Frequentist Estimation

Frank Schorfheide
Professor of Economics, University of Pennsylvania

September 14, 2021

Popular Frequentist Inference Methods

- **Minimum Distance (MD) Estimation:** minimize discrepancy between sample statistics $\hat{m}_T(Y)$ and model-implied population statistics $\mathbb{E}[\hat{m}_T(\tilde{Y})|\theta, M_1]$:

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

- **Maximum likelihood (ML) Estimation:**

$$\hat{\theta}_{ml} = \text{argmax}_{\theta\in\Theta}\ \log p(Y|\theta, M_1).$$

- Impulse Response Function Matching.
- Generalized Method of Moments (GMM) Estimation.

Simulated Minimum Distance (MD) Estimation

- From now on: drop "tilde" from $\tilde{Y}$ in $\mathbb{E}[\hat{m}_T(\tilde{Y})|\cdot]$.
- Minimize discrepancy between sample moments of the data $\hat{m}_T(Y)$ and model-implied moments $\mathbb{E}[\hat{m}_T(Y)|\theta, M_1]$:

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

- **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 2:**
  - $\hat{m}_T(Y)$ is OLS estimator of, say, a VAR(1).
  - Not feasible to compute $\mathbb{E}[\hat{m}_T(Y)|\theta, M_1]$ directly.
  - Use simulation approximation for $\hat{\mathbb{E}}[\hat{m}_T(Y)|\theta, M_1]$.
  - Or, replace by probability limit as $T \longrightarrow \infty$ which is $\left(\mathbb{E}[y_{t-1}y_{t-1}'|\theta, M_1]\right)^{-1}\mathbb{E}[y_{t-1}y_t'|\theta, M_1]$.

Simulated Minimum Distance Estimation: Illustration

- Choose a set of "true" parameters.
- Fix all parameters except for the Calvo parameter ζ_p at their "true" values and use the MD 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:

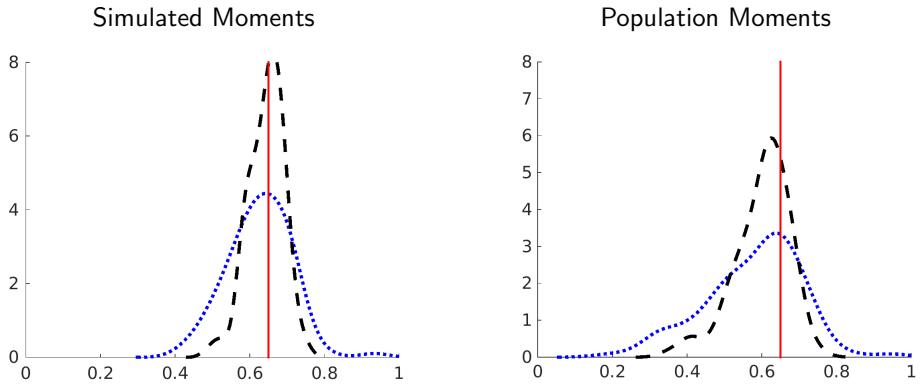
$$Q_T(\theta|Y) = \left\| \hat{m}_T(Y) - \hat{\mathbb{E}}[\hat{m}_T(Y)|\theta, M_1] \right\|_{W_T}.$$

- 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."

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:

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

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

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

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).

State-space model:

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

All conditional distributions are Normal, track means and variances:

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

Note: 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.

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:

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

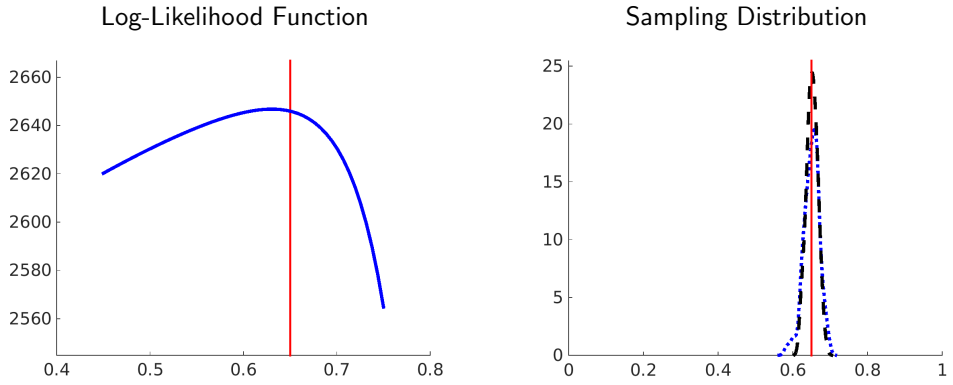
- Maximum likelihood (ML) estimator

$$\hat{\theta}_{ml} = \text{argmax}_{\theta \in \Theta} \ \log p(Y|\theta).$$

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

- 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:

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

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:

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