LECTURE 2.4

Estimating DSGE Models

ECON42240 Advanced Macroeconomics

Sam Deegan
University College Dublin

Spring 2027

Where We Are

1. Time Series Methods

1.1 Time Series and Macro

1.2 Vector Autoregressions

1.3 VARs: Applications

1.4 The Kalman Filter

1.5 Solving RE Models

2. Structural Models

2.1 The RBC Model

2.2 The Phillips Curve

2.3 The New Keynesian Model

2.4 Estimating DSGE Models

2.5 The Smets-Wouters Model

3. Finance and Policy

3.1 Credit and Banking Crises

3.2 Fiscal Policy, and What Next

Last Week's Lecture

1. **The Three Equations.**
   IS, Phillips curve, policy rule.

2. **The Taylor Principle.**
   What makes the equilibrium unique.

3. **Optimal Policy Under Commitment.**
   A promise to pull the price level back.

4. **Discretion and the Model Appraised.**
   Re-optimising every period, and the model's record.

Lecture Plan

1. **From Calibration to a Likelihood.**
   Why estimate, and what a likelihood needs.

2. **Singularity and the Filter.**
   Too few shocks, and the filter as the likelihood.

3. **Priors and Posterior.**
   What a prior adds, and how a sampler draws the posterior.

4. **Judging the Estimate.**
   Reading a posterior, and holding the model against the VAR.

SECTION 1

From Calibration to a Likelihood

Five Weeks of Parts Fit Together Today

Estimating a DSGE model uses three pieces already built, and one new one:

- **The Solution:** Lecture 1.5 turned a linear model into a policy function.
- **The State Space:** Lecture 1.4 wrote unobservables and data as two equations.
- **The Kalman Filter:** Lecture 1.4 produced one forecast error each period.

> **Today Adds the Prior**
> The likelihood is Lecture 1.4's filter, run again. The new piece is a prior over the parameters.

Early DSGE Models Were Calibrated, Not Estimated

Early researchers set DSGE parameters from outside evidence rather than estimating them:

- **Steady-State Ratios:** Labour share and capital-output ratio match history.
- **Micro Evidence:** Risk aversion and depreciation rates from micro studies.

> **Calibration Computes No Likelihood**
> Calibration fixes each parameter from steady-state ratios or micro studies. No likelihood is ever computed.

Matching Moments Uses Part of the Data

Rotemberg and Woodford chose parameters to match estimated responses to monetary policy shocks.

> **Indirect Inference Targets Chosen Features**
> Indirect inference picks parameters so the model reproduces selected features of the data.

Policy shocks explain little variation, so most of the sample's information is unused.

Modern Estimation Uses the Whole Sample

Most current papers estimate DSGE models by Bayesian methods, which raise four issues:

1. **Observables:** Which model variables have data, and which have none.
2. **Shock Count:** Whether enough shocks exist to explain the observed series.
3. **Kalman Filtering:** The likelihood of a state-space model, from Lecture 1.4.
4. **Bayesian Methods:** Priors over the parameters, and a posterior distribution.

Estimation Starts from the Solved Model

Estimation begins from the linear model in Lecture 1.5's solver form:

$$KZ_t = AZ_{t-1} + B\,E_t Z_{t+1} + GX_t, \qquad X_t = \Phi X_{t-1} + \epsilon_t \tag{1}$$

where:

- G: loading on exogenous variables

Lecture 1.5 solved this form.

> **The Solver's Loading Is Written G**
> Lecture 1.4 gave H to the measurement matrix, which returns later today. Whelan and Lecture 1.5 write this loading H.

What Estimation Has to Recover

The solved model maps the structural matrices into two reduced-form matrices:

$$Z_t = \overbrace{C Z_{t-1} + \Pi X_t}^{\text{both matrices fixed by the structure}} \tag{2}$$

The Data Speak Only About C and Π

The model's restrictions carry the estimate back to the structural matrices. Only those are assumed to survive a policy change.

Cross-Equation Restrictions Bind the Solution

With every variable observed, the model predicts each period's data exactly:

$$Z_t = CZ_{t-1} + \Pi X_t \tag{3}$$

One Parameter Fixes Many Coefficients
A cross-equation restriction ties coefficients in different equations to the same structural parameters. Only very particular C and Π are possible.

A Model with No Errors Cannot Fit the Data

No structural matrices make $Z_t = CZ_{t-1} + \Pi X_t$ hold exactly in real data.

> **An Errorless Model Leaves Nothing to Maximise**
> Maximum likelihood asks how probable the data are under the model. Here the data's probability is zero.

Nothing in the model absorbs what the structure cannot explain.

Adding Errors Restores a Likelihood

Adding an error to each endogenous equation gives the model room to miss:

$$Z_t = CZ_{t-1} + \Pi X_t + u_t, \qquad u_t \sim N(0, \Sigma_u) \tag{4}$$

> **These Errors Have No Microeconomic Foundation**
> No household or firm decision produces u_t. A large estimated Σ_u still signals a badly fitting model.

The Likelihood When Everything Is Observed

Each period's error is squared and scaled by the error variance:

$$\log L_Z = -\frac{T}{2}\log|\Sigma_u| \; - \; \frac{1}{2}\sum_{t=1}^{T} u_t'\Sigma_u^{-1}u_t \; + \; \text{constant} \tag{5}$$

The Determinant Term Stops the Variance Absorbing Errors
The quadratic term rewards small errors. The determinant term stops Σ_u from absorbing them.

The Exogenous Block Adds Its Own Likelihood

The exogenous block has its own likelihood, and independence lets the two add:

$$\log L = \underbrace{\log L_X(\Phi, \Sigma_\epsilon)}_{\text{exogenous block}} + \underbrace{\log L_Z(C, \Pi, \Sigma_u)}_{\text{endogenous block}} \tag{6}$$

where:

- Σ_ϵ: exogenous innovation covariance

> **The Restrictions Make the Estimate Structural**
> Estimation maximises the sum over structure and variances, subject to the restrictions. Without them the sum is a VAR's likelihood.

Most Models Mix Observable and Unobservable

Most DSGE models carry variables with no directly measured counterpart.

> **A Mixed Model Is a State-Space Model**
> The unobserved variables form the state, and the observed ones are the data. Lecture 1.4's filter computes its likelihood.

Technology is a residual, and a capital stock needs an assumed depreciation rate.

Output Splits Between Consumption and Investment

Output (y_t) is consumption (c_t) plus investment (i_t), at their steady-state shares:

$$y_t = \left(1 - \frac{\alpha\delta}{\beta^{-1} + \delta - 1}\right) c_t + \underbrace{\frac{\alpha\delta}{\beta^{-1} + \delta - 1}}_{\text{steady-state investment share}} i_t \tag{7}$$

Lecture 2.1 derived these shares from the steady state.

Output Uses Capital, Hours and Technology

Output uses last period's capital (k_{t-1}), current hours (n_t) and technology (a_t):

$$y_t = a_t + \alpha k_{t-1} + (1 - \alpha) n_t \tag{8}$$

Lecture 2.1 log-linearised this production function.

Capital Accumulates from Investment

Capital this period equals the undepreciated stock plus new investment:

$$k_t = \delta i_t + (1 - \delta) k_{t-1} \tag{9}$$

Lecture 2.1 log-linearised this capital accumulation identity.

> **An Unknown Depreciation Rate Hides Capital**
> Measuring capital needs a depreciation rate (δ). With δ unknown, k_t has no direct measure.

Hours Rise with Output and Fall with Consumption

The labour condition sets hours from output and consumption:

$$n_t = y_t - \eta c_t \tag{10}$$

Lecture 2.1 derived this labour condition.

Expected Consumption Growth Rises with the Return

Consumption today is expected consumption tomorrow, less the expected return over η:

$$c_t = E_t c_{t+1} - \frac{1}{\eta} E_t r_{t+1} \tag{11}$$

Lecture 2.1 derived this Euler equation.

The Return Rises with Output Relative to Capital

The return to capital (r_t) rises with output relative to last period's capital:

$$r_t = \big(1 - \beta(1 - \delta)\big)(y_t - k_{t-1}) \tag{12}$$

Lecture 2.1 derived this coefficient from the steady state.

> **The Model's Return Has No Market Quote**
> The model's r_t is no quoted interest rate, and depends on the unknown δ.

Technology Is the Model's Only Shock

Technology follows its own process, and carries the model's only random term:

$$a_t = \rho a_{t-1} + \epsilon_t \tag{13}$$

Lecture 2.1 set technology as an AR(1) in logs.

Six Endogenous Variables Share One Shock
Seven equations govern six endogenous variables and technology. Only ϵ_t is random.

Can One Shock Explain Four Series?

Output, consumption, investment and hours are observed. Technology is the only shock.

> **Question**
> Can the likelihood of those four series be computed? What does the model predict about them?

Take answers from the room.

SECTION 2

Singularity and the Filter

Four Observed Series Against Three Unobserved

Four of the RBC model's seven variables have data, and three do not:

- **Observed:** Output, consumption, investment and hours, each HP-filtered.
- **Not Observed:** Technology, the capital stock, and the return to capital.

> **The Observable Set Is a Choice**
> Hours or investment could be left out of the estimation. The modeller decides which series enter.

Four Series Cannot Come from One Shock

The unobserved technology series gives the model a reason to miss the data:

- **A Micro-Foundation:** All four series depend on unobserved technology.
- **The Gain:** A reason the model misses the data, with a behavioural story.
- **The Limit:** No likelihood, because all four series share one shock.

What a Singular Model Predicts

A model with fewer shocks than observed series makes an exact prediction:

- **The Shocks Are Multiples:** Every equation's innovation is a multiple of one shock.
- **The Prediction:** Given the lagged state, some combinations of the series hold exactly.

> **Stochastic Singularity Means Too Few Shocks**
> Fewer shocks than observed series make a model stochastically singular. Real data break the exact combinations, so the likelihood is zero.

One Shock for Every Observed Series

Every observed series needs at least one unobserved shock.

> **The Shock Count Is Necessary and Not Sufficient**
> Too few shocks make estimation impossible. Enough shocks do not make the estimate good.

Each shock is either measurement error or a structural shock with an interpretation.

Adding a Shock or Admitting a Failure

A one-shock model of four series needs at least three more shocks.

> **Question**
> You add measurement error to hours worked. What have you just assumed about the data, and about the model?

Take answers from the room.

Two Ways to Break the Singularity

Two fixes exist, and they make different admissions about the model:

- **Measurement Error:** Assumes the series is measured badly, and the equation is right.
- **More Structural Shocks:** Each added shock names a new source of fluctuation.
- **What Gets Hidden:** Measurement error can absorb a genuine failure of the model.

> **Lecture 2.5 Adds Structural Shocks**
> Lecture 2.5's model carries seven structural shocks, each with an economic interpretation.

A Solved DSGE Is a State-Space Model

A solved model with unobservables is the state-space model of Lecture 1.4:

$$S_t = FS_{t-1} + u_t, \qquad Z_t^o = HS_t + v_t \tag{14}$$

where:

- Z_t^o: observed model variables

> **The Observed Vector Gets a Superscript**
> Lecture 1.5 uses Z_t for the endogenous vector. Lecture 1.4's observed Z_t is Z_t^o here.

The RBC Solution Has Four Coefficients

The RBC solution without labour has four coefficients, each fixed by the structure:

$$k_t = a_{kk} k_{t-1} + a_{kz} z_t, \qquad c_t = a_{ck} k_{t-1} + a_{cz} z_t \tag{15}$$

Technology (z_t) follows its own first-order process:

$$z_t = \rho z_{t-1} + \epsilon_t \tag{16}$$

- a_{kk}, a_{kz}, a_{ck}, a_{cz}: solution coefficients

Capital and Consumption Are Observed with Error

Each observed series is the model's variable plus its own measurement error:

$$k_t^* = k_t + u_t^k, \qquad c_t^* = c_t + u_t^c \tag{17}$$

where:

- k_t^*, c_t^*: measured capital and consumption
- u_t^k, u_t^c: measurement errors

> **A Star Here Means Measured**
> In Lecture 2.1 a star marked a steady state. Here a star marks the data.

The RBC Model Fills the State Equation

The state is yesterday's capital and today's technology; each piece has an RBC match:

$$S_t = F\, S_{t-1} + u_t \tag{18}$$

Capital's rule a quarter back fills row one, and technology fills row two:

$$S_t = \begin{pmatrix} k_{t-1} \\ z_t \end{pmatrix}, \qquad F = \begin{pmatrix} a_{kk} & a_{kz} \\ 0 & \rho \end{pmatrix}, \qquad u_t = \begin{pmatrix} 0 \\ \epsilon_t \end{pmatrix} \tag{19}$$

> **Row One Is the Capital Rule, Lagged**
> One quarter back the rule reads $k_{t-1} = a_{kk} k_{t-2} + a_{kz} z_{t-1}$.

The Transition Equation Carries One Shock

Dropping the three matches into the state equation gives the RBC transition:

$$\begin{pmatrix} k_{t-1} \\ z_t \end{pmatrix} = \begin{pmatrix} a_{kk} & a_{kz} \\ 0 & \rho \end{pmatrix} \begin{pmatrix} k_{t-2} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \epsilon_t \end{pmatrix} \tag{20}$$

The State Shock Covariance Has Rank One

Only the second row carries a shock, so Σ_u has rank one. The filter runs anyway.

The Measured Series Fill the Measurement Equation

Lecture 1.4's measurement equation has three pieces; the RBC example fills each:

$$Z_t^o = H\, S_t + v_t \tag{21}$$

Measured capital is dated to match the state; consumption's rule fills row two:

$$Z_t^o = \begin{pmatrix} k_{t-1}^* \\ c_t^* \end{pmatrix}, \qquad H = \begin{pmatrix} 1 & 0 \\ a_{ck} & a_{cz} \end{pmatrix}, \qquad v_t = \begin{pmatrix} u_{t-1}^k \\ u_t^c \end{pmatrix} \tag{22}$$

> **Row Two Is Consumption's Decision Rule**
> Consumption is $a_{ck} k_{t-1} + a_{cz} z_t$, so H carries solution coefficients.

The Measurement Equation Adds Two Errors

Dropping the three matches into the measurement equation gives the RBC version:

$$\begin{pmatrix} k^*_{t-1} \\ c^*_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ a_{ck} & a_{cz} \end{pmatrix} \begin{pmatrix} k_{t-1} \\ z_t \end{pmatrix} + \begin{pmatrix} u^k_{t-1} \\ u^c_t \end{pmatrix} \tag{23}$$

Three Random Terms Answer Two Observables

One shock and two measurement errors give three random terms for two series. The counting rule holds.

The State Vector Shifts the Dates

Every standard DSGE rearranges into this form, and capital enters the state lagged:

- **The Data Follow:** Measured capital is dated $t-1$, and measured consumption t.
- **The Shift Is Harmless:** A state is any set making tomorrow depend on today alone.

The Filter Returns as the Likelihood

One filter pass builds the likelihood from forecast errors (ε_t) and their variances (Ω_t):

$$\log L(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left[\log|\Omega_t| + \varepsilon_t'\Omega_t^{-1}\varepsilon_t\right] + \text{constant} \tag{24}$$

Structural Parameters Now Fix *F* and *H*
In Lecture 1.4, θ held F and H themselves. Here θ holds structural parameters, and the solution fixes F and H.

Estimating a DSGE by Maximum Likelihood

Software runs five steps, repeating the first four at each candidate parameter vector:

1. **Solve:** Find C and Π at the candidate parameter values.
2. **Arrange:** Sort the solution into the state and measurement equations.
3. **Filter:** Run the Kalman filter, collecting each forecast error and variance.
4. **Score:** Add the period-by-period log-likelihoods into a single number.
5. **Search:** Propose new parameters, and report the best value found.

> **Each Stage Can Fail**
> Solution can fail at a candidate. The search can stop at a local maximum.

What Would Make You Distrust the Maximum?

The search returns the parameter vector with the highest likelihood it found.

> **Question**
> What could make that vector a poor estimate, even with the code correct?

Take answers from the room.

SECTION 3

Priors and Posterior

High Dimension and Sparse Data Defeat Maximum Likelihood

This section replaces one best parameter vector with a whole distribution. The failures come first (Fernández-Villaverde, 2009):

- **High Dimension:** A dozen parameters in simple models, near a hundred in rich ones.
- **Sparse Data:** Quarterly series give few observations beside a microeconomic panel.

> **Maximising Is Harder Than Integrating**
> Maximising a DSGE likelihood is harder than integrating it. Bayesian estimation integrates.

Flat Surfaces and Local Maxima

Two features of DSGE likelihoods defeat an optimiser:

- **Local Maxima:** The surface carries many peaks, and a search stops at the first.
- **Flat Regions:** Different parameter combinations produce nearly the same behaviour.

> **Flatness Means Weak Identification**
> Flexible models let parameters substitute for each other. The data then barely tell the values apart.

A Point Estimate Discards the Surface

A maximum likelihood estimate reports one point on the surface and discards the rest.

> **A Point Cannot Express Parameter Uncertainty**
> A distribution over parameters reports how plausible every value is, not only the best.

The estimate’s standard errors are hard to compute and poorly approximated in small samples.

A Prior Turns the Likelihood into a Posterior

Bayesian estimation attaches a distribution to the parameters before the data are seen:

$$\underbrace{p\left(\theta \mid Z^{o}_{1:T}\right)}_{\text{posterior}} \propto \underbrace{L\left(Z^{o}_{1:T} \mid \theta\right)}_{\text{likelihood}} \times \underbrace{p(\theta)}_{\text{prior}} \tag{25}$$

The Dropped Term Involves No Parameter
Bayes' rule divides by the probability of the data, which involves no θ.

What Bayesian Estimation Adds

The posterior uses the whole likelihood surface, not only its peak.

> **Calibration Returns as a Prior Mean**
> Values from other studies enter as prior means with a spread. The data may disagree with them.

Reported intervals come from the posterior, not from an asymptotic formula.

The Parameter's Range Picks the Prior Family

A prior distribution has to put probability only where the parameter can go:

- **Unrestricted:** A normal prior, for a parameter free to take either sign.
- **Positive Only:** A gamma prior, for a standard deviation or risk aversion.
- **Between Zero and One:** A beta prior, for a persistence parameter or a share.

> **The Prior's Support Comes First**
> A normal prior on persistence puts probability above one. There a shock's effect grows forever.

Three Prior Families with the Same Mean

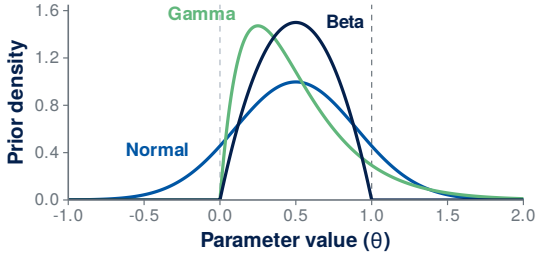

Figure 1: Normal, gamma and beta prior densities, each with mean 0.5; dashed lines at 0 and 1.

> **Only the Beta Stays in the Unit Interval**
> The normal puts 11 per cent below zero and 11 above one. The gamma puts 9 per cent above one.

Nobody Can Integrate the Posterior Directly

The posterior has no formula, because the likelihood comes from a filter run:

- **No Closed Form:** Each evaluation at one θ solves the model and runs the filter.
- **What Is Wanted:** Means, intervals and marginal densities, each one an integral.

> **Averaging Draws Estimates Each Integral**
> Given draws from the posterior, their average estimates any posterior mean.

The Metropolis–Hastings Random Walk

A random walk through parameter space produces draws using only posterior ratios:

1. **Start Anywhere:** Begin at a parameter vector with non-zero posterior density.
2. **Propose:** Add a normal draw to the current vector, giving a candidate.
3. **Accept:** Accept the candidate with a probability, or keep the current vector.

> **Metropolis–Hastings Builds a Chain That Visits the Posterior**
> The chain lingers where the posterior is high. After warm-up, its kept draws approximate the posterior.

The Acceptance Rule for a Random-Walk Step

The candidate (θ^*) replaces the current draw ($\theta^{(s)}$) with probability α:

$$\alpha = \min\left\{1,\ \frac{L\left(Z^o_{1:T} \mid \theta^*\right) p(\theta^*)}{L\left(Z^o_{1:T} \mid \theta^{(s)}\right) p\left(\theta^{(s)}\right)}\right\} \tag{26}$$

Downhill Steps Keep the Chain Exploring

Uphill candidates are always accepted, downhill ones sometimes. The unknown normalising constant cancels from the ratio.

Tuning the Proposal's Step Size

The step size decides whether the chain explores the surface or stalls:

- **Steps Too Large:** Almost every candidate is rejected, and the chain stays put.
- **Steps Too Small:** Almost every candidate is accepted, and the chain crawls.

> **Adaptation Belongs in the Warm-Up**
> Software tunes the step during warm-up, then freezes it. Adapting throughout can distort the draws.

Both Chains Reach One Level Within Warm-Up

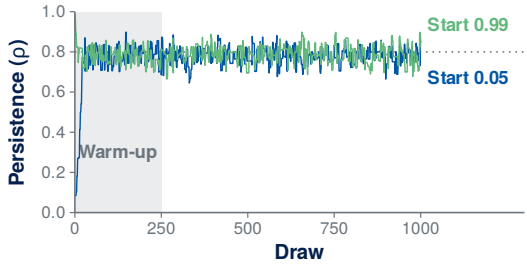

Figure 2: Random-walk Metropolis draws of AR(1) persistence: two chains, first 1,000 draws.

> **Simulated Data, True Persistence 0.8**
> Both chains leave their starting values within warm-up. Dotted: the true 0.8.

Running the Sampler Yourself

Three pieces make the estimation route, and two are already written:

1. **Solve:** The generalised Schur routine written for Lecture 1.5.
2. **Filter:** The state-space code written for Lecture 1.4, unchanged.
3. **Sample:** An adaptive random-walk Metropolis–Hastings loop, in about twenty lines.

> **A Package Only Checks Your Sampler**
> A package can estimate the same model in one call. The problem set assesses the sampler you write.

Choosing a Prior for a Persistence Parameter

A shock's persistence is assumed to lie between zero and one. Its prior is beta.

> **Question**
> How wide should that beta prior be, and what justifies the width?

Take answers from the room.

SECTION 4

Judging the Estimate

A Tight Prior Outweighs 200 Observations

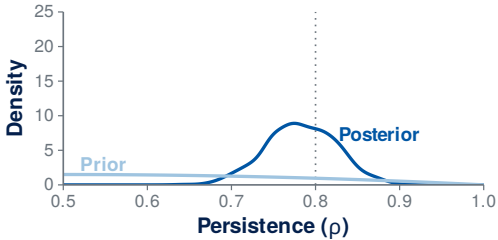

Figure 3: Wide prior: beta(2, 2), mean 0.5.

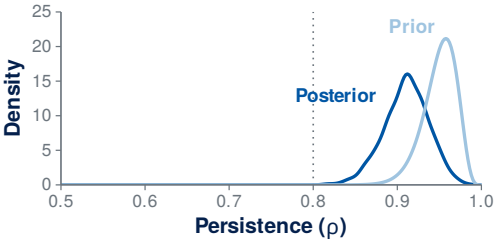

Figure 4: Tight prior: mean 0.95, s.d. 0.02.

Same Data, Two Priors
AR(1) data simulated with persistence 0.8, dotted. The posteriors centre on 0.78 and 0.91.

Reading a Posterior from Its Draws

A posterior is reported through a few summaries of the kept draws:

- **The Posterior Mean:** The average of the draws, 0.78 in the wide-prior panel.
- **Credible Interval:** Holds a stated share of draws; 0.71 to 0.85 at 90 per cent.
- **The Shape:** A skewed or two-peaked posterior makes the mean misleading.

> **A Credible Interval Is Not a Confidence Interval**
> A credible interval holds a stated share of posterior probability. No repeated sampling is involved.

A Posterior That Equals Its Prior

A parameter the data cannot speak to keeps the prior it was given:

- **The Symptom:** The posterior density sits on top of the prior density.
- **The Danger:** A posterior mean gets reported as though the data produced it.

By Bayes' rule, a flat likelihood leaves the prior unchanged.

> **Identification Fails When the Likelihood Is Flat**
> A parameter is identified when different values imply different distributions for the observables.

Two Meanings of Identification

Identification has two meanings here, and both ask what data alone cannot settle:

Identification in a VAR
Recovering structural shocks from residuals, using outside assumptions.

Identification in a DSGE
Whether the data distinguish one parameter vector from another at all.

A proper prior always yields a posterior, even when the data said nothing.

Reporting an Estimated DSGE

Papers table prior and posterior means side by side, with posterior intervals.

> **A Posterior Alone Hides the Prior's Work**
> Without the prior mean beside it, a posterior mean cannot be judged.

A reader checks which parameters left their prior, and which kept it.

A Posterior Is Not a Verdict

Estimation delivered a posterior over one model's parameters, and asked nothing more.

> **The Likelihood Never Scored an Impulse Response**
> The likelihood scores one-step-ahead forecast errors. An impulse response is a different claim.

Whether the model reproduces Lecture 1.3's impulse responses is still open.

Holding the Model Against the VAR

The comparison the module has been building toward has three fixed parts:

- **The Same Shock:** Monetary policy, in the model and in the identified VAR.
- **The Same Variables:** Output, inflation and the policy rate, transformed alike.
- **The Same Horizon:** Twenty quarters, with bands around the VAR's responses.

> **The Two Identifications Must Match**
> The VAR's shock rests on a timing assumption. The model's responses must impose the same timing.

Model Responses Against the Identified VAR

Figure 5: Responses to a monetary policy shock: estimated model against identified VAR.

> **The Model Either Enters the Bands or Does Not**
> Posterior model responses are drawn against the VAR's bands. Agreement is not guaranteed.

A Mismatch Is Evidence About Assumed Behaviour

The VAR summarises the data without using the model's behavioural assumptions:

- **What a Miss Shows:** Assumed behaviour cannot produce the measured response.
- **What Can Differ:** The response's size, its sign, or the timing of its peak.

> **A Mismatch Points at the Assumption, Not the Estimate**
> A well-estimated model can still miss a response. The miss implicates assumed behaviour rather than parameter values.

What a Modeller Does About a Mismatch

Four responses to a mismatch are available, in rising order of cost:

1. **Check Comparability:** Same sample, transformation and identification.
2. **Check Observables:** A series left out of the estimation can produce the mismatch.
3. **Add Mechanisms:** A friction or a shock that generates the measured response.
4. **Report Failure:** State the margin on which the model fails, and stop there.

> **Mechanisms Added After the Fact Weaken Agreement**
> A friction chosen after seeing the mismatch makes agreement less informative.

Fitting the Sample Is Not Reproducing the Evidence

Three kinds of agreement with data look alike and carry rising weight:

1. **Fitted:** The likelihood was computed on the sample, so some agreement is built in.
2. **Not Targeted:** An impulse response the estimation never scored is separate evidence.
3. **Held Out:** A series left out of the observables, predicted well anyway.

> **A Model Is Judged on Evidence It Was Never Fitted To**
> Agreement with the estimation sample is cheap. Agreement with untargeted evidence is hard to get.

What the Comparison Cannot Settle

A model matching the VAR responses has still not been shown true.

> **Agreement Does Not Single Out One Model**
> Matching is necessary for credibility and nowhere near sufficient. Agreement removes a model from suspicion and no more.

Different structures can produce the same responses, and VAR bands admit many models.

Reporting a Model That Misses One Response

Your estimated model matches output and inflation, and misses the policy rate's peak.

> **Question**
> Do you add a mechanism, or report the failure? What would justify either choice?

Take answers from the room.

Before Lecture 2.5

1. **Play:**
   Draw a beta, a gamma and a normal prior, changing each spread.

2. **Apply:**
   Write a random-walk Metropolis sampler for one persistence parameter.

3. **Read:**
   Smets and Wouters in full, before Lecture 2.5 (Smets & Wouters, 2007).

Next Lecture

1. **The Frictions Block.**
   What Smets and Wouters add, and why.

2. **Shocks and Estimation.**
   Seven shocks, seven series, one posterior.

3. **What the Model Explains.**
   Forecasts, decompositions and responses.

4. **Our Model Against Smets and Wouters.**
   What the simplifications cost.

5. **What It Missed.**
   The model, and 2008.

Thank you

sam.deegan@ucdconnect.ie
sam-deegan.com

BACKUP

Appendix

Notation

Symbol	What it is	First met
G	loading on exogenous variables	Estimation Starts From the Solved Model
L_Z, L_X	the two blocks' likelihoods	The Likelihood When Everything Is Observed
Σ_ϵ	exogenous innovation covariance	The Exogenous Block Adds Its Own Likelihood
Z_t^o	observed model variables	A Solved DSGE Is a State-Space Model
a_{kk}, a_{kz}, a_{ck}, a_{cz}	solution coefficients	The RBC Solution Has Four Coefficients
z_t	technology, in the solved RBC model	The RBC Solution Has Four Coefficients
k_t^*, c_t^*	measured capital and consumption	Capital and Consumption Are Observed With Error
u_t^k, u_t^c	measurement errors	Capital and Consumption Are Observed With Error
$Z_{1:T}^o$	the observed sample	A Prior Turns the Likelihood Into a Posterior
$p(\theta)$	the prior density	A Prior Turns the Likelihood Into a Posterior
$p(\theta \mid Z_{1:T}^o)$	the posterior density	A Prior Turns the Likelihood Into a Posterior
$L(Z_{1:T}^o \mid \theta)$	the likelihood of the sample	A Prior Turns the Likelihood Into a Posterior
θ^*	the candidate parameter vector	The Acceptance Rule for a Random-Walk Step
$\theta^{(s)}$	the chain's current draw	The Acceptance Rule for a Random-Walk Step
α (second use)	the acceptance probability	The Acceptance Rule for a Random-Walk Step
T	the number of periods in the sample	Lecture 1.1

Notation

Symbol	What it is	First met
ρ	the persistence of technology	Lecture 1.1
S_t, F	the state vector and its transition matrix	Lecture 1.4
H, v_t	the measurement matrix and measurement error	Lecture 1.4
u_t, Σ_u	the state shock and its covariance	Lecture 1.4
ε_t, Ω_t	the forecast error and its variance	Lecture 1.4
θ	the parameter vector	Lecture 1.4
K, A, B	the solver's contemporaneous, lag and expectation matrices	Lecture 1.5
Z_t, X_t	the endogenous and the exogenous vectors	Lecture 1.5
Φ, ϵ_t	the exogenous process's matrix and innovation	Lecture 1.5
C, Π	the solution's reduced-form matrices	Lecture 1.5
E_t	the expectation using information at t	Lecture 1.5
y_t, c_t, i_t	log deviations of output, consumption, investment	Lecture 2.1
k_t, n_t, a_t	log deviations of capital, hours, technology	Lecture 2.1
r_t	the log deviation of the return to capital	Lecture 2.1
α	the capital share	Lecture 2.1
β, δ	the discount factor and the depreciation rate	Lecture 2.1
η	how fast the value of extra consumption falls	Lecture 2.1

References

Fernández-Villaverde, J. (2009). *The econometrics of DSGE models* (NBER Working Paper No. 14677). National Bureau of Economic Research.

Smets, F., & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, *97*(3), 586–606.