

# 15.472: GMM

Daniel Greenwald

Fall 2020

## Why Use GMM?

- ▶ Ideal estimation procedures require maximizing the likelihood function over the data  $\{x_t\}$ .
- ▶ Often, we don't know the likelihood, but can write down some **moment conditions**:

$$E[f(x_t; b)] = 0$$

that should be satisfied from theory (e.g., equilibrium conditions).

- ▶ Key application in finance:

$$E_t [M_{t+1} R_{t+1}] = 1.$$

for stochastic discount factor  $M_{t+1}$ .

- ▶ Unconditional version:

$$E \left[ \left( M_{t+1} R_{t+1} - 1 \right) z_t \right] = 0, \quad \forall z_t$$

# Why Use GMM?

- ▶ Using moment conditions, GMM theory provides:
  1. Efficient procedures for estimating parameters  $b$
  2. Standard errors and hypothesis tests
- ▶ Unified theory nesting huge set of standalone econometric procedures:

$$\text{OLS : } E \left[ \left( y_t - \beta' x_t \right) x_t \right] = 0$$

$$\text{IV : } E \left[ \left( y_t - \beta' x_t \right) z_t \right] = 0$$

$$\text{MLE : } E \left[ \frac{\partial L(x; b)}{\partial b} \right] = 0$$

- ▶ Especially useful for jointly estimating related equations.

# Implementation

- ▶ In practice, don't observe  $E[f(x_t; b)]$ , but can compute sample analogue:

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

- ▶ Basic idea is to choose  $b$  so that  $g_T$  is “close” to zero.
- ▶ Formal version:

$$\hat{b} = \arg \min_b \frac{1}{2} g_T(b)' W g_T(b)$$

for some symmetric weighting matrix  $W$

# Review: Matrix Calculus

- ▶ Derivative of  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  is  $k \times n$  matrix

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k(x)}{\partial x_1} & \cdots & \frac{\partial f_k(x)}{\partial x_n} \end{bmatrix}$$

- ▶ Product rule: if  $h(x) = f(x)'g(x)$  then

$$\frac{\partial h(x)}{\partial x} = f(x)' \frac{\partial g(x)}{\partial x} + g(x)' \frac{\partial f(x)}{\partial x}.$$

- ▶ For constant matrix  $A$ :

$$\frac{\partial (Af(x))}{\partial x} = A \frac{\partial f(x)}{\partial x}$$

$$\frac{\partial (f(x)'A)}{\partial x} = A' \frac{\partial f(x)}{\partial x}$$

# GMM Solution

- ▶ GMM optimality condition:

$$d_T(\hat{b})' W g_T(\hat{b}) = 0$$

for  $d_T(b) \equiv \nabla g_T(b)$ .

- ▶ Derivation: let  $Q(b)$  be the objective function, then

$$\frac{\partial Q(b)}{\partial b} = \frac{\partial}{\partial b} \left[ \frac{1}{2} (W' g_T(b))' g_T(b) \right] = \frac{1}{2} g_T(b)' W \frac{\partial g_T(b)}{\partial b} + \frac{1}{2} g_T(b)' W' \frac{\partial g_T(b)}{\partial b}.$$

To finish, recall that  $W$  is symmetric. Transpose to get  $n \times 1$  vector.

- ▶ Generalization: solve

$$a_T(\hat{b}) g_T(\hat{b}) = 0.$$

directly without specifying minimization problem.

# GMM Solution

- ▶ GMM optimality condition:

$$d_T(\hat{b})' W g_T(\hat{b}) = 0$$

for  $d_T(b) \equiv \nabla g_T(b)$ .

- ▶ Derivation: let  $Q(b)$  be the objective function, then

$$\frac{\partial Q(b)}{\partial b} = \frac{\partial}{\partial b} \left[ \frac{1}{2} (W' g_T(b))' g_T(b) \right] = \frac{1}{2} g_T(b)' W \frac{\partial g_T(b)}{\partial b} + \frac{1}{2} g_T(b)' W' \frac{\partial g_T(b)}{\partial b}.$$

To finish, recall that  $W$  is symmetric. Transpose to get  $n \times 1$  vector.

- ▶ Generalization: solve

$$a_T(\hat{b}) g_T(\hat{b}) = 0.$$

directly without specifying minimization problem.

# GMM Asymptotics

- ▶ Mean-value expansion:

$$g_T(\hat{b}) = g_T(b) + d_T(\bar{b})(\hat{b} - b)$$

where  $b$  are true parameters, and  $\bar{b} = \alpha\hat{b} + (1 - \alpha)b$  for some  $\alpha \in [0, 1]$ .

- ▶ Multiply by  $a_T(\hat{b})$  to get

$$0 = a_T(\hat{b})g_T(b) + a_T(\hat{b})d_T(\bar{b})(\hat{b} - b).$$

- ▶ If  $d_T$  is invertible, rearrange into

$$\sqrt{T}(\hat{b} - b) = -(a_T(\hat{b})d_T(\bar{b}))^{-1}a_T(\hat{b})\sqrt{T}g_T(b)$$

- ▶ In the limit:

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

where  $a = \text{plim } a_T$ ,  $d = \text{plim } d_T$ ,  $S = \text{avar } g_T$ .



# GMM Asymptotics

- ▶ Mean-value expansion:

$$g_T(\hat{b}) = g_T(b) + d_T(\bar{b})(\hat{b} - b)$$

where  $b$  are true parameters, and  $\bar{b} = \alpha\hat{b} + (1 - \alpha)b$  for some  $\alpha \in [0, 1]$ .

- ▶ Multiply by  $a_T(\hat{b})$  to get

$$0 = a_T(\hat{b})g_T(b) + a_T(\hat{b})d_T(\bar{b})(\hat{b} - b).$$

- ▶ If  $d_T$  is invertible, rearrange into

$$\sqrt{T}(\hat{b} - b) = -(a_T(\hat{b})d_T(\bar{b}))^{-1}a_T(\hat{b})\sqrt{T}g_T(b)$$

- ▶ In the limit:

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

where  $a = \text{plim } a_T$ ,  $d = \text{plim } d_T$ ,  $S = \text{avar } g_T$ .

# Efficient Weighting Matrix

- ▶ GMM always consistent, but asymptotic variance of  $\hat{b}$  depends on weighting matrix  $a$ . Denote it  $V(a)$ .
- ▶ Define  $a^*$  to be **efficient** if  $V(a) - V(a^*)$  is positive semi-definite for any alternative  $a$ , so that

$$w'(V(a) - V(a^*))w \geq 0, \quad \forall w.$$

- ▶ This is equivalent to saying that for any weights  $w$ , the asymptotic variance of  $w'\hat{b}(a^*)$  is smaller than that of  $w'\hat{b}(a)$ .
- ▶ Efficient  $a$  turns out to be  $a^* = d'S^{-1}$  or, alternatively,  $W^* = S^{-1}$ .
- ▶ Intuition: for diagonal  $S$  this puts more weight on moments with less noise. For non-diagonal  $S$ , use correlations to reduce noise further.

# Efficient Weighting Matrix

- ▶ GMM always consistent, but asymptotic variance of  $\hat{b}$  depends on weighting matrix  $a$ . Denote it  $V(a)$ .
- ▶ Define  $a^*$  to be **efficient** if  $V(a) - V(a^*)$  is positive semi-definite for any alternative  $a$ , so that

$$w'(V(a) - V(a^*))w \geq 0, \quad \forall w.$$

- ▶ This is equivalent to saying that for any weights  $w$ , the asymptotic variance of  $w'\hat{b}(a^*)$  is smaller than that of  $w'\hat{b}(a)$ .
- ▶ Efficient  $a$  turns out to be  $a^* = d'S^{-1}$  or, alternatively,  $W^* = S^{-1}$ .
- ▶ Intuition: for diagonal  $S$  this puts more weight on moments with less noise. For non-diagonal  $S$ , use correlations to reduce noise further.

# The Recipe

- ▶ Three step procedure:
  1. Estimate  $\hat{b}$  using some initial  $a_T$  ( $W = I$  or  $W = E[xx']^{-1}$ ).
  2. Estimate the spectral density matrix  $\hat{S}_T$  and compute efficient  $a_T^*$ .
  3. Re-estimate  $\hat{b}^*$  using the efficient matrix, as in Step 1.
- ▶ For “iterated GMM” continue until  $W$  converges.
- ▶ Estimation steps either by exact solution or numerical equation solver.
- ▶ For  $W = I$  choose moment conditions to have similar scale and be as close to uncorrelated as possible (i.e.,  $S \simeq I$ ).

## Estimation of $S$ : Nonparametric

- ▶ Define  $u_t = f(x_t, \hat{b}) - g_T(x_t, \hat{b})$  to be the demeaned GMM residuals. Then

$$S = \text{Cov}\left(\sqrt{T}g_T(b)\right) = \sum_{j=-\infty}^{\infty} E\left[u_t u'_{t-j}\right].$$

- ▶ If errors uncorrelated for all  $|j| > J$ , then we can compute this directly (Hansen-Hodrick estimator). Otherwise, need another approach.
- ▶ General nonparametric estimator uses weights  $w$  to compute

$$\hat{S} = \sum_{j=-k}^k w_j \left( \frac{1}{T} \sum_{t=1}^T u_t u'_{t-j} \right)$$

- ▶ Typically use the “Bartlett kernel”  $w_j = (k - |j|)/k$ , which defines the **Newey-West estimator**.

## Estimation of $S$ : Parametric

- ▶ **Parametric alternative #1:** for complex autocovariance structure, can assume that  $u_t$  follows a VAR, so that

$$\begin{aligned}u_t &= \Phi u_{t-1} + e_t, & e_t &\sim N(0, Q) \\ \Sigma &= \Phi \Sigma \Phi' + Q\end{aligned}$$

where  $\Sigma \equiv \text{Var}(u_t)$ .

- ▶ “Lyapunov” fixed point equation can be solved by numerical packages (e.g., `dlyap` in Matlab) or doubling algorithm.
- ▶ Use the relation

$$E \left[ u_t, u'_{t-j} \right] = \Phi^j \Sigma$$

to obtain

$$S = \Sigma + (I - \Phi)^{-1} \Phi \Sigma + \Sigma (I - \Phi')^{-1} \Phi'$$

# Estimation of $S$ : Parametric

- ▶ **Parametric alternative #2:** for large  $n$  (many equations), can assume that  $u_t$  has a factor structure so that

$$u_t = Bf_t + e_t, \quad f_t \sim N(0, I), \quad e_t \sim N(0, \text{diag}(v))$$
$$S = BB' + Q$$

- ▶ Could combine approaches: VAR with factor errors.

# Inverting $\hat{S}$

- ▶ Once  $\hat{S}$  has been estimated, need to invert to get  $W^*$ .
  - Problem:  $\hat{S}$  can be close to singular, inverse unreliable.
  - Solution: take “pseudoinverse” (e.g., `pinv`)
- ▶ To compute, take **singular value decomposition**:  $\hat{S} = UDV'$ .
  - $UU' = V'V = I$  and  $D$  is diagonal.
  - Aside: principal components are  $Z = XV$  when  $X = UDV'$ . (Take z-scores first).
- ▶ Stable inverse: for some tolerance  $\varepsilon$  (e.g.,  $1e-10$ ), compute

$$\tilde{D}_{ii} = \begin{cases} D_{ii}^{-1} & \text{if } D_{ii} > \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

- ▶ Pseudoinverse:  $\hat{S}^{-1} = V\tilde{D}U'$ .



## Inference: Standard Errors

- ▶ Define  $V = \left(0, (ad)^{-1}aSa'(d'a')^{-1}\right)$ , or in  $W = \hat{S}^{-1}$  case,  $V^* = (d'S^{-1}d)^{-1}$ .
- ▶ Standard errors:  $SE(\hat{b}_i) = \sqrt{V_{ii}/T}$ .
- ▶ Can compute standard errors of objects depending on  $b$  using the delta method (e.g., regression  $R^2$ ). For function  $\phi(\hat{b})$ :

$$\sqrt{T}(\phi(\hat{b}) - \phi(b)) \xrightarrow{d} N\left(0, \frac{\partial\phi(b)'}{\partial b} V \frac{\partial\phi(b)}{\partial b}\right)$$

where the derivative  $\partial\phi(b)/\partial b$  is represented as a column vector.

- ▶ Use **numerical derivatives** if analytical derivatives not known.

# Inference: Hypothesis Testing

- ▶ General idea: if a  $n \times 1$  vector  $x \sim N(0, \Sigma)$ , then  $x' \Sigma^{-1} x \sim \chi^2(n)$ .
  - In practice, estimation procedure may use up d.f., so we may end up with  $\chi^2(n - k)$ .
- ▶ From our GMM estimates,  $(\hat{b} - b) \sim N(0, V/T)$  for  $V = \text{avar}(\hat{b})$ .
- ▶ Assume we are testing a  $k \times 1$  restriction  $r(b) = 0$ . Then under the null we have

$$\sqrt{T}r(\hat{b}) \xrightarrow{d} N\left(0, \frac{\partial r(b)'}{\partial b} V \frac{\partial r(b)}{\partial b}\right)$$

and so

$$T \cdot r(\hat{b})' \left[ \frac{\partial r(b)'}{\partial b} V \frac{\partial r(b)}{\partial b} \right]^{-1} r(\hat{b}) \sim \chi^2(n - k).$$

- ▶ Alternatives: likelihood ratio test, Lagrange multiplier test (these require you to estimate the restricted model).

# Inference: Hypothesis Testing

- ▶ Under the optimal GMM weighting matrix we have some special hypothesis tests.
- ▶ Assuming moment condition is correct, then under  $W = \hat{S}^{-1}$ :

$$Tg_T(\hat{b})'S^{-1}g_T(\hat{b}) \xrightarrow{d} \chi^2(\#\text{moments} - \#\text{parameters})$$

where correction for degrees of freedom because we are using fitted value  $\hat{b}$  instead of  $b$ .

- ▶ Can also compare unrestricted model with restricted model that requires  $r(\hat{b}) = 0$ :

$$T\left(g_T(\hat{b}_R)'S_U^{-1}g_T(\hat{b}_R) - g_T(\hat{b}_U)'S_U^{-1}g_T(\hat{b}_U)\right) \xrightarrow{d} \chi^2(\#\text{restrictions})$$

or compare “big” set of moments  $g'_{B,T} = (g'_{1,T}, g'_{2,T})$  with “small” set  $g'_{S,T} = (g'_{1,T}, 0)$

$$T\left(g_{B,T}(\hat{b}_B)'S_B^{-1}g_{B,T}(\hat{b}_B) - g_{S,T}(\hat{b}_S)'S_B^{-1}g_{S,T}(\hat{b}_S)\right) \xrightarrow{d} \chi^2(\#\text{extra moments})$$

## Example: C-CAPM with Mimicking Factors

- ▶ You are trying to estimate the consumption CAPM moment condition

$$E \left[ R_t \cdot \left( \frac{c_t}{c_{t-1}} \right)^{-\gamma} - 1 \right] = 0$$

for a country without reliable consumption data.

- ▶ Have data on factors related to consumption  $x_t$ , and both consumption and factor data in the US ( $c_t^{US}, x_t^{US}$ ).
- ▶ Consumption growth follows

$$\begin{aligned} \Delta \log c_t &= \phi' x_t + e_t & e_t &\sim N(0, \omega) \\ \Delta \log c_t^{US} &= \phi' x_t^{US} + e_t^{US} & e_t^{US} &\sim N(0, \omega) \end{aligned}$$

- ▶ Goal: compute exact standard errors for  $\gamma$  (assuming  $\omega$  known).

## Example: C-CAPM with Mimicking Factors

- ▶ Jointly estimate US consumption regression and C-CAPM.
- ▶ Moment conditions:

$$E \begin{bmatrix} \exp\{\log R_t - \gamma\phi'x_t + \gamma^2\omega/2\} - 1 \\ (\Delta \log c_t^{US} - \phi'x_t^{US}) x_t^{US} \end{bmatrix} = 0.$$

- ▶ Population equivalent:

$$g_T = \frac{1}{T} \sum_{t=1}^T \begin{bmatrix} \exp\{\log R_t - \gamma\phi'x_t + \gamma^2\omega/2\} - 1 \\ (\Delta \log c_t^{US} - \phi'x_t^{US}) x_t^{US} \end{bmatrix}$$

- ▶ Derivatives w.r.t. parameters  $\gamma, \phi$ :

$$d_T = \frac{1}{T} \sum_{t=1}^T \begin{bmatrix} -\exp\{\cdot\} (\phi'x_t - \gamma\omega) & \exp\{\cdot\} \gamma x_t' \\ 0 & x_t^{US} (x_t^{US})' \end{bmatrix}$$

# Linear GMM: Warm Up

- ▶ Linear GMM moment condition (single equation):

$$E \left[ \underbrace{z_t}_{m \times 1} \left( \underbrace{y_t}_{1 \times 1} - \underbrace{x_t'}_{1 \times n} \underbrace{b}_{n \times 1} \right) \right] = 0.$$

- ▶ Sample moment condition:

$$g_T(b) = \frac{1}{T} \sum_{t=1}^T z_t (y_t - x_t' b) = \underbrace{Z'}_{m \times T} \left( \underbrace{y}_{T \times 1} - \underbrace{X}_{T \times n} \underbrace{b}_{n \times 1} \right) / T$$

where  $Z' = (z_1, \dots, z_T)$ ,  $X' = (x_1, \dots, x_T)$ .

- ▶ Easy to solve, apply formulas. But what about multiple equation (e.g., pricing multiple assets with multiple factors)? Need vector  $b$  but have matrix  $B$ :

$$E \left[ \underbrace{Z_t}_{m \times k} \left( \underbrace{y_t}_{k \times 1} - \underbrace{B}_{k \times n} \underbrace{x_t}_{n \times 1} \right) \right] = 0.$$

## Tools: Vectorization and Kronecker Products

- ▶ Kronecker product  $\otimes$  defines the operation

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$$

- ▶ Bilinear, associative:

$$A \otimes (B + C) = A \otimes B + A \otimes C$$

$$(A + B) \otimes C = A \otimes C + B \otimes C$$

$$(kA) \otimes B = A \otimes (kB) = k(A \otimes B)$$

$$(A \otimes B) \otimes C = A \otimes (B \otimes C)$$

## Tools: Kronecker Products and Vectorization

- ▶ Properties of Kronecker product:

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

$$(A \otimes B)' = (A' \otimes B')$$

- ▶ Vectorization stacks columns to form vector:  $\text{vec}(X) = (x'_1, \dots, x'_n)'$ .

- ▶ Key equation:

$$\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$$

- ▶ Application: how do we solve  $X = AXA' + Q$  in one step?



# Linear GMM

- ▶ General GMM moment condition:

$$E \left[ \underbrace{Z_t}_{m \times k} \left( \underbrace{y_t}_{k \times 1} - \underbrace{B}_{k \times n} \underbrace{x_t}_{n \times 1} \right) \right] = 0.$$

- ▶ Convert  $B$  to a vector using Kronecker product. Solution:

- ▶ Sample moment condition:

$$g_T(b) = \frac{1}{T} \sum_{t=1}^T Z_t (y_t - X_t' b) = \underbrace{Z'}_{m \times kT} \left( \underbrace{y}_{kT \times 1} - \underbrace{X}_{kT \times nk} \underbrace{b}_{nk \times 1} \right) / T$$

where  $Z' = (Z_1, \dots, Z_T)$ ,  $X' = (X_1, \dots, X_T)$ .

# Linear GMM

- ▶ General GMM moment condition:

$$E \left[ \underbrace{Z_t}_{m \times k} \left( \underbrace{y_t}_{k \times 1} - \underbrace{B}_{k \times n} \underbrace{x_t}_{n \times 1} \right) \right] = 0.$$

- ▶ Convert  $B$  to a vector using Kronecker product. Solution:

$$Bx_t = \text{vec}(IBx_t) = (x_t' \otimes I_k) \text{vec}(B) \equiv \underbrace{X_t'}_{k \times nk} \underbrace{b}_{nk \times 1}$$

- ▶ Sample moment condition:

$$g_T(b) = \frac{1}{T} \sum_{t=1}^T Z_t (y_t - X_t' b) = \underbrace{Z'}_{m \times kT} \left( \underbrace{y}_{kT \times 1} - \underbrace{X}_{kT \times nk} \underbrace{b}_{nk \times 1} \right) / T$$

where  $Z' = (Z_1, \dots, Z_T)$ ,  $X' = (X_1, \dots, X_T)$ .

# Linear GMM

- ▶ Derivative  $d_T(b)$  defined by

$$d_T(b) = Z'X/T$$

- ▶ GMM solution:  $\hat{b} = (X'ZWZ'X)^{-1}X'ZWZ'y$

- ▶ Ordinary least squares:  $Z_t = X_t$ .

- Solution for any weighting matrix (just identified):  $\hat{b}_{OLS} = (X'X)^{-1}X'y$
- Asymptotic variance:  $\hat{V}_{OLS} = (X'X/T)^{-1}S(X'X/T)^{-1}$
- Serially uncorrelated errors:  $S_{OLS} = E[X_t\varepsilon_t\varepsilon_t'X_t'] = E[x_t x_t' \otimes \varepsilon_t \varepsilon_t']$
- Serially uncorrelated + homoskedasticity:  $S_{OLS} = E[x_t x_t'] \otimes E[\varepsilon_t \varepsilon_t']$ ,  $V_{OLS} = E[x_t x_t']^{-1} \otimes E[\varepsilon_t \varepsilon_t']$

- ▶ Generalized least squares:  $Z_t = X_t \Sigma^{-1}$ , where  $\Sigma = E[\varepsilon_t \varepsilon_t']$ .

- Solution:  $\hat{b}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$

# Nonlinear GMM

- ▶ Want to solve system of equations

$$q(b) = d_T(b)'Wg_T(b) = 0.$$

- ▶ Best numerical equation solvers require Jacobian:

$$J = d(b)'Wd(b) + \sum_i H_i(b)W_i g(b)$$

where  $H_i(b)$  is the Hessian of  $g^i(b)$ , and  $W_i$  is the  $i$ th row of  $W$ .

- ▶ Alternative formulation:

$$\text{vec}(J) = \text{vec}(d'Wd) + \mathcal{H}Wg$$

where the  $i$ th column of  $\mathcal{H}$  is  $\text{vec}(H_i)$ .

# Efficiency of Maximum Likelihood

- ▶ Often, you have many moments to choose from. Which is best? Turns out to be maximum likelihood (MLE).
- ▶ Can implement MLE given density  $\varphi$  as GMM:

$$0 = \frac{1}{T} \nabla_b L(b) = \frac{1}{T} \sum_{t=1}^T \nabla_b \log \varphi(x; b).$$

where asymptotic variance is  $V_{MLE} = E[ss']^{-1}$  where  $s' = \nabla_b \log \varphi(x; b)$ .

- ▶ Proof of efficiency (see Newey and McFadden (1994)):
  - First step: show  $d'Wd = E[ms']$  for  $m = d'Wg$ .
  - Second step: show  $V_{GMM} - V_{MLE} = z'UU'z$ :

$$V_{GMM} - V_{MLE} = E[ms']^{-1} E[mm'] E[sm']^{-1} - E[ss']^{-1} = E[ms']^{-1} E[UU'] E[sm']$$
$$U \equiv m - E[ms'] E[ss']^{-1} s$$

# Estimating Structural Models

- ▶ Often estimate structural models using simulated method of moments:

$$\hat{b}_{SMM} = \arg \min_b \frac{1}{2} (h(\hat{x}; b) - \bar{h}(\bar{x}; b))' W (h(\hat{x}; b) - \bar{h}(\bar{x}; b)).$$

where  $\hat{x}$  is simulated data and  $\bar{x}$  is true data.

- ▶ If  $h(x; b)$  can be written in moment condition form

$$h(x; b) = \frac{1}{T} \sum_{t=1}^T \tilde{h}(x_t; b)$$

then this is just GMM for  $g(\hat{x}; b) = h(\hat{x}; b) - \bar{h}(\bar{x}; b)$ .

- ▶ Can simulate anything from the model, but don't know the likelihood. What moments should we pick?

## Estimating Structural Models

- ▶ Gallant and Tauchen (1996)'s Efficient Method of Moments (EMM) procedure: pick moments that look a lot *like* the moments implied by MLE:

$$g(x; b) \simeq \frac{1}{T} \sum_{t=1}^T \nabla_b \log \varphi(x; b)$$

- ▶ Approach: pick an **auxiliary model** that provides a good fit of the data and has a known density  $\varphi^A(x; \theta)$ .
  - Examples: VAR, GARCH, Gallant's Semi Non-Parametric (SNP) model.
- ▶ Fit auxiliary model parameters on *true* data, then estimate parameters of structural model on *simulated* data using score of auxiliary model at ML estimate as moments:

$$\hat{\theta}_{ML} = \arg \min_{\theta} \frac{1}{T} \sum_{t=1}^T \nabla_{\theta} \log \varphi^A(\bar{x}_t; \theta)$$

$$g_{EMM}(\hat{x}; b) = \frac{1}{T} \sum_{t=1}^T \nabla_{\theta} \log \varphi^A(\hat{x}; \hat{\theta}_{ML})$$

# Recap

- ▶ GMM: procedure for estimating moment conditions  $E[f(x; b)] = 0$ .
  - Solution:  $d(\hat{b})'Wg(\hat{b}) = 0$
  - Asymptotics can provide exact standard errors for many exotic cases.
- ▶ Efficient GMM procedure:
  - Estimate  $\hat{b}$  given initial weighting matrix  $W$
  - Compute  $\hat{S} = \text{avar}(g_T)$
  - Re-estimate  $\hat{b}$  using  $\hat{S}$ .
- ▶ Most efficient moments are score functions from MLE, approximate these if you can!
- ▶ Tools:
  - Singular value decomposition
  - Kronecker product, vectorization