Estimating and evaluating DSGE models without filtering to recover states

Tom Holden

http://www.tholden.org/

PhD Macroeconomics, Semester 2

Outline of today's talk

- A reminder of a baseline RBC model.
- Calibration.
- Wedge accounting and ML estimation of simple models with at least as many observables as shocks.
- GMM/SMM.

Reading for today

- Canova: “Methods for applied macroeconomic research”.
  - Chapter 5 covers GMM/SMM.
  - Chapter 7 covers calibration.

- Chari, Kehoe, McGrattan (2007):
  - Covers wedge accounting.
  - Another classic paper which you ought to be familiar with.
  - http://www.econ.umn.edu/~pkehoe/papers/CKMeconometrica2007.pdf

- Gali and Gertler (1999)
  - Early paper applying GMM to the NK Phillips curve.
  - https://www.sciencedirect.com/science/article/pii/S0304393299000239

- Holden (2014): “Estimating non-linear models”
  - I’ll email this to you.

Taking the model seriously

- We saw in the penultimate lecture that it is incredibly hard to obtain reliable identification of shocks via “theory-free” techniques.
  - Identification of shocks in a VAR requires imposing restrictions from theory, restrictions which may only be justified by a model.

- This begs the question of why the model is not directly taken to the data instead. I.e., why we don’t “take the model seriously”.

- This week we shall look at techniques for doing this that do not involve econometric techniques for recovering the state variables of a model.
  - Either we will not need to work out these states, or they will be directly observable.

Baseline RBC model: Household's Problem

- The representative household maximises:

$$U_t(K_{t-1}, B_{t-1}) = u(C_t, L_t) + \beta \mathbb{E}_t U_{t+1}(K_t, B_t),$$

- subject to the budget constraint:

$$C_t + (1 + \tau_{I,t}) I_t + B_t = (1 - \tau_{L,t}) W_t L_t + r_t \omega_t K_{t-1} + R_{t-1} B_{t-1} + T_t,$$

- where:
  - K_t is capital, which evolves according to $K_t = (1 - \delta(\omega_t)) K_{t-1} + I_t$,
  - ω_t is capacity utilisation,
  - B_t is holdings of a zero net supply real bond, which pays an interest rate R_t,
  - C_t is consumption, L_t is labour supply, I_t is investment,
  - W_t is the market wage, r_t is the rental rate of capital,
  - T_t is net lump-sum transfers,
  - $\tau_{I,t}$ is an exogenous investment tax, and $\tau_{L,t}$ is an exogenous labour tax.

Baseline RBC model: Household's FOCs

- Lazy approach to solving this problem. Just sub-in constraints:

$$U_t(K_{t-1}, B_{t-1})$$
$$= u\Big(\big(1 - \tau_{L,t}\big)W_t L_t + r_t \omega_t K_{t-1} + R_{t-1} B_{t-1} + T_t$$
$$- \big(1 + \tau_{I,t}\big)\big(K_t - \big(1 - \delta(\omega_t)\big)K_{t-1}\big) - B_t, L_t\Big) + \beta \mathbb{E}_t U_{t+1}(K_t, B_t),$$

- FOCs:
  - $\big(1 - \tau_{L,t}\big)W_t u_1(C_t, L_t) + u_2(C_t, L_t) = 0.$
  - $\beta \mathbb{E}_t \Big(r_{t+1}\omega_{t+1} + \big(1 + \tau_{I,t+1}\big)\big(1 - \delta(\omega_{t+1})\big)\Big) u_1(C_{t+1}, L_{t+1}) = \big(1 + \tau_{I,t}\big)u_1(C_t, L_t).$
  - $r_t = \delta_1(\omega_t)\big(1 + \tau_{I,t}\big).$
  - $\beta \mathbb{E}_t R_t u_1(C_{t+1}, L_{t+1}) = u_1(C_t, L_t).$

Baseline RBC model: Firms

- The representative price taking firm chooses how much capital and labour to hire in order to maximise profits, where profits from a firm hiring are given by:

$$Y_t - r_t \widetilde{K}_t - W_t \widetilde{L}_t,$$

- where:
  - $Y_t = A_t F(\widetilde{K}_t, \widetilde{L}_t)$ is firm output,
  - $\widetilde{K}_t$ is capital demand, $\widetilde{L}_t$ is labour demand, and
  - A_t is an exogenous productivity process.

- So:

$$A_t F_1(\widetilde{K}_t, \widetilde{L}_t) = r_t,$$
$$A_t F_2(\widetilde{K}_t, \widetilde{L}_t) = W_t,$$

Baseline RBC model: Market Clearing

- Output goods market:

$$Y_t = C_t + I_t + G_t,$$

  - where G_t is an exogenous government consumption process.

- Labour market:

$$\tilde{L}_t = L_t.$$

- Capital market:

$$\tilde{K}_t = \omega_t K_{t-1}.$$

Baseline RBC model: Special case

- Suppose $u(C,L)=\frac{C^{1-\rho}-1}{1-\rho}-\frac{1}{1+\nu}L^{1+\nu}$, $F\left(\widetilde{K},\widetilde{L}\right)=\widetilde{K}^{\alpha}\widetilde{L}^{1-\alpha}$, and $\delta(\omega)=\delta_0+\delta_2\omega^2$, then the equilibrium conditions simplify to:
    - $\alpha\frac{Y_t}{\omega_t K_{t-1}}=r_t$,
    - $(1-\alpha)\frac{Y_t}{L_t}=W_t$,
    - $\left(1-\tau_{L,t}\right)W_t C_t^{-\rho}=L_t^{\nu}$.
    - $\beta\mathbb{E}_t\left(r_{t+1}\omega_{t+1}+\left(1+\tau_{I,t+1}\right)(1-\delta_0-\delta_2\omega_{t+1}^2)\right)C_{t+1}^{-\rho}=\left(1+\tau_{I,t}\right)C_t^{-\rho}$.
    - $r_t=2\delta_2\omega_t\left(1+\tau_{I,t}\right)$.
    - $\beta\mathbb{E}_t R_t C_{t+1}^{-\rho}=C_t^{-\rho}$.
    - $Y_t=A_t(\omega_t K_{t-1})^{\alpha}L_t^{1-\alpha}=C_t+I_t+G_t$.
    - $K_t=(1-\delta_0-\delta_2\omega_t^2)K_{t-1}+I_t$.

Calibration

- The early (modern) macro literature relied heavily on calibration.
- Calibration is what you do when you wish to take the model seriously, but not *too* seriously…
  - Macro-calibrators tend to view it as inappropriate to apply econometric techniques to macro models when they are no more than fables.
- Rather than expecting the model to generate historical time paths, you instead require it to approximately match certain stylized facts of the data.
- For example, in macro we often seek to match correlations between standard macro time-series and the leads and lags of output.
- Or to match the broad shape of certain impulse responses.

Calibrating the RBC model (1/3)

- One common approach to calibration is to choose parameters in order to match the long-run means of certain great ratios.
- The ratios to look at are directly suggested by the equilibrium conditions of the model.
  - E.g. from the labour FOC: $\alpha = 1 - \frac{W_t L_t}{Y_t}$, so α may be calibrated using one minus the observed long-run mean of $\frac{W_t L_t}{Y_t}$.
  - Note that in the model these relationships hold exactly, though they never will in the data. In this sense we are relaxing the restrictions of the model.

- Other parameters may be estimated via regression.
  - E.g. assuming that we observe taxes, $\log\left(1 - \tau_{L,t}\right) W_t = \nu \log L_t + \rho \log C_t$ gives a regression we can run to calibrate ν and ρ.
  - Note of caution though: if we reinterpret $\tau_{L,t}$ as reflecting any of the many distortions which may enter the labour FOC, then it is not observable, and will appear as the error on the RHS, however it will be correlated with the RHS variables, rendering the estimates inconsistent.
  - Similar endogeneity problems may affect even parameters estimated via long-run ratios.

Calibrating the RBC model (2/3)

- If we are working with a first order approximation to a model in levels, then there's no difference between the simulated means of our model variables, and their steady-state.
- This gives other avenues for calibration.
  - For example, in steady-state $\beta R = 1$, so we may calibrate β as the observed long-run mean of $\frac{1}{R_t}$.

- Alternatively, we may simulate the model many times, with different values of the parameters, and keep the ones that come closest to matching the relevant moments of the data.
  - E.g. we choose β, so that R_t matches the return on inflation indexed bonds.

- Since practitioners are often reluctant to use NIPA measures of stocks, and so most RBC models have one or more unobservable states, in practice many or all parameters might be calibrated through simulation, or through the steady-state.
  - One difficulty here is that there are often many moments that could be used to calibrate each parameter, and the choice of calibration target can heavily influence results.

Calibrating the RBC model (3/3)

- A final approach to calibration is to use micro estimates of the relevant parameters.

- This is perhaps the ideal approach, as if parameters are estimated from micro-data, then the model's performance on macro data gives a clean measure of the model's performance.

- The issue here is that what the micro-economists measure may not map neatly to the relevant macro quantity.
  - For example, micro estimates of the Frisch elasticity of labour supply are often well below 0.5. However, these are often estimates of the intensive margin for the main wage earner in a household, whereas in macro we are interested in the combination of intensive and extensive margins for the entire population.

How seriously should we take calibration?

- Canova’s textbook takes it remarkably seriously, giving quite sophisticated ways of quantifying uncertainty of the “estimates”, and various ways of comparing the performance of differing calibrated models.
  - Some macro-calibrators might be sceptical of such techniques on philosophical grounds though.
  - You might also legitimately wonder whether there is any point applying sophisticated techniques to fix unsophisticated ones, when there are other alternatives out there.

- One note of caution: be wary of the difference between calibration and parameterisation.
  - I have seen big name economists stand-up and shout at the speaker for using the former when they meant the latter, so some people take calibration very seriously!

Wedge accounting: Introduction

- The four shocks in our business cycle model were not chosen accidentally. Rather, they are the shocks used by Chari, Kehoe and McGrattan in their 2007 wedge accounting paper.

- However, in that paper those four shocks are stripped of their interpretation as e.g. tax shocks.
- Instead, the four shocks are just interpreted as reduced form “wedges”, which, much like the shocks in a reduced form VAR, may be a combination of many structural shocks.
  - Indeed, the four wedges are allowed to follow an arbitrary $\mathrm{VAR}(1)$ in that paper.

- They estimate the model on output, consumption, investment and hours, and so they have four shocks and four observables.
  - With as many shocks as observables, there are time series for the wedges which will exactly generate the data.
  - They then perform experiments shutting off all but one wedge in order to see their relative importance, to guide future modelling.

Wedge accounting: Equivalencies

- One of the main results of the Chari, Kehoe, McGrattan paper is that almost all macro models are reducible to this RBC+4 wedge set-up, as long as the wedges are appropriately defined.
  - For example, financial frictions might show up as an investment wedge, and sticky wages might show up as a labour one.

- The RBC model we presented initially had variable capacity utilisation (which their model doesn't). As an illustration of the idea, we show that this is observationally equivalent to a model without it, but with redefined wedges (denoted by $*$).
  - We need $A_t^* {K_{t-1}^*}^{\alpha} = A_t(\omega_t K_{t-1})^{\alpha}$.
  - Note: $r_t^* = \alpha \frac{Y_t}{K_{t-1}^*} = \alpha \frac{Y_t}{\omega_t K_{t-1}} \frac{\omega_t K_{t-1}}{K_{t-1}^*} = r_t \frac{\omega_t K_{t-1}}{K_{t-1}^*}$.
  - Then we just define $\tau_{I,t}^*$ as the solution to the equation:

$$\tau_{I,t}^* - \tau_{I,t}$$

$$= \beta \mathbb{E}_t \left(r_{t+1} \left(\frac{\omega_{t+1} K_t}{K_t^*} - \omega_{t+1} \right) + \left(1 + \tau_{I,t+1}^*\right)(1 - \delta) - \left(1 + \tau_{I,t+1}\right)(1 - \delta_0 - \delta_2 \omega_{t+1}^2) \right) \left(\frac{C_{t+1}}{C_t} \right)^{-\rho}$$

Maximum likelihood estimation of non-linear models when states are inferable without solving the model (1/2)

- The Chari, Kehoe, McGrattan paper estimates the wedges in a linearised version of the model, using the Kalman filtering techniques we'll cover in the next topic.
- But is linearization really necessary here? Suppose we knew K_0, and all of the model's parameters. Could we then work out the time-paths of the wedges?
  - If we could, then we can work out the likelihood, and so we could estimate K_0 and the model's parameters via ML (or penalized ML). Details on next slide…
  - It is trivial to infer the government wedge, from $G_t = Y_t - C_t - I_t$.
  - From the equations $(1-\alpha)\frac{Y_t}{L_t} = W_t$ and $\left(1-\tau_{L,t}\right)W_t C_t^{-\rho} = L_t^{\nu}$, we can also infer the labour wedge very easily.
  - Suppose that capacity utilisation is observable. Then, from the law of motion for capital we can infer capital in all periods without solving the model. Once we know capital, from the production function we can infer the productivity wedge.
  - This just leaves the investment wedge. With observable capacity utilisation, this may be inferred from $\alpha \frac{Y_t}{\omega_t K_{t-1}} = r_t = 2\delta_2 \omega_t \left(1+\tau_{I,t}\right)$.
    - Will mean that investment may not be matched exactly, however, if the model is misspecified.
    - Perhaps we ought to have included a capacity utilisation wedge manifesting as varying δ_2?

Maximum likelihood estimation of non-linear models when states are inferable without solving the model (2/2)

- Note the following minimal requirements for the above procedure:
    1. There are at most as many shocks/wedges as there are observables.
    2. When equations are combined so that the only variables are observables, states, or shocks/wedges, then there are at least as many purely backwards looking equations left as there are shocks/wedges.

- Once we have processes for the shocks/wedges, we can then estimate time-series for them using standard methods.
    - If we are viewing the stochastic processes as wedges, then we will allow for interdependence between processes, so we will estimate a VAR type model.
    - If we are viewing the stochastic processes as structural, then we will estimate independent AR type models.
    - Maximising the log-likelihood of the VAR model, or the sum of the log-likelihoods of the AR models, would give a "non-linear least squares" estimate of the original model.
    - More properly, we should adjust the likelihood for the non-linear transformation of the model. See https://en.wikipedia.org/wiki/Integration_by_substitution#Application_in_probability and the course notes chapter.

Maximum likelihood estimation of non-linear models when states are *mostly* inferable without solving the model

- Suppose we did not have variable capacity utilisation in the model.
- Then we would be forced to infer the investment wedge from:
$$\beta \mathbb{E}_t \left(r_{t+1} + \left(1 + \tau_{I,t+1}\right)(1-\delta) \right) C_{t+1}^{-\rho} = \left(1 + \tau_{I,t}\right) C_t^{-\rho}.$$
- Since the LHS is a function of the entire distribution of $\tau_{I,t+1}$, differing stochastic processes for $\tau_{I,t}$ will lead to differing decision rules when we solve the model.
- This is further complicated as $\mathbb{E}_t \left(r_{t+1} + \left(1 + \tau_{I,t+1}\right)(1-\delta) \right) C_{t+1}^{-\rho}$ may be increasing, decreasing, or non-monotonic in $\tau_{I,t}$.
- In this case, we must solve the model for the current parameters, and then use simulation or quadrature techniques in order to evaluate the integral.
  - This may be a bit slow as we will be solving a non-linear equation involving an integral.

The Generalized Method of Moments (GMM) Approach: Motivation

- The first order conditions to a DSGE model can always be written in the form:
$$\mathbb{E}_t F(x_{t-1}, x_t, x_{t+1}, \theta) = 0,$$
  - where x_t collects the shocks, state and control variables of the model (possibly augmented by lags), and θ collects the parameters of the model.

- Since this expectation is conditional on the time t information set, for any variable z_t known at t:
$$\begin{aligned}\mathbb{E}[F(x_{t-1}, x_t, x_{t+1}, \theta) z_t'] &= \mathbb{E}\big[\mathbb{E}_t[F(x_{t-1}, x_t, x_{t+1}, \theta) z_t']\big] \\ &= \mathbb{E}[\mathbb{E}_t[F(x_{t-1}, x_t, x_{t+1}, \theta)] z_t'] \\ &= \mathbb{E}[0 z_t'] = 0.\end{aligned}$$
- Thus given z_t (the "instruments"), the model implies a set of moment restrictions.
- The usual choice for $z_t = [1 \quad x_t \quad x_{t-1} \quad \cdots \quad x_{t-l}]'$, for some $l > 0$.

The GMM Approach: Idea behind the estimate

- Let us define $g(Y_t, \theta) = \operatorname{vec} F(x_{t-1}, x_t, x_{t+1}, \theta) z_t'$.
  - Possible by stacking the leads and lags of x_t along with z_t in Y_t.
- The basic idea of GMM is to replace the theoretical moment constraint:
$$\mathbb{E}[g(Y_t, \theta_0)] = 0,$$
- with its sample equivalent:
$$\widehat{m}(\theta) \coloneqq \frac{1}{T} \sum_{t=1}^{T} g(Y_t, \theta) = 0.$$
- We then use nonlinear minimization to choose θ to minimize the discrepancy between the LHS and RHS of this equation.
  - Note that for the moment we are assuming all of the components of x_t are observable.
- In practice, some moment restrictions will be more or less informative than others, so we use a weighting matrix W and solve:
$$\widehat{\theta} = \arg\min_{\theta} \widehat{m}(\theta)' W \widehat{m}(\theta).$$

The GMM Approach : Choice of W

- In the absence of serial correlation in $g(Y_t, \theta)$, the optimal W is (given certain assumptions) proportional to:

$$\left[\mathbb{E}\frac{1}{T}\sum_{t=1}^{T} g(Y_t, \theta_0) g(Y_t, \theta_0)'\right]^{-1}.$$

- In practice, W is usually estimated from the data.
- One obvious algorithm is to first fix $\widehat{W}_0$ (e.g. take $\widehat{W}_0 = I$), then produce GMM estimates of θ, say $\hat{\theta}_1$, using $W = \widehat{W}_0$. Then we may derive an asymptotically efficient W using:

$$\widehat{W}_1 \coloneqq \left[\frac{1}{T}\sum_{t=1}^{T} g\left(Y_t, \hat{\theta}_1\right) g\left(Y_t, \hat{\theta}_1\right)'\right]^{-1}.$$

- We can then produce a new GMM estimate of θ, say $\hat{\theta}_2$, using $W = \widehat{W}_1$.
- Obviously this procedure may be repeated, and doing so tends to increase small-sample efficiency, however asymptotically even $\hat{\theta}_2$ is efficient.

The GMM Approach : Dealing with serial correlation

- In macro contexts, $g(Y_t, \theta)$ will often be serially correlated however.

- In this case, the optimal W is given by the variance at frequency zero of $g(Y_t, \theta)$.

- As we mentioned last week, the variance at frequency zero is an ill-posed problem, since it requires an infinite amount of data. (See Potscher (2002).)
- However, it turns out that this ill-posedness miraculously does not impact our ability to get efficient GMM estimates. (See Perron and Ren (2010).)

- Most standard software for GMM will use a Newey-West type estimator.
  - Details are in Canova’s textbook if you’re interested.

The GMM Approach : J-tests

- Since we may include arbitrarily many lags in z_t, it is easy to produce set-ups in which there are more moment constraints than parameters.
- Such a situation is said to be "over-identified".

- The J-test uses this over-identification to test the model's restrictions.

- It uses the fact that $T\widehat{m}\left(\hat{\theta}\right)'\widehat{W}\widehat{m}\left(\hat{\theta}\right)$ has a known distribution with finite variance under the null that the model is correct, but is otherwise unbounded.

The GMM Approach : Applications in macro

- Canova's textbook gives an application to the estimation of an RBC model, which is worth reading.
    - There are also extensive examples in my course notes.
- Another nice application is the estimation of the NKPC.
    - This will also illustrate how shocks in equations may be dealt with.
- Suppose:

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \kappa y_t + \varepsilon_t,$$

    - where π_t is inflation, y_t is some observable measure of real marginal costs, such as an output gap, and ε_t is an i.i.d. cost push shock.
- Then, taking expectations at $t-1$:

$$\mathbb{E}_{t-1}[\beta \pi_{t+1} + \kappa y_t - \pi_t] = 0,$$

- So if z_{t-1} is known at $t-1$:

$$\mathbb{E} z_{t-1}[\beta \pi_{t+1} + \kappa y_t - \pi_t] = 0.$$

- Careful construction of the instrument set can allow this to be generalized to the case when ε_t is serially correlated.

The Simulated Method of Moments (SMM)

- The Simulated Method of Moments is applicable even when GMM is not, due, for example, to unobserved variables.
  - The idea is that rather than matching some theoretical moments, which we may evaluate analytically, we instead choose parameters so that the sample average of some moment over many simulations of our model is equal to the sample average of the same moment in the data.
- SMM can be thought of as calibration 2.0. Just as when calibrating we might aim to match cross correlations between a range of variables, we can match exactly the same cross correlations via SMM.
  - Unlike calibration, SMM has rigorous econometric theory behind it, and it permits the derivation of standard errors, and the performance of J-tests.
  - The estimators are very similar to the corresponding GMM ones.
  - And SMM is even in Dynare nowadays!
- In the course notes I discuss more sophisticated alternatives to SMM which rely on analytic moment calculations, rather than simulation.

Conclusion and recap

- Inversion of observables to find shocks may be possible if there are at least as many observables as shocks.

- It can be helpful to add additional non-structural shocks to capture missing features of your model.

- Calibration can be helpful for getting vaguely plausible parameters.

- GMM/SMM is calibration “done right”.