Likelihood Function and Frequentist Inference

Frank Schorfheide
University of Pennsylvania

Econ 722 – Part 1

February 13, 2019

Likelihood Function

- Likelihood function plays a key role in frequentist and Bayesian inference.
- We will spend some time on how to evaluate this function.

Recall: State-Space Representation of DSGE Model

State-space representation:

$$
\begin{aligned}
y_t &= \Psi_0(\theta) + \Psi_1(\theta)s_t \\
s_t &= \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t
\end{aligned}
$$

System matrices:

$$
\Psi_0(\theta) = M_y' \begin{bmatrix} \log\gamma \\ \log(lsh) \\ \log\pi^* \\ \log(\pi^*\gamma/\beta) \end{bmatrix}, \quad x_\phi = -\frac{\kappa_p\psi_p/\beta}{1-\psi_p\rho_\phi}, \quad x_\lambda = -\frac{\kappa_p\psi_p/\beta}{1-\psi_p\rho_\lambda}, \quad x_z = \frac{\rho_z\psi_p}{1-\psi_p\rho_z}, \quad x_{\epsilon_R} = -\psi_p\sigma_R
$$

$$
\Psi_1(\theta) = M_y' \begin{bmatrix} x_\phi & x_\lambda & x_z + 1 & x_{\epsilon_R} & -1 \\ 1+(1+\nu)x_\phi & (1+\nu)x_\lambda & (1+\nu)x_z & (1+\nu)x_{\epsilon_R} & 0 \\ \frac{\kappa_p}{1-\beta\rho_\phi}(1+(1+\nu)x_\phi) & \frac{\kappa_p}{1-\beta\rho_\lambda}(1+(1+\nu)x_\lambda) & \frac{\kappa_p}{1-\beta\rho_z}(1+\nu)x_z & +\kappa_p(1+\nu)x_{\epsilon_R} & 0 \\ \frac{\kappa_p/\beta}{1-\beta\rho_\phi}(1+(1+\nu)x_\phi) & \frac{\kappa_p/\beta}{1-\beta\rho_\lambda}(1+(1+\nu)x_\lambda) & \frac{\kappa_p/\beta}{1-\beta\rho_z}(1+\nu)x_z & (\kappa_p(1+\nu)x_{\epsilon_R}/\beta + \sigma_R) & 0 \end{bmatrix}
$$

$$
\Phi_1(\theta) = \begin{bmatrix} \rho_\phi & 0 & 0 & 0 & 0 \\ 0 & \rho_\lambda & 0 & 0 & 0 \\ 0 & 0 & \rho_z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ x_\phi & x_\lambda & x_z & x_{\epsilon_R} & 0 \end{bmatrix}, \quad \Phi_\epsilon(\theta) = \begin{bmatrix} \sigma_\phi & 0 & 0 & 0 \\ 0 & \sigma_\lambda & 0 & 0 \\ 0 & 0 & \sigma_z & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}
$$

M_y' is an $n_y \times 4$ selection matrix that selects rows of Ψ_0 and Ψ_1.

State-Space Representation and Likelihood

- Measurement:

$$y_t = \Psi_0(\theta) + \Psi_1(\theta)t + \Psi_2(\theta)s_t \underbrace{+u_t}_{\text{optional}}$$

- State transition:

$$s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t$$

- Joint density for the observations and latent states:

$$\begin{aligned} p(Y_{1:T}, S_{1:T}|\theta) &= \prod_{t=1}^{T} p(y_t, s_t|Y_{1:t-1}, S_{1:t-1}, \theta) \\ &= \prod_{t=1}^{T} p(y_t|s_t, \theta)p(s_t|s_{t-1}, \theta). \end{aligned}$$

- Problem: we need the marginal $p(Y_{1:T}|\theta)$.

Filtering - General Idea

- State-space representation of linearized DSGE model

$$\begin{aligned} y_t &= \Psi_0(\theta) + \Psi_1(\theta)t + \Psi_2(\theta)s_t(+u_t) \quad \text{measurement} \\ s_t &= \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t \quad \text{state transition} \end{aligned}$$

- Likelihood function:

$$p(Y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|Y_{1:t-1},\theta)$$

- A filter generates a sequence of conditional distributions $s_t|Y_{1:t}$.
- Iterations:
  - Initialization at time $t-1$: $p(s_{t-1}|Y_{1:t-1},\theta)$
  - Forecasting t given $t-1$:
    1. Transition equation: $p(s_t|Y_{1:t-1},\theta) = \int p(s_t|s_{t-1},Y_{1:t-1},\theta)p(s_{t-1}|Y_{1:t-1},\theta)ds_{t-1}$
    2. Measurement equation: $p(y_t|Y_{1:t-1},\theta) = \int p(y_t|s_t,Y_{1:t-1},\theta)p(s_t|Y_{1:t-1},\theta)ds_t$
  - Updating with Bayes theorem. Once y_t becomes available:

$$p(s_t|Y_{1:t},\theta) = p(s_t|y_t,Y_{1:t-1},\theta) = \frac{p(y_t|s_t,Y_{1:t-1},\theta)p(s_t|Y_{1:t-1},\theta)}{p(y_t|Y_{1:t-1},\theta)}$$

Kalman Filter (Linear+Gaussian) and Particle Filter (Fully Nonlinear)

- If the DSGE model is log-linearized and the errors are Gaussian, then the Kalman filter can be used to construct the likelihood function (see summary below).
- Alternatively, one can compute the likelihood by sequential numerical integration which is done for DSGE models that have been solved nonlinearly. The algorithm is called particle or sequential Monte Carlo filter (See Chapter 8 of Herbst and Schorfheide (2015) for details).

Kalman Filter

- We consider the state-space model

$$\begin{aligned} y_t &= \Psi s_t + u_t \\ s_t &= \Phi s_{t-1} + \epsilon_t \end{aligned}$$

where $\epsilon_t \sim iidN(0, \Sigma)$ and $u_t \sim iidN(0, H)$.

- Initialization: if
  1. s_t is stationary we can initialize the filter with the unconditional distribution of s_t. Covariance matrix:

$$\mathbb{E}[s_t s_t'] = \Phi \mathbb{E}[s_t s_t'] \Phi' + \Sigma;$$

  2. otherwise, could assume that $s_t = 0$ for $t = -\tau$ or treat s_0 as parameter.

- In linear Gaussian state-space model all distributions are Gaussian. Thus, Kalman filter only tracks means and covariance matrices.

Kalman Filter – Initialization

$y_t = \Psi s_t + u_t$, $s_t = \Phi s_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, \Sigma)$ and $u_t \sim N(0, H)$.

- Write $\mathbb{E}[s_0] = \hat{s}_{0|0}$ and $\mathbb{V}[s_0] = P_{0|0}$.
- Prior distribution for initial state s_0: $s_0 \sim \mathcal{N}(\bar{s}_{0|0}, P_{0|0})$.

Kalman Filter – Forecasting

$y_t = \Psi s_t + u_t$, $s_t = \Phi s_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, \Sigma)$ and $u_t \sim N(0, H)$.

- At $(t-1)^+$, that is, after observing y_{t-1}, the belief about the state vector has the form $s_{t-1}|Y_{1:t-1} \sim \mathcal{N}(\bar{s}_{t-1|t-1}, P_{t-1|t-1})$.
- "Posterior" from period $t-1$ turns into a prior for $(t-1)^+$.
- Since s_{t-1} and ϵ_t are independent multivariate normal random variables, it follows that

$$s_t|Y_{1:t-1} \sim \mathcal{N}(\bar{s}_{t|t-1}, P_{t|t-1})$$

  where

$$\begin{aligned} \bar{s}_{t|t-1} &= \Phi \bar{s}_{t-1|t-1} \\ P_{t|t-1} &= \Phi P_{t-1|t-1} \Phi' + \Sigma \end{aligned}$$

Kalman Filter – Forecasting and Likelihood Function

$y_t = \Psi s_t + u_t$, $s_t = \Phi s_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, \Sigma)$ and $u_t \sim N(0, H)$.

- The conditional distribution of $y_t | s_t, Y_{1:t-1}$ is of the form

$$y_t | s_t, Y_{1:t-1} \sim \mathcal{N}(\Psi s_t, H)$$

- Since $s_t | Y_{1:t-1} \sim \mathcal{N}(\bar{s}_{t|t-1}, P_{t|t-1})$, we can deduce that the marginal distribution of y_t conditional on $Y_{1:t-1}$ is of the form

$$y_t | Y_{1:t-1} \sim \mathcal{N}(\bar{y}_{t|t-1}, F_{t|t-1})$$

  where $\bar{y}_{t|t-1} = \Psi \bar{s}_{t|t-1}$ and $F_{t|t-1} = \Psi P_{t|t-1} \Psi' + H$.

- Likelihood Function:

$$\begin{aligned} & p(Y_{1:T} | \Psi, \Phi, \Sigma, H) \\ = \; & (2\pi)^{-nT/2} \left(\prod_{t=1}^{T} |F_{t|t-1}| \right)^{-1/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T} (y_t - \bar{y}_{t|t-1})' F_{t|t-1}^{-1} (y_t - \bar{y}_{t|t-1}) \right\} \end{aligned}$$

Kalman Filter – Updating

$y_t = \Psi s_t + u_t$, $s_t = \Phi s_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, \Sigma)$ and $u_t \sim N(0, H)$.

- To obtain the posterior distribution of $s_t | y_t, Y_{1:t-1}$ note that

$$\begin{aligned} s_t &= \bar{s}_{t|t-1} + (s_t - \bar{s}_{t|t-1}) \\ y_t &= \bar{y}_{t|t-1} + \Psi(s_t - \bar{s}_{t|t-1}) + u_t \end{aligned}$$

- and the joint distribution of s_t and y_t is given by

$$\left[\begin{array}{c} s_t \\ y_t \end{array} \right] \Bigg| Y^{t-1} \sim \mathcal{N}\left(\left[\begin{array}{c} \bar{s}_{t|t-1} \\ \bar{y}_{t|t-1} \end{array} \right], \left[\begin{array}{cc} P_{t|t-1} & P_{t|t-1}\Psi' \\ \Psi P'_{t|t-1} & F_{t|t-1} \end{array} \right] \right)$$

Kalman Filter – Updating

$y_t = \Psi s_t + u_t$, $s_t = \Phi s_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, \Sigma)$ and $u_t \sim N(0, H)$.

- Applying Bayes theorem, i.e., calculating a conditional distribution based on a joint...

$$s_t | y_t, Y_{1:t-1} \sim \mathcal{N}(\bar{s}_{t|t}, P_{t|t})$$

where

$$\begin{aligned} \bar{s}_{t|t} &= \bar{s}_{t|t-1} + P_{t|t-1}\Psi' F_{t|t-1}^{-1}(y_t - \bar{y}_{t|t-1}) \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1}\Psi' F_{t|t-1}^{-1}\Psi P_{t|t-1} \end{aligned}$$

- The conditional mean and variance $\bar{y}_{t|t-1}$ and $F_{t|t-1}$ were given above.
- This completes one iteration of the algorithm. The posterior $s_t | Y_{1:t}$ is the prior for the next iteration.

Summary: Conditional Distributions for Kalman Filter

	Distribution	Mean and Variance
$s_{t-1}\|(Y_{1:t-1},\theta)$	$N\big(\bar{s}_{t-1\|t-1}, P_{t-1\|t-1}\big)$	Given from Iteration $t-1$
$s_t\|(Y_{1:t-1},\theta)$	$N\big(\bar{s}_{t\|t-1}, P_{t\|t-1}\big)$	$\bar{s}_{t\|t-1} = \Phi_1 \bar{s}_{t-1\|t-1}$ $P_{t\|t-1} = \Phi_1 P_{t-1\|t-1}\Phi_1' + \Phi_\epsilon \Sigma_\epsilon \Phi_\epsilon'$
$y_t\|(Y_{1:t-1},\theta)$	$N\big(\bar{y}_{t\|t-1}, F_{t\|t-1}\big)$	$\bar{y}_{t\|t-1} = \Psi_0 + \Psi_1 t + \Psi_2 \bar{s}_{t\|t-1}$ $F_{t\|t-1} = \Psi_2 P_{t\|t-1}\Psi_2' + \Sigma_u$
$s_t\|(Y_{1:t},\theta)$	$N\big(\bar{s}_{t\|t}, P_{t\|t}\big)$	$\bar{s}_{t\|t} = \bar{s}_{t\|t-1} + P_{t\|t-1}\Psi_2' F_{t\|t-1}^{-1}(y_t - \bar{y}_{t\|t-1})$ $P_{t\|t} = P_{t\|t-1} - P_{t\|t-1}\Psi_2' F_{t\|t-1}^{-1}\Psi_2 P_{t\|t-1}$

- A DSGE model may or may not have measurement errors.
- Without measurement errors it is important that there are at least as many structural shocks ϵ_t as observables y_t. If not, the forecast error covariance matrix $F_{t|t-1}$ is non-invertible.
- In practice, measurement errors are a bit of a misnomer, as they tend to capture model misspecification.

Maximum Likelihood Estimation

- Recall definition of likelihood function: $p(Y|\theta)$ as function of θ given Y. It's convenient to take logs and work with $\ell_T(\theta|Y) = \log p(Y|\theta)$.

- Decomposition:

$$\ell_T(\theta|Y) = \sum_{t=1}^{T} \log p(y_t|Y_{1:t-1},\theta) = \sum_{t=1}^{T} \log \int p(y_t|s_t,\theta)p(s_t|Y_{1:t-1})ds_t.$$

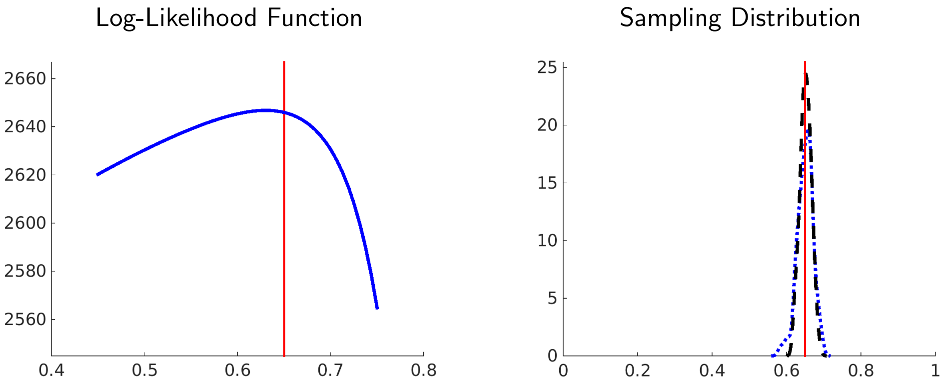
Parameters for Stylized DSGE Model

Parameter	Value	Parameter	Value
β	1/1.01	γ	exp(0.005)
λ	0.15	π^*	exp(0.005)
ζ_p	0.65	ν	0
ρ_ϕ	0.94	ρ_λ	0.88
ρ_z	0.13		
σ_ϕ	0.01	σ_λ	0.01
σ_z	0.01	σ_R	0.01

Maximum Likelihood Estimation: Experiment

- Treat values in Table as "true" parameters.
- Fix all parameters except for the Calvo parameter ζ_p at their "true" values and use the ML approach to estimate ζ_p.
- Use Kalman filter to evaluate likelihood function.
- Data: output growth, labor share, inflation, and interest rate data.

Log-Likelihood Function and Sampling Distribution of $\hat{\zeta}_{p,ml}$



Notes: Left panel: log-likelihood function $\ell_T(\zeta_p|Y)$ for a single data set of size $T = 200$. Right panel: We simulate samples of size $T = 80$ (dotted) and $T = 200$ (dashed) and compute the ML estimator for the Calvo parameter ζ_p. All other parameters are fixed at their "true" value. The plot depicts densities of the sampling distribution of $\hat{\zeta}_p$. The vertical lines in the two panels indicate the "true" value of ζ_p.

Maximum Likelihood Estimation: Asymptotics

- Sampling distribution of MLE can be approximated based on a Central Limit Theorem (CLT):

$$T^{-1/2}\nabla_\theta \ell_T(\theta|Y) \Longrightarrow N(0, \mathcal{I}(\theta_0)),$$

where $\mathcal{I}(\theta_0)$ is the Fisher information matrix.

- Standard error estimates for t-tests and confidence intervals for elements of the parameter vector θ can be obtained from the diagonal elements of the inverse Hessian:

$$[-\nabla^2_\theta \ell_T(\theta|Y)]^{-1}$$

of the log-likelihood function evaluated at the ML estimator.

Maximum Likelihood Estimation: Stochastic Singularity

- Imagine removing all shocks except for the technology shock from the stylized DSGE model, while maintaining that y_t comprises output growth, the labor share, inflation, and the interest rate.

- $\Longrightarrow$ one exogenous shock and four observables.

- DSGE model places probability one on

$$\beta \log R_t - \log \pi_t = \beta \log(\pi^* \gamma / \beta) - \log \pi^*.$$

$\Longrightarrow$ Not consistent with actual data!

- Remedies:
  - "measurement error" approach;
  - "more structural shocks" approach.

Maximum Likelihood Estimation: Lack of Strong Identification

- In many applications it is quite difficult to maximize the likelihood function:
  - local extrema and/or weak curvature in some directions of the parameter space;
  - may be a manifestation of identification problems.
  - Fix some parameters?
- Identification robust-inference, e.g.:
  - ϕ is (identifiable) reduced-form parameter. Model implies $\phi = f(\theta)$.
  - $H_0: \ \theta = \theta_0$ can be translated into $H_0: \ \phi = f(\theta_0)$. Likelihood ratio (LR) statistic is

$$LR(Y|\theta_0) = 2\big[\log p(Y|\hat{\phi}, M_1^{\phi}) - \log p(Y|f(\theta_0), M_1^{\phi})\big] \Longrightarrow \chi^2_{\dim(\phi)}.$$

  - Confidence interval:

$$CS^{\theta}(Y) = \big\{\theta \;\big|\; LR(Y|\theta) \le \chi^2_{crit}\big\},$$

Simulated Minimum Distance (MD) Estimation

- Minimize discrepancy between sample moments of the data $\hat{m}_T(Y)$ and model-implied moments $\mathbb{E}[\hat{m}_T(Y)|\theta, M_1]$:

$$\hat{\theta}_{md} = \text{argmin}_{\theta \in \Theta} \; Q_T(\theta|Y) = \Big\| \hat{m}_T(Y) - \mathbb{E}[\hat{m}_T(Y)|\theta, M_1] \Big\|_{W_T},$$

- **Example 1:**
  - $\hat{m}_T(Y) = \frac{1}{T}\sum y_t y'_{t-1}$.
  - Derive $\mathbb{E}[\hat{m}_T(Y)|\theta, M_1] = \frac{1}{T}\sum \mathbb{E}[y_t y'_{t-1}|\theta, M_1] = \mathbb{E}[y_2 y'_1|\theta, M_1]$ from state-space representation of DSGE.
- **Example 1:**
  - $\hat{m}_T(Y)$ is OLS estimator of a VAR(1).
  - Not feasible to compute $\mathbb{E}[\hat{m}_T(Y)]$ directly.
  - Replace by $\hat{\mathbb{E}}[\hat{m}_T(Y)] = \big(\mathbb{E}[y_{t-1}y'_{t-1}|\theta, M_1]\big)^{-1}\mathbb{E}[y_{t-1}y'_t|\theta, M_1]$,
  - or use simulation approximation.

Simulated Minimum Distance Estimation: Illustration

- Treat values in Table as "true" parameters.
- Fix all parameters except for the Calvo parameter ζ_p at their "true" values and use the ML approach to estimate ζ_p.
- Definition of $\hat{m}_T(Y)$:
  - $y_t = [\log(X_t/X_{t-1}), \pi_t]'$
  - Use VAR(2) in output growth and inflation:

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \Phi_0 + u_t.$$

  - Let $\hat{m}_T(Y) = \hat{\Phi}$ be the OLS estimate of $[\Phi_1, \Phi_2, \Phi_0]'$.

Simulated Minimum Distance Estimation: Implementation

- Objective Function:

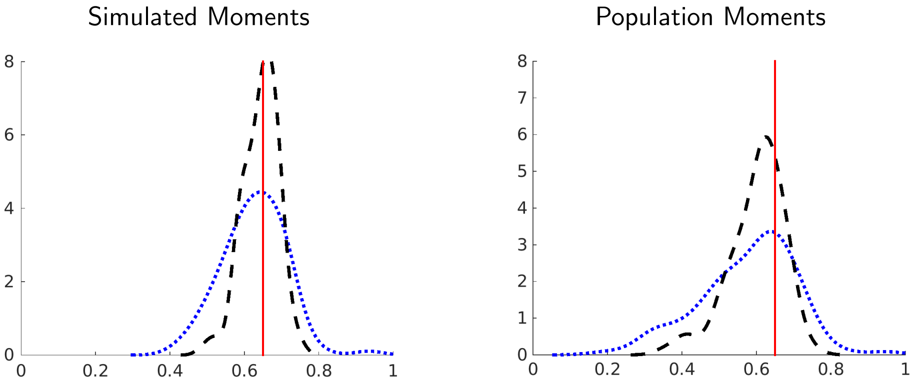
$$Q_T(\theta|Y) = \left\| \hat{m}_T(Y) - \hat{\mathbb{E}}[\hat{m}_T(Y)|\theta, M_1] \right\|_{W_T}.$$

- Given a θ, we simulate $N = 100$ trajectories of length $T + T_0$, discard the first T_0 observations, and define:

$$\hat{\mathbb{E}}[\hat{m}_T(Y)|\theta, M_1] \frac{1}{N} \sum_{i=1}^{N} \hat{m}_T(Y^{(i)}(\theta)).$$

- (Optimal) weight matrix: $W_T = \hat{\Sigma}^{-1} \otimes X'X$, where X is the matrix of VAR (2) regressors and $\hat{\Sigma}$ estimates covariance matrix of the VAR innovations.
- Use same random number seed for simulation as minimization routine varies θ.
- Alternative: use population moments $\left(\mathbb{E}[x_t x_t'|\theta, M_1]\right)^{-1} \mathbb{E}[x_t y_t'|\theta, M_1]$ in objective function.

Sampling Distribution of $\hat{\zeta}_{p,md}$



Notes: We simulate samples of size $T = 80$ (dotted) and $T = 200$ (dashed) and compute two versions of an MD estimator for the Calvo parameter ζ_p. All other parameters are fixed at their "true" value. The plots depict densities of the sampling distribution of $\hat{\zeta}_{p,md}$. The vertical line indicates the "true" value of ζ_p.

Simulated Minimum Distance Estimation: Asymptotics

- Sampling distribution of MD estimator can be approximated based on a Central Limit Theorem.
- Calculations are a bit more complicated because asymptotic variance has to reflect simulation approximation.
- Will lead to standard errors that can be used for t-tests and confidence intervals.
- Sampling distribution can be derived under assumption that:
  - DSGE model M_1 is “true” or
  - a reference model M_0, e.g., VAR, is “true.”

Impulse Response Function (IRF) Matching Estimator

- Special case of minimum distance estimator.
- Attractive if DSGE model is incomplete in the sense that not all structural shocks are specified.
- Mismatch between IRFs can provide insights in misspecification of propagation mechanism.
- However, there are some complications...

Impulse Response Function (IRF) Matching Estimator

- IRFs are based on finite-order VARs.
- Linearized DSGEs are linear state-space models. Three cases – DSGE model solution can be expressed as
  1. finite-order VAR(p);
  2. as VARMA with invertible MA polynomial and rewritten as VAR(∞) in terms of ϵ_t;
  3. as VARMA with non-invertible MA polynomial, cannot be written as VAR(∞) in terms of ϵ_t.

Recall: State-Space Representation of DSGE Model

State-space representation:

$$
\begin{aligned}
y_t &= \Psi_0(\theta) + \Psi_1(\theta) s_t \\
s_t &= \Phi_1(\theta) s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t
\end{aligned}
$$

System matrices:

$$
\Psi_0(\theta) = M_y' \begin{bmatrix} \log\gamma \\ \log(lsh) \\ \log\pi^* \\ \log(\pi^*\gamma/\beta) \end{bmatrix}, \quad x_\phi = -\frac{\kappa_p\psi_p/\beta}{1-\psi_p\rho_\phi}, \quad x_\lambda = -\frac{\kappa_p\psi_p/\beta}{1-\psi_p\rho_\lambda}, \quad x_z = \frac{\rho_z\psi_p}{1-\psi_p\rho_z}, \quad x_{\epsilon_R} = -\psi_p\sigma_R
$$

$$
\Psi_1(\theta) = M_y' \begin{bmatrix} x_\phi & x_\lambda & x_z + 1 & x_{\epsilon_R} & -1 \\ 1+(1+\nu)x_\phi & (1+\nu)x_\lambda & (1+\nu)x_z & (1+\nu)x_{\epsilon_R} & 0 \\ \frac{\kappa_p}{1-\beta\rho_\phi}(1+(1+\nu)x_\phi) & \frac{\kappa_p}{1-\beta\rho_\lambda}(1+(1+\nu)x_\lambda) & \frac{\kappa_p}{1-\beta\rho_z}(1+\nu)x_z & +\kappa_p(1+\nu)x_{\epsilon_R} & 0 \\ \frac{\kappa_p/\beta}{1-\beta\rho_\phi}(1+(1+\nu)x_\phi) & \frac{\kappa_p/\beta}{1-\beta\rho_\lambda}(1+(1+\nu)x_\lambda) & \frac{\kappa_p/\beta}{1-\beta\rho_z}(1+\nu)x_z & (\kappa_p(1+\nu)x_{\epsilon_R}/\beta+\sigma_R) & 0 \end{bmatrix}
$$

$$
\Phi_1(\theta) = \begin{bmatrix} \rho_\phi & 0 & 0 & 0 & 0 \\ 0 & \rho_\lambda & 0 & 0 & 0 \\ 0 & 0 & \rho_z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ x_\phi & x_\lambda & x_z & x_{\epsilon_R} & 0 \end{bmatrix}, \quad \Phi_\epsilon(\theta) = \begin{bmatrix} \sigma_\phi & 0 & 0 & 0 \\ 0 & \sigma_\lambda & 0 & 0 \\ 0 & 0 & \sigma_z & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}
$$

M_y' is an $n_y \times 4$ selection matrix that selects rows of Ψ_0 and Ψ_1.

Impulse Response Function (IRF) Matching Estimator

Example:

- Two MA processes that represent the DSGE models:

$$M_1 \quad : \quad y_t = \epsilon_t + \theta \epsilon_{t-1} = (1+\theta L)\epsilon_t, \quad 0 < \theta < 1, \quad \epsilon_t \sim iidN(0,1)$$
$$M_2 \quad : \quad y_t = \theta \epsilon_t + \epsilon_{t-1} = (\theta + L)\epsilon_t,$$

- M_1 and M_2 are observationally equivalent, because they are associated with the same autocovariance sequence.

- M_1: MA polynomial is invertible. Thus,

$$AR(\infty) \text{ for } M_1: \; y_t = -\sum_{j=1}^{\infty} (-\theta)^j y_{t-j} + \epsilon_t.$$

and

$$\frac{\partial y_t}{\partial \epsilon_t} = 1, \quad \frac{\partial y_{t+1}}{\partial \epsilon_t} = \theta, \quad \frac{\partial y_{t+h}}{\partial \epsilon_t} = 0 \text{ for } h > 1 \quad \Longrightarrow \quad \text{reproduces } M_1 \text{ IRFs}$$

Impulse Response Function (IRF) Matching Estimator

Example:

- Two MA processes that represent the DSGE models:

$$\begin{aligned} M_1 &: \quad y_t = \epsilon_t + \theta\epsilon_{t-1} = (1+\theta L)\epsilon_t, \quad 0<\theta<1, \quad \epsilon_t \sim iidN(0,1) \\ M_2 &: \quad y_t = \theta\epsilon_t + \epsilon_{t-1} = (\theta + L)\epsilon_t, \end{aligned}$$

- M_1 and M_2 are observationally equivalent, because they are associated with the same autocovariance sequence.

- M_2: MA polynomial is NOT invertible. Thus,

$$AR(\infty) \text{ for } M_2: \; y_t = -\sum_{j=1}^{\infty}(-\theta)^j y_{t-j} + u_t, \quad \frac{\theta+L}{1+\theta L}\epsilon_t.$$

- AR does not reproduce IRFs of M_2. IRF matching will be misleading.

Impulse Response Function (IRF) Matching Estimator: Practical Considerations

- Identification of structural shocks in VAR should be “consistent” with DSGE model. Might require to adjust DSGE model.
- Computing IRFs
  - directly from DSGE model
  - versus from VAR approximation of DSGE model
- Weight matrix for the impulse response discrepancies.

Impulse Response Function (IRF) Matching Estimator: Illustration

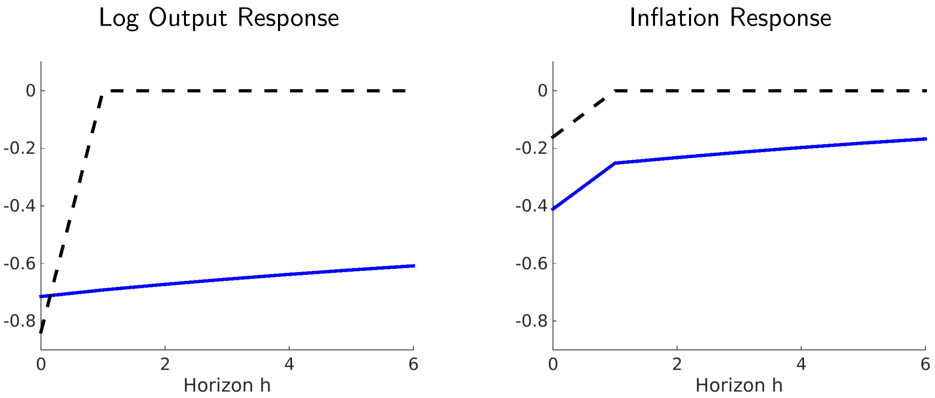
- Treat values in Table as "true" parameters.
- Fix all parameters except for the Calvo parameter ζ_p at their "true" values and use the ML approach to estimate ζ_p.
- $\hat{m}_T(Y)$ contains IRFs from an estimated VAR(p) for

$$y_t = \left[R_t - \pi_t/\beta, \log(X_t/X_{t-1}), \pi_t \right]'.$$

- First equation in VAR corresponds to policy rule. Innovation to this equation is MP shock.
- VAR approximation of DSGE model:

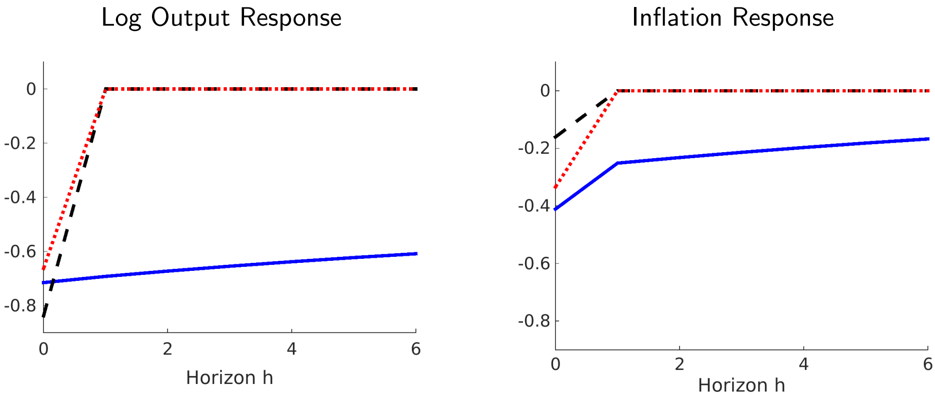
$$\begin{aligned} \Phi_*(\theta) &= \left(\mathbb{E}[x_t x_t' | \theta, M_1]\right)^{-1} \left(\mathbb{E}[x_t y_t' | \theta, M_1]\right), \\ \Sigma^*(\theta) &= \mathbb{E}[y_t y_t' | \theta, M_1] - \mathbb{E}[y_t x_t' | \theta, M_1] \left(\mathbb{E}[x_t x_t' | \theta, M_1]\right)^{-1} \mathbb{E}[x_t y_t' | \theta, M_1]. \end{aligned}$$

DSGE Model and VAR Impulse Responses to a Monetary Policy Shock



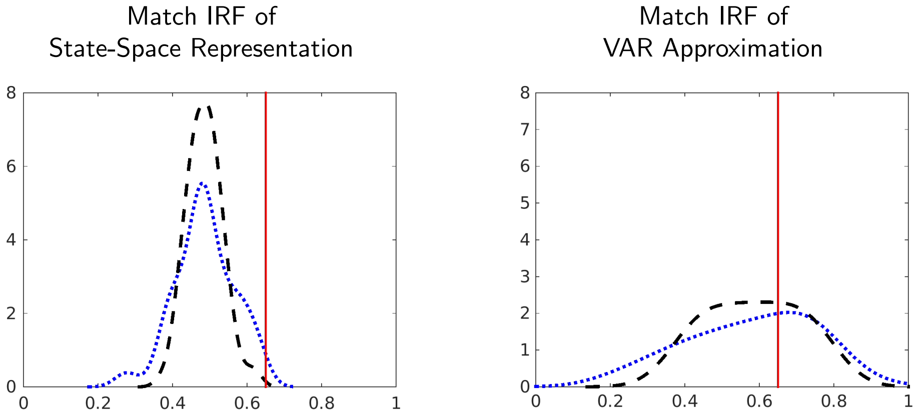
Notes: The figure depicts impulse responses to a monetary policy shock computed from the state-space representation of the DSGE model (dashed) and the VAR(1) approximation of the DSGE model (solid).

Sensitivity of IRF to ζ_p



Notes: The solid lines indicate IRFs computed from the VAR approximation of the DSGE model. The other two lines depict DSGE model-implied IRFs based on $\zeta_p = 0.65$ (dashed) and $\zeta_p = 0.5$ (dotted).

Sampling Distribution of $\hat{\zeta}_{p,irf}$



Notes: IRF matching estimators for ζ_p. Left panel: we use the IRFs from the state-space representation of the DSGE model. Right panel: we use the IRF from the VAR approximation of the DSGE model. The plot depicts densities of the sampling distribution of $\hat{\zeta}_p$ for $T = 80$ (dotted) and $T = 200$ (dashed). The vertical line indicates the "true" value of ζ_p.

Generalized Method of Moments (GMM) Estimation

- Derive moment conditions of the form

$$\mathbb{E}[g(y_{t-p:t}|\theta, M_1)] = 0$$

for $\theta = \theta_0$ from the DSGE model equilibrium.

- **Example:**

$$g(y_{t-p:t}|\theta, M_1) = \left[\begin{array}{c} \big(-\log(X_t/X_{t-1}) + \log R_{t-1} - \log \pi_t - \log(1/\beta) \big) Z_{t-1} \\ \big(\log R_t - \log(\gamma/\beta) - \psi \log \pi_t - (1-\psi) \log \pi^* \big) Z_{t-1} \end{array} \right].$$

- Identifiability of θ requires that the moments be different from zero whenever $\theta \neq \theta_0$.
- A GMM estimator is obtained by replacing population expectations by sample averages:

$$Q_T(\theta|Y) = G_T(\theta|Y)' W_T G_T(\theta|Y), \quad G_T(\theta|Y) = \frac{1}{T} \sum_{t=1}^{T} g(y_{t-p:t}|\theta, M_1).$$

- Model does not have to be solved during the estimation phase.
- It's not straightforward to use equilibrium conditions that contain latent variables, e.g., λ_t in the Phillips curve.