$$

where $k'(k,z)$ indicates $k'(k,z;\mathbf{S})$, p indicates $p(\mathbf{S})$, p' indicates $p(\mathbf{S}')$, $k''(k',z)$ indicates $k^*(k^*(k,z;\mathbf{S});\mathbf{S}')$, $(X^Z)'$ indicates $X^Z(\mathbf{S})$, and z' indicates $\rho^z\log z+\sigma\varepsilon'$ and the expectation is with respect to $(X^{Z\prime}, X^{\beta\prime}, X^{\phi\prime})$. Therefore, we can characterize the equilibrium using this first-order condition, rather than the Bellman equation.

2.4.4 Aggregate state

Let $\lambda_t = \lambda_t(k,z)$ denote the distribution of firms over capital k and productivity z in period t. It is clear from Section 2.4.2 that the firms' policy at time t is a function of the quantity p_t, which is (through the resource constraint $Y = I+C$) in turn a function of the the distribution of firms over capital and productivity at time t:

$$\begin{aligned}
p_t &= u'(C_t) \\
C_t &= Y_t - I_t \\
&= \int y_{j,t} d\lambda(k_{j,t}, z_{j,t}) - \int [k_t^*(k_{j,t}, z_{j,t}) - (1-\delta) k_{j,t}] d\lambda(k_{j,t}, z_{j,t})
\end{aligned}$$

Thus it is clear that the distribution of firms over capital and productivity, $\lambda_t(k,z)$ is an aggregate state of the model. Let $\mathbf{S}_t = (\lambda_t, \mathbf{X}_t)$ denote the aggregate state, where $\mathbf{X}_t$ denotes the aggregate shocks $\mathbf{X}_t = \left(X_t^Z, X_t^\beta, X_t^\phi\right)$.

2.5 Equilibrium

A recursive competitive equilibrium of this model consists of the firms' policy function $k^*(k, z; \mathbf{S})$, a household consumption function $C(\mathbf{S})$, a distribution $\lambda(k, z; \mathbf{S})$, a price function $p(\mathbf{S})$ and a law of motion for the aggregate state $\mathbf{S}' = (\mathbf{X}', \mu')$, such that

1. *(Firm optimization) Taking $p(\mathbf{S})$ and $\mathbf{S}'(Z'; \mathbf{S})$ as given, $k^*(k, z; \mathbf{S})$ solves the firm's optimization problem (17).*

2. *(Household optimization) The function $C(\mathbf{S})$ solves the household's problem (7) and $p(\mathbf{S})$ is the corresponding price function:*
$$p(\mathbf{S}) = u'(C(\mathbf{S})).$$

3. *(Law of motion for distribution) The law of motion for the distribution Ψ is generated by the stochastic process for idiosyncratic productivity and the firm policy function $k^*(k, z; \mathbf{S})$, i.e. for all measurable sets $\mathcal{K} \times \mathcal{Z}$,*
$$\lambda'(\mathcal{K} \times \mathcal{Z}) = \Psi_{(\mathcal{K} \times \mathcal{Z})}(\mathbf{S}) = \int_{\mathcal{K} \times \mathcal{Z}} Q((k,z), \mathcal{K} \times \mathcal{Z}; \mathbf{S}) d\lambda(k,z)$$
where Q is the transition function when the aggregate state is $\mathbf{S}$ and is defined by
$$Q((k,z), \mathcal{K} \times \mathcal{Z}; \mathbf{S}) = \mathbf{1}\{k^*(k,z;\mathbf{S}) \in \mathcal{K}\} \sum_{z' \in \mathcal{Z}} P(z, z')$$
where $\mathbf{1}$ is the indicator function and $P(z, z')$ is the Markov transition probability generated by the stochastic process for idiosyncratic productivity.

4. *(Law of motion for aggregate state) The law of motion for the aggregate state $\mathbf{S}'$ is consistent with Ψ and the stochastic processes for the aggregate shocks (2), (4) and (6).*

3 Solution method

I solve for the equilibrium in the three steps:

1. Define a discretized version of the model, choosing finite-dimensional approximation schemes for the infinite-dimensional policy function and the infinite-dimensional distribution of firms, and re-defining the equilibrium conditions in terms of these finite-dimensional approximations.

2. Solve for the aggregate steady state of the discretized model (the steady state of the model if aggregate shocks are shut off).

3. Solve for the approximate dynamics of the discretized model by computing a first-order perturbation of the discretized models around its aggregate steady state in all variables.

Each of these three steps bears some discussion.

3.1 Discretized model

Defining a discretized version of the model entails choosing an approximation scheme for the value functions and/or policy functions as well the infinite-dimensional distribution of firms. The researcher has considerable leeway in this step, since many potential discretization schemes exist.

As mentioned in Section 2.4.3, for this model the first-order condition as a functional equation of the firm's policy function is sufficient as an equilibrium condition. In other applications, it may not be possible to characterize the equilibrium in this way; it is often necessary to include the firm's Bellman equation (a functional equation of the firm's *value* function) as an equilibrium condition. Whether it is the value function or policy function that is being discretized, there are many finite approximation schemes available to the researcher. In what follows I approximate the policy function using cubic splines. See Miranda and Fackler (2002) and Judd (1998) for discussions of other options and their relative advantages and disadvantages.

3.1.1 Policy

I approximate the policy function $k^*(k,z;\mathbf{S})$ using a functional approximation scheme in the idiosyncratic states k and z, for a given level of $\mathbf{S}$ (dependence on $\mathbf{S}$ will come through the perturbation discussed in Section 3.3).

I first construct a discretization of the idiosyncratic state space. The true state space is $\mathbb{R}^2_+$, but I constrain the firms' capital to lie on the interval $\left[0,\bar{k}\right]$ and productivity to lie in the interval $[\underline{z},\bar{z}]$. I check *ex post* that it is never optimal for a firm to choose a capital stock $k' > \bar{k}$ and I choose a discretization of the idiosyncratic productivity process such that it lies in the interval $[\underline{z},\bar{z}]$. I define a set of n_k discrete points $X_k \equiv \{k_i | i = 1,\ldots,n_k, k_i < k_{i+1}\}$ such that $k_1 = 0$ and $k_{n_k} = \bar{k}$.

Similarly for idiosyncratic productivity, I choose an upper bound $\bar{z}$ and lower bound $\underline{z}$, and define a discrete set of n_z points X_z on the interval $[\underline{z},\bar{z}]$. I construct the Markov transition matrix $P(z,z')$ to approximate the stochastic process $\log z' = \rho_z \log z_t + \sigma_z \varepsilon'$, using the method described in Rouwenhorst (1995).

As noted by Miranda and Fackler (2002), it is typically convenient to work with function approximants which can be expressed as a linear combination of a set of known linearly independent basis functions. For this illustration, I approximate the firms' policy function using a cubic B-spline, with $n \equiv n_k \times n_z$ interpolation nodes, i.e. I approximate $k^*(k,z;\mathbf{S})$ by $\hat{k}(k,z;\mathbf{S})$, where

$$\hat{k}(k,z;\mathbf{S}) = \sum_{j=1}^{n} \phi_j(k,z)\, c_j(\mathbf{S})$$

where $c_j(\mathbf{S})$ are the approximating coefficients, given aggregate state $\mathbf{S}$. The definition of ϕ_j is standard for cubic B-splines.[1] Let $\Phi(k,z)$ denote the $n \times 1$ vector of basis functions $[\phi_1(k,z),\ldots,\phi_n(k,z)]$ evaluated at

[1] See Judd (1998) for details.

(k, z), and let $\mathbf{c}(\mathbf{S})$ denote the $n \times 1$ vector of approximating coefficients $\mathbf{c}(\mathbf{S}) = [c_1(\mathbf{S}), \ldots, c_n(\mathbf{S})]$:

$$\hat{k}(k, z; \mathbf{S}) = \Phi(k, z)\, \mathbf{c}(\mathbf{S}).$$

Given a choice of approximant, the next step is to solve for the basis coefficients $\{c_j\}_{j=1}^{n}$ that best approximate $k^*(k, z; \mathbf{S})$.

I do this using the collocation method, that is, by requiring the discretized version of the firm's first-order condition to hold on the n interpolation nodes $\{(k_i, z_j)\,|k_i \in X_k, z_j \in X_z\}$. For a given value of the aggregate state $\mathbf{S}$, solving for the basis coefficient vector $\mathbf{c}(\mathbf{S})$ using the collocation method entails solving for the elements of $\mathbf{c}(\mathbf{S})$ such that the approximate first-order condition holds for all $k_i \in X_k$ and for all $z_j \in X_z$:

$$\begin{aligned}
& p(\mathbf{S}) \left(1 + 2\phi \left(\frac{\Phi(k_i, z_j)\, \mathbf{c}(\mathbf{S})}{k_i} + (1-\delta)\right)\right) \\
= \; & \beta X^{\beta}(\mathbf{S})\, \mathbb{E}\left\{ p(\mathbf{S}') \sum_{z' \in Z} P(z_j, z') \left\{ X^{Z}(\mathbf{S}')\, z' \alpha \left[\Phi(k_i, z_j)\, \mathbf{c}(\mathbf{S})\right]^{\alpha-1} + 1 - \delta \right\} \right\} \\
& + \beta X^{\beta}(\mathbf{S})\, \mathbb{E}\left\{ p(\mathbf{S}') \sum_{z' \in Z} P(z_j, z') \left\{ X^{\phi}(\mathbf{S}')\, \phi \left[\left(\frac{\Phi(\Phi(k_i, z_j)\, \mathbf{c}(\mathbf{S}), z')\, \mathbf{c}(\mathbf{S}')}{\Phi(k_i, z_j)\, \mathbf{c}(\mathbf{S})} \right)^2 - (1-\delta)^2 \right] \right\} \right\}
\end{aligned}$$

I denote this equilibrium condition $g(k, z, \mathbf{c}; \mathbf{S}) = 0$, $\forall k \in X_k$, $z \in X_z$.

3.1.2 Distribution of firms over capital and productivity

I approximate the distribution of firms using a two-dimensional histogram defined over $\left[0, \bar{k}\right]$ and $[\underline{z}, \bar{z}]$. I approximate the distribution by assuming all mass in the distribution lies on discrete points in this space. For simplicity, I choose for this purpose the same set of points X that I do for the policy function approximation, although in general these need not be the same.

I approximate the law of motion of the distribution using the method described by Young (2010). If a mass m of firms choose k' as their capital stock next period, where $k_i < k' < k_{i+1}$ for some $k_i \in X_k$, then the mass is split between gridpoints k_i and k_{i-1} according to the relative distance of k' from each neighboring gridpoint: a fraction ω_i of the mass m gets assigned to gridpoint k_i and and the remaining $1 - \omega_i$ gets assigned to gridpoint k_{i+1}, where ω_i is given by

$$\omega_i \equiv \frac{k_{i+1} - k'}{k_{i+1} - k_i}.$$

As a result, the discretized law of motion for the distribution takes the following form. Let $\mu(k_i, z_j; \mathbf{S})$ indicate the density associated with point (k_i, z_j) in a period when the aggregate state is $\mathbf{S}$ and let $\mu'(k_{i'}, z_{j'}; \mathbf{S}')$ indicate the density associated with point $(k_{i'}, z_{j'})$ in the following period. Then the law of motion for the distribution can be expressed as:

$$
\begin{aligned}
\mu'\left(k_{i'}, z_{j'}; \mathbf{S}'\right) &= \sum_{k_i \in X_k} \sum_{z_j \in X_z} Q\left(k_{i'}, z_{j'}, k_i, z_j, \mathbf{c}; \mathbf{S}\right) \mu\left(k_i, z_j; \mathbf{S}\right) \\
Q_t\left(k_{i'}, z_{j'}, k_i, z_j, \mathbf{c}; \mathbf{S}\right) &= P\left(z_{j'}, z_j\right) \times \begin{cases} \frac{\Phi(k_i, z_j)\mathbf{c}(\mathbf{S}) - k_{i'-1}}{k_{i'} - k_{i'-1}} & \text{if } \Phi\left(k_i, z_j\right) \mathbf{c}\left(\mathbf{S}\right) \in \left[k_{i'-1}, k_{i'}\right], \\ \frac{k_{i'+1} - \Phi(k_i, z_j)\mathbf{c}(\mathbf{S})}{k_{i'+1} - k_{i'}} & \text{if } \Phi\left(k_i, z_j\right) \mathbf{c}\left(\mathbf{S}\right) \in \left[k_{i'}, k_{i'+1}\right], \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{18}
$$

3.1.3 Prices

The household's optimality condition determines that $p\left(\mathbf{S}\right) = u'\left(C\left(\mathbf{S}\right)\right)$, where $C\left(\mathbf{S}\right)$ is the consumption of the representative household, given aggregate state $\mathbf{S}$, and can be computed from the economy's resource constraint $C = Y - I$, where Y is aggregate output and I is aggregate net investment. Both Y and I can be computed by integrating over the distribution of firms, giving the following expression for aggregate consumption:

$$
\begin{aligned}
C\left(\mathbf{S}\right) &= Y\left(\mathbf{S}\right) - I\left(\mathbf{S}\right) \\
&= \sum_{i=1}^{n_k} \sum_{j=1}^{n_z} X^Z\left(\mathbf{S}\right) z_j k_i^{\alpha} \mu\left(k_i, z_j; \mathbf{S}\right) - \sum_{i=1}^{n_k} \sum_{j=1}^{n_z} \left[\Phi\left(k_i, z_j\right) \mathbf{c}\left(\mathbf{S}\right) - (1-\delta) k_i\right] \mu\left(k_i, z_j; \mathbf{S}\right)
\end{aligned}
$$

3.1.4 Approximate equilibrium conditions

I summarize here the full set of equilibrium conditions in terms of the finite representation of the model state. An equilibrium of the discretized model is a set of approximating coefficients for the policy function $\mathbf{c}\left(\mathbf{S}\right) = \left\{c^j\left(\mathbf{S}\right)\right\}_{j=1}^{n}$, a set of approximating coefficients for the distribution $\Lambda\left(\mathbf{S}\right) = \left\{\lambda\left(k_i, z_j; \mathbf{S}\right)\right\}_{k_i \in X_k, z_j \in X_z}$, a price function $p\left(\mathbf{S}\right)$ and a law of motion for the approximate distribution, such that:

1. (Policy function). For all $k_i \in X_k$ and $z_j \in X_z$,

$$
g\left(k_i, z_j, \mathbf{c}; p, \mathbf{S}\right) = 0 \tag{19}
$$

2. (Law of motion). For all $k' \in X_k$ and $z' \in X_z$,

$$
\lambda\left(k', z'; \mathbf{S}'\right) = \sum_{k_i \in X_k} \sum_{z_j \in X_z} Q\left(k', z', k_i, z_j, \mathbf{c}\right) \lambda\left(k_i, z_j; \mathbf{S}\right) \tag{20}
$$

   where Q is defined in equation (18) above.

3. (Stochastic discount factor.)

$$
p\left(\mathbf{S}\right) = u'\left(C\left(\mathbf{S}\right)\right)
$$

where

$$C(\mathbf{S}) = \sum_{k_i \in X_k} \sum_{z_j \in X_z} \left(X^Z(\mathbf{S}) z_j k_i^\alpha - \left[\Phi(k_i, z_j) \mathbf{c}(\mathbf{S}) - (1-\delta) k_i \right] \right) \lambda(k_i, z_j; \mathbf{S}) \tag{21}$$

3.2 Aggregate steady state of the discretized model

Given the set of equilibrium conditions of the discretized model defined in equations (19), (20) and (21) above, the next stop is to solve for the aggregate steady state of the model, which is the stationary solution of equations when aggregate shocks are identically zero. It is straightforward to see that the aggregate steady state consists of a policy function $\bar{k}(k,z)$, a distribution $\bar{\lambda}(k,z)$ and a price $\bar{p}$, such that conditions (19), (20) and (21) are satisfied when $p(\bar{\mathbf{S}}) = \bar{p}$, when $\lambda(k, z; \bar{\mathbf{S}}) = \bar{\lambda}(k, z; \bar{\mathbf{S}})$, where $\bar{\mathbf{S}} = (\bar{\lambda}, \bar{\mathbf{X}})$ and $\bar{\mathbf{X}} = (1,1,1)$.

Solving for the steady state involves solving a one-dimensional fixed point problem in $\bar{p}$. The algorithm is as follows:

1. Guess the price $\bar{p}^*$, solve for the corresponding policy function $\bar{k}^*$ and stationary distribution $\bar{\lambda}^*$, and resulting aggregate consumption $\bar{C}^*$
2. solve for the value of $\bar{p}^{**}$ which is consistent with $\bar{C}^*$
3. If $\bar{p}^{**} \neq \bar{p}^*$, I modify the guess and redo the procedure.

This is a fixed point problem in $\bar{p}$, which can be solved using bisection. For each guess of $\bar{p}^*$, solving for aggregate consumption $\bar{C}^*$ involves solving for the policy function of the firms. I do this using the collocation method described in Section 3.1.1.

3.3 Dynamics of the discretized model

Having solved the approximate model in steady state, I now proceed to solve the dynamics of the model. Solving the dynamic model involves computing a first-order perturbation of the model equations around their values at the steady state in each variable, in order to cast the model as a first-order difference equation in t. I then use standard methods to obtain a solution in of the model in linear state-space form, $\xi_t = A\xi + B\mathcal{E}_t$, where ξ_t is a vector of model state variables. In what follows, I switch from the recursive formulation presented above to a representation with explicit time dependence. Note that policy functions etc. still depend on the aggregate state through their dependence on time t.

3.3.1 System of equations

Before taking the first-order perturbation, I make some modifications to the system of equations to conform more closely with the canonical form described in Sims (2002), which deals with a model of the form

$$\Gamma_0 \xi_t = \Gamma_1 \xi_{t-1} + \Psi \mathcal{E}_t + \Pi \eta_t$$

where ξ_t is a vector of state variables, $\mathcal{E}_t$ is a vector of exogenous Gaussian disturbances and η_t is a vector of expectational errors. To cast the system in this form, I expand the set of state variables to include variables representing expected values at time t of the "jump" variables at time $t+1$, $\left\{c_{t+1}^j\right\}_{j=1}^n$ and p_t, denoted $\left\{E_t c_{t+1}^j\right\}_{j=1}^n$ and $E_t p_{t+1}$, respectively, and include in the system definitional equations for these new variables:

$$E_{t-1} c_{t+1}^j = c_t^j + \eta_t^j, \quad \forall j \in \{1, \ldots, n\},$$

and

$$E_{t-1}p_t = p_t + \eta_t^p,$$

where the η_t are mean-zero expectational errors. Substituting in these new expectational variables, the first order conditions in (19) become, for all $k_i \in X_k$ and $z_j \in X_z$:

$$\begin{aligned} & p_t\left(1+2\phi\left(\frac{\Phi\left(k_i,z_j\right)\mathbf{c}_t}{k_i}+(1-\delta)\right)\right) \\ = & \beta X_t^\beta\left\{E_t p_{t+1}\sum_{z'\in Z}P\left(z_j,z'\right)\left\{E_t X_{t+1}^Z z'\alpha\left[\Phi\left(k_i,z_j\right)\mathbf{c}_t\right]^{\alpha-1}+1-\delta\right\}\right\} \\ & +\beta X_t^\beta\left\{E_t p_{t+1}\sum_{z'\in Z}P\left(z_j,z'\right)\left\{E_t X_{t+1}^\phi\phi\left[\left(\frac{\Phi\left(\Phi\left(k_i,z_j\right)\mathbf{c}_t,z'\right)E_t\mathbf{c}_{t+1}}{\Phi\left(k_i,z_j\right)\mathbf{c}_t}\right)^2-(1-\delta)^2\right]\right\}\right\} \end{aligned} \tag{22}$$

Note that the expectation of the aggregate shocks do not need to be added to the state vector, since they are all $E_t X_{t+1}^Z$ is straightforwardly a function of X_t^Z:

$$E_t \log X_{t+1}^Z = \rho \log X_t^Z.$$

I collect the variables of the discretized model into a "state" vector ξ_t. The state vector ξ_t is an $m\times 1$ vector given by

$$\xi_t = \left[\left\{c_t^j\right\}_{j=1}^n, \left\{E_t c_{t+1}^j\right\}_{j=1}^n, \left\{\mu_t\left(k_i,z_j\right)\right\}_{k_i\in X_k, z_j\in X_z}, p_t, E_t p_{t+1}, \log\mathbf{X}_t\right]'$$

where $m = 3n+5$.

I collect the exogenous innovations to $\log\mathbf{X}_t$ into a shock vector $\mathcal{E}_t$, where $\mathcal{E}_t = \left[\varepsilon_t^Z, \varepsilon_t^\beta, \varepsilon_t^\phi\right]'$ is a 3×1 vector of exogenous disturbances distributed according to $\mathcal{E}_t \sim N\left(\mathbf{0},\mathbf{I}_3\right)$. I collect the expectational errors $\{\eta_t^v\}$ into a $n_\eta \times 1$ vector η_t. Denote by F the system of discrete equilibrium conditions given by (22), (20) and (21), then F can be expressed as a function of ξ_t, ξ_{t-1}, $\mathcal{E}_t$ and η_t:

$$F\left(\xi_t,\xi_{t-1},\mathcal{E}_t,\eta_t\right) = \mathbf{0}.$$

3.3.2 Perturbation

Section 3.3.1 showed how the equilibrium of the discrete model can be expressed as a system of equations F of a finite state vector ξ_t, its lag ξ_{t-1}, exogenous innovations $\mathcal{E}_t$ and expectational errors η_t. The next step is to construct a linear approximation to the system F. I take a first-order Taylor series approximation of $F\left(\xi_t,\xi_{t-1},\mathcal{E}_t,\eta_t\right)=\mathbf{0}$ around the deterministic steady state $\left(\bar\xi,\bar\xi,\mathbf{0},\mathbf{0}\right)$, the value of which is solved for nonlinearly in Section 3.2. The first-order approximation can be written:

$$\begin{aligned} 0 = & F\left(\bar\xi,\bar\xi,\mathbf{0},\mathbf{0}\right)+F_1\left(\bar\xi,\bar\xi,\mathbf{0},\mathbf{0}\right)\left(\xi_t-\bar\xi\right)+F_2\left(\bar\xi,\bar\xi,\mathbf{0},\mathbf{0}\right)\left(\xi_{t-1}-\bar\xi\right) \\ & +F_3\left(\bar\xi,\bar\xi,\mathbf{0},\mathbf{0}\right)\mathcal{E}_t+F_4\left(\bar\xi,\bar\xi,\mathbf{0},\mathbf{0}\right)\eta_t \end{aligned}$$

where F_j, $j=1,2,3,4$ denotes the Jacobian of F with respect to its jth argument. This becomes

$$0 = F_1\left(\bar\xi,\bar\xi,\mathbf{0},\mathbf{0}\right)\hat\xi_t + F_2\left(\bar\xi,\bar\xi,\mathbf{0},\mathbf{0}\right)\hat\xi_{t-1} + F_3\left(\bar\xi,\bar\xi,0,0\right)\mathcal{E}_t + F_4\left(\bar\xi,\bar\xi,0,0\right)\eta_t \tag{23}$$

where $\hat{\xi}_t = \xi_t - \bar{\xi}$ indicates deviation from steady state. We can write this as

$$\Gamma_0 \hat{\xi}_t = \Gamma_1 \hat{\xi}_{t-1} + \Psi \mathcal{E}_t + \Pi \eta_t$$

where $\Gamma_0 \equiv -F_1\left(\bar{\xi}, \bar{\xi}, 0, 0\right)$, $\Gamma_1 = F_2\left(\bar{\xi}, \bar{\xi}, 0, 0\right)$, $\Psi = F_3\left(\bar{\xi}, \bar{\xi}, 0, 0\right)$ and $\Pi = F_4\left(\bar{\xi}, \bar{\xi}, 0, 0\right)$. Equations of this form are common in macroeconomics, and many methods for solving them exist.[2] One such method is described by Sims (2002), which I use here. Solving this equation means obtaining a linear expression for the state $\hat{\xi}_t$ as a function of last period's state $\hat{\xi}_{t-1}$ and the vector of exogenous disturbances $\mathcal{E}_t$:

$$\hat{\xi}_t = A \hat{\xi}_{t-1} + B \mathcal{E}_t. \tag{24}$$

3.3.3 Derivatives

There are several methods for computing the derivatives involved in equation (23) above. In some cases it may be possible, to compute the derivatives analytically. However, when this is not possible, the derivatives may be computed numerically using automatic differentiation or other methods. Reiter (2009) computes these derivatives using first-differencing and Winberry (2016) makes use of MATLAB's symbolic differentiation toolbox. Automatic differentiation is superior to both of these methods because it can efficiently compute exact derivatives (to within machine precision).[3]

3.3.4 Additional state variables

In order to facilitate estimation, I also include in the state vector ξ_t some additional variables, denoted by $\mathbf{m}_t$, which are functions of the aggregate state and the firms' policy function. In particular, I include aggregate output Y_t and aggregate investment I_t, defined as

$$Y_t = X_t^Z \sum_{k_i \in X_k} \sum_{z_j \in X_z} x_j k_j^{\alpha} \mu_{t-1}\left(k_i, z_j\right)$$

and

$$I_t = \sum_{k_i \in X_k} \sum_{z_j \in X_z} \left\{\Phi\left(k_i, z_j\right) \mathbf{c} - (1-\delta) k_i\right\} \mu_{t-1}\left(k_i, z_j\right)$$

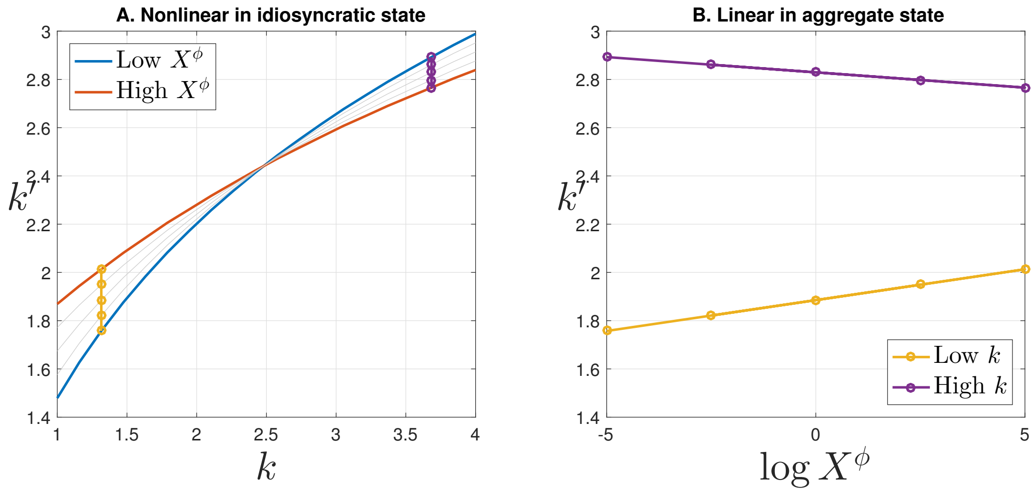
These additional variables are not required for the model solution, but are used in the estimation step, so are included here for simplicity.

An advantage of likelihood-based estimation of heterogeneous agent models in particular, is that we can use as observable time series in the estimation any function of the cross-sectional distribution of agents. Since the solution method entails incorporating the cross-sectional distribution of agents into the latent state variable, it is natural to include elements of that distribution, or linear functions of elements of the distribution, as observables. In order to illustrate this, I include as an observable in the estimation procedure a variable which corresponds to total investment by small firms, which I define as

$$I_t^{\text{small}} = \sum_{k_i \in X_k} \sum_{z_j \in X_k} \left\{\Phi\left(k_i, z_j\right) \mathbf{c} - (1-\delta) k_i\right\} \mu_{t-1}\left(k_i, z_j\right) \mathbf{1}\left\{k_i \leq k^*\right\}$$

[2]See Fernández-Villaverde et al. (2016) for a discussion.
[3]See Judd (1998) for a discussion.

Figure 1: The firm's policy function.



Panel A plots the policy function of a firm with median idiosyncratic productivity ($z = 1$) against the firm's capital stock k, for differing values of the aggregate shock to adjustment costs X^ϕ. "Low X^ϕ" corresponds to a value of $\log X^\phi = -5$ and "High X^ϕ" corresponds to $\log X^\phi = 5$. Panel B plots the policy function of a firm with median idiosyncratic productivity ($z = 1$) against the log of the adjustment cost shock X^ϕ, for differing values of capital. Note that the values for $\log X^\phi$ considered here are more extreme than is typically seen given the "true" parameterization, and are used here only for illustrative purposes.

where k^* is the median value of k in steady state. As with the aggregates Y_t and I_t, I include I_t^{small} in the state vector ξ_t at this step.

3.3.5 Form of the solution

Because the dynamics of the model are solved for using the linearization technique described in Section 3.3, the firm's policy function, along with the other equilibrium objects, is linear in the aggregate state. Due to the projection step for solving the idiosyncratic problem in steady state, the firm's policy function is nonlinear in the firm's own idiosyncratic state. I demonstrate how these two characteristics interact, in Figure 1, where I plot the policy of a firm against one dimension of the firm's idiosyncratic state (current period capital k) and against one dimension of the economy's aggregate state (the level of the aggregate shock to adjustment costs X^ϕ).

In Panel A of Figure 1, I plot the policy of a firm with the median idiosyncratic productivity $\bar{z}$ against the firm's current capital stock, for different values of the aggregate adjustment cost shock X^ϕ. Panel A shows how the firm's policy is nonlinear in k, for a given level of the aggregate state.

In Panel B of Figure 1, I plot the policy of a firm with the median idiosyncratic productivity $\bar{z}$ against the log of the aggregate adjustment cost shock X^ϕ, for different levels of capital k. Panel B demonstrates the linearity of the policy with respect to the aggregate state, for a given level of the idiosyncratic state.

Note that the policy is linear with respect to the aggregate state only for a given level of the idiosyncratic state. The slope of the "low k" line is different from the slope of the high k line: this demonstrates how the slope of the policy function with respect to the aggregate state differs for different values of the idiosyncratic state. This last point bears repeating: linearization with respect to aggregate states does not mean that fluctuations in the aggregate state merely add a component to the firm's nonlinear steady-state policy. Rather, the entire nonlinear policy function varies linearly with the aggregate state.

4 Estimation

4.1 Method

The Bayesian estimation procedure consists of evaluating, for every parameter vector θ, the likelihood of the observed data, given the parameterized model, and combining that with the prior distribution of the parameters. In order to proceed, I augment equation (24), which describes the economic model, with a measurement equation, which describes how the observed data series $\{Y_t\}_{t=1}^{T}$ (where Y_t is potentially vector-valued) relate to the variables of the economic model. When the measurement equation is of the form

$$Y_t = C\hat{\xi}_t + D\zeta_t$$

where ζ_t is an $n_\zeta \times 1$ vector of normally distributed innovations, then the combined equations

$$\begin{aligned} \hat{\xi}_t &= A(\theta)\hat{\xi}_{t-1} + B(\theta)\mathcal{E}_t && (25) \\ Y_t &= C\hat{\xi}_t + D\zeta_t && (26) \end{aligned}$$

is a linear Gaussian state-space representation of the model, and the likelihood of a dataset $\{Y_t\}_{t=1}^{T}$ can be computed recursively using the Kalman filter (see An and Schorfheide (2007) a full description of the Kalman filter).

For a given parameter vector θ, let $p\left(\{Y_t\}_{t=1}^{T}|\theta\right)$ denote the likelihood of the observed data. Then the posterior can be computed by combining the likelihood with the prior, to give

$$p\left(\theta|\{Y_t\}_{t=1}^{T}\right) \propto p\left(\{Y_t\}_{t=1}^{T}|\theta\right)p(\theta)$$

Given this procedure, it is possible to compute maximum a posteriori estimate of θ using a numerical optimization procedure over $p\left(\theta|\{Y_t\}_{t=1}^{T}\right)$ with respect to θ. However, in order to quantify uncertainty about the parameter estimate, I characterize the full posterior distribution, by drawing from this distribution using Markov chain Monte Carlo (MCMC). I draw from the posterior distribution using a random-walk Metropolis-Hastings algorithm (see An and Schorfheide (2007)). In order to check the convergence of the procedure, I compute the Gelman-Rubin statistic.

Observables. An advantage of likelihood-based estimation of heterogeneous agent models in particular, is that we can use as observable time series in the estimation functions of the cross-sectional distribution of agents. Since the solution method entails incorporating the cross-sectional distribution of agents into the latent state variable, it is natural to include elements of that distribution, or linear functions of elements of the distribution, as observables. The increasing availability of firm-based and household-based panel data in economics is likely to render this approach appealing.

Alternatives. A variety of methods have been used in the literature for full-information estimation of models in linear Gaussian state-space form. For instance, Ireland (2004) estimates using maximum likelihood, Sargent (1989) adds measurement error to each measurement equation and allows for more time series than structural shocks. Moreover, for Bayesian estimation, the random-walk Metropolis-Hastings is only one of many algorithms for drawing from the posterior. Other methods, such as Hamiltonian Monte Carlo, have also been used for models such as these. Fernández-Villaverde et al. (2016) provide a more comprehensive survey of the literature. Once the model is in the form given in equations (25) and (26), many applicable methods may be used, at the discretion of the researcher.

4.2 An estimation exercise

To demonstrate the efficacy of the method, I estimate the parameters of the model using data generated by simulating from the model for a given set of known parameters θ^*. The model features three aggregate shocks (the aggregate TFP shock, the discount factor shock and the adjustment cost shock), so I estimate the model using simulated data for three time series of "observables". For the three series, I choose aggregate output Y_t, aggregate investment I_t, and investment by small firms I_t^{small}, which are defined in the discretized model as follows:

$$Y_t = \sum_{k_i \in X_k} \sum_{z_j \in X_z} X_t^Z z_j k_t^\alpha \mu_{t-1}(k_i, z_j)$$

$$I_t = \sum_{k_i \in X_k} \sum_{z_j \in X_z} \left[\Phi(k_i, z_j)\, \mathbf{c}_t - (1-\delta) k_i\right] \mu_{t-1}(k_i, z_j)$$

and

$$I_t^{\text{small}} = \sum_{k_i \in X_k} \sum_{z_j \in X_z} \left[\Phi(k_i, z_j)\, \mathbf{c}_t - (1-\delta) k_i\right] \mu_{t-1}(k_i, z_j)\, \mathbf{1}\{k_i \leq k^*\}$$

where k^* is the median level of k in the steady-state distribution $\bar{\mu}(k_i, z_j)$. As described in Section 3.3.4 above, I include these three variables in the state variable $\hat{\xi}_t$, so the coefficient C of the measurement equation is of the form

$$C = \begin{bmatrix} & 1 & 0 & 0 & \\ \dots & 0 & 1 & 0 & \dots \\ & 0 & 0 & 1 & \end{bmatrix}$$

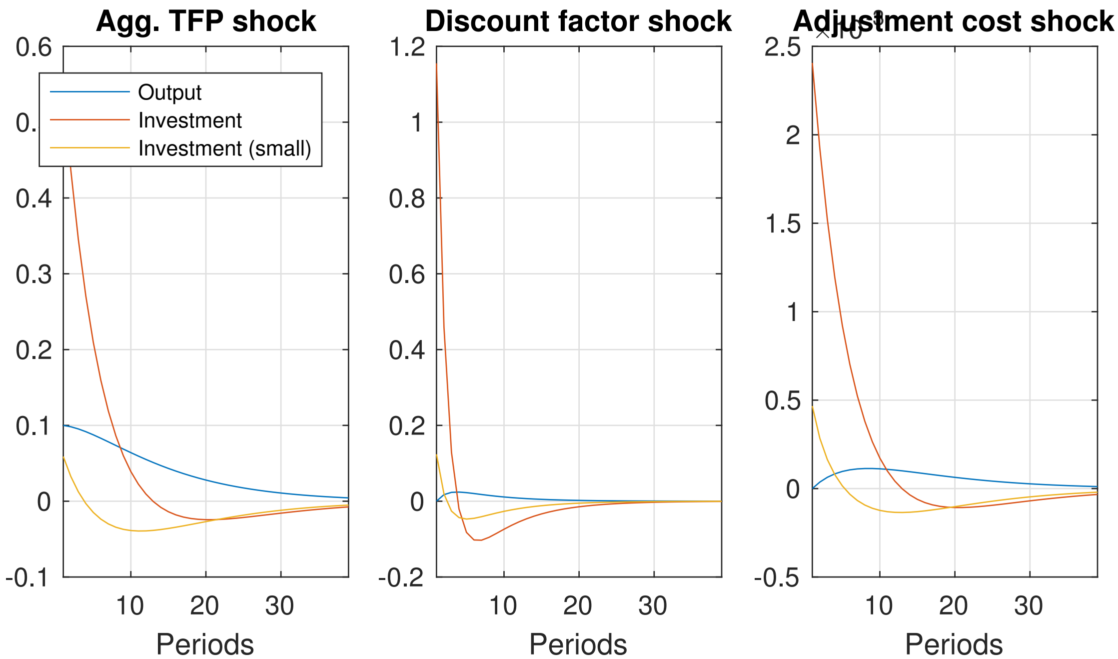
and picks out the elements of $\hat{\xi}_t$ which correspond to Y_t, I_t and I_t^{small}, respectively—or, more precisely, their deviations from steady state: recall that the state equation is expressed in terms of deviations from steady state, so the observation equation is expressed in terms of deviation from steady state too. I assume no measurement error on the observable time series, so the matrix D is set to 0.

I simulate the model for 200 time periods, in order to corresponds to a typical length of time series confronted by researchers in macroeconomics.

4.2.1 Parameters

I select an arbitrary set of true parameters, reported in Table 1. Figure 2 shows the impulse responses of the three observables with respect to three aggregate shocks in the model. They show the percentage response of Y_t, I_t and I_t^{small} to a positive aggregate TFP shock, a positive discount factor shock and a negative adjustment cost shock, respectively, where the shock is a 1 standard deviation shock in all cases. It is worth noting that the discount factor shock and the adjustment cost shock have similar effects on the three observables, in terms of the sign and relative magnitudes of their effects. As Section 4.2.3 demonstrates, however, this does not prevent the estimation procedure from producing well-identified estimates of the parameters of the stochastic processes for these two shocks.

Figure 2: Impulse responses.



4.2.2 Priors

I fix three of the parameters: α, which determines the degree of decreasing returns in production, β, which is the household's discount factor, and δ which is the depreciation rate of capital. In the literature on estimating DSGE models, these parameters (or their close counterparts) are often calibrated using out-of-dataset information, such as, for α, the average profit rate (e.g. Eisfeldt and Muir (2014)); for β, the average real interest rate, and for δ, the average investment-to-capital ratio (e.g. in Khan and Thomas (2008)).

In the choice of priors, I follow the literature on Bayesian estimation of DSGE models. I assume that all priors are a priori independent. I follow An and Schorfheide (2007) and others, in using Beta distributions for the priors of persistence parameters in AR(1) processes, inverse Gamma distributions for the corresponding shock standard deviations and Gamma distributions for parameters on the positive real line (which, in this case consists of ϕ, the adjustment cost parameter).

Because the data is simulated, I know the true parameters in this case, so I can ensure that the choice of priors will not be at odds with the data. In a real-world application, a researcher could check for this using prior predictive checks (see Geweke (2005)).

4.2.3 Results

Table 2 summarizes the results of the Bayesian estimation procedure. The posterior statistics are computed using 600,000 draws from the posterior distribution using the MCMC method described in Section 4.1. For the marginal posterior of each parameter in θ, I compute the median, mode, standard deviation and the 90% credible interval. Figures 3 and 4 give graphical representations of the prior and posterior distributions, where I plot the posterior density by kernel-smoothing the MCMC draws. Both distributions are scaled for visual comparability, and the prior is only partially plotted for all but the autocorrelation parameters.

The parameter vector θ consists of two sets of parameters, the "steady state" parameters $\theta^{\text{ss}} = (\phi, \rho_z, \sigma_z)$,

Table 1: Estimation using simulated data: true parameters and prior distributions.

Parameter		True	Prior	P1*	P2*	Mean	StdDev
Decreasing returns to scale	α	0.75					
Discount factor	β	0.96					
Depreciation	δ	0.10					
Adjustment cost	ϕ	0.05	Gamma	2.00	1.00	1.00	1.00
Persistence (idio. productivity)	ρ_z	0.90	Beta	2.00	2.00	0.50	0.22
Volatility idio. productivity shocks	$10 \times \sigma_z$	0.10	Inv. Gamma	3.00	0.50	0.25	0.25
Autocorrelation TFP shock	ρ^Z	0.90	Beta	2.00	2.00	0.50	0.22
Volatility TFP shock	σ^Z	0.10	Inv. Gamma	3.00	0.50	0.25	0.25
Autocorrelation discount rate shock	ρ^β	0.50	Beta	2.00	2.00	0.50	0.22
Volatility discount rate shock	σ^β	0.10	Inv. Gamma	3.00	0.5	0.25	0.25
Autocorrelation adj. cost shock	ρ^ϕ	0.90	Beta	2.00	2.00	0.50	0.22
Volatility adj. cost shock	σ^ϕ	0.10	Inv. Gamma	3.00	0.50	0.25	0.25

"True" indicates the parameters chosen for the simulation from which data was drawn. For Gamma-distributed and Inverse-Gamma-distributed priors, P1 is the shape parameter and P2 is the rate parameter. For Beta-distributed priors, P1 and P2 are the first and second shape parameters, respectively.

those which have an effect on the steady state of the model, and the aggregate shock parameters $\theta^{\text{shock}} = \left(\rho^Z, \sigma^Z, \rho^\beta, \sigma^\beta, \rho^\phi, \sigma^\phi\right)$, those which affect only the aggregate dynamics of the model.

As Table 2 shows, the estimation procedure is capable of recovering the true parameters of the model. The point estimate for each parameter, given by the mode of the posterior distribution, is close to the true value in each case and the true value lies within the 90% credible interval in each case. Moreover, the estimates are quite precise, in that the standard deviations of the posterior is significantly smaller than the standard deviation of the prior in each case.

It is not particularly surprising that the parameters which affect the steady state are less well-identified than those which govern the aggregate shock process. In fact, it is a success of the estimation procedure that it can identify these parameters from this dataset at all. The estimation procedure uses data on the deviation of three variables of the model from their steady state values so it doesn't explicitly incorporate any data on the model's steady state (recall that the model state equation $\hat{\xi}_t = A\hat{\xi}_{t-1} + B\mathcal{E}_t$ is in terms of $\hat{\xi}_t$, which represents deviations from steady state). The estimation procedure identifies these parameters purely through their effect on the dynamics of the model.

For instance, the parameter ϕ is an adjustment cost, which affects the responsiveness of firms to shocks (both aggregate and idiosyncratic). Therefore its value has an affect on the steady state of the model (when aggregate shocks are eliminated but idiosyncratic shocks are still in place), and some studies estimate or calibrate ϕ (and other parameters which play a similar role to ϕ in their models) using features of the stationary distribution of agents in their models. But ϕ also has an affect on the dynamics of the model, since the response of aggregate investment, say, to a shock depends on ϕ via individual firms' investment decisions. As these estimation results show, ϕ can be identified using data on the dynamics of the model only,

Figure 3: Estimation using simulated data: priors and posteriors for steady-state parameters.

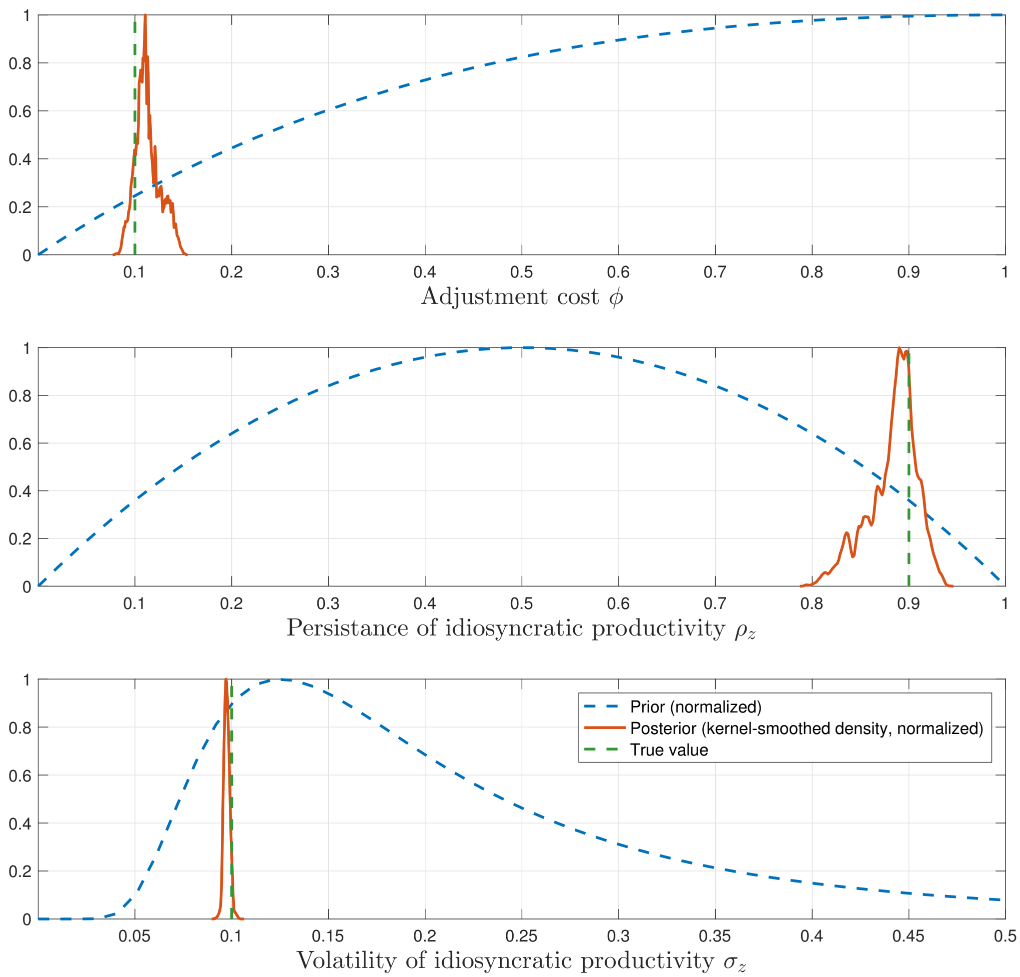


Figure 4: Estimation on simulated data: priors and posteriors for shock process parameters.

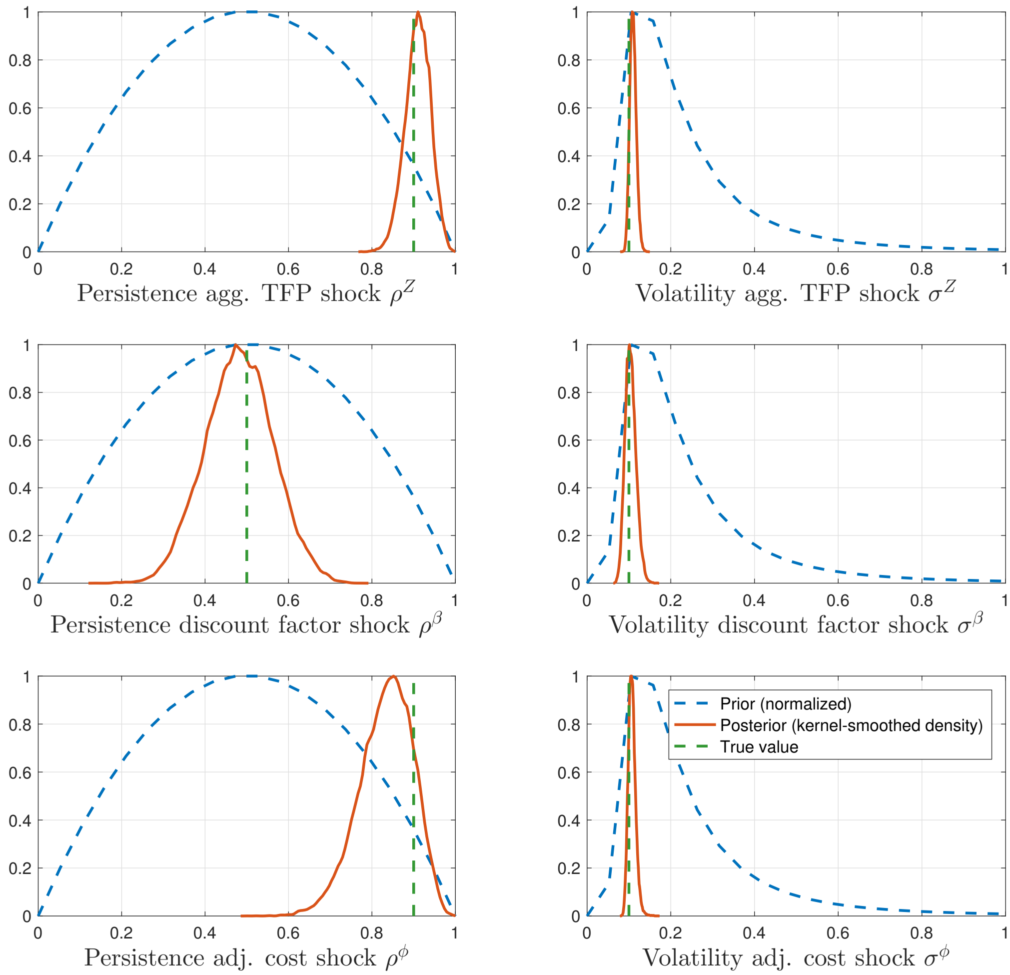


Table 2: Estimation using simulated data: prior and posterior distributions.

Parameter		True	Prior Mean	Prior SD	Posterior Median	Posterior Mode	Posterior SD	5%	95%
Adjustment cost	ϕ	0.10	1.00	1.00	0.109	0.111	0.009	0.092	0.138
Persistence idio. productivity	ρ_z	0.90	0.50	0.22	0.893	0.891	0.017	0.833	0.918
Volatility idio. productivity	σ_z	0.10	0.25	0.25	0.097	0.097	0.001	0.095	0.100
Autocorrelation TFP shock	ρ^Z	0.90	0.50	0.22	0.911	0.909	0.030	0.858	0.958
Volatility TFP shock	σ^Z	0.10	0.25	0.25	0.110	0.111	0.008	0.098	0.123
Autocorrelation discount rate	ρ^β	0.50	0.50	0.22	0.482	0.472	0.081	0.346	0.615
Volatility discount rate shock	σ^β	0.10	0.25	0.25	0.103	0.096	0.013	0.084	0.127
Autocorrelation adj. cost shock	ρ^ϕ	0.90	0.50	0.22	0.838	0.856	0.064	0.719	0.929
Volatility adj. cost shock	σ^ϕ	0.10	0.25	0.25	0.107	0.104	0.009	0.094	0.123

and this suggests that studies which attempt to estimate ϕ and other adjustment-related parameters using data on stationary distributions only are ignoring informative features of the data and possibly producing biased estimates.

4.2.4 Convergence of estimation

For the estimates presented in Section 4.2.3 to be credible, it is important to check that the MCMC procedure which generated them is actually drawing from the posterior distribution. For the Metropolis-Hastings algorithm, one convenient way of testing for convergence involves running the algorithm in a finite number of separate "chains", and testing whether the distributions produced by each of these chains are sufficiently similar to one another. Gelman and Rubin (1992) describe how to construct a metric of similarity across chains, the Gelman-Rubin $\hat{R}$ statistic. Table 3 shows gives the $\hat{R}$ statistics for the estimation procedure, where the $\hat{R}$ is computed separately for each parameter. A value of $\hat{R}$ closer to 1 indicates better convergence, so we see from Table 3 that the MCMC procedure has converged to within acceptable limits for all parameters, except for ϕ and ρ_z, which we saw in the preceding section, were less well-estimated than the remaining parameters.

Another check of convergence is given by a visual inspection of the "trace" of the MCMC chains, that is a plot of the chains overlaid over time. I plot the trace for each parameter in Figures 5 and 6. These figures indicate that the MCMC draws are stationary for all parameters except ϕ and ρ_z, consistent with the results in Table 3.

5 Comparison with method of moments estimator

I now compare the estimation procedure to estimates obtained using simulated method of moments. For this exercise, I fix the steady state parameters at their true values, as given in Table 2. I also remove the adjustment cost shock from the model, for computational reasons, and remove its parameters ρ^ϕ and σ^ϕ from the estimation, purely for computational reasons. I estimate only the four parameters which govern the

Table 3: Gelman-Rubin $\hat{R}$ statistic.

Parameter	ϕ	ρ_z	σ_z	ρ^Z	σ^Z	ρ^β	σ^β	ρ^ϕ	σ^ϕ
$\hat{R}$	1.0640	1.0668	1.0079	1.0007	1.0003	1.0019	1.0019	1.0004	1.0020

The Gelman-Rubin $\hat{R}$ statistic is given by $\hat{R} = ([(1-1/T)\,W + (1/T)\,B]\,/W)^{1/2}$ where B is the between-chain variance, W is the average within-chain variance and T is the length of chain. I use 10 chains of 100,000 draws each and discard the first 40,000 draws.

Figure 5: Estimation on simulated data: trace and histogram of posterior draws for steady-state parameters.

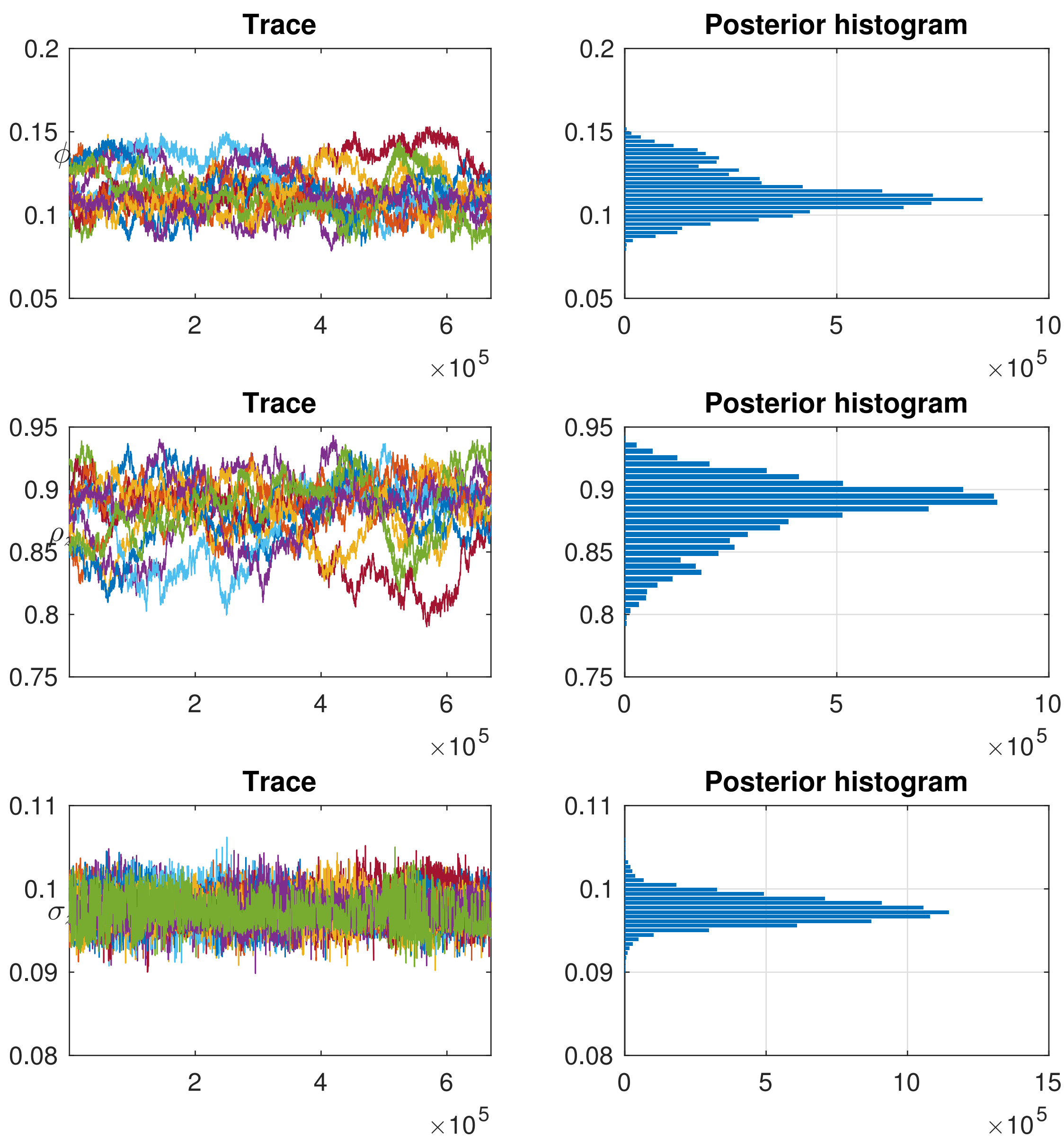


Figure 6: Estimation on simulated data: trace and histogram of posterior draws for shock process parameters.

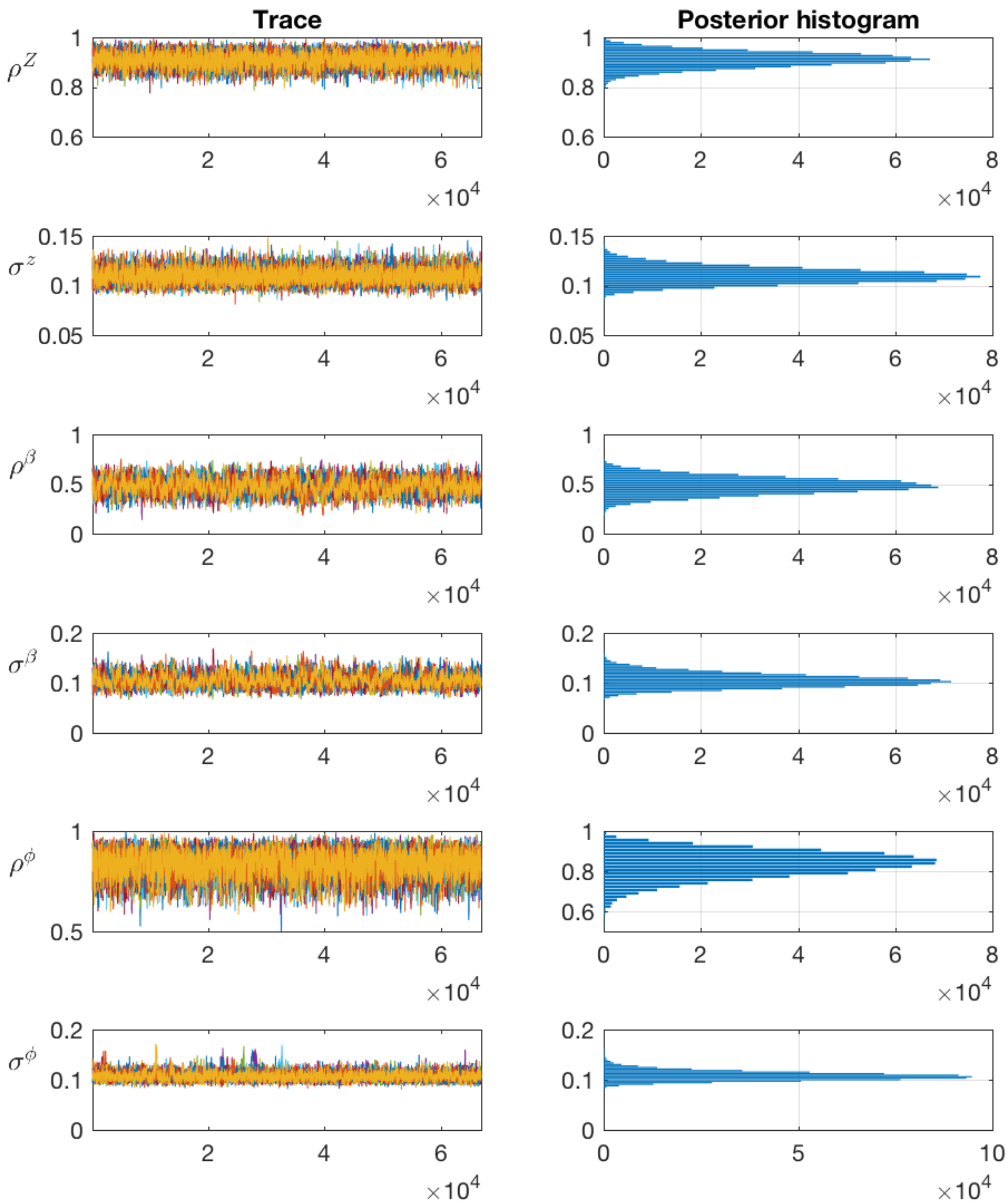


Table 4: Comparison with simulated method of moments

		SMM		MLE	
Parameter	True	Estimate	S.E.	Estimate	S.E.
ρ^Z	0.80	0.797	0.100	0.815	0.041
σ^Z	0.20	0.196	0.017	0.215	0.016
ρ^β	0.80	0.744	0.218	0.837	0.048
σ^β	0.10	0.120	0.071	0.100	0.013

remaining aggregate shock processes, that is, the autocorrelation and volatility parameters of the aggregate TFP and discount factor shocks, respectively.

5.1 Simulated method of moments

To obtain the simulated method of moments estimator, $\hat{\theta}$, I obtain the value of θ which minimizes the distance between a set of moments of the observed data, and moments of simulated data. I first compute the moments of the true data, denoted by $g(\theta^*)$. Then, for each candidate value θ', I solve the model, and simulate data from the model corresponding to the true data. I then compute some moments of the simulated data $g(\theta')$ and I compute the distance between moments, weighted by some weighting matrix W. I minimize the weighted distance, with respect to θ, by iterating over θ. I use a total of nine target moments in the estimation: the three variances of each variable, the three first-order autocorrelations, and the three covariances of each variable with one another.

Columns 3 and 4 of Table 4 shows the point estimates and estimated standard errors obtained using simulated method of moments. These standard errors are computed using Monte Carlo methods. That is, I run the estimation procedure repeatedly (10,000 times, in this case) using a different observed dataset each time. Since the observed data was computed by simulation in the first place, I can simply generate each sample of data by re-simulating, using a different random seed, so there is no need to resample with replacement from the original dataset, as is typically done when bootstrapping standard errors.

Columns 5 and 6 of Table 4 shows the point estimates and estimated standard errors obtained when using maximum likelihood estimation. The maximum likelihood estimation procedure is identical to the Bayesian procedure outlined in Section 4.2, except that the priors for all parameters are set to be uniform over the support in each case. Table 4 shows that both methods obtain estimates for the four parameters which are close to their true values. As columns 4 and 6 show, the standard errors for the MLE estimates are smaller than those obtained using SMM, which illustrates the advantage of the full-information estimation.

6 Conclusion

In this chapter, I have outlined a new approach for estimating heterogeneous agent DSGE models. The approach combines the projection and perturbation solution method with Bayesian estimation techniques in order to feasibly estimate the parameters of DSGE with nontrivial heterogeneity. The primary advantage of the approach is that it allows us to estimate the parameters of a heterogeneous agents model using an observable the time series behavior of any arbitrary feature of the distribution of agents. I illustrate this benefit in the context of a simple firm dynamics model with three shocks. I estimate the parameters of the model on simulated data, using as observable variables in the estimation the time series of two aggregate variables and a variable obtained from the distribution of agents (namely, the quantity invested by small firms). As I have shown, this procedure succeeds in retrieving the true parameters of the model from which the data were simulated. I also show that the standard error associated with these estimates is smaller than those one would obtain by estimating the model parameters using a moment-based method.

Bibliography

Stéphane Adjemian, Houtan Bastani, Fréderic Karamé, Michel Juillard, Junior Maih, Ferhat Mihoubi, George Perendia, Johannes Pfeifer, Marco Ratto, and Sébastien Villemot. Dynare: Reference Manual Version 4. Dynare Working Papers 1, CEPREMAP, April 2011.

Yann Algan, Olivier Allais, Wouter J. Den Haan, and Pontus Rendahl. Chapter 6 - solving and simulating models with heterogeneous agents and aggregate uncertainty. In Karl Schmedders and Kenneth L. Judd, editors, *Handbook of Computational Economics Vol. 3*, volume 3 of *Handbook of Computational Economics*, pages 277 – 324. Elsevier, 2014. doi: http://dx.doi.org/10.1016/B978-0-444-52980-0.00006-2.

Sungbae An and Frank Schorfheide. Bayesian analysis of dsge models. *Econometric Reviews*, 26(2-4):113–172, 2007. doi: 10.1080/07474930701220071.

Lawrence J. Christiano, Martin Eichenbaum, and Charles Evans. Nominal rigidities and the dynamic effects of a shock to monetary policy. Working Paper 8403, National Bureau of Economic Research, July 2001.

Andrea L. Eisfeldt and Tyler Muir. Aggregate Issuance and Savings Waves. NBER Working Papers 20442, National Bureau of Economic Research, Inc, September 2014.

Jesús Fernández-Villaverde, Juan F. Rubio Ramírez, and Frank Schorfheide. Solution and estimation methods for dsge models. Working Paper 21862, National Bureau of Economic Research, January 2016.

Andrew Gelman and Donald B. Rubin. Inference from iterative simulation using multiple sequences. *Statist. Sci.*, 7(4):457–472, 11 1992. doi: 10.1214/ss/1177011136.

J. Geweke. *Contemporary Bayesian Econometrics and Statistics*. John Wiley, 2005.

Lars Peter Hansen and James J. Heckman. The Empirical Foundations of Calibration. *Journal of Economic Perspectives*, 10(1):87–104, Winter 1996.

Peter N. Ireland. Technology Shocks in the New Keynesian Model. *The Review of Economics and Statistics*, 86(4):923–936, November 2004.

Kenneth L. Judd. *Numerical Methods in Economics*. MIT Press, 1998.

Alejandro Justiniano, Giorgio Primiceri, and Andrea Tambalotti. Investment Shocks and the Relative Price of Investment. *Review of Economic Dynamics*, 14(1):101–121, January 2011.

Aubhik Khan and Julia K. Thomas. Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics. *Econometrica*, 76(2):395–436, 03 2008.

Mario J. Miranda and Paul L. Fackler. *Applied Computational Economics and Finance*. MIT Press, Cambridge, MA, USA, 2002. ISBN 0262134209.

Michael Reiter. Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control*, 33(3):649–665, March 2009.

K. G. Rouwenhorst. Asset pricing implications of equilibrium business cycle models. In T. F. Cooley, editor, *Frontiers of Business Cycle Research*. Princeton University Press, 1995.

Francisco J. Ruge-Murcia. Methods to estimate dynamic stochastic general equilibrium models. *Journal of Economic Dynamics and Control*, 31(8):2599–2636, August 2007.

Thomas J. Sargent. Two models of measurements and the investment accelerator. *Journal of Political Economy*, 97(2):251–287, 1989. doi: 10.1086/261603.

Ali Shourideh and Ariel Zetlin-Jones. External financing and the role of financial frictions over the business cycle: Measurement and theory. *Working paper*, 2013.

Christopher A. Sims. Solving linear rational expectations models. *Computational Economics*, 20(1):1–20, 2002. ISSN 1572-9974. doi: 10.1023/A:1020517101123.

Frank Smets and Rafael Wouters. Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. CEPR Discussion Papers 6112, C.E.P.R. Discussion Papers, February 2007.

Thomas Winberry. A toolbox for solving and estimating heterogeneous agent macro models. *Working paper*, 2016.

Eric R. Young. Solving the incomplete markets model with aggregate uncertainty using the krusell–smith algorithm and non-stochastic simulations. *Journal of Economic Dynamics and Control*, 34(1):36 – 41, 2010. ISSN 0165-1889. doi: http://dx.doi.org/10.1016/j.jedc.2008.11.010. Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty.