Statistics Review

Lecture 1

ECON5345, Spring 2026

Byoungchan Lee

January 1, 2026

Contents

Basic Asymptotics I

Law of Large Numbers and Central Limit Theorem
For a covariance-stationary stochastic process $\{x_t\}$, under some regularity conditions,

$$
\begin{aligned}
LLN: \quad & \frac{1}{T}\sum_{t=1}^{T} x_t && \overset{p/a.s.}{\rightarrow} \quad \mu_x, \\
CLT: \quad & \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} x_t - \mu_x\right) && \overset{d}{\rightarrow} \quad \mathcal{N}(0, LRV),
\end{aligned}
$$

where the long-run variance $LRV = \cdots + \Gamma_{-1} + \Gamma_0 + \Gamma_1 + \ldots$ and $\Gamma_j = cov(x_t, x_{t-j})$.

Basic Asymptotics II

Some asymptotic theory
Let g be a continuous/differentiable function (e.g., a mapping from VAR coefficients to impulse response functions / forecast error variance decompositions).

Continuous Mapping Theorem

$$X_T \overset{p/d/a.s.}{\rightarrow} X \quad \Rightarrow \quad g(X_T) \overset{p/d/a.s.}{\rightarrow} g(X)$$

ex) $\sqrt{T}(\hat{\theta}_T - \theta_0) \overset{d}{\rightarrow} \mathcal{N}(0, V) \quad \Rightarrow$
$\sqrt{T}(\hat{\theta}_T - \theta_0)' V^{-1} \sqrt{T}(\hat{\theta}_T - \theta_0) \overset{d}{\rightarrow} ||\mathcal{N}(0, I)||^2 = \chi^2(Rank(V))$.

Basic Asymptotics III

Slutsky's Theorem

If $X_T \xrightarrow{d} X$ and $Y_T \xrightarrow{p} y$ (constant),

$$X_T + Y_T \xrightarrow{d} X + y$$

$$X_T Y_T \xrightarrow{d} Xy$$

$$X_T / Y_T \xrightarrow{d} X/y \quad \text{if } y \neq 0.$$

ex) $\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, V), \quad \hat{V} \xrightarrow{p} V \quad \Rightarrow$

$\sqrt{T}\hat{V}^{-1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, I)$

$\sqrt{T}(\hat{\theta}_T - \theta_0)'\hat{V}^{-1}\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} ||\mathcal{N}(0, I)||^2 = \chi^2(Rank(V))$.

Basic Asymptotics IV

Delta Method

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, V)$$
$$\Rightarrow \quad \sqrt{T}(g(\theta_T) - g(\theta_0)) \xrightarrow{d} \mathcal{N}(0, G'VG),$$

where $G' = \frac{\partial g}{\partial \theta'}|_{\theta=\theta_0}$.

ex) Consider an AR(1) process. $x_t = \rho x_{t-1} + e_t$, where $e_t \sim i.i.d.\mathcal{N}(0, \sigma^2)$. Given $\sqrt{T}(\hat{\rho} - \rho) \xrightarrow{d} \mathcal{N}(0, V)$, what can we say about the IRF $\{1, \rho, \rho^2, \dots\}$?

ans) $\sqrt{T}([\hat{\rho}]^h - \rho^h) \xrightarrow{d} \mathcal{N}(0, (h[\hat{\rho}]^{h-1})^2 V)$.

OLS I

- Model:

$$y_t = x_t'\beta + e_t$$

- Estimation: $\min_\beta \sum_t (y_t - x_t'\beta)^2$
- Identification assumption: $\mathbb{E}[x_t e_t] = 0$, $\Sigma_{xx} = \mathbb{E}[x_t x_t']$ is non-singular.
- Threats to identification
  - Measurement errors in x_t
    - It usually matters when x_t is a generated regressor.
    - If the null hypothesis is $\beta = 0$, we do not need to adjust inferences although there exist measurement errors (Pagan, 1984).
  - Omitted variable bias
  - Reverse-causality

OLS II

- Estimator:

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\sum_{t=1}^{T} x_t y_t\right)$$
$$= \beta + \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} x_t e_t\right)$$

- Consistency: $\hat{\beta} \xrightarrow{p} \beta + \Sigma_{xx}^{-1} 0 = \beta$
- Asymptotic Normality: $\sqrt{T}(\hat{\beta} - \beta) = \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t e_t\right) \xrightarrow{d} \Sigma_{xx}^{-1}\mathcal{N}(0, LRV(x_t e_t))$
- For statistical inference, we need to consistently estimate $LRV(x_t e_t)$.

Coibion and Gorodnichenko (2015, AER) I

$$\pi_{t+3,t} - F_t\pi_{t+3,t} = c + \beta(F_t\pi_{t+3,t} - F_{t-1}\pi_{t+3,t}) + error_t$$

- $\pi_{t+3,t}$: average inflation rate over the current and next three quarters
- F_t: average time t forecast across agents

Theoretical Predictions

- FIRE: $\beta = 0$
- Sticky information (Mankiw and Reis, 2002): $\beta = \lambda/(1-\lambda)$, where λ=prob. of no information update.

Coibion and Gorodnichenko (2015, AER) II

Table 1—Tests of the Inflation Expectations Process

Forecast error $\pi_{t+3,t} - F_t\pi_{t+3,t}$	Additional control: z_{t-1} None (1)	Inflation (2)	Average quarterly 3-month Tbill rate (3)	Quarterly change in the log of the oil price (4)	Average unemployment rate (5)
Panel A. $\pi_{t+3,t} - F_t\pi_{t+3,t} = c + \gamma F_t\pi_{t+3,t} + \delta z_{t-1} + error_t$					
Constant	−0.181	−0.045	−0.091	−0.181	1.449**
	(0.248)	(0.223)	(0.236)	(0.221)	(0.676)
$F_t\pi_{t+3,t}$	0.059	−0.299**	0.210*	0.045	0.095
	(0.085)	(0.148)	(0.111)	(0.078)	(0.085)
Additional control: z_{t-1}		0.318**	−0.125*	1.603**	−0.281**
		(0.147)	(0.066)	(0.763)	(0.117)
Observations	178	178	178	178	178
R^2	0.010	0.109	0.054	0.046	0.148
Panel B. $\pi_{t+3,t} - F_t\pi_{t+3,t} = c + \beta(F_t\pi_{t+3,t} - F_{t-1}\pi_{t+3,t}) + \delta z_{t-1} + error_t$					
Constant	0.002	−0.074	0.151	−0.021	1.134**
	(0.144)	(0.174)	(0.175)	(0.146)	(0.546)
$F_t\pi_{t+3,t} - F_{t-1}\pi_{t+3,t}$	1.193**	1.141**	1.196**	1.125**	1.062**
	(0.497)	(0.458)	(0.504)	(0.499)	(0.465)
Additional control: z_{t-1}		0.021	−0.029	0.576	−0.178**
		(0.050)	(0.031)	(0.608)	(0.076)
Observations	173	173	173	173	173
R^2	0.195	0.197	0.201	0.200	0.249

Notes: The table reports coefficient estimates for the specified equations at the top of each panel. The additional controls (z) are lagged by one quarter. Newey-West standard errors are in parentheses.

$\beta \approx 1$, $\lambda \approx 0.5$, Information is updated every 2 quarters on avg.

IV I

- Model:

$$y_t = x_t'\beta + e_t,$$
$$(\text{stacked}) : Y = X\beta + e$$

- Identification assumption: $\mathbb{E}[z_t e_t] = 0$, $\Sigma_{zx} = \mathbb{E}[z_t x_t']$ has a full rank, and $\Sigma_{zz} = \mathbb{E}[z_t z_t']$ is non-singular.
- Estimator:

$$\hat{\beta}^{IV} = (X'P_Z X)^{-1}(X'P_Z Y) = \beta + (X'P_Z X)^{-1}(X'P_Z e),$$

where $P_Z = Z(Z'Z)^{-1}Z$.
Note that $\frac{1}{T}X'Z = \frac{\sum_{t=1}^{T} x_t z_t'}{T} \xrightarrow{p} \mathbb{E}[x_t z_t']$.

IV II

- Consistency: $\hat{\beta}^{IV} \xrightarrow{p} \beta + (\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})^{-1}(\Sigma_{xz}\Sigma_{zz}^{-1}\mathbb{E}\,[z_t e_t]) = \beta$
- Asymptotic Normality:
$\sqrt{T}(\hat{\beta}^{IV} - \beta) \approx (\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})^{-1}(\Sigma_{xz}\Sigma_{zz}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t e_t)$
$\xrightarrow{d} (\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx})^{-1}(\Sigma_{xz}\Sigma_{zz}^{-1})\mathcal{N}(0, LRV(z_t e_t))$
- For statistical inference, we need to consistently estimate $LRV(z_t e_t)$.

- Finding a proper IV is not easy.
- Weak IV: What if $P_Z X \approx 0$?
$\hat{\beta}^{IV} = (X'P_Z X)^{-1}(X'P_Z Y) \approx 0^{-1}(X'P_Z Y)$.
  - $\hat{\beta}^{IV}$ can easily blow up. Its sign can be easily flipped.
  - Sampling distribution of $\hat{\beta}^{IV}$ may be very different from the asymptotic normal distribution above.
  - (First-stage.) We can directly check whether $P_Z X \approx 0$ or not.
  - What if your IV is weak? See Andrews, Stock and Sun (2019).

Coibion and Gorodnichenko (2012, AEJ: Macro) I

$$i_t = c + \phi_\pi E_{t-} \pi_{t+h_\pi} + \phi_{dy} E_{t-} dy_{t+h_{dy}} + \phi_x E_{t-} x_{t+h_x} + \sum_{k=1}^{K} \rho_{i,k} i_{t-k} + u_t,$$
$$u_t = \sum_{j=1}^{J} \rho_{u,j} u_{t-j} + \varepsilon_t,$$

- The target FFR i_t is very persistent.
- Two explanations for the persistence
  - Interest rate smoothing: $\rho_{i,k} > 0$.
  - Persistent monetary policy shocks: $\rho_{u,j} > 0$.
- How to test the two hypotheses?
  If $\rho_{u,j} > 0$ is the primary reason for the persistence in i_t, the response of i_t to nonmonetary policy shocks should be less persistent.
- Use nonmonetary shocks as IVs.

Coibion and Gorodnichenko (2012, AEJ: Macro) II

Table 5—Instrumental Variable Estimation of the Taylor Rule

	1987:IV–2006:IV			1983:I–2006:IV		
	Least squares (1)	IV inertia only (2)	IV nested case (3)	Least squares (4)	IV inertia only (5)	IV nested case (6)
ϕ_π: $E_{t-}\pi_{t+2,t+1}$	0.40***	0.60***	0.71***	0.35***	0.62***	0.549***
	(0.06)	(0.19)	(0.18)	(0.06)	(0.23)	(0.09)
ϕ_x: $E_{t-}x_t$	0.14***	0.18	0.22***	0.08***	0.18***	0.11***
	(0.02)	(0.13)	(0.07)	(0.02)	(0.07)	(0.02)
ϕ_{dy}: $E_{t-}dy_t$	0.19***	0.27***	0.23***	0.22***	0.31***	0.27***
	(0.03)	(0.06)	(0.03)	(0.02)	(0.05)	(0.03)
ρ_i: i_{t-1}	0.83***	0.79***	0.70***	0.87***	0.81***	0.82***
	(0.03)	(0.15)	(0.10)	(0.03)	(0.05)	(0.03)
ρ_u: u_{t-1}			0.14			0.10
			(0.11)			(0.09)
R^2	0.986	0.979	0.979	0.980	0.961	0.978
s.e.e.	0.266	0.325	0.323	0.354	0.497	0.367
AIC	0.251			0.811		
SIC	0.404			0.946		

Notes: The table presents least squares and instrumental variable (IV) estimates of the Taylor rule in equation (3) in the text. In columns 2, 3, 5, and 6, instruments include a constant and two lags of technology shocks from Gali (1999), TFP residuals from Basu, Fernald, and Kimball (2004), oil supply shocks from Kilian (2009), news shocks from Beaudry and Portier (2006), and fiscal shocks from Romer and Romer (2010). Newey-West HAC standard errors are reported in parentheses. See Section III for details.

$\rho_i >> 0$, $\rho_u \approx 0$.

Nakamura and Steinsson (2014, AER) I

Want to estimate G military spending multiplier β:

$$\frac{Y_{i,t} - Y_{i,t-2}}{Y_{i,t-2}} = \alpha_i + \gamma_t + \beta \frac{G_{i,t} - G_{i,t-2}}{Y_{i,t-2}} + \varepsilon_{it},$$

where Y_{it} and G_{it} are per capital output and military spending in state i in year t.

- There exists an obvious endogeneity issue here. (why?)

IV: $\{D_i \frac{G_t - G_{t-2}}{Y_{t-2}}\}$.

- Justification: (1) National military spending G_t is mostly driven by geopolitical events. (2) Sensitivity of G_{it} to G_t varies across states.
- Identifying assumption: "the United States does not embark on a military buildup because states that receive a disproportionate amount of military spending are doing poorly relative to other states."

Nakamura and Steinsson (2014, AER) II

TABLE 2—THE EFFECTS OF MILITARY SPENDING

	Output		Output defl. state CPI		Employment		CPI	Population
	States	Regions	States	Regions	States	Regions	States	States
Prime military contracts	1.43	1.85	1.34	1.85	1.28	1.76	0.03	−0.12
	(0.36)	(0.58)	(0.36)	(0.71)	(0.29)	(0.62)	(0.18)	(0.17)
Prime contracts plus military compensation	1.62	1.62	1.36	1.44	1.39	1.51	0.19	0.07
	(0.40)	(0.84)	(0.39)	(0.96)	(0.32)	(0.91)	(0.16)	(0.21)
Observations	1,989	390	1,989	390	1,989	390	1,763	1,989

Notes: Each cell in the table reports results for a different regression with a shorthand for the main regressor of interest listed in the far left column. A shorthand for the dependent variable is stated at the top of each column. The dependent variable is a two-year change divided by the initial value in each case. Output and employment are per capita. The regressor is the two-year change divided by output. Military spending variables are per capita except in Population regression. Standard errors are in parentheses. All regressions include region and time fixed effects, and are estimated by two-stage least squares. The sample period is 1966–2006 for output, employment, and population, and 1969–2006 for the CPI. Output is state GDP, first deflated by the national CPI and then by our state CPI measures. Employment is from the BLS payroll survey. The CPI measure is described in the text. Standard errors are clustered by state or region.

Open economy relative multiplier $>$ closed-econ multiplier:
the responses of monetary policy stance and aggregate taxation are differenced out by the time-fixed effects.

MoM I

- Moment conditions: there uniquely exists θ_0 such that
$$g(\theta_0) = 0,$$
where $g(\theta) = \mathbb{E}\left[h(\theta, x_t)\right]$, g and θ are $r \times 1$ vectors.
- (just identified) # of parameters = # of moment conditions
- Estimation: Let $g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} h(\theta, x_t)$.
$$g_T(\hat{\theta}_T) = 0.$$
- Example:
$h(\mu, \sigma^2, x) = \begin{pmatrix} x - \mu \\ \sigma^2 - (x-\mu)^2 \end{pmatrix} \Rightarrow \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} \bar{x} \\ \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2 \end{pmatrix}$.
$h(\beta, x_t, y_t) = x_t y_t - x_t x_t' \beta \Rightarrow$ OLS

MoM II

- From conditional moments to unconditional moments:
  Consider the Euler equation, $\mathbb{E}_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} (1 + r_{t+1}) - 1 \right] = 0.$
  By the L.I.E.,
  $\mathbb{E} \left[\left\{ \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} (1 + r_{t+1}) - 1 \right\} z_t \right] = 0$ for any $z_t \in \mathcal{F}_t$.
  We may consider $z_t = (1, C_t / C_{t-1})'$ to estimate β and γ.

- Moments in g may include impulse response coefficients, forecast error variance decompositions, mean, variance, auto-covariances, cross-correlations, cross-sectional distributions, etc.

MoM III

- Heuristic derivation of the asymptotic dist.:

$$0 = g_T(\hat{\theta}_T) \approx g_T(\theta_0) + G'_T(\hat{\theta}_T - \theta_0),$$

where $G'_T = \frac{\partial g_T}{\partial \theta'}|_{\theta_0}$. Then,

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \approx -(G')^{-1}\sqrt{T} g_T(\theta_0),$$

where $G'_T = \frac{\partial g_T}{\partial \theta'}|_{\theta_0} \xrightarrow{p} \frac{\partial g}{\partial \theta'}|_{\theta_0} = G'$, which is assumed to be invertible. Note that

$$\sqrt{T} g_T(\theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} h(\theta_0, x_t) \xrightarrow{d} \mathcal{N}(0, LRV(h(\theta_0, x_t))),$$

$$\therefore \sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, (G')^{-1} LRV(h(\theta_0, x_t))(G)^{-1}).$$

Kocherlakota (1996, JEL) I

Equity Premium Puzzle based on Aggregate Consumption.

$$\mathbb{E}_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} (1 + r^i_{t+1}) \right] = 1 \text{ for } i = \text{ bond, stock.}$$

Take the difference, divide by β, and apply the L.I.E.:

$$\mathbb{E} \left[\left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} (r^s_{t+1} - r^b_{t+1}) \right] = 0.$$

When

$$e_t = \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} (r^s_{t+1} - r^b_{t+1}),$$

$\bar{e}$ should be close to zero.

Kocherlakota (1996, JEL) II

THE EQUITY PREMIUM PUZZLE

a	$\bar{e}$	t-stat
0.0	0.0594	3.345
0.5	0.0577	3.260
1.0	0.0560	3.173
1.5	0.0544	3.082
2.0	0.0528	2.987
2.5	0.0512	2.890
3.0	0.0496	2.790
3.5	0.0480	2.688
4.0	0.0464	2.584
4.5	0.0449	2.478
5.0	0.0433	2.370
5.5	0.0418	2.262
6.0	0.0403	2.153
6.5	0.0390	2.044
7.0	0.0372	1.934
7.5	0.0357	1.824
8.0	0.0341	1.715
8.5	0.0326	1.607
9.0	0.0310	1.501
9.5	0.0295	1.395
10.0	0.0279	1.291

In this table, $\bar{e}$ is the sample mean of $e_t = (C_{t+1}/C_t)^{-\alpha}$ $(R^s_{t+1} - R^b_{t+1})$ and α is the coefficient of relative risk

Toda and Walsh (2015, JPE) I

Sometimes, LLN and CLT may not work!

Consumption Euler equation at the individual level

$$1 = \mathbb{E}_{i,t-1}\left[\beta(c_{i,t}/c_{i,t-1})^{-\gamma}(1+r_t^j)\right] \qquad \forall\ i \text{ and } j$$

$$\overset{LIE}{\Rightarrow} 1 = \mathbb{E}\left[\beta\int\left(\frac{c_{i,t}}{c_{i,t-1}}\right)^{-\gamma} di\ (1+r_t^j)\right]$$

$$\Rightarrow 0 = \mathbb{E}\left[\int\left(\frac{c_{i,t}}{c_{i,t-1}}\right)^{-\gamma} di\ (r_t^{stock} - r_t^{bond})\right]$$

Thus, we can consider a MoM estimator of γ by solving:

$$0 = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{I}\sum_{i=1}^{I}\left(\frac{c_{i,t}}{c_{i,t-1}}\right)^{-\gamma}(r_t^{stock} - r_t^{bond}).$$

Toda and Walsh (2015, JPE) II

$$0 = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{I}\sum_{i=1}^{I}\left(\frac{c_{i,t}}{c_{i,t-1}}\right)^{-\gamma}(r_t^{stock} - r_t^{bond}).$$

However, what if $\int \left(\frac{c_{i,t}}{c_{i,t-1}}\right)^{-\gamma} di = \infty$ for the true value of γ?

- Toda and Walsh show that the cross-sectional distribution of consumption growth features fat upper and lower tails.
- In this case, the lower tail may make $\int \left(\frac{c_{i,t}}{c_{i,t-1}}\right)^{-\gamma} di = \infty$ for reasonable values of γ.
- As a result, $\hat{\gamma} \approx 0$.

GMM I

- Moment conditions: there uniquely exists θ_0 such that

$$g(\theta_0) = 0,$$

where $g(\theta) = \mathbb{E}\left[h(\theta, x_t)\right]$, $dim(g) = r \geq dim(\theta) = a$.
- (over identified) if $r > a$
- Estimation: Let $g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} h(\theta, x_t)$.

$$\min_{\theta} g_T(\theta)' W_T g_T(\theta),$$

where $\{W_T\}$ is a sequence of p.d. weighting matrix such that $W_T \xrightarrow{p} W$.
- F.O.C. $\left[\frac{\partial g_T(\hat{\theta}_T)}{\partial \theta'}\right]' W_T g_T(\hat{\theta}_T) = 0$.

GMM II

- Heuristic derivation of the asymptotic dist.:

$$
\begin{aligned}
0 &= \left[\frac{\partial g_T(\hat{\theta}_T)}{\partial \theta'}\right]' W_T g_T(\hat{\theta}_T) \\
&\approx \left[\frac{\partial g_T(\hat{\theta}_T)}{\partial \theta'}\right]' W_T \left[g_T(\theta_0) + G_T'(\hat{\theta}_T - \theta_0)\right],
\end{aligned}
$$

where $G_T' = \frac{\partial g_T}{\partial \theta'}|_{\theta_0}$. Then, for $\Omega = LRV(h(\theta_0, x_t))$

$$
\begin{aligned}
&\sqrt{T}(\hat{\theta}_T - \theta_0) \approx -(GWG')^{-1} GW \sqrt{T} g_T(\theta_0) \\
&\Rightarrow \sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} (GWG')^{-1} GW \mathcal{N}(0, \Omega).
\end{aligned}
$$

- Optimal weighting matrix $W = \Omega^{-1}$. Then, $var(\sqrt{T}(\hat{\theta}_T - \theta_0)) = (G\Omega^{-1}G')^{-1}$.

GMM III

- How to estimate the optimal weighting matrix?
  - **(Two-step approach.)** $\hat{\theta}_T$ is consistent for any W_T. Thus, start with $W_T = I$.
  - Estimate $\Omega = LRV(h(\theta_0, x_t))$ using $\{h(\hat{\theta}_T, x_t)\}$. Denote the estimate by $\hat{\Omega}$.
  - Set W_T at $\hat{\Omega}^{-1}$, which is a consistent estimate of Ω^{-1}. Run GMM again.
  - **(Iterated GMM.)** iterate until $\hat{W}$ converges.

- Overidentification tests (Hansen's J-test)
  - $\sqrt{T} g_T(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Omega)$
    $\Rightarrow \quad \sqrt{T} g_T(\theta_0)' \Omega^{-1} \sqrt{T} g_T(\theta_0) \xrightarrow{d} \chi^2(r)$, where $r = dim(g)$.
  - Based on consistent estimators $\hat{\theta}_T$ and $\hat{\Omega}$,
    $J = \sqrt{T} g_T(\hat{\theta}_T)' \hat{\Omega}^{-1} \sqrt{T} g_T(\hat{\theta}_T) \xrightarrow{d} \chi^2(r - a)$, where $r = dim(g)$ and $a = dim(\theta)$.
  - In theory, we can use this test to check whether our moment conditions are consistent with the data.

GMM IV

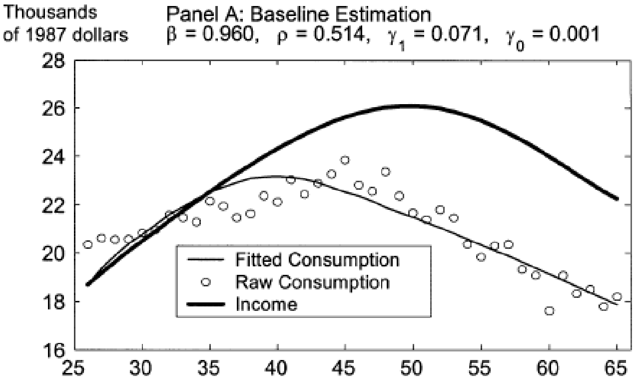
- IV exogeneity can be tested.
- In practice, it is well-known that the finite-sample performances of both the optimal GMM and J-tests are not so good...
- It is advisable to supplement the results by trying several specifications when you use GMM.

Gourinchas and Parker (2002, ECTA) I

Using a Method of Simulated Moments (or SMM) to estimate a life-cycle model

- A model of a typical household working from age 25 to 65 and retire thereafter.
- Uninsurable idiosyncratic earnings risk.
- MSM
  - Pick parameter values.
  - Solve the model.
  - Simulate the moments of interest (20,000 income processes over 40 years in the paper).
  - Compare the moments based on the simulated data to the empirical counterpart.
  - Iterate until the difference is minimized.

Gourinchas and Parker (2002, ECTA) II



Given the estimated income process, $\beta = 0.96$ and $RRA = 0.514$ replicate the life-cycle pattern in consumption, derived from CEX, pretty well. When the optimal weighting matrix is used, $RRA = 1.4$.

MLE I

- When the previous methods focus on a subset of the population properties (i.e. moments), they are considered as the limited information approach.
- In contrast, MLE is a full information approach in the sense that it requires us to specify the whole data-generating process, and we use the model structure in estimation.
- ex) suppose that the economy is driven by two shocks (say, a productivity shock and a monetary policy shock). If we estimate the model by matching the impulse responses to only monetary policy shocks, it is a limited information approach. For this approach, we do not need to fully specify how the productivity shocks propagate in the model. In contrast, to use MLE, we need a fully specified DGP.
- Because we use the "full" information, it is (asymptotically) efficient when the model is correctly specified.

MLE II

- PDF: $X^T = \{x_T, x_{T-1}, \ldots, x_1\}$ is given by $f(X^T|\theta)$.
- Likelihood Ftn: $\mathcal{L}(\theta|X^T) \equiv f(X^T|\theta)$
- Log Likelihood Ftn: $l(\theta|X^T) = \log(\mathcal{L}(\theta|X^T))$

- Estimation: $\max_\theta \mathcal{L}(\theta|X^T)$ or $l(\theta|X^T)$.

- Asymptotics:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^{-1}),$$

  where $\mathcal{I}$ is the Fisher information matrix.
- Two methods of estimating the information matrix
  - $\mathcal{I} = -\mathbb{E}\left[\frac{\partial^2 l(\theta_0)/T}{\partial\theta\partial\theta'}\right] \quad \Rightarrow \quad \hat{\mathcal{I}}_1 = -\frac{1}{T}\frac{\partial^2 l(\hat{\theta}_T)}{\partial\theta\partial\theta'}$.
  - $\mathcal{I} = \mathbb{E}\left[\frac{\partial \log(f(x_t|x^{t-1},\theta_0))}{\partial\theta}\frac{\partial \log(f(x_t|x^{t-1},\theta_0))}{\partial\theta'}\right]$

    $\Rightarrow \quad \hat{\mathcal{I}}_2 = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial \log(f(x_t|x^{t-1},\hat{\theta}_T))}{\partial\theta}\frac{\partial \log(f(x_t|x^{t-1},\hat{\theta}_T))}{\partial\theta'}\right]$.

MLE III

- If the model is correctly specified, $\hat{\mathcal{I}}_1$ and $\hat{\mathcal{I}}_2$ should be similar.
- If not, the model might be misspecified.
- White (1982) proposes the "quasi-maximum likelihood" standard error that is valid sometimes:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \overset{approximately}{\sim} \mathcal{N}(0, (\hat{\mathcal{I}}_1 \hat{\mathcal{I}}_2^{-1} \hat{\mathcal{I}}_1)^{-1})$$

Laubach and Williams (2003, REStat) I

MLE of the unobservable variables, such as the natural rate of interest and the trend growth rate of output

- Transition Equations:

$$
\begin{aligned}
y_t^* &= y_{t-1}^* + g_t + \epsilon_{4,t}, \\
g_t &= g_{t-1} + \epsilon_{5,t}, \\
z_t &= z_{t-1} + \epsilon_{3,t}, \\
r_t^* &= c g_t + z_t.
\end{aligned}
$$

- Measurement Equations:

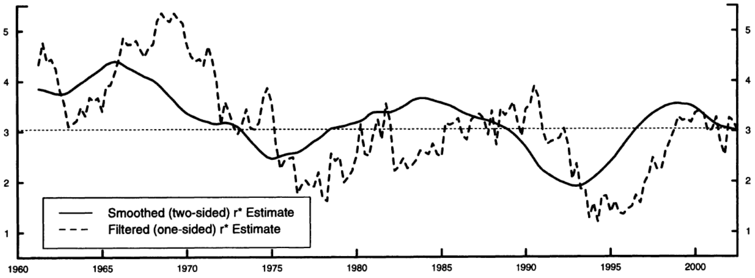
$$
\tilde{y}_t = a_{y,1}\tilde{y}_{t-1} + a_{y,2}\tilde{y}_{t-2} + \frac{a_r}{2}\sum_{j=1}^{2}(r_{t-j} - r_{t-j}^*) + \epsilon_{t,1},
$$

$$
\pi_t = B(L)\pi_{t-1} + b_y\tilde{y}_{t-1} + b_i(\pi_t^I - \pi_t) + b_o(\pi_{t-1}^o - \pi_{t-1}) + \epsilon_{2,t}.
$$

- MLE using the Kalman filter

Laubach and Williams (2003, REStat) II

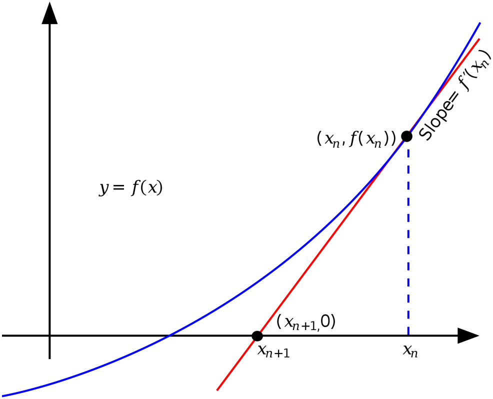
FIGURE 1.—ONE- VS. TWO-SIDED ESTIMATES OF THE NATURAL RATE OF INTEREST (BASELINE MODEL)



We can compute the MLE of the unobservable variables, such as y^*, g, z, and r^*.

Newton-Raphson root-finding algorithm

Want. find x s.t. $f(x) = 0$.



Given x_n,

- linearly approximate f at x_n: $f_n = f(x_n) + f'(x_n)(x - x_n)$
- find x_{n+1} s.t. $f_n(x_{n+1}) = 0$
- $x_{n+1} = x_n - [f'(x_n)]^{-1} f(x_n)$.

Figure source: Wikipedia

Newton's algorithm in optimization I

Want. find argmin $f(x)$ $\Leftrightarrow$ find x s.t. $f'(x) = 0$.

Apply the N-R algorithm to $f'(x) = 0$. Given x_n,

- linearly approximate f' at x_n: $g_n = f'(x_n) + f''(x_n)(x - x_n)$
- find x_{n+1} s.t. $g_n(x_{n+1}) = 0$
- $x_{n+1} = x_n - [f''(x_n)]^{-1} f'(x_n)$.

Alternative interpretation. Given x_n,

- quadratic approximation of f at x_n:
  $g_n = f(x_n) + f'(x_n)(x - x_n) + \frac{1}{2} f''(x_n)(x - x_n)^2$
- find x_{n+1} s.t. $x_{n+1} =$ argmin g_n.
- f.o.c.: $0 = g_n'(x_{n+1}) = f'(x_n) + f''(x_n)(x_{x+1} - x_n)$.
- $x_{n+1} = x_n - [f''(x_n)]^{-1} f'(x_n)$.

Newton's algorithm in optimization II

Scalar case. $x_{n+1} = x_n - [f''(x_n)]^{-1} f'(x_n)$.

Multivariate case. $f : \mathbb{R}^k \to \mathbb{R}$.

$$\underbrace{x_{n+1}}_{k\times 1} = \underbrace{x_n}_{k\times 1} - [\underbrace{D^2 f(x_n)}_{\text{hessian, } k\times k}]^{-1} \underbrace{Df(x_n)}_{\text{gradient, } k\times 1}$$

Computational issues.

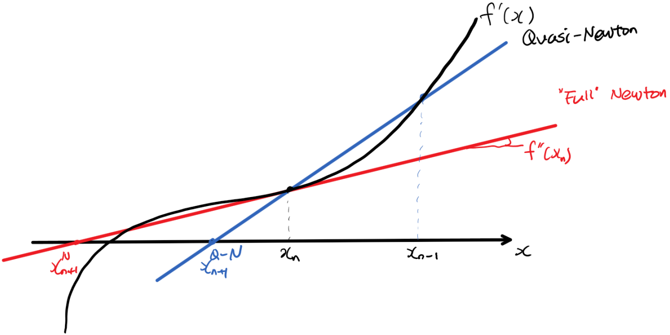
- Computing $H_n = D^2 f(x_n)$ and inverting it is numerically costly.
- Note that at $x^* = \text{argmin } f$, $H = D^2 f(x^*)$ is p.s.d.
- However, there is no guarantee that H_n is also p.s.d. Thus, we are not sure whether

$$x_{n+1} = \text{argmin } g_n(x), \qquad \text{where}$$
$$g_n(x) = f(x_n) + Df(x_n)(x - x_n) + \frac{1}{2}(x - x_n)' H_n (x - x_n)$$

or whether x_{n+1} is a better approximation of x^* than x_n.

Quasi-Newton method I

Want. find x s.t. $f'(x) = 0$.



- "Full" Newton:
$y = f'(x_n) + f''(x_n)(x - x_n) \qquad \Rightarrow x_{n+1} = x_n - [f''(x_n)]^{-1} f'(x_n)$

- Quasi-Newton (secant method):
$y = f'(x_n) + \frac{f'(x_n) - f'(x_{n-1})}{x_n - x_{n-1}}(x - x_n) \Rightarrow x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f'(x_n) - f'(x_{n-1})} f'(x_n)$

Quasi-Newton method II

Scalar case. $x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f'(x_n) - f'(x_{n-1})} f'(x_n)$.

- Note that $f''(x_n) \approx \frac{f'(x_n) - f'(x_{n-1})}{x_n - x_{n-1}} \equiv B_n$.
- We are approximating $[f''(x_n)]^{-1}$ by using x_n, x_{n-1}, $f'(x_n)$, and $f'(x_{n-1})$
- w/o directly computing $f''(x_n)$ and w/o inverting it.

Remarks.

- Multivariate cases are more complicated.
- Fast (no 2nd order numerical diff. and matrix inversion).
- The approximated H_n^{-1} can be restricted to be p.d. (e.g., BFGS algorithm). Thus, $g_n(x_{n+1}) < g_n(x_n)$.
- The approximated hessian may not be very accurate. That is, $\lim_n B_n$ may not be close to the true hessian $H = D^2 f(x^*)$ even if $x_n \to x^*$.
- MATLAB: If you need an estimate of the true hessian, use fminunc.
  - fminunc returns $D^2 f(\lim_n x_n)$, numerical hessian around x^*.
  - fmincon returns $\lim_n B_n$.