GMM, SMM, and Indirect Inference

Constantino Hevia
UTDT

Introduction

- Estimating DSGE models is difficult:
  - Rational expectations: mathematical expectations of future state and control variables.
  - Nonlinear policy functions.
- To do maximum likelihood, we need to:
  - Find the equilibrium allocation and policy functions.
  - Impose distributional assumptions about shocks.
  - Evaluate the likelihood function.
- Difficult issues...

Introduction

- We showed how to linearize a model and estimate its parameters using full information maximum likelihood and the Kalman filter.

- But linearization could be problematic:
  - We could be dropping important information (e.g. certainty equivalence).
  - With non-linear methods cannot use the Kalman filter to evaluate the likelihood function (e.g. particle filter).

- You may trust some aspects of the model more than others. May want to use only some information contained in the model $\Rightarrow$ **Limited information estimation procedures.**

- **GMM** is one such limited information estimation method that does not *necessarily* requires linearizing the model.

- **Often, GMM allows us to estimate structural parameters without the need to actually solve the model (i.e. without finding the policy functions)**.

Introduction

- **Generalized method of moments (GMM)**: *minimum distance estimator* that can be used to estimate parameters using only moment conditions.

- Need to write the relevant equations as:

$$E(\text{something}) = 0.$$

- If we cannot evaluate the expectation because, for example, there are unobserved variables, we may use simulation:

  - **Simulated method of moments (SMM)**: approximates the expectation by simulating the model and using a Law of Large Numbers.

  - **Indirect inference**, **matching impulse responses**, etc.: generalizations of GMM.

The sample average of a vector process

- Key idea is estimating population means by sample means.
- We have a sample of size T of an n-dimensional stationary vector process $\{x_t\}$

$$
\begin{aligned}
E\left[x_t\right] &= \mu \\
E\left[\left(x_t - \mu\right)\left(x_{t-j} - \mu\right)'\right] &= \Gamma_j .
\end{aligned}
$$

- Assume that autocovariances are absolutely summable:

$$\sum_{j=-\infty}^{\infty} \left|\Gamma_j\right| < \infty .$$

The sample average of a vector process

- Consider the sample mean

$$\bar{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t.$$

  Clearly, $E\left[\bar{x}_T\right] = \mu$.

- The covariance matrix of $\bar{x}_T$ satisfies

$$T\ E\left[\left(\bar{x}_T - \mu\right)\left(\bar{x}_T - \mu\right)'\right] = \sum_{j=-(T-1)}^{T-1}\left(1 - \frac{|j|}{T}\right)\Gamma_j.$$

  Key result used in GMM asymptotics (see Hamilton, p. 279 or my first set of notes)

- The asymptotic covariance matrix of the sample mean $\bar{x}_T$ satisfies

$$\lim_{T\to\infty} T\ E\left[\left(\bar{x}_T - \mu\right)\left(\bar{x}_T - \mu\right)'\right] = \sum_{j=-\infty}^{\infty}\Gamma_j.$$

Generalized method of moments estimator (GMM)

- x_t : $n \times 1$ vector of variables observed at time t.

- θ : $k \times 1$ vector of parameters, where $\theta \in \Theta \subseteq R^k$.
    - θ_0 is the true parameter that we want to estimate.

- $g(x_t, \theta)$: $m \times 1$ vector valued function.

- When evaluated at the true value θ_0, $g(x_t, \theta_0)$ satisfies the orthogonality condition

$$E\left[g(x_t, \theta_0)\right] = 0_{m\times 1}. \tag{1}$$

- GMM: choose $\hat{\theta}$ so that the sample analog of equation (1) is as close to zero as possible.

- The sample analog of equation (1) is

$$g_T(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T} g(x_t, \theta)$$

  where $g_T : \Theta \Rightarrow R^m$

Generalized method of moments estimator (GMM)

Suppose that there are k parameters and m moment conditions. Then

- If $k > m$ **the model is not identified**. More parameters than moment conditions. There could be multiple values of $\hat{\theta}$ that yield $g_T(\hat{\theta}) = 0$.

- If $k = m$ **the model is just-identified**. $g_T(\hat{\theta}) = 0$ is a system of m equations with m unknowns. In principle, we may find a unique $\hat{\theta}$ such that $g_T(\hat{\theta}) = 0$.

- If $k < m$ **the model is overidentified**. More moment conditions than parameters.

  - Choose $\hat{\theta}$ to make $g_T(\hat{\theta})$ as close to zero as possible using some criterion.

  - Overidentified case enables formal tests of the null hypothesis that the orthogonality conditions (1) are satisfied.

Generalized method of moments estimator (GMM)

Minimize the distance to zero of the sample moments using the criterion

$$\hat{\theta} = \arg\min_{\theta \in \Theta} g_T(\theta)' W_T g_T(\theta)$$

- W_T is an $m \times m$ positive-definite weighting matrix that determines the relative importance of each component of $g_T(\theta)$.
- One first order condition for each parameter: system of k equations with k unknowns.
- Hansen (1982): the optimal weighting matrix satisties $W_T \xrightarrow{p} S^{-1}$, where S is the asymptotic variance of $\sqrt{T} g_T(\theta_0)$ evaluated at the true parameter θ_0:

$$S = \lim_{T \to \infty} T\, E\left[\left(\frac{1}{T} \sum_{t=1}^{T} g(x_t, \theta_0) \right) \left(\frac{1}{T} \sum_{t=1}^{T} g(x_{t-j}, \theta_0) \right)' \right] = \sum_{j=-\infty}^{+\infty} \Gamma_j$$

where $\Gamma_j = E\left[g(x_t, \theta_0) g(x_{t-j}, \theta_0)' \right]$ is the j^{th} autocovariance of $g(x_t, \theta_0)$.

Generalized method of moments estimator (GMM)

- Let $\hat{\theta}$ be a consistent estimator of θ_0.

- Newey and West (1987): a consistent estimator of S is given by

$$\hat{S} = \sum_{j=-L}^{L} \left(1 - \frac{|j|}{L+1}\right) \hat{\Gamma}_j = \hat{\Gamma}_0 + \sum_{j=1}^{L} \left(1 - \frac{|j|}{L+1}\right) \left[\hat{\Gamma}_j + \hat{\Gamma}_j'\right]$$

  where $\hat{\Gamma}_j = \frac{1}{T} \sum_{t=j+1}^{T} g(x_t, \hat{\theta}) g(x_{t-j}, \hat{\theta})'$ is the sample autocovariance matrix.

- Criteria to choose the lag-length L (Andrews, 1991; Newey and West, 1994). One that is sometimes used is $L = 0.75T^{1/3}$ [integer part].

- The continuous mapping theorem implies $\hat{S}^{-1} \xrightarrow{p} S^{-1}$.

- Circular argument?
  - To estimate θ_0 using optimal GMM we need an estimate of $\hat{S}^{-1}$ which, in turn, requires an estimate of θ_0.

Generalized method of moments estimator (GMM)

Two ways to proceed:

1. Iterative procedure:
   1.1 Choose an arbitrary initial (positive definite) weighting matrix W_T, usually the identity matrix. Perform inefficient GMM to obtain a consistent initial estimate $\hat{\theta}_1$.

   1.2 Estimate of the covariance matrix $\hat{S}(\hat{\theta}_1)$ and use it to perform a new GMM estimation to obtain $\hat{\theta}_2$. The estimate $\hat{\theta}_2$ is asymptotically efficient.

   1.3 Iterating between (a) and (b) until convergence may have better small sample properties.

2. Continuously updating GMM (simultaneous estimation):

$$\min_{\theta} g_T(\theta)' \hat{S}^{-1}(\theta) g_T(\theta).$$

Asymptotic distribution of GMM: general formulas

- **General GMM estimate**: choose $\hat{\theta}$ so that a linear combination of sample means is zero

$$\underbrace{a_T(\hat{\theta})}_{k\times m}\underbrace{g_T(\hat{\theta})}_{m\times 1} = 0.$$

  $a_T(\hat{\theta})$ is a $k \times m$ matrix of "weights".

- The minimization discussed above is a special case: For a weighting matrix W_T,

$$a_T(\hat{\theta}) = \underbrace{\frac{\partial g_T'(\hat{\theta})}{\partial \theta}}_{k\times m}\underbrace{W_T}_{m\times m} \xrightarrow{p} \underbrace{a}_{k\times m}$$

- Let the Jacobian of the moment conditions be

$$d_T(\hat{\theta}) = \frac{\partial g_T(\hat{\theta})}{\partial \theta'} \xrightarrow{p} E\left[\frac{\partial g(x_t,\theta)}{\partial \theta'}\right] = \underbrace{d}_{m\times k}$$

  where the entry (i,j) is the derivative of the ith component of $g_T(\theta)$ with respect to the jth parameter: d is an $m \times k$ matrix.

Asymptotic distribution of GMM

- $g_T(\theta_0)$ is a sample mean: under regularity conditions a Central Limit Theorem applies:

$$\sqrt{T} g_T(\theta_0) \xrightarrow{d} N(0, S)$$

- The mean-value theorem and the continuous mapping theorem implies

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, (ad)^{-1} aSa' (ad)^{-1\prime}\right)$$

- In practical terms, we use

$$\mathrm{var}(\hat{\theta}) = \frac{1}{T}(ad)^{-1} aSa' (ad)^{-1\prime}$$

Details

The mean value theorem implies

$$g_T(\hat{\theta}) = g_T(\theta_0) + d_T(\tilde{\theta})(\hat{\theta} - \theta_0)$$

where $\tilde{\theta} = \lambda\hat{\theta} + (1-\lambda)\theta_0$ for $\lambda \in [0, 1]$. Multiplying both sides by $a_T(\hat{\theta})$ gives

$$a_T(\hat{\theta})g_T(\hat{\theta}) = a_T(\hat{\theta})g_T(\theta_0) + a_T(\hat{\theta})d_T(\tilde{\theta})(\hat{\theta} - \theta_0).$$

But the GMM estimate solves $a_T(\hat{\theta})g_T(\hat{\theta}) = 0$, hence

$$0 = a_T(\hat{\theta})g_T(\theta_0) + a_T(\hat{\theta})d_T(\tilde{\theta})(\hat{\theta} - \theta_0).$$

Multiplying by $\sqrt{T}$ and rearranging gives

$$a_T(\hat{\theta})d_T(\tilde{\theta})\sqrt{T}(\hat{\theta} - \theta_0) = -a_T(\hat{\theta})\sqrt{T}g_T(\theta_0) \text{ or .}$$

$$\sqrt{T}(\hat{\theta} - \theta_0) = -\left(a_T(\hat{\theta})d_T(\tilde{\theta})\right)^{-1} a_T(\hat{\theta})\sqrt{T}g_T(\theta_0)$$

Now, we know that

$$\sqrt{T}g_T(\theta_0) \xrightarrow{d} N(0, S)$$

$$a_T(\hat{\theta}) \xrightarrow{p} a$$

$$d_T(\tilde{\theta}) \xrightarrow{p} d.$$

Therefore, using the Central Limit Theorem and the Continuous Mapping Theorem it follows that

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, (ad)^{-1} aSa'(ad)^{-1\prime}\right).$$

Asymptotic distribution of GMM: Efficient GMM

- Hansen (1982) showed that $W = S^{-1}$ is optimal.
- In this case, $a = \frac{\partial g_T'}{\partial \theta} S^{-1}$ and

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, \left(d' S^{-1} d\right)^{-1}\right)$$

  or

$$\text{var}(\hat{\theta}) = \frac{1}{T}\left(d' S^{-1} d\right)^{-1}.$$

- In practice, one inserts $\hat{\theta}$ for θ_0 in the expressions above to form estimates $\hat{d}$ and $\hat{S}$.
- This gives the asymptotic standard errors of the estimated parameters.
- This also allows the construction of various *tests*.

Test of overidentifying restrictions

- When $k < m$, the model is overidentified: more orthogonality conditions than parameters and we can test the model.

- Hansen's test of the null hypothesis that the orthogonality conditions (1) are satisfied.
  - This is a **test of misspecification**. If the model is correctly specified, $g_T(\theta)$ would be different from zero only due to sampling uncertainty.

- **Under the null** that the model is correctly specified, $\sqrt{T} g_T(\theta_0) \sim N(0, S)$.

- Replacing (θ_0, S) by their estimated values $(\hat{\theta}, \hat{S})$, **under the null**, the test statistic

$$TJ_T = \left[\sqrt{T} g_T(\hat{\theta})\right]' \hat{S}^{-1} \left[\sqrt{T} g_T(\hat{\theta})\right] = T g_T(\hat{\theta})' \hat{S}^{-1} g_T(\hat{\theta}) \sim \chi^2(m-k).$$

- Degrees of freedom: number of moment conditions minus number of parameters.

Test of overidentifying restrictions and a t-test

- J_T is a measure of "how far the sample is from satisfying the overidentifying restrictions". If the J_T statistic is too large, the model is misspecified.

- **Remark**: this is a joint test of all the restrictions. If we reject the null, the test doesn't tell us which moment condition is wrong.

A standard t-test: Suppose that you want to test $\theta_j = \bar{\theta}_j$ for some j. You may compute a t statistic

$$t_j = \frac{\hat{\theta}_j - \bar{\theta}_j}{\hat{\sigma}_{\theta_j}}$$

where $\hat{\sigma}_{\theta_j}$ is the standard error of $\hat{\theta}_j$–the square root of the j^{th} diagonal element of $\frac{1}{T}\left(\hat{d}'\hat{S}^{-1}\hat{d}\right)^{-1}$. Under the null, t_j has a t distribution with $T-1$ degrees of freedom.

Example 1: Model of consumption and saving

$$\max_{c_t, s_{t+1}} E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma}}{1-\gamma} \text{ subject to}$$

$$c_t + a_{t+1} = w_t + (1 + r_{t+1}) a_t$$

where w_t is labor income, a_{t+1} is assets, and r_t is a stochastic interest rate.

- Euler equation

$$c_t^{-\gamma} = E_t \left[\beta c_{t+1}^{-\gamma} (1 + r_{t+1}) \right].$$

- The Euler equation is a (conditional) orthogonality condition of the form:

$$E_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + r_{t+1}) - 1 \right] = 0. \tag{2}$$

- Two parameters to estimate, (β, γ), and one *conditional* orthogonality condition: **At first sight, model is underidentified.**

Example 1: Model of consumption and saving

- How can we estimate this model with GMM?

- FOC is a conditional moment restriction. The law of iterated expectations implies

$$E\left[E_t\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}(1+r_{t+1})-1\right]\right]=E\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}(1+r_{t+1})-1\right]=0.$$

- Standard approach to "create" more orthogonality conditions in rational expectations models.
  - Let $\mathcal{F}_t$ denote the "information set" of the investor at time t: this typically includes previous observations of economic variables like lagged consumption, returns, etc.

  - Then, any variable z_t in $\mathcal{F}_t$, is a valid instrument in the sense that

$$E\left[E_t\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}(1+r_{t+1})-1\right]z_t\right]=E\left[\left(\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}(1+r_{t+1})-1\right)z_t\right]=0.$$

Example 2: Model with Two Assets

Agents choose consumption and two assets: a risky asset which pays a return r^e_{t+1} and a riskless asset with a known return at time t of r^f_{t+1}.

- The Euler equations of the consumer's problem can be written as

$$\begin{aligned} c_t^{-\gamma} =& E_t\left[\beta c_{t+1}^{-\gamma}(1+r^e_{t+1})\right] \\ c_t^{-\gamma} =& E_t\left[\beta c_{t+1}^{-\gamma}(1+r^f_{t+1})\right] \end{aligned}$$

  This gives two (conditional) orthogonality conditions which can be written as

$$\begin{aligned} E_t\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}\left(r^e_{t+1}-r^f_{t+1}\right)\right] =& 0 \\ E_t\left[\beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}\left(1+r^f_{t+1}\right)-1\right] =& 0. \end{aligned}$$

- We have two parameters (β, γ) and two orthogonality conditions: **just identified case**.
- But if we assume, say, $\beta = 0.95$, we have one parameter and two orthogonality conditions: **overidentified case**.

Example 3: GMM estimation of DSGE models

Consider the RBC model

$$\max E_0 \sum_{t=0}^{\infty} \beta^t U(c_t, n_t) \text{ subject to}$$

$$c_t + k_{t+1} = A_t F(k_t, n_t) + (1-\delta)k_t$$

- The Euler equation for capital accumulation is

$$E_t \left[\beta \frac{U_{c,t+1}}{U_{c,t}} \left(A_{t+1} F_{k,t+1} + 1 - \delta \right) - 1 \right] = 0.$$

- One conditional orthogonality condition and *at least* two parameters to estimate: β and δ (probably more from the utility and production functions).

- How do we add orthogonality conditions to estimate the model?

Example 3: GMM estimation of DSGE models

- We can use the law of iterated expectations in the Euler equation and add variables that belong in the agent's information set to create orthogonality conditions.

- For example

$$E\left[E_t\left[\beta\frac{U_{c,t+1}}{U_{c,t}}\left(A_{t+1}F_{k,t+1}+1-\delta\right)-1\right]\frac{c_t}{c_{t-1}}\right]=0$$
$$E\left[E_t\left[\beta\frac{U_{c,t+1}}{U_{c,t}}\left(A_{t+1}F_{k,t+1}+1-\delta\right)-1\right]\frac{c_{t-1}}{c_{t-2}}\right]=0$$
$$E\left[E_t\left[\beta\frac{U_{c,t+1}}{U_{c,t}}\left(A_{t+1}F_{k,t+1}+1-\delta\right)-1\right]A_tF_{k,t}\right]=0$$
$$E\left[E_t\left[\beta\frac{U_{c,t+1}}{U_{c,t}}\left(A_{t+1}F_{k,t+1}+1-\delta\right)-1\right]A_{t-1}F_{k,t-1}\right]=0$$

etc.

Example 3: GMM estimation of DSGE models

- GMM can also be applied to static optimality conditions, for example, MRS=MPL

$$\frac{U_{Nt}}{U_{ct}} = A_t F_{N,t}$$

- Since this holds for any t, it trivially holds on average, so

$$E\left[\frac{U_{Nt}}{U_{ct}} - A_t F_{N,t}\right] = 0.$$

- GMM can also be applied to identities. For example, the feasibility constraint:

$$E\left[c_t + k_{t+1} - (1-\delta)k_t - F(k_t, n_t)\right] = 0.$$

Conclusion: Dynamic rational expectation models produce orthogonality conditions as the result of optimization and equilibrium conditions. Parameters entering these conditions can be estimated by GMM.

Example 3: GMM estimation of DSGE models

- It is also possible to match **unconditional moments** derived from the model with those based on actual data:

- We observe x_t and construct a function of the data $\mathbf{m}_t = h(x_t)$, where $h(\bullet)$ is a $p \times 1$ vector of empirical observations on variables whose moments are of interest. (to construct variances, correlations, etc).
  - Let $\mathbf{m}_t^M(\theta) = h^M(y_t, \theta)$ be the model-based counterpart of $\mathbf{m}_t$ whose elements are computed based on the solution of the log-linearized model. In that case, there may be analytical expressions for the unconditional moments $E(\mathbf{m}_t^M(\theta))$.

  - GMM can be applied using the orthogonality condition

$$g_T(\theta) = \left[\frac{1}{T} \sum_{t=1}^{T} \mathbf{m}_t - E(\mathbf{m}_t^M(\theta)) \right] = 0.$$

- Stochastic singularity can also be an issue with GMM. E.g. covariances of different variables could be linearly dependent.

Example: GMM and asset pricing

- The model is

$$E_t\left[\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} R^e_{t+1}\right] = 0$$

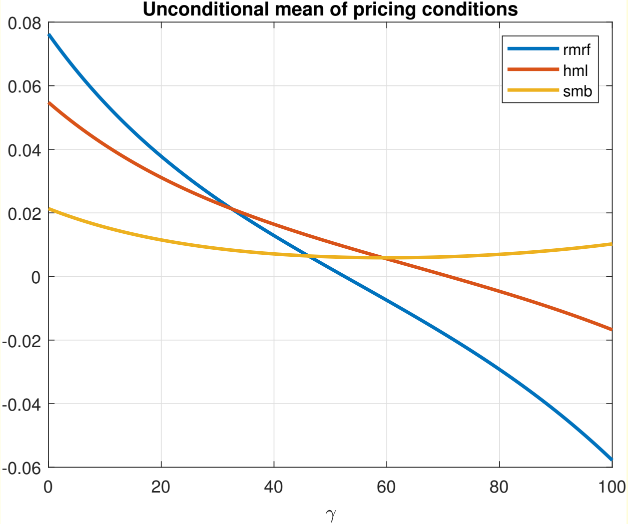
  where R^e_{t+1} is an excess return and $\gamma > 0$ is the coefficient of risk aversion.

- Excess returns are
  - *rmrf*: market portfolio
  - *hml*: high minus low (value - growth portfolio)
  - *smb*: small minus big portfolio

- Objective is to estimate the coefficient of risk aversion γ using GMM.

Example: GMM and asset pricing

- Sample moments using the rmrf, hml, and smb excess returns as a function of γ



Example: GMM and asset pricing

- GMM using only one excess return (unconditional)

- Model exactly identified: cannot do test of overidentifying restrictions

Return/parameter	$\hat{\gamma}$	Std Err
RMRF	52.52	26.87
HML	71.03	31.42
SMB	59.95	3.9E+7

Example: GMM and asset pricing

- GMM using two and three excess returns (unconditional).
- Model overidentified: 1 parameter, 2 or 3 moments.

Return/parameter	$\hat{\gamma}$	Std Err	J-stat	p-val
RMRF and HML ($W = I$)	56.33	26.5	-	-
RMRF and HML (2-stage)	63.57	24.6	0.67	0.41
HML and SMB ($W = I$)	52.77	26.8		
HML and SMB (2-stage)	61.16	24.8	0.82	0.36
RMRF, HML, and SMB ($W = I$)	56.43	26.5		
RMRF, HML, and SMB (2-stage)	70.72	22.5	1.66	0.43

Example: estimating fiscal and monetary rules

$$d_t = a_0 + a_1 d_{t-1} + a_2 DD_t gap_t + a_3(1 - DD_t) gap_t + a_4 debt_t + e_t$$
$$i_t = b_0 + b_1 \pi_t + b_2 gap_t + b_3 i_{t-1} + u_t$$

- DD_t: dummy variable equal to one if output gap is positive (expansion) and zero if gap is negative (contraction). d_t: deficit (% of GDP). i_t: nominal interest rate

- Treat e_t and u_t as expectational errors: $E(e_t|\mathcal{F}_t) = E(u_t|\mathcal{F}_t) = 0$.

Issues of interest:

- Is monetary policy active? $\Rightarrow$ $b_1 > 1$.

- Is fiscal policy active? $\Rightarrow$ $a_4 < 0$.

- Does fiscal policy react to output symmetrically over the cycle? $\Rightarrow$ $a_2 = a_3$.

Use US data over 1967:Q1-2004:Q2. Output gap is actual output minus potential output as reported by FRED-II dataset: series GDPPOT.

Example: estimating fiscal and monetary rules

- Orthogonality conditions (using unconditional moments)

$$E\left[d_t - a_0 - a_1 d_{t-1} - a_2 DD_t gap_t - a_3(1 - DD_t) gap_t - a_4 debt_t\right] = 0$$
$$E\left[i_t - b_0 - b_1 \pi_t - b_2 gap_t - b_3 i_{t-1}\right] = 0$$

- Sample counterparts

$$\frac{1}{T}\sum_{t=1}^{T}\left[d_t - a_0 - a_1 d_{t-1} - a_2 DD_t gap_t - a_3(1 - DD_t) gap_t - a_4 debt_t\right] = 0$$
$$\frac{1}{T}\sum_{t=1}^{T}\left[i_t - b_0 - b_1 \pi_t - b_2 gap_t - b_3 i_{t-1}\right] = 0$$

- Parameters to estimate: $\{a_0, a_1, a_2, a_3, a_4, b_0, b_1, b_2, b_3\}$

- 2 equations, 9 unknowns. Need to use instrumental variables (GIV)

Example: estimating fiscal and monetary rules

- **Case 1. Instruments are:**
  - For fiscal equation: a constant, 1 lag of deficit, inflation, debt, and the interacted variables $DD_t gap_t$ and $(1 - DD_t) gap_t$.
  - For monetary equation: a constant, one lag of output gap, inflation, and the interest rate.

Coef.	Value	Std. Err	t-stat
a_0	0.550	0.224	2.447
a_1	0.855	0.042	20.344
a_2	-0.080	0.036	-2.208
a_3	-0.175	0.076	-2.285
a_4	-0.0047	0.0034	-1.376
b_0	-0.093	0.283	-0.329
b_1	1.364	0.537	2.540
b_2	0.015	0.069	0.230
b_3	0.782	0.086	9.015

Test	J-test	equality test ($a_2 = a_3$)
statistic	1.112	1.206
p-value	0.292	0.272

Equality test is test of symmetric reaction of deficit to the cycle.

Example: estimating fiscal and monetary rules

- **Case 2. Instruments are:**
  - For fiscal equation: same as Case 1.
  - For monetary equation: two lags of each instrument.

Coef.	Value	Std. Err	t-stat
a_0	0.494	0.207	2.386
a_1	0.865	0.0397	21.778
a_2	-0.086	0.0346	-2.496
a_3	-0.157	0.0741	-2.126
a_4	-0.0044	0.0032	-1.372
b_0	-0.141	0.187	-0.755
b_1	0.740	0.227	3.249
b_2	0.0674	0.0418	1.613
b_3	0.9005	0.0248	36.214

Test	J-test	equality test ($a_2 = a_3$)
statistic	4.031	0.710
p-value	0.401	0.399

Example: estimating fiscal and monetary rules

- **Case 3.** As Case 1 but starting sample in 1983.

- Estimates unstable. It likely comes from the interest rate equation.

Coef.	Value	Std. Err	t-stat
a_0	1.291	0.434	2.970
a_1	0.794	0.0564	14.093
a_2	-0.0424	0.0424	-0.999
a_3	-0.509	0.146	-3.479
a_4	-0.012	0.0057	-2.111
b_0	2.426	2.031	1.194
b_1	-4.907	4.390	-1.117
b_2	0.0267	0.124	0.215
b_3	1.214	0.257	4.714

Test	J-test	equality test ($a_2 = a_3$)
statistic	0.304	9.248
p-value	0.581	0.0024

Problems with GMM

- How do we choose the instruments z_t? How many? How many lags?
- Choice of orthogonality conditions? Which one do we select? Sometimes if we rewrite the same condition in a different form, we obtain different estimates.
- Too many instruments: while OK asymptotically, causes problems in small samples.
- **What if there are unobservable variables in the orthogonality conditions?**

Simulated Method of Moments (SMM)

- Often we cannot compute the orthogonality conditions of the model analytically.
  - There may be unobservable variables in the orthogonality conditions, like preference shocks.

- One solution is using simulation methods.
  - Replace analytical expectations by Monte Carlo expectations simulating the model.

- This applies to moment conditions of the form

  $$E_t\left[g(x_t, \theta, \nu_t)\right] = 0,$$

  where ν_t is unobservable.
  - In this case g_T cannot be evaluated and GMM cannot be applied.

Example: model with unobserved preference shocks

- Social planner maximizes

$$\max_{c_t, n_t, k_{t+1}} E_0 \left[\sum_{t=0}^{\infty} \beta^t \nu_t U(c_t, n_t) \right] \text{ subject to}$$

$$c_t + k_{t+1} = A_t F(k_t, n_t) + (1 - \delta) k_t.$$

- Euler equation

$$E_t \left[\beta \frac{U_{c,t+1}}{U_{c,t}} \frac{\nu_{t+1}}{\nu_t} \left(A_{t+1} F_{k,t+1} + 1 - \delta \right) - 1 \right] = 0.$$

- Here,

$$g(x_t, \theta, \nu_t) = \beta \frac{U_{c,t+1}}{U_{c,t}} \frac{\nu_{t+1}}{\nu_t} \left(A_{t+1} F_{k,t+1} + 1 - \delta \right) - 1.$$

- Problem: Can't construct g_T because ν_t is unobservable.

Models with unobservable variables

- This problem could also be present in standard RBC models:
  - Technology shocks may be unobservable (may use Solow residuals as a proxy)
  - Stock of capital may be unobservable (may construct capital using the perpetual inventory method).
- General result: if there are unobserved shocks (such as ν_t) or unobserved variables (such as the stock of capital) that enter the orthogonality conditions, GMM cannot be used to estimate the parameters.
- What to do then?
  - Simulated method of moments.

Models with unobservable variables

- Orthogonality condition is $E_t\left[g_t(x_t,\theta,\nu_t)\right]=0$.

- Shocks ν_t are unobservable, but with a known distribution.

- Draw shocks $\left\{v_t^l\right\}$ for $l=1,2,...,L$, where L is "large".

- Construct $g_t^l(x_t,\theta,\nu_t^l)$ for each draw l.

- Under regularity conditions, a Law of Large number applies and

$$\frac{1}{L}\sum_{l=1}^{L} g_t^l \xrightarrow{p} E\left[g_t(x_t,\theta,\nu_t)\right].$$

- **If there are unobserved variables but with a known distribution, simulate the model, construct g_t using simulated data, and apply GMM to the "simulated" orthogonality conditions.**

Simulated method of moments

- We observe x_t and construct moments $\boldsymbol{m}_t = h(x_t)$, where $h(\bullet)$ is a $p \times 1$ vector of empirical observations on variables whose moments are of interest. (e.g. variances, correlations, etc).

- Let $\boldsymbol{m}_i(\theta)$ be the synthetic counterpart of $\boldsymbol{m}_t$ whose elements are computed based on simulated data generated by a model given some parameter values $\theta \in R^k$ and shocks v_t. Let the number of observations in artificial time series be $i = 1, 2, ..., \tau T$.

- The SMM estimator, $\tilde{\theta}_{SMM}$ is the value of θ that solves

$$\tilde{\theta}_{SMM} = \arg\min_{\theta} \left[\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{m}_t - \frac{1}{\tau T} \sum_{i=1}^{\tau T} \boldsymbol{m}_i(\theta) \right]' S^{-1} \left[\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{m}_t - \frac{1}{\tau T} \sum_{i=1}^{\tau T} \boldsymbol{m}_i(\theta) \right],$$

  - S determines the optimal weighting matrix and is given by

$$S = \lim_{T \to \infty} Var \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{m}_t \right).$$

Simulated method of moments

- Under regularity conditions (Duffie and Singleton, 1993), if θ_0 is the true value,

$$\sqrt{T}(\tilde{\theta}_{SMM} - \theta_0) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{\tau}\right)\left(dS^{-1}d'\right)^{-1}\right)$$

  where $d = E\left[\partial m_i(\theta)/\partial\theta\right]$ is a $k \times p$ matrix assumed to be finite and of full rank.

- A specification test of overidentifying restrictions when $p > k$ is given by

$$T(1 + \frac{1}{\tau})\left[G(\tilde{\theta}_{SMM})'\tilde{S}^{-1}G(\tilde{\theta}_{SMM})\right] \xrightarrow{d} \chi^2(p-k).$$

  where

$$G(\theta) = \left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{m}_t - \frac{1}{\tau T}\sum_{i=1}^{\tau T}\mathbf{m}_i(\theta)\right].$$

- Implementation detail: use the same sequence of shocks $\{\nu_t\}$ during the minimization iterations. Otherwise, we don't know if the objective function changes because parameters change or because shocks change.

Indirect Inference (or Extended Method of Simulated Moments)

- **SMM** constructs an estimate of θ by minimizing the distance between the unconditional moments of the data and those of an artificial series simulated using parameter values.

- **Indirect Inference** is more general: it constructs an estimate of θ by minimizing the distance between some continuous function of the data and the equivalent function estimated from an artificial time series simulated for some parameter values:
  - The functions need not be moments.

- Proposed by Smith (1993). Can be interpreted as a generalization of the simulated method of moments.

- Use of an auxiliary model to capture aspects of data upon which to base estimation.

- Auxiliary model is a continuous function of the data.

- **Indirect Inference:** choose parameters of the model so that the actual and simulated data look the same through the lens of the auxiliary model.

Indirect Inference

Examples:

- **Matching VAR Coefficients** (Smith, 1993): Run a VAR using actual data and run the same VAR in the model. Indirect Inference chooses θ to minimize the distance between the estimated VAR coefficients using actual data and those using simulated data.
  - In this case the auxiliary model is a VAR.

- **Matching Impulse Responses** (Christiano, Eichenbaum, and Evans, 2005): estimate impulse responses in the data (using appropriate identifying assumptions) and in the model. Choose θ to minimize the distance between the IRs.
  - In this case, the auxiliary model are the impulse responses of a structural VAR.

Indirect Inference

- Let $\boldsymbol{\eta} \equiv \eta(\{x_t\})$ denote a $p \times 1$ vector with the parameters of the auxiliary model. For example, the estimates of the VAR representation of the data.

- Let $\boldsymbol{\eta}(\theta)$ denote the synthetic counterpart of $\boldsymbol{\eta}$. For example, with the estimates of a VAR representation of the artificial data generated by the model.

- Let T denote the sample size of the actual data and τT the sample size of the simulated data

- The indirect inference estimator, $\tilde{\theta}_{II}$, solves

$$\tilde{\theta}_{II} = \arg\min_{\theta} \left[\boldsymbol{\eta} - \boldsymbol{\eta}(\theta)\right]' V \left[\boldsymbol{\eta} - \boldsymbol{\eta}(\theta)\right]$$

  where V is the $p \times p$ optimal weighing matrix.

  - Smith suggests using the inverse of the variance covariance matrix of the estimate η as an estimator of V.

Indirect Inference

- Under regularity conditions (Smith, 1993),

$$\sqrt{T}(\tilde{\theta}_{II} - \theta_0) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{\tau}\right)(J'VJ)^{-1}\right)$$

  where $J = E\left[\partial \boldsymbol{\eta}(\theta)/\partial\theta\right]$ is a $k \times p$ matrix assumed to be finite and of full rank.

- A test of overidentifying restrictions can be constructed using

$$T(1 + \frac{1}{\tau})\left[\left(\boldsymbol{\eta} - \boldsymbol{\eta}(\tilde{\theta}_{II})\right)' V\left(\boldsymbol{\eta} - \boldsymbol{\eta}(\tilde{\theta}_{II})\right)\right] \xrightarrow{d} \chi^2(p-k).$$

Putting it all together: Ruge-Murcia (2007)

- Ruge-Murcia (2007) compares the estimation of a standard RBC model using four different methods:
    - Maximum likelihood, with and without measurement errors and incorporating priors.
    - Generalized Method of Moments.
    - Simulated Method of Moments.
    - Indirect Inference.

- The idea of the paper is to study the small sample properties of the different estimation techniques using a controlled Monte Carlo experiment.

- In the process, he discusses issues of identification, stochastic singularity, small sample properties, etc.

- Nice and pedagogical paper. Read it.

Ruge-Murcia: The artificial economy

- The artificial economy is a one-shock RBC model:

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \left(\ln c_t + \psi(1 - n_t)\right) \text{ subject to}$$

$$\begin{aligned} c_t + k_{t+1} =& y_t + (1-\delta)k_t; \;\; y_t = z_t k_t^{\alpha} n_t^{1-\alpha} \\ \ln z_{t+1} =& \rho \ln z_t + \varepsilon_{t+1} \end{aligned}$$

  where $\varepsilon_t \sim N(0, \sigma^2)$.

- The model is stochastically singular: a single productivity shock drives the evolution of several endogenous variables, like output y_t, consumption c_t, and employment, n_t.

- The model is solved using a log-linearization around the steady state.

- Artificial data is constructed using known parameter values. Experiment consists of estimating the parameters $\theta = (\beta, \rho, \sigma)$.

- Data used for estimation is output, consumption, and labor: y_t, c_t, n_t.

Ruge-Murcia: maximum likelihood

- Two approaches to deal with stochastic singularity (one shock and three observable variables)
    1. Estimate the model using just one observable at a time: y_t, n_t, or c_t.
    2. Add measurement errors (see paper).

- Incorporating priors centered around true values.

Ruge-Murcia: method of moments (GMM and SMM)

- Using second moments between pair of variables to avoid stochastic singularity

$$m_t = \left\{ \hat{y}_t^2, \hat{c}_t^2, \hat{y}_t \hat{c}_t, \hat{c}_t \hat{c}_{t-1}, \hat{y}_t \hat{y}_{t-1} \right\}$$
$$m_t = \left\{ \hat{y}_t^2, \hat{n}_t^2, \hat{y}_t \hat{n}_t, \hat{n}_t \hat{n}_{t-1}, \hat{y}_t \hat{y}_{t-1} \right\}$$
$$m_t = \left\{ \hat{n}_t^2, \hat{c}_t^2, \hat{c}_t \hat{n}_t, \hat{n}_t \hat{n}_{t-1}, \hat{c}_t \hat{c}_{t-1} \right\}$$

- Note: we are not using first order conditions or Euler equations, but second moments of the data to do GMM and SMM.

- In SMM, use different sample sizes for simulation $\tau = 5, 10, 20$.

Ruge-Murcia: indirect inference

- Auxiliary model is a bivariate VAR(1) (cannot have more variables due to stochastic singularity).

- Matching VAR coefficients using data with the equivalent coefficients estimated using simulated data.

- Use pair of variables: $(\hat{y}_t, \hat{c}_t)$; $(\hat{y}_t, \hat{n}_t)$; $(\hat{c}_t, \hat{n}_t)$.

- Use different sample sizes for simulation: $\tau = 5, 10, 20$.

Table 1
Maximum likelihood under the null hypothesis

Experiment no.	*Var.*	β			ρ			σ		
		Mean (median)	ASE (SD)	Size (SE)	Mean (median)	ASE (SD)	Size (SE)	Mean (median)	ASE (SD)	Size (SE)
A. Using as many variables as shocks										
1	$\hat{y}_t$	.7896	.2205	.0900	.8080	.0600	.0620	.0495	.0173	.2300
		.8202	.1721	.0128	.8206	.0583	.0108	.0505	.0104	.0188
2	$\hat{c}_t$	.9156	.0847	.1660	.8127	.1201	.0980	.0534	.0375	.1040
		.9489	.1359	.0166	.8585	.1370	.0133	.0376	.0371	.0137
3	$\hat{n}_t$	.9403	.0623	.1280	.8406	.0890	.0580	.0433	.0225	.1440
		.9425	.0477	.0149	.8575	.0609	.0105	.0441	.0157	.0157
B. Incorporating priors										
4	$\hat{y}_t$	.9501	.0250	.0000	.8434	.0334	.0280	.0398	.0037	.0000
		.9502	.0010	.0000	.8464	.0295	.0074	.0397	.0023	.0000
5	$\hat{c}_t$	.9494	.0230	.0020	.8482	.0522	.0060	.0397	.0088	.0000
		.9499	.0067	.0020	.8500	.0286	.0035	.0396	.0027	.0000
6	$\hat{n}_t$	.9498	.0225	.0000	.8450	.0440	.0020	.0397	.0082	.0000
		.9495	.0068	.0000	.8473	.0257	.0020	.0399	.0026	.0000

Notes: The true values are $\beta = .95$, $\rho = .85$, and $\sigma = .04$. Mean is the arithmetic average of the estimated

Method of moments under the null hypothesis

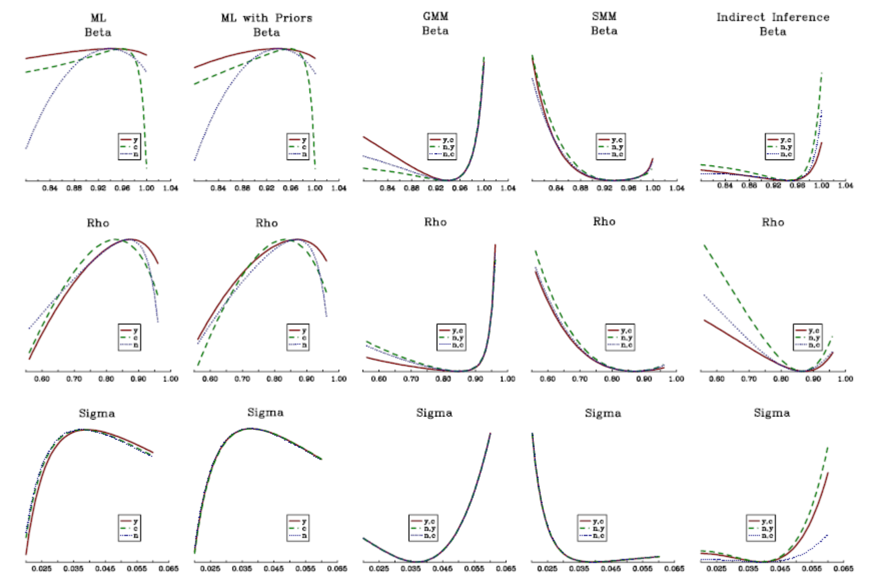
Experiment no.	Var.	β Mean (median)	β ASE (SD)	β Size (SE)	ρ Mean (median)	ρ ASE (SD)	ρ Size (SE)	σ Mean (median)	σ ASE (SD)	σ Size (SE)	OI (SE)
A. Simulated method of moments with $\tau = 5$											
1	$\hat{y}_t, \hat{c}_t$	.9503	.0128	.0300	.8344	.0410	.0920	.0396	.0041	.0280	.0060
		.9504	.0094	.0076	.8426	.0489	.0129	.0396	.0031	.0074	.0035
2	$\hat{n}_t, \hat{y}_t$	.9489	.0140	.0980	.8415	.0314	.1520	.0399	.0037	.0640	.0000
		.9497	.0159	.0133	.8455	.0440	.0161	.0398	.0038	.0109	.0000
3	$\hat{n}_t, \hat{c}_t$	.9506	.0107	.0420	.8403	.0297	.1320	.0395	.0032	.0940	.0020
		.9508	.0098	.0090	.8460	.0413	.0151	.0393	.0033	.0131	.0020
B. Simulated method of moments with $\tau = 10$											
4	$\hat{y}_t, \hat{c}_t$	.9504	.0124	.0280	.8342	.0400	.1100	.0393	.0040	.0340	.0160
		.9511	.0097	.0074	.8426	.0499	.0140	.0393	.0030	.0081	.0056
5	$\hat{n}_t, \hat{y}_t$	.9495	.0134	.0680	.8425	.0298	.1700	.0397	.0036	.0700	.0000
		.9507	.0140	.0113	.8477	.0423	.0168	.0395	.0035	.0114	.0000
6	$\hat{n}_t, \hat{c}_t$	.9508	.0102	.0400	.8410	.0281	.1580	.0393	.0030	.0960	.0000
		.9510	.0094	.0088	.8468	.0400	.0163	.0391	.0033	.0132	.0000
C. Simulated method of moments with $\tau = 20$											
7	$\hat{y}_t, \hat{c}_t$	.9497	.0122	.0160	.8351	.0388	.1240	.0395	.0039	.0300	.0080
		.9503	.0084	.0056	.8462	.0508	.0147	.0394	.0028	.0076	.0040
8	$\hat{n}_t, \hat{y}_t$	.9485	.0131	.0800	.8395	.0295	.1480	.0395	.0034	.0680	.0000
		.9497	.0145	.0121	.8450	.0407	.0159	.0394	.0034	.0113	.0000
9	$\hat{n}_t, \hat{c}_t$	.9506	.0100	.0600	.8432	.0272	.1280	.0395	.0030	.0940	.0020
		.9509	.0098	.0106	.8479	.0368	.0149	.0394	.0032	.0131	.0020
D. Generalized method of moments											
10	$\hat{y}_t, \hat{c}_t$	.9502	.0117	.0080	.8366	.0375	.0940	.0395	.0038	.0200	.0080
		.9505	.0082	.0040	.8423	.0446	.0131	.0394	.0027	.0063	.0040
11	$\hat{n}_t, \hat{y}_t$	.9485	.0127	.0700	.8383	.0288	.1520	.0394	.0033	.0520	.0000
		.9506	.0141	.0114	.8436	.0393	.0161	.0392	.0034	.0099	.0000
12	$\hat{n}_t,\hat{c}_t$	.9503	.0098	.0400	.8390	.0270	.1380	.0394	.0029	.0980	.0020
		.9506	.0093	.0088	.8420	.0362	.0154	.0395	.0030	.0133	.0020

Table 3
Indirect inference under the null hypothesis

Experiment no.	*Var*	β			ρ			σ			OI (SE)
		Mean (median)	ASE (SD)	Size (SE)	Mean (median)	ASE (SD)	Size (SE)	Mean (median)	ASE (SD)	Size (SE)	
A. $\tau = 5$											
1	$\hat{y}_t, \hat{c}_t$	.9514	.0249	.3360	.8182	.0795	.3580	.0399	.0004	.7360	.0800
		.9506	.0325	.0211	.8521	.1425	.0214	.0399	.0022	.0197	.0121
2	$\hat{n}_t, \hat{y}_t$	.9489	.0250	.2300	.8119	.0920	.2380	.0402	.0002	.8520	.1740
		.9502	.0298	.0188	.8508	.1466	.0190	.0401	.0023	.0159	.0170
3	$\hat{n}_t, \hat{c}_t$	.9492	.0267	.0020	.8428	.0982	.0020	.0398	.0006	.6240	.0020
		.9489	.0111	.0020	.8458	.0425	.0020	.0399	.0023	.0217	.0020
B. $\tau = 10$											
4	$\hat{y}_t, \hat{c}_t$	.9535	.0235	.3240	.8266	.0708	.3740	.0401	.0003	.7180	.0820
		.9513	.0329	.0209	.8548	.1333	.0216	.0401	.0020	.0201	.0123
5	$\hat{n}_t, \hat{y}_t$	.9463	.0251	.2100	.7995	.0955	.2120	.0401	.0002	.8260	.1760
		.9472	.0302	.0182	.8393	.1543	.0182	.0401	.0020	.0170	.0170
6	$\hat{n}_t, \hat{c}_t$	.9488	.0256	.0060	.8414	.0944	.0000	.0401	.0005	.6380	.0000
		.9487	.0111	.0035	.8452	.0423	.0000	.0401	.0021	.0215	.0000
C. $\tau = 20$											
7	$\hat{y}_t, \hat{c}_t$	.9566	.0246	.2920	.8377	.0691	.3700	.0399	.0003	.7460	.1040
		.9558	.0334	.0203	.8704	.1308	.0216	.0397	.0020	.0195	.0137
8	$\hat{n}_t, \hat{y}_t$	.9399	.0237	.2800	.7672	.0982	.2480	.0398	.0002	.8520	.2180
		.9416	.0307	.0201	.8167	.1726	.0193	.0398	.0019	.0159	.0185
9	$\hat{n}_t, \hat{c}_t$	.9480	.0248	.0040	.8391	.0928	.0020	.0400	.0005	.5940	.0000
		.9475	.0101	.0028	.8406	.0396	.0020	.0400	.0020	.0220	.0000

Notes: For Experiments 1, 4, and 7, the VAR consists of $\hat{y}_t$ and $\hat{c}_t$; for Experiments 2, 5, and 8, the VAR consists of $\hat{y}_t$ and $\hat{n}_t$; and for Experiments 3, 6, and 9, the VAR consists of $\hat{n}_t$, and $\hat{c}_t$. In all cases a VAR of order one is used. See the notes to Table 1.

Plot of objective functions: identification issues



Empirical distributions of estimated parameters

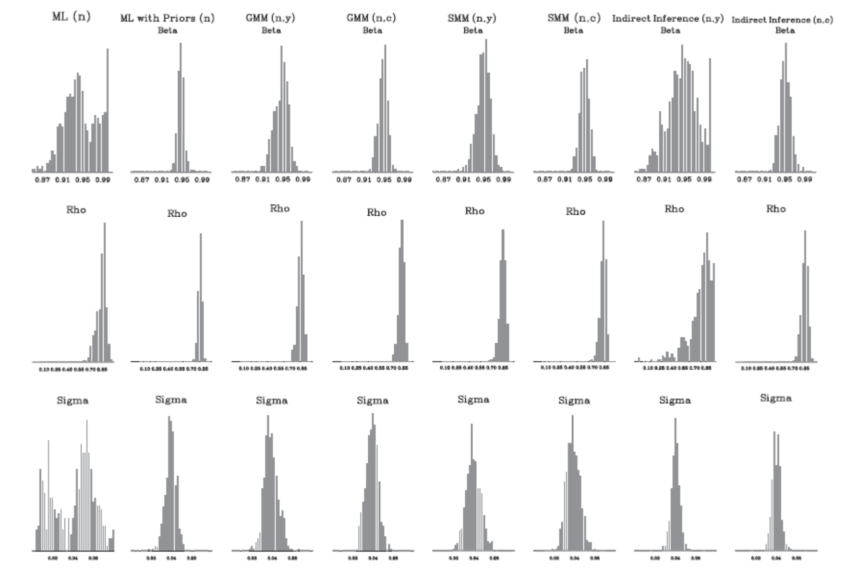


Fig. 2. Empirical distributions under the null.

References

1. Andrews, D. W. K. (1991), “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation”, Econometrica, 59, 817-858.
2. Canova, Fabio. (2007), “Methods for Applied Macroeconomic Research”, Chapter 5
3. Cochrane, John. (2005). Asset Pricing, 2nd Edition, Chapters 10 and 11.
4. Dejong, D. and C. Dave. (2011), “Structural Macroeconometrics”, Chapter 12.
5. Duffie, D. and K. J. Singleton. (1993), “Simulated Moments Estimation of Markov Models of Asset Prices”, Econometrica, 61, 929=952.
6. Hamilton, J. (1994). “Time Series Analysis”. Chapter 14.
7. Hansen, L. P. (1982). “Large Sample Properties of Generalized Method of Moments Estimators”, Econometrica, 50, 1029-1054.
8. Newey, W. K. and West, L. D. (1987), “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance matrix”, Econometrica, 55, 703-708.
9. Newey, W. K. and West, L. D. (1994), “Automatic Lag Selection in Covariance Matrix Estimation”, Review of Economic Studies, 61, 631-653.
10. Ruge-Murcia, Francisco J. (2007), “Methods to Estimate Dynamic Stochastic General Equilibrium Models”, Journal of Economic Dynamics and Control, 31, 2599-2636
11. Smith, A. (1993), “Estimating Nonlinear Time-Series Models using Simulated Vector Autoregressions”, Journal of Applied Econometrics 8, 63-84.