# MA Advanced Macroeconomics: 10. Estimating DSGE Models

Karl Whelan

School of Economics, UCD

Spring 2016

# Early Approaches to Parameterising DSGE Models

- Because DSGE models are relatively complex, early researchers did not attempt to use econometrics to estimate their parameters.
- Instead the early models were "calibrated" by picking parameter values that matched certain steady-state values (labour share of income, capital-output ratio and so on) with historical average values or else by using estimates of parameters from microeconomic studies (coefficient of relative risk aversion, labour supply elasticities, depreciation rates).
- A more formal approach was "indirect inference"– choosing parameters to match certain moments of the data. For example, Rotemberg and Woodford (1997) chose parameters that delivered impulse responses to monetary policy shocks that came closest to matching the data.
- This approach has been developed to be considerably more sophisticated than the Rotemberg-Woodford paper (see the Hall et al paper on the website) but it still falls well short of using all the information in data.
- For example, monetary policy shocks typically only account for a small percentage of the variation in the sample, so why focus only on this?

# Modern Approaches

- Most state-of-the-art papers estimating DGSE models now use Bayesian econometric techniques that are similar to (but not the same as) the methods used for estimating VARs that we discussed earlier.
- To understand these techniques, we will need to cover a number of new issues.

  1. Breaking our model into observable and unobservable variables.
  2. The role played by the number of shocks in DSGE models.
  3. Kalman filter estimation of state-space models.
  4. Bayesian methods for DSGE.

# Starting Point: A Solved Model

- The modern approach to estimation starts with the solved version of the log-linearised model. Let's recall what is meant by that.
- Suppose we have a model described by

$$KZ_t = AZ_{t-1} + BE_t Z_{t+1} + HX_t$$

where $Z_t$ is a set of $n$ endogenous variables and $X_t$ is a set of $k$ exogenous variables that evolve according to

$$X_t = DX_{t-1} + \epsilon_t$$

- Then we showed before that the model has a solution of the form

$$Z_t = CZ_{t-1} + PX_t$$

where $C$ depends on the coefficients in $A$ and $B$ and $P$ depends on the coefficients in $A$, $B$, $H$ and $D$.

- This can be simulated to establish properties of the model. But how do we go from observable data back to obtain the "'best" (however defined) estimates of the coefficients in $A$, $B$, $H$ and $D$? How this works depends on the kind of model and the kind of data that we have.

# All Variables Observable

- Suppose that all variables in $X_t$ and $Z_t$ are observable.
- Then the model makes a clear prediction that, given any set of structural parameters, $A$, $B$, $H$ and $D$, the data will be given by $Z_t = CZ_{t-1} + PX_t$.
- The "cross-equation restrictions" in DSGE models tend to be very limiting. In other words, given any values for the $A$, $B$, $H$ and $D$ matrices, there are very particular patterns that must be obeyed by the $C$ and $P$ matrices.
- Most likely, there is no set of $A$, $B$, $H$ and $D$ matrices that will allow $Z_t = CZ_{t-1} + PX_t$ to perfectly fit the data.
- In this case, maximum likelihood methods do not work. These methods ask "how likely" it is that a model might be able to explain the data. But here we know for sure that the model does not fit the data.
- One way to address this issue is to add error terms, $u_t$ and then apply maximum likelihood to estimate $A$, $B$, $H$ and $D$ as those matrices that give the best fitting model of the form $Z_t = CZ_{t-1} + PX_t + u_t$.
- Though the $u_t$ don't have a microeconomic foundation, the size of the error terms for the best-fitting model gives us a sense of how well this model fits reality.

# Maximum Likelihood Estimation with Observable Variables

- We can use maximum likelihood to estimate the $A$, $B$, $H$ and $D$ coefficients that deliver the best-fitting joint model.

$$\begin{aligned} Z_t &= CZ_{t-1} + PX_t + u_t \\ X_t &= DX_{t-1} + \epsilon_t \end{aligned}$$

  where it is assumed that $u_t \sim N(0, \Sigma_u)$ and $\epsilon_t \sim N(0, \Sigma_\epsilon)$.

- Suppose we observe data $Z_1, Z_2, ..., Z_T$ for our endogenous variables and $X_1, X_2, ..., X_T$ for our exogenous variables.
- The log-likelihood function for the $X$ data is

$$\log L_X = -\frac{T}{2}\log 2\pi - T\log\left|\Sigma_\epsilon^{-1}\right| - \frac{1}{2}\sum_{k=1}^{T}(X_i - DX_{i-1})'\,\Sigma_\epsilon^{-1}(X_i - DX_{i-1})$$

- And the log-likelihood function for the $Z$ data is

$$\log L_Z = -\frac{T}{2}\log 2\pi - T\log\left|\Sigma_u^{-1}\right| - \frac{1}{2}\sum_{k=1}^{T}(Z_i - CZ_i - PX_i)'\,\Sigma_u^{-1}(Z_i - CZ_i - PX_i)$$

# Maximum Likelihood Estimation with Observable Variables

- The likelihood for the full model multiplies the likelihood of the $X$ data and the likelihood of the $Z$ data, so the combined log-likelihood is the sum of the two log-likelihoods.
- So the maximum likelihood estimates of $A$, $B$, $H$, $D$, $\Sigma_\epsilon$ and $\Sigma_u$ are those that maximise the log-likelihood

$$
\begin{aligned}
-T \log 2\pi - T\left(\log\left|\Sigma_\epsilon^{-1}\right| + \log\left|\Sigma_u^{-1}\right|\right) \\
-\frac{1}{2}\sum_{k=1}^{T}\left(X_i - DX_{i-1}\right)'\Sigma_\epsilon^{-1}\left(X_i - DX_{i-1}\right) \\
-\frac{1}{2}\sum_{k=1}^{T}\left(Z_i - CZ_i - PX_i\right)'\Sigma_u^{-1}\left(Z_i - CZ_i - PX_i\right)
\end{aligned}
$$

subject to the restrictions that map $A$ and $B$ into $C$ and map $A$, $B$, $H$ and $D$ into $P$.

# A Mix of Observables and Unobservables

- We have been discussing the case in which we can see all of the variables – both endogenous and exogenous – in our DSGE model.
- In fact, most DSGE models are not like this. Instead, these models tend to mix observable and unobservable variables.
- Consider again the log-linearised RBC model that we solved earlier. The equations of this model are listed on the next page.
  - This model features 7 equations in six endogenous variables, $y_t, c_t, i_t, k_t, n_t, r_t$ and one exogenous variable, $a_t$.
  - We can observe $y_t, c_t, i_t$ and $n_t$ (or at least the HP-filtered version of them that we are likely to use to estimate the model). But we don't observe $a_t$ and since we don't really know depreciation rates, this means we don't observe $k_t$ or $n_t$.
  - So this model mixes four observable variables with three unobservable variables.
- Estimation of these kinds of models requires special techniques to handle unobservable variables.

# The Linearised RBC Model

$$
\begin{aligned}
y_t &= \left(1 - \frac{\alpha\delta}{\beta^{-1} + \delta - 1}\right) c_t + \left(\frac{\alpha\delta}{\beta^{-1} + \delta - 1}\right) i_t \\
y_t &= a_t + \alpha k_{t-1} + (1-\alpha) n_t \\
k_t &= \delta i_t + (1-\delta) k_{t-1} \\
n_t &= y_t - \eta c_t \\
c_t &= E_t c_{t+1} - \frac{1}{\eta} E_t r_{t+1} \\
r_t &= (1 - \beta(1-\delta))(y_t - k_{t-1}) \\
a_t &= \rho a_{t-1} + \epsilon_t
\end{aligned}
$$

# The Stochastic Singularity Problem

- Models like the one on the previous page provide a micro-foundation for why we cannot find a perfect fitting model with the observed data: There is an unobservable technology series and all of the observed series depend on this.
- However, it is still not possible to estimate this joint model by maximum likelihood. This is because the same unobserved series shows up all the reduced-form solution equations.
- So while the model features stochastic shocks, it has a feature that is known as a *stochastic singularity*: The shocks in all the equations are just multiples of each other.
- The model thus predicts that certain ratios of the observed variables (e.g. current and lagged consumption, current and lagged investment) will be constant. In practice, these predictions will not hold in the data so there is no chance that this model can fit the data.
- In general, for a model to have well-defined econometric estimates, it is necessary that for every observable variable there be at least one unobservable shock. This can either take the form of a "measurement error" or else involve a shock in each equation with a clear structural interpretation.

# DSGEs are State-Space Models

- Log-linearised DSGE models with a mix of observable and unobservable variables are an example of **state-space models**. Recall that these models can be described using two equations.
- The first, known as the **state** or **transition equation**, describes how a set of unobservable state variables, $S_t$, evolve over time as follows:

  $$S_t = FS_{t-1} + u_t$$

  The term $u_t$ can include either normally-distributed errors or perhaps zeros if the equation being described is an identity. We will write this as $u_t \sim N\left(0, \Sigma^u\right)$ though $\Sigma^u$ may not have a full matrix rank.
- The second equation in a state-space model, which is known as the **measurement equation**, relates a set of observable variables, $Z_t$, to the unobservable state variables

  $$Z_t = HS_t + v_t$$

  Again, the term $w_t$ can include either normally-distributed errors or perhaps zeros if the equation being described is an identity. We will write this as $v_t \sim N\left(0, \Sigma^v\right)$ though $\Sigma^v$ may not have a full matrix rank.

# Example: An RBC Model

- The solution to the basic RBC model without labour input can be summarised as

$$
\begin{aligned}
k_t &= a_{kk} k_{t-1} + a_{kz} z_t \\
c_t &= a_{ck} k_{t-1} + a_{cz} z_t \\
z_t &= \rho z_{t-1} + \epsilon_t
\end{aligned}
$$

- Now let's assume that consumption and capital are only observed with error so that the two observable variables are

$$
\begin{aligned}
k_t^* &= a_{kk} k_{t-1} + a_{kz} z_t + u_t^k \\
c_t^* &= a_{ck} k_{t-1} + a_{cz} z_t + u_t^c
\end{aligned}
$$

# Example: An RBC Model

- This can be written in state-space form as follows.
- The transition equation is

$$\begin{pmatrix} k_{t-1} \\ z_t \end{pmatrix} = \begin{pmatrix} a_{kk} & a_{kz} \\ 0 & \rho \end{pmatrix} \begin{pmatrix} k_{t-2} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \epsilon_t \end{pmatrix}$$

- And the measurement equation is

$$\begin{pmatrix} k^*_{t-1} \\ c^*_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ a_{ck} & a_{cz} \end{pmatrix} \begin{pmatrix} k_{t-1} \\ z_t \end{pmatrix} + \begin{pmatrix} u^k_{t-1} \\ u^c_t \end{pmatrix}$$

- Note that a little bit of jiggery-pokery had to be done to get the model in state-space form and the timing conventions associated with this representations are not quite the same as in the original model, i.e. we have $S_t = \begin{pmatrix} k_{t-1} \\ z_t \end{pmatrix}$ and $X_t = \begin{pmatrix} k^*_{t-1} \\ c^*_t \end{pmatrix}$.
- Still, all standard DSGE models can be re-arranged to be put in this format.

# MLE for DSGE Models via Kalman Filter

- So Kalman filter provides a way to do maximum likelihood estimation of DSGE models that mix observable and unobservable variables.
- You may have found the lecture on the Kalman filter complicated but the good news is that software packages such as Dynare can do this for you with a minimum of effort from you once you have specified your model.
- In other words, computer packages can now
  1. Sort your model into state-space methods.
  2. Search across a wide range of possible parameter values.
  3. For each of these, apply the Kalman filter/smoother.
  4. Then, for each possible set of parameters, it can sum up each of the period-by-period likelihoods.
  5. Then it can decide what the best parameters are and use standard MLE-related methods to calculate asymptotically valid standard errors.
- That's cool but if you think this is a complicated process where things might go wrong, then you'd be right.

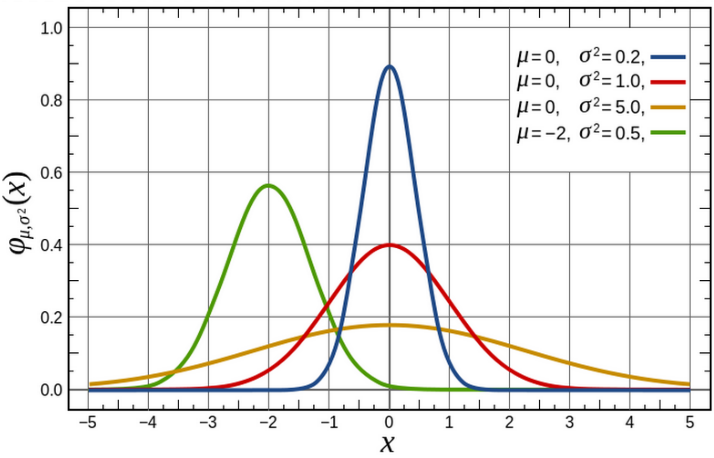
# A Drawback of MLE

- See the paper on the website by Jesus Fernandez-Villaverde. He discusses some of the problems associated with MLE for DSGE models and explains why a Bayesian approach of calculating the full posterior distribution may be preferable.
- "maximizing a complicated, highly dimensional function like the likelihood of a DSGE model is actually much harder than it is to integrate it, which is what we do in a Bayesian exercise. First, the likelihood of DSGE models is, as I have just mentioned, a highly dimensional object, with a dozen or so parameters in the simplest cases to close to a hundred in some of the richest models in the literature. Any search in a high dimensional function is fraught with peril. More pointedly, likelihoods of DSGE models are full of local maxima and minima and of nearly flat surfaces. This is due both to the sparsity of the data (quarterly data do not give us the luxury of many observations that micro panels provide) and to the flexibility of DSGE models in generating similar behavior with relatively different combination of parameter values .... Moreover, the standard errors of the estimates are notoriously difficult to compute and their asymptotic distribution a poor approximation to the small sample one."

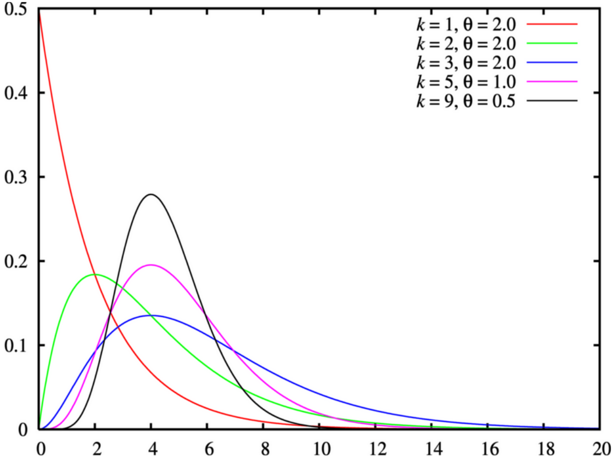
# Bayesian DSGE

- For these reasons, Bayesian approaches to estimating DGSE models have become the standard approach in recent years.
- A prior distribution for the parameters is specified and then this is combined with the full likelihood function to produce an estimate of the posterior distribution. This posterior distribution can be integrated using numerical methods to produce means and confidence intervals of various sorts.
- Importantly, because you are using an estimate of the full likelihood function, you are less likely to fall victim to the major errors that can occur from using an incorrect MLE, which uses only one point of the function.
- Dynare allows you to specify priors and to estimate a DSGE model directly.
- Researchers generally specify prior means for parameters using values considered "reasonable" from other studies with the form of the distributions usually being of a form that fits with a "common sense" view of the potential range of outcomes.
- The estimation results are generally reported by comparing the posterior means with the prior means as well as reporting the "confidence intervals" from the posterior distributions.

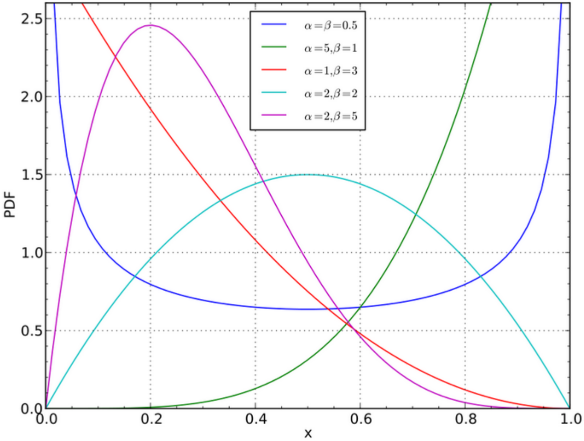
# Choosing Priors: Normal Distribution



# Choosing Priors: Gamma Distribution (Parameter Restricted to be Positive)



# Choosing Priors: Beta Distribution (Parameter Restricted to between Zero and One)



# Some Readings

Three useful papers for more details and discussion all available on the website.

1. Francisco J. Ruge-Murcia, "Methods to Estimate Dynamic Stochastic General Equilibrium Models." A nice discussion of non-Bayesian estimation methods for DSGE model with a particularly clear focus on the stochastic singularity issue.
2. Peter Ireland, "A Method for Taking Models to the Data." A clear presentation of how to use the Kalman Filter to do MLE for DSGE models with a fully-worked example.
3. Jesus Fernandez-Villaverde, "The Econometrics of DGSE Models." A detailed (and fairly advanced) discussion of Bayesian methods for estimating DSGE models and a nice example of how the methods are used.