Discussion of Estimating Linearized Heterogeneous Agent Models Using Panel Data by Tamas Papp and Michael Reiter

Wouter J. Den Haan

Abstract

The techniques proposed in Papp and Reiter (2020) allow the use of cross-sectional and aggregate data observed at different frequencies in the estimation of dynamic stochastic macroeconomic models. However, the question is whether technique is getting ahead of what is sensible in terms of currently available empirical strategies to estimate macroeconomic models which are – without exception – misspecified.

Keywords: Solution techniques, heterogeneous agents, misspecification.

1. Introduction

During the nineties, the first algorithms were developed to solve models with heterogeneous agents and aggregate uncertainty. There are now several algorithms.[1] The method proposed in Reiter (2009) stands out in being much faster than other algorithms. As discussed below, the "Reiter approach" is not suitable for all models, but if it is then it comes with a massive computational advantage. Papp and Reiter (2020) extend the Reiter approach to make it suitable for estimation when cross-sectional information is not available at the same frequency as aggregate data, a situation that is quite typical.

In section 2, I describe the Reiter method in the most straightforward way. This may be of some educational value, since the Reiter method is often combined with additional

*Den Haan: Centre for Macroeconomics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK and CEPR, London, UK. E-mail: wjdenhaan@gmail.com. I would like to thank Miguel Faria-e-Castro, Michael Reiter, and other participants at the 2019 JEDC conference in Gerzenzee. This research is supported by the Economic and Social Research Council (ESRC).

[1] See Algan et al. (2014) for an overview.

bells and whistles which obscure its intrinsic ability to obtain a numerical solution at low computational cost. This detailed discussion will make it easy to illustrate a weakness of the Reiter method and that is dealing with occasionally binding constraints. In the presence of occasionally binding constraints, it is important for the Reiter method – and more so than for other perturbation methods – that the fluctuations in aggregate uncertainty are really small. This weakness can be easily overcome by using penalty functions instead of inequality constraints, which often are actually more realistic than inequality constraints.

In section 3, I discuss Papp and Reiter (2020) in detail and I will refer to my discussion in section 2 and reiterate that the complexity can be reduced if penalty functions are used instead of inequality constraints.

The last section touches upon a more fundamental issue. Maximum Likelihood (ML), a full-information estimation method, is used by Papp and Reiter (2020) to estimate the model. ML and especially its Bayesian version are by far the most popular estimation procedure for stochastic dynamic macroeconomic models. Are these methods the right ones when we know that macroeconomic models are misspecified in at least some nontrivial dimensions? And the question should be asked whether the sophisticated extensions proposed in Papp and Reiter (2020) alleviate this fundamental problem or make it worse? That is, are computational technique to solve models perhaps ahead of available empirical methodologies?

2. The Reiter method

In models with heterogeneous agents, the policy rules depend on the cross-sectional distribution of agents' characteristics, a high-dimensional and typically complex object. It does so through a limited set of prices such as the wage rate and the rental rate of capital. Without aggregate uncertainty these prices are constant. At a given set of prices, simple solution techniques can be used to obtain approximations to the individual policy rules. Given those policy rules one can then check whether equilibrium conditions hold. Iteration procedures or an equation solver can be used to find equilibrium prices.

The situation is quite different in the presence of aggregate uncertainty. The cross-sectional distribution is then time-varying and its role no longer characterized by a finite set of constant prices. Thus, the true solutions of the model are functions of this high-dimensional object with an unknown functional form. Since the functional forms of the policy functions are also unknown, we are faced with a challenging numerical problem.

Reiter (2009) proposes a procedure which is computationally *much* faster than alternatives. Recall from the discussion above, that solving the model without aggregate uncertainty is not that challenging even if one uses accurate global projection methods. The starting point of the Reiter approach is the assumption that aggregate uncertainty is relatively small and its impact on the economic outcomes relatively minor. This means that economies with aggregate uncertainty display relatively minor fluctuations around the economy without aggregate uncertainty. It is obviously true that aggregate uncertainty is much smaller than idiosyncratic uncertainty. And aggregate uncertainty does have only a limited impact in many existing models.[2] Reiter (2009) points out that the consequence of this observation is that one can use a fast numerical technique, namely perturbation, to deal with that part of the solution problem that changes the dimensionality of the problem so drastically, that is, aggregate uncertainty.

The underlying principles of the Reiter method are straightforward and very sensible. Nevertheless, I struggled to find a way to explain to my students what programming of the Reiter method actually entails. Without understanding the structure of the program to be written, students will also not understand how powerful the Reiter method is.

In this section, I present what I have found to be a useful way to teach the Reiter method. In addition, I will use the exposition to make one fundamental point regarding the way borrowing restrictions are incorporated into our models. Reiter (2009) and Papp and Reiter (2020) follow the literature and capture such restrictions by limiting the maximum amount on can borrow with an inequality constraint. At interesting parameter values, this means that agents are at times not constrained (and the consumption/savings decisions is determined

[2] It is not true in all models. In the model of Den Haan et al. (2017), aggregate risk has an important quantitative impact on individual behavior and the economy with aggregate risk cannot be considered as a perturbation of the corresponding economy without aggregate risk.

by the usual Euler equation and the budget constraint) and are at times constrained (and the borrowing constraint and the budget constraint are the relevant equations).

2.1. The Model

The economy used to explain the algorithm is a Krusell-Smith type economy modified to make the exposition easier. The firm problem is identical to the one of the standard neoclassical growth model. That is, firms hire N_t effective units of labor at wage rate w_t and hire K_t capital units at rental rate r_t. Specifically, the firm problem is given by

$$\begin{gathered}\max_{K_t,N_t} K_t^{\alpha} N_t^{1-\alpha} \\ \text{s.t.} \\ K_t r_t + N_t w_t = z_t K_t^{\alpha} + N_t^{(1-\alpha)},\end{gathered} \tag{1}$$

where z_t is a productivity shock. The law of motion for z_t is given by

$$\ln z_t = \rho \ln z_{t-1} + e_{z,t}, \tag{2}$$

with $\mathbb{E}_t[e_{z,t+1}] = 0$. The Reiter method uses first-order perturbation to deal with the fluctuations due to $e_{z,t}$ which means that we do not need to make any additional distributional assumptions regarding $e_{z,t}$ to derive the policy functions.[3]

The first-order conditions are given by

$$r_t = \alpha z_t K_t^{1-\alpha} N_t^{1-\alpha}, \text{and} \tag{3}$$

$$w_t = (1-\alpha) z_t K_t^{\alpha} N_t^{-\alpha}. \tag{4}$$

Firm size is not determined with constant returns to scale. Thus, we can work with a

[3] If second-order perturbation would be used, then we need to know the standard deviation to derive the policy functions. Of course, one would need to make distributional assumptions if one wants to generate simulated time paths.

representative firm.

There is a unit mass of workers whose behavior is characterized by the following optimization problem

$$\max_{c_{i,t},k_{i,t}} \ln c_{i,t} - G(k_{i,t})$$
$$\text{s.t.} \tag{5}$$
$$c_{i,t} + k_{i,t} = r_t k_{i,t-1} + w_t e_{i,t} + (1-\delta)k_{i,t-1},$$

where $c_{i,t}$ is the consumption of worker i, $k_{i,t}$ is the capital level chosen in period t available for production in period $t{+}1$, and $e_{i,t}$ is an idiosyncratic worker-specific productivity disturbance.

Instead of the popular inequality constraint to limit how much on can go short in capital (borrowing), i.e. $k_{i,t} \geq \overline{k}$ with $\overline{k} \leq 0$, there is a penalty function, $G(k_{i,t})$, with the following properties

$$\frac{\partial G(k)}{\partial k} \leq 0, \tag{6}$$
$$\frac{\partial^2 G(k)}{\partial k^2} \geq 0, \tag{7}$$

that is, the penalty gets smaller (or remains the same) as capital increases and the shape is convex. A simple example would be[4]

$$G(k) = k^2 \text{ if } k < 0$$
$$\text{and} \tag{8}$$
$$G(k) = 0 \text{ if } k \geq 0. \tag{9}$$

Below, I will discuss the relationship between penalty functions and occasionally binding borrowing constraints in more detail and also highlight how the choice between these two

[4]An alternative is $G(k) = (k-\overline{k})^{-\phi}$ with $\overline{k} < 0$ which converges to the inequality constraint specification, $k \geq \overline{k}$, as ϕ goes to infinity.

options affect the suitability of the Reiter method.

Workers' first-order conditions are given by

$$c_{i,t} + k_{i,t} = r_t k_{i,t-1} + w_t e_{i,t} + (1-\delta)k_{i,t-1}, \tag{10a}$$

$$\frac{1}{c_{i,t}} + \frac{\partial G(k_{i,t})}{\partial k_{i,t}} = \beta \mathbb{E}_t \left[\frac{1}{c_{i,t+1}} (r_{t+1} + 1 - \delta) \right]. \tag{10b}$$

An increase in $k_{i,t}$ has the usual cost of giving up a unit of consumption, but has the benefit of reducing the penalty unless $k_{i,t}$ is such that $\partial G(k_{i,t})/\partial k_{i,t} = 0$.

Assumptions to help the exposition. To shorten the equations, and for this reason only, I assume that

$$\delta = 1. \tag{11}$$

Furthermore, I assume that the idiosyncratic shock, $e_{i,t}$, is not only *i.i.d.* across individuals but also across time. This means that the only cross-sectional distribution one has to keep track of is the cross-sectional distribution of individual capital, $k_{i,t-1}$. By contrast, if $e_{i,t}$ is a first-order Markov process, as is usually the case, then one would need to keep track of the cross-sectional *joint* distribution of $k_{i,t-1}$ and $e_{i,t}$. This is not problematic for the Reiter method, but would introduce more notation, variables, and equations.

Finally, I assume that $e_{i,t}$ can take on two values, ε_L and ε_H with equal probability, $\varepsilon_L < \varepsilon_H$, and $\mathbb{E}[e_{i,t}] = 1$. This means that expectations over $e_{i,t+1}$ can be replaced by a simple sum. The Reiter method can easily deal with continuous support of $e_{i,t}$ by using Gaussian quadrature which also boils down to representing the conditional expectation with a finite sum, but this would introduce additional notation.

Equilibrium. Let f_{t-1} denote the cross-sectional distribution of end-of-period $t-1$ capital holdings, which is equal to the beginning-of-period t distribution. We use period $t-1$ subscripts because what matters for period t are the capital stocks chosen in period $t-1$

and carried over into period t.[5] Since there are a continuum of workers, this is fully pinned down by f_{t-2} and the value of z_{t-1}. The set of aggregate state variables consists of f_{t-1} and z_t and will be denoted by s_t. The set of state variables relevant for the individual are s_t as well as $k_{i,t-1}$ and $e_{i,t}$. [6]

The equilibrium is given by

$$\begin{aligned}
&\frac{1}{r_t k_{i,t-1} + w_t e_{i,t} - k_{i,t}} + \frac{\partial G(k_{i,t})}{\partial k_{i,t}} \\
&= \\
&\beta \mathbb{E}_t \left[\sum_{\tilde{j}=1}^{2} \frac{1}{2} \frac{r_{t+1}}{r_{t+1} k_{i,t} + w_{t+1} \varepsilon_{\tilde{j}} - k_{i,t+1}} \right] \qquad (12a)
\end{aligned}$$

$$r_t = \alpha z_t K_t^{1-\alpha} N_t^{1-\alpha}, \qquad (12b)$$

$$w_t = (1-\alpha) z_t K_t^{\alpha} N_t^{-\alpha}, \qquad (12c)$$

$$N_t = 1, \qquad (12d)$$

$$K_t = \int_k k f_{t-1}(k) dk, \qquad (12e)$$

$$f_t = \Gamma(f_{t-1}, z_t). \qquad (12f)$$

Note that the expectations operator in equation (12) is only over $e_{z,t+1}$. The summation takes care of the idiosyncratic shock.

2.2. Numerical approximation

Before we start with a description of the Reiter method, let's recall what we would do if we approximate an individual policy function using projection methods. We would use a flexible functional form such as a polynomial or a spline to approximate individual behavior and the inputs are the state variables. That is,

$$k_{i,t+1} = P(e_{i,t}, k_{i,t-1}, z_t, f_{t-1}; \eta), \qquad (13)$$

[5] That is, we use notation consistent with Dynare convention.

[6] If $e_{i,t}$ is an *i.i.d.* variable, then a sufficient state variable to characterize the individual is their cash-on-hand level, i.e., $r_t k_{i,t-1} + w_t e_{i,t}$. This could be used in the solution method, but is not done here since it makes the equations less transparent.

where the vector η contains the coefficients of the approximating function. Since $e_{i,t}$ can take on only two values, we can solve for two separate policy functions. That is,

$$k_{i,t+1} = P(k_{i,t-1}, z_t, f_{t-1}; \eta_L) \qquad \text{if } e_{i,t} = \varepsilon_L, \tag{14}$$

$$k_{i,t+1} = P(k_{i,t-1}, z_t, f_{t-1}; \eta_H) \qquad \text{if } e_{i,t} = \varepsilon_H. \tag{15}$$

To implement the Reiter method we will write this as

$$k_{i,t+1} = P(k_{i,t-1}; \eta_L(z_t, f_{t-1})) \qquad \text{if } e_{i,t} = \varepsilon_L, \tag{16}$$

$$k_{i,t+1} = P(k_{i,t-1}; \eta_H(z_t, f_{t-1})) \qquad \text{if } e_{i,t} = \varepsilon_H. \tag{17}$$

For example, if $P(\cdot)$ is a second-order polynomial in $k_{i,t-1}$, then its three coefficients are the elements of η and each coefficient is a function of the aggregate state variables.

For the Reiter method, it is important that

1. one characterizes the whole cross-sectional distribution and
2. one can write down computer code to describe its law of motion given the policy function of the agent.

Reiter (2009) and Papp and Reiter (2020) use a histogram, that is, a grid with a vector that contains the mass of agents at each grid point. Given the histogram of this period's beginning-of-period capital holdings, f_{t-1}, z_t, and the individual policy rule one can calculate next period's histogram.[7] Here, I will use a version of the Reiter method as it is implemented in Winberry (2018) because it simplifies the equations and, thus, makes it easier to understand the structure of the Reiter method. Instead of a histogram, Winberry (2018) uses a flexible functional form from the class of functions proposed in Algan et al. (2008) to characterize the cross-sectional distribution. If one knows the parameters of the approximating density, then one knows everything there is to know about the distribution. An example of a second-

[7] If a capital choice falls between grid points one has to allocate the associated mass over the two adjacent grid points.

order approximation would be the Normal density. For this application we will use as the parameters of the Normal, the mean, $\mu_{k,t-1}$ and the un-centered second moment, $\Omega_{k,t-1}$. Knowing these two statistics, one can immediately calculate the standard deviation and use the standard expression of the Normal density.

Now, let's turn to the second requirement. Is it possible to write down exact equations for the law of motion of this (approximate) distribution if one has this period's distribution and the policy function of the agent. Specifically, can we write down computer code that relates $\mu_{k,t}$ and $\Omega_{k,t}$ to $\mu_{k,t-1}$ and $\Omega_{k,t-1}$? We know that

$$\mu_{k,t} = \int_k \begin{pmatrix} \frac{1}{2}P(k;\eta_L(z_t,\mu_{k,t-1},\Omega_{k,t-1}))+ \\ \frac{1}{2}P(k;\eta_H(z_t,\mu_{k,t-1},\Omega_{k,t-1})) \end{pmatrix} \frac{1}{\sigma_{k,t-1}\sqrt{2\pi}} e^{-\frac{(k-\mu_{k,t-1})^2}{2\sigma^2_{k,t-1}}} dk, \tag{18}$$

$$\Omega_{k,t} = \int_k \begin{pmatrix} \frac{1}{2}P(k,\eta_L(z_t,\mu_{k,t-1},\Omega_{k,t-1})^2 \\ +\frac{1}{2}P(k,\eta_H(z_t,\mu_{k,t-1},\Omega_{k,t-1})^2 \end{pmatrix} \frac{1}{\sigma_{k,t-1}\sqrt{2\pi}} e^{-\frac{(k-\mu_{k,t-1})^2}{2\sigma^2_{k,t-1}}} dk, \tag{19}$$

where $\sigma_{k,t-1} = \sqrt{\Omega_{k,t-1} - \mu^2_{k,t-1}}$. To turn this into computer code, we approximate the integral using Gaussian-Hermite quadrature. That is,

$$\mu_{k,t} = \sum_{j^*=1}^{J^*} \begin{pmatrix} \frac{1}{2}P(\mu_{k,t-1}+\sqrt{2}\sigma_{k,t-1}\zeta_{j^*};\eta_L(z_t,\mu_{k,t-1},\Omega_{k,t-1}))+ \\ \frac{1}{2}P(\mu_{k,t-1}+\sqrt{2}\sigma_{k,t-1}\zeta_{j^*};\eta_H(z_t,\mu_{k,t-1},\Omega_{k,t-1})) \end{pmatrix} \frac{1}{\sqrt{\pi}}\Omega_{j^*} \tag{20}$$

$$\Omega_{k,t} = \sum_{j^*=1}^{J^*} \begin{pmatrix} \frac{1}{2}P(\mu_{k,t-1}+\sqrt{2}\sigma_{k,t-1}\zeta_{j^*};\eta_L(z_t,\mu_{k,t-1},\Omega_{k,t-1}))^2+ \\ \frac{1}{2}P(\mu_{k,t-1}+\sqrt{2}\sigma_{k,t-1}\zeta_{j^*};\eta_H(z_t,\mu_{k,t-1},\Omega_{k,t-1}))^2 \end{pmatrix} \frac{1}{\sqrt{\pi}}\Omega_{j^*}, \tag{21}$$

where ζ_{j^*} and Ω_{j^*} are the Gauss-Hermite nodes and weights and J^* the number of quadrature nodes.

The actual algorithm. We are now all set to spell out the actual algorithm. To solve for the individual policy rule we are going to use a projection method. Thus, we need a grid for individual capital. Let κ_j with $j \in \{1, \cdots, J\}$ denote the J grid points for the worker's capital holdings, where J equals the number of coefficients of the approximating function.

i.e., the number of elements of η. For example, if $P(\cdot;\eta)$ is a second-order polynomial, then $J=3$. We want the Euler equation for the agents at the two productivity levels to hold at all $2\times J$ grid points.

We have done quite a bit of preparatory work, but we are now ready to write down a perturbation system that can be solved using a program like Dynare to generate both the nonlinear projection-based individual policy rule as well as the perturbation-based responses to aggregate disturbances.

That system of equations is given by

$$\frac{1}{r_t\kappa_j+w_t\varepsilon_L-k_{L,j}}+G'(k_{L,j})=\beta\mathbb{E}_t\left[\sum_{\widetilde{j}=1}^{2}\frac{1}{2}\frac{r_{t+1}}{r_{t+1}k_{L,j}+w_{t+1}\varepsilon_{\widetilde{j}}-P(k_{L,j};\eta_{\widetilde{j}}(\mu_{k,t},\Omega_{k,t})}\right] \tag{22a}$$

$$k_{L,j}=P(\kappa_j;\eta_L(z_t,\mu_{k,t-1},\Omega_{k,t-1})) \tag{22b}$$

$$\frac{1}{r_t\kappa_j+w_t\varepsilon_L-k_{H,j}}+G'(k_{H,j})=\beta\mathbb{E}_t\left[\sum_{\widetilde{j}=1}^{2}\frac{1}{2}\frac{r_{t+1}}{r_{t+1}k_{H,j}+w_{t+1}\varepsilon_{\widetilde{j}}-P(k_{H,j};\eta_{\widetilde{j}}(\mu_{k,t},\Omega_{k,t})}\right] \tag{22c}$$

$$k_{H,j}=P(\kappa_j;\eta_H(z_t,\mu_{k,t-1},\Omega_{k,t-1})) \tag{22d}$$

$$r_t=\alpha z_t\mu_{k,t-1}^{1-\alpha}, \tag{22e}$$

$$w_t=(1-\alpha)z_t\mu_{k,t-1}^{\alpha}, \tag{22f}$$

$$\mu_{k,t}=\sum_{j^*=1}^{J^*}\begin{pmatrix}\frac{1}{2}P(\mu_{k,t-1}+\sqrt{2}\sqrt{\Omega_{k,t-1}-\mu_{k,t-1}^2}\zeta_{j^*};\eta_L(z_t,\mu_{k,t-1},\Omega_{k,t-1}))+\\ \frac{1}{2}P(\mu_{k,t-1}+\sqrt{2}\sqrt{\Omega_{k,t-1}-\mu_{k,t-1}^2}\zeta_{j^*};\eta_H(z_t,\mu_{k,t-1},\Omega_{k,t-1}))\end{pmatrix}\frac{1}{\sqrt{\pi}}\Omega_{j^*} \tag{22g}$$

$$\Omega_{k,t}=\sum_{j^*=1}^{J^*}\begin{pmatrix}\frac{1}{2}P(\mu_{k,t-1}+\sqrt{2}\sqrt{\Omega_{k,t-1}-\mu_{k,t-1}^2}\zeta_{j^*};\eta_L(z_t,\mu_{k,t-1},\Omega_{k,t-1}))^2+\\ \frac{1}{2}P(\mu_{k,t-1}+\sqrt{2}\sqrt{\Omega_{k,t-1}-\mu_{k,t-1}^2}\zeta_{j^*};\eta_H(z_t,\mu_{k,t-1},\Omega_{k,t-1}))^2\end{pmatrix}\frac{1}{\sqrt{\pi}}\Omega_{j^*}, \tag{22h}$$

where we have used $N_t=1$. The two Euler equations hold for $j=\{1,\cdots,J\}$ and thus represent $2J$ equations.

For a choice of $G(k)$ and numerical values for the parameter values, this is a system that can be solved with perturbation methods. How is it possible that Dynare can solve for a higher-order projection-based individual policy rule even when it uses first-order perturbation? And how is it possible that this simple solution method can solve a model with heterogeneous agents in the presence of aggregate uncertainty? And – conditional on having reasonable initial values for its steady state – does so almost instantly.

Finding the perturbation solution of this system consists of two steps.

1. Find the steady state.

2. Determine how this steady-state solution changes if the stochastic variable changes. In this system, the only stochastic variable is z_t. The realizations for $e_{i,t}$ are parameters in the perturbation system.

At the steady state, z_t is constant. Solving for the steady state means solving for the constant values of the J constant values of η, the constant values of μ_k and Ω_k, and the constant implied values of r and w. The steady state version of the system in equation (22) gives the system of equations to solve for these values. The solution for μ_k implies values for r and w which imply a partial-equilibrium solution to the nonlinear policy function of the individual characterized by the η coefficients. In equilibrium, μ_k is such that given the value of Ω_k and the distributional form assumption, the distribution remains constant. Solving for the steady state is typically the hardest part of deriving a perturbation solution. And one may need good initial conditions.[8] Note that this "steady state" solution takes into account precautionary savings, since the $2J$ Euler equations take into account the nonlinearities of the original problem are are not an approximation around the steady-state solution in which $e_{i,t}$ is also constant.

The perturbation solution of this system tells us how this steady state solution consisting

[8]This system is actually well-behaved and not hard to solve. Specifically, given the monotonicity properties an iteration procedure that adjusts the value of r depending on whether there is excess demand or excess supply works very well. This can be used to solve the system for some parameter values and these can be used as initial conditions in a program like Dynare if the system needs to be solved for several parameter values as is the case when one estimates the model.

of a nonlinear individual policy rule and a cross-sectional distribution responds to aggregate uncertainty. To fully understand this, we should know what is known and what is not known in the system. The state variables in the system above are z_t, $\mu_{k,t-1}$, and $\Omega_{k,t-1}$ and a perturbation solution consists of functions of these three variables. Now consider the $2J$ equations of the Euler equation at the $2J$ grid points. Imagine we substitute out this period's and next period's interest rate and the wage rate on both sides of the Euler equation. Then we are left with $2J+2$ equations in the $2J+2$ unknowns. The perturbation approach will tell us what these functions are as a function of z_t, $\mu_{k,t-1}$, and $\Omega_{k,t-1}$.[9]

Occasionally binding constraints, the penalty function, and the Reiter method. With the specifics of the Reiter method spelled out, I can make a fundamental point about the use of occasionally binding constraints as used in Reiter (2009) and in Papp and Reiter (2020). The Reiter method can deal with occasionally binding constraints. But it has one practical and one fundamental problem. With an inequality constraint, optimal behavior is characterized by the two-part Kuhn-Tucker conditions. That is, an agent's capital choice may be determined by the Euler equation (when the agent is not constrained) *or* by the constraint. In terms of the perturbation system given in equation (22), this means that for some js, i.e., for some grid points, the Euler equations remains the right equation, but for others it has to be replaced with the borrowing constraint. The practical problem is that one has to find out at which grid points the constraint is binding and at which grid points it is not binding. One has to do this before one obtains the perturbation solution. This would not be problematic if one solves the model for one set of parameter values, but would be problematic if one has to solve it many times at different parameter values, which would be the case if one estimates the model.

There is a more fundamental problem. Implicit in the perturbation approach is that the equations do not change if there are changes in the aggregate disturbance, z_t. That

[9]A first-order perturbation solution will give a first-order approximation for the laws of motion for $\mu_{k,t}$ and $\Omega_{k,t}$. Since we actually have analytical (nonlinear) expressions for these state variables for given policy functions, we could use these instead of the first-order approximations when generating model data.

is, if an agent is at the constraint at the steady state value of z_t, then the perturbation solution assumes that this remains the case if z_t changes. This may be plausible if one has just a few grid points with large gaps in between them. The cut-off level of capital at which the constraint becomes binding may then always be in between the same two grid points. However, with nontrivial idiosyncratic risk one may very well need a fine grid to get an accurate individual policy rule. Then this property is unlikely to remain true unless fluctuations in z_t are really small.

These problems are not present if penalty functions are used instead of borrowing constraint. Moreover, in many cases penalty functions actually make more economic sense than inequality constraints.[10] Note that an inequality constraint is also a penalty function and a very peculiar one. That is, the penalty is equal to 0 as long as $k_{i,t} \geq \overline{k}$ and infinite if $k_{i,t} < \overline{k}$. It seems more sensible that it becomes gradually more difficult to borrow and only as one borrows more and more it will be met with prohibitive cost increases.[11] In the model above, I put the penalty function in the utility function, but that was mainly because such a penalty function only enters the model equations through the Euler equation. A more realistic implementation would consist of letting a penalty function affect the interest rate paid on short positions with the penalty increasing in the amount borrowed.

3. The contributions of Papp and Reiter (2020)

The motivation of Papp and Reiter (2020) is the recent surge in the use of micro-level data to estimate macroeconomic models. Estimation will require evaluation of an objective function such as the Likelihood or the posterior at many parameter values. Consequently, one would need a very fast solution algorithm. The Reiter approach has the potential to do exactly this. If the cross-sectional information used is part of the perturbation system,

[10] The natural borrowing limit is a theoretical justification for an inequality constraint. In practice, however, natural borrowing constraints are not used because they turn out to be so weak that they do not constrain agents.

[11] Den Haan and De Wind (2012) show that models with smooth penalty functions can generate behavior that is very similar to that obtained in models with occasionally binding inequality constraints.

such as $\mu_{k,t}$ or $\Omega_{k,t}$ in the system above, then one can use the Reiter approach without any modification. However, cross-sectional information is usually only available at an annual frequency whereas economic aggregates such as GDP are typically available at a quarterly frequency.

One cannot derive the law of motion of annual dispersion by using only the law of motion of quarterly dispersion. One would also have to use the individual laws of motion. This means we will need some type of panel simulation. The idea of Papp and Reiter (2020) is to use the idea underlying the Reiter approach to do this.

Specifically, suppose one wants to derive a perturbation-based law of motion for the annual dispersion of capital holdings when the other data is quarterly. It involves the following steps.

1. Start with a large number of workers, N, with capital holdings and productivity levels distributed consistent with the distribution of the economy without aggregate uncertainty and with idiosyncratic uncertainty.[12]

2. Write down the equations that describe the choices of each agent for the first four periods. Exactly as one would in an actual simulation, the idiosyncratic productivity level changes over time according to its law of motion.

3. Write down the expression for the cross-sectional dispersion of annual capital holdings.

This is a perturbation system in which the object of interest, the cross-sectional dispersion of annual capital holdings will be a function of the aggregate state variables over the last four periods.

Why use a panel simulation of a finite number of individuals and not integrate over all possible outcomes for each agent? With only two possible realizations for $e_{i,t}$ and four periods, there are sixteen possible outcomes to consider and this alternative might still be doable. But is unlikely to be feasible if more periods are involved and/or more realizations.

[12]As shown in Algan et al. (2014), simulating an economy with a histogram is more accurate than simulating with a large number of agents, since it completely eliminates sampling variation. But a quarterly time series of histograms does not allow us to construct a time series for the dispersion of *individual* annual capital holdings. The latter requires keeping track of individual workers and the cross-sectional histograms do not do that.

Papp and Reiter (2020) implement this idea in a model with discrete choice, namely the agent has to choose beteen working and not working. This leads to an additional complication. With a discrete choice decision, there is a cut-off level which will depend not only on the worker's idiosyncratic productivity level, but also on the aggregate state variables. This creates a problem for the Papp-Reiter perturbation approach if a finite number of individuals is considered. The reason is that the perturbation approach is based on partial derivatives and these will not induce changes in the discrete choice employment decisions of the workers included in the panel.[13] Papp and Reiter (2020) propose a smoothing technique to deal with this problem.

If penalty functions are used instead of inequality constraints, then there is an alternative solution which would also be able to deal with discrete choice. This would consist of simulating the economy with N agents for T periods with a realistic volatility for z_t. When using a realistic volatility for z_t, then one would capture changes in the cut-off level. To obtain the (linearized) law of motion for the cross-sectional statistic of interest, one simply runs a regression. T would have to be sufficiently large to avoid sampling variation. As explained above, this doesn't work with inequality constraints since the individual policy rules are derived under the assumption that the area where the inequality constraint binds does not change if z_t changes which is only guaranteed for arbitrarily small changes in z_t.

The paper documents that the proposed procedure is feasible. Specifically, it estimates second moments of the shock processes and it documents that the estimate of the variance of the idiosyncratic shock is substantially smaller if cross-sectional information is used.

4. Evaluation: Technique ahead of empirical methodology

The Papp-Reiter implementation of the Reiter approach is a clever one and may very well stimulate the use of cross-sectional information in the estimation of business cycle models.

I wonder, however, whether it makes sense to increase the sophistication of our estima-

[13] The probability that a worker is exactly at the cut-off level happens with probability zero. At the cut-off level the worker is indifferent. Thus even then, there is an original discrete choice which is the same as the preferred one if z_t changes.

tion techniques without dealing with a much more fundamental problem. Prescott (1986) writes in his "Theory Ahead of Business cycle measurement paper," that *"The models constructed within this theoretical framework are necessarily highly abstract. Consequently, they are necessarily false, and statistical hypothesis testing will reject them."* Prescott argued that economic variables are not always measured in conformity with economic theory. It is true that we now have more and better data. Moreover, our theories have become richer and including heterogeneity has been an important contributor in doing so. Nevertheless, I think that it is still the case that all macroeconomic models are wrong and that we should take this into account when we estimate them.

Empirical estimation of structural macroeconomic models is difficult. Over time, the profession has used quite different strategies that were disgarded after criticism or disappointing performance. In the tradition of Tinbergen, the first approach consisted of estimating time-invariant relationships between macroeconomic aggregates without taking into account cross-equation restrictions. The Lucas critique revealed the flaws of this approach. In the eighties, the Generalized Methods of Moments (GMM) approach of Hansen (1982) became popular. The advantage of GMM is that one can estimate model parameters without writing down the full model. Consequently, one can focus the estimation on only those parts of the model that one is more confident about. Unfortunately, the small-sample properties of the GMM estimator turned out to be quite bad.[14] Calibration also has the advantage that parameters are chosen using only the dimensions of the model in which one has confidence. The disadvantage of calibration is that one does not have a statistical measure of goodness of fit. That is, one has to eyeball the results and decide whether the predicted model properties are close to their sample counterparts. Simulated Method of Moments (SMM) is like calibration but with proper statistical inference.

Although Simulated Method of Moments has sporadically been used in macroeconomics, it has now been completely taken over by full information methods such as Maximum Likelihood or the Bayesian version of Maximum Likelihood. Maximum Likelihood is rarely feasible unless

[14]See, for example, the 1996 special issue of the Journal of Business and Economic Statistics, vol. 14(3).

– as is the case in Papp and Reiter (2020) – one estimates just a few parameters. The reason is that finding the global maximum of the Likelihood function is extremely difficult given that it often has a highly erratic shape. By combining the Likelihood with an informative prior, the problem becomes much more tractable, especially with the help of Markov Chain Monte Carlo techniques to trace the posterior. And this has become the dominant empirical technique to estimate macroeconomic models. And if the Papp-Reiter approach is incorporated into a non-trivial empirical project, then this is most likely the empirical strategy that will be used.

Whereas the small-sample properties of GMM were scrutinized extensively in the eighties and nineties, we do not have such knowledge of the properties of MCMC techniques for typical macroeconomic applications. The Brooks-Gelman statistics are diagnostics to evaluate whether the generated sequence of parameter values resembles the posterior distribution one tries to recover. But there are no proper metrics to judge what are good and bad outcomes for these statistics. Not that it matters, because the profession does not expect authors to report these statistics and discuss whether the posterior is characterized accurately!

An even more striking omission from the empirical literature is dealing with the problem that Prescott already pointed out in 1986 and that is that all macroeconomic models are wrong. Calibration, GMM, and SMM are attempts to deal with this although in a limited way. But the literature has moved away from these methods in favor of full-information methods. Papp and Reiter (2020) is just one example of many. There are some papers that do explicitly deal with misspecification when estimating structural models.[15] But most do not. Unfortunately, even the consequences of minor misspecifications can be substantial as is documented in Den Haan and Drechsel (2019).

This detour through the history of empirical macroeconomics was done for a reason. So let's get back to the empirical excercise in Papp and Reiter (2020). Couldn't one argue that the Papp-Reiter approach reduces the problems of misspecification and data limitations and is, thus, a step in the right direction? I am sure this will be true in some applications. But the increased complexity and sophistication of the procedure also makes it much harder to

[15] See Den Haan and Drechsel (2019) and the discussion therein.

understand what happens in the estimation procedure. How does the smoothing technique to deal with the discrete choice of the agent affect small sample properties? How does it interact with misspecification of the model? Are richer models necessarily closer to reality or do they make misspecification worse because they have to take a stand on more issues? Are the small-sample properties of larger models better or worse than those of smaller models. Do MCMC techniques work better for posteriors of complex models? We don't know the answer to any of these questions.

It is great to see numerical and estimation techniques evolve over time. But it has to go hand in hand with more research on the question how one should do empirical research on how these currently popular methods perform in samples of realistic length. Both when the empirical model is correctly specified and when it is not. Perhaps we should before we estimate our models using these methods.

References

Algan, Y., Allais, O., Den Haan, W.J., 2008. Solving heterogeneous-agent models with parameterized cross-sectional distributions. Journal of Economic Dynamics and Control 32, 875–908.

Algan, Y., Allais, O., Den Haan, W.J., Rendahl, P., 2014. Solving and simulating models with heterogeneous agents and aggregate uncertainty, in: Judd, K.L., Schmedders, K. (Eds.), Handbook of Computational Economics. Elsevier, Amsterdam, The Netherlands.

Den Haan, W.J., De Wind, J., 2012. Nonlinear and stable perturbation-based approximations. Journal of Economic Dynamics and Control 36, 1477–1797.

Den Haan, W.J., Drechsel, T., 2019. Agnostic structural disturbances (asds): Detecting and reducing misspecification in empirical macroeconomic models. Unpublished manuscript, London School of Economics.

Den Haan, W.J., Rendahl, P., Riegler, M., 2017. Unemployment (fears) and deflationary spirals. Journal of the European Economic Association 16, 1281–1349.

Hansen, L.P., 1982. Large sample properties of genegeneral method of moments. Econometrica 50, 1029–1054.

Papp, T., Reiter, M., 2020. Estimating linearized heterogeneous agent models using panel data. Journal of Economic Dynamics and Control .

Prescott, E.C., 1986. Theory ahead of business cycle measurement. Federal Reserve Bank of Minneapolis Quarterly Review 10, 9–22.

Reiter, M., 2009. Solving heterogenous-agent models by projection and perturbation. Journal of Economic Dynamics and Control 33, 649–665.

Winberry, T., 2018. A method for solving and estimating heterogenous agent macro models. Quantitative Economics 9, 1123–1151.