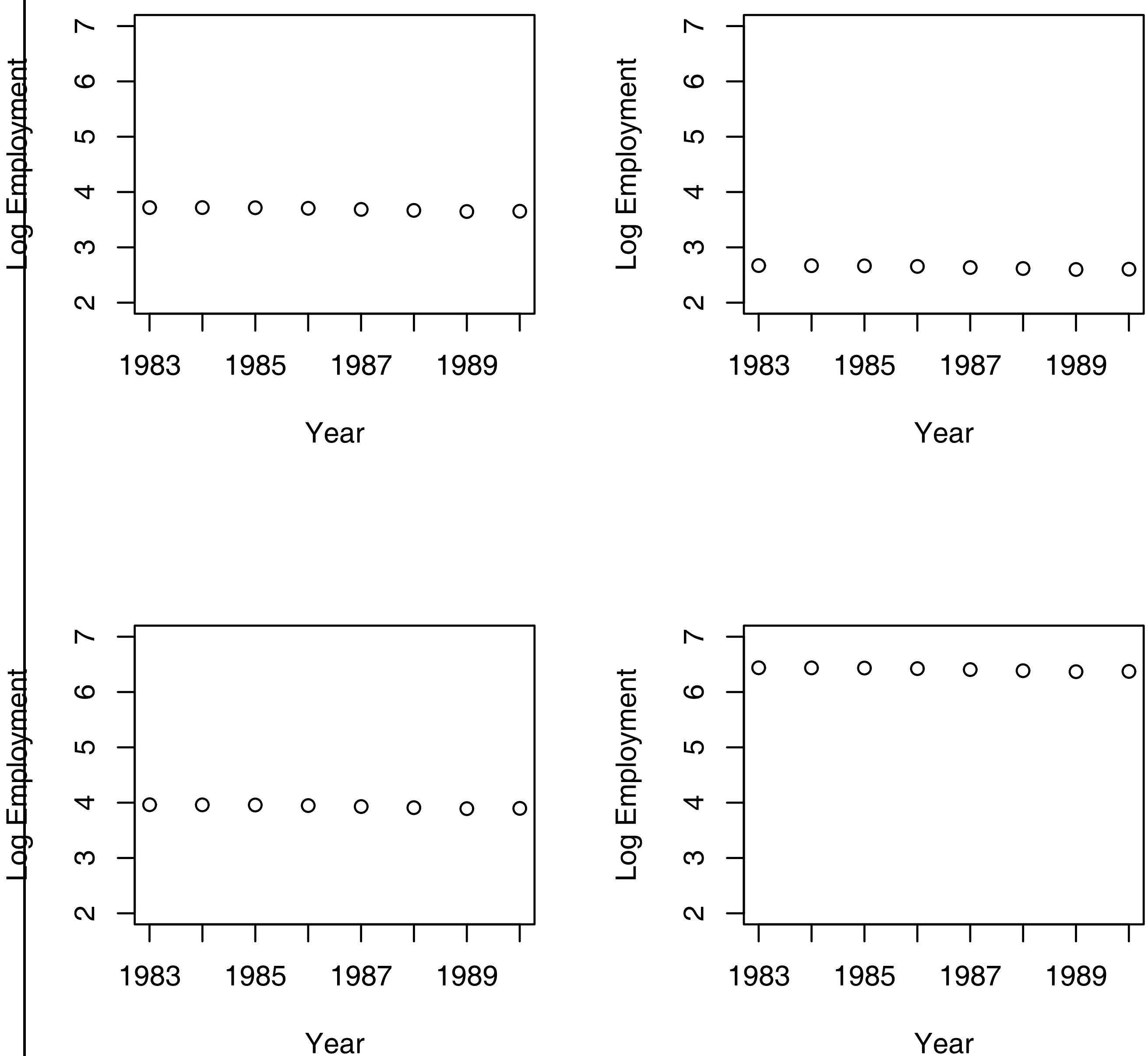
# Moment Condition Selection for Dynamic Panel Data Models

Dylan Small
Department of Statistics
The Wharton School, University of Pennsylvania

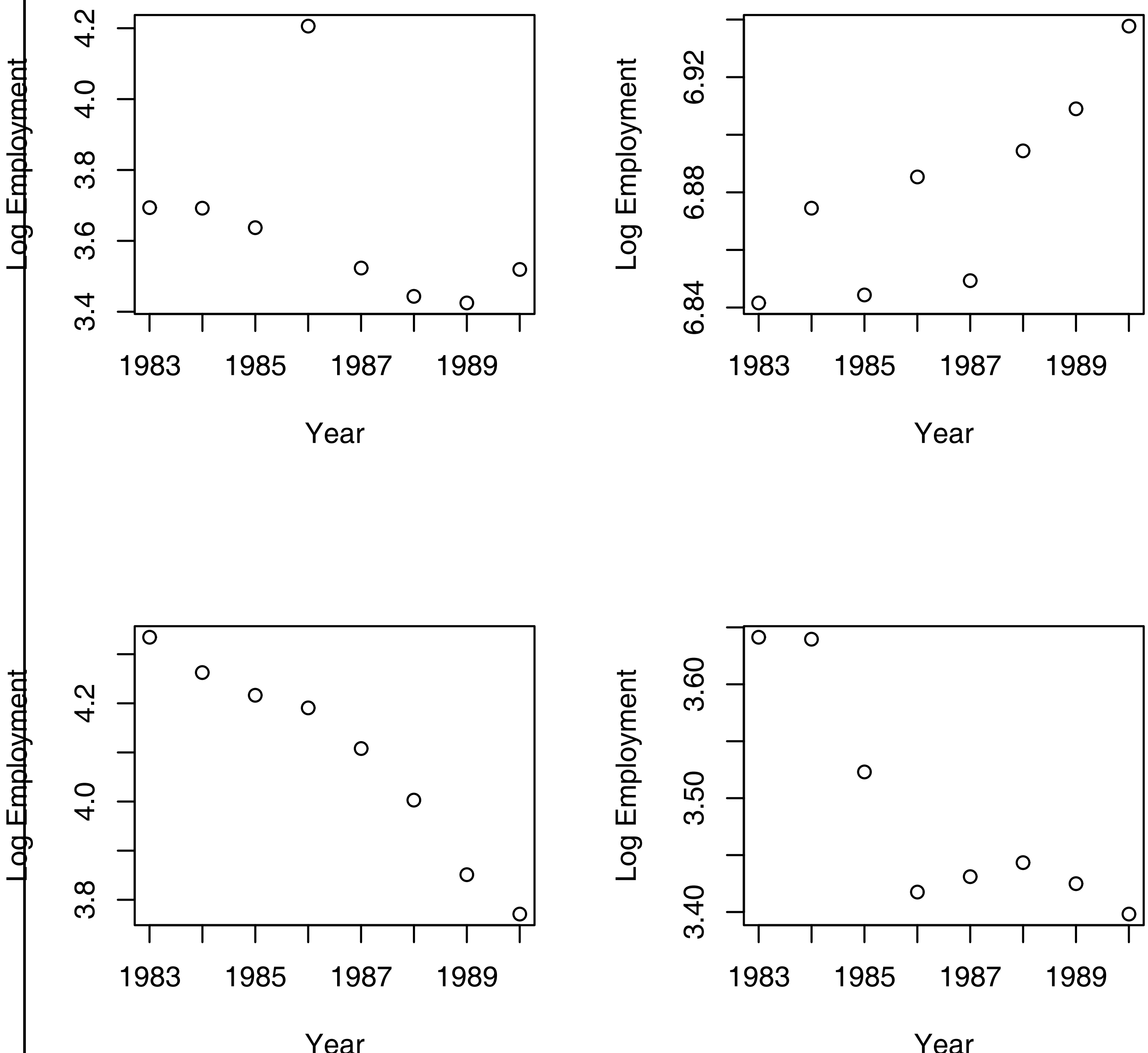
Joint work with Tze Leung Lai and Jia Liu

# Dynamic Panel Data

- Panel (longitudinal) data set: Follows a sample of units over time, providing a time series for each unit in the sample.
- In many economic settings, the number of units $N$ is relatively large but the time series for each unit is relatively short (let length $=T$).
- Example: Panel of 738 Spanish manufacturing firms, observed yearly from 1983-1990.
  - We are interested in understanding the dynamics of the firms' employment levels.
  - We remove a time trend and transform employment to log employment.
  - Source: Alonso-Borrego and Arellano (1999).



Log Employment
2
3
4
5
6
7
1983
1985
1987
1989
Year
Log Employment
2
3
4
5
6
7
1983
1985
1987
1989
Year
Log Employment
2
3
4
5
6
7
1983
1985
1987
1989
Year
Log Employment
2
3
4
5
6
7
1983
1985
1987
1989
Year


Log Employment
4.2
4.0
3.8
3.6
3.4
1983
1985
1987
1989
Year
Log Employment
6.92
6.88
6.84
1983
1985
1987
1989
Year
Log Employment
4.2
4.0
3.8
1983
1985
1987
1989
Year
Log Employment
3.60
3.50
3.40
1983
1985
1987
1989
Year

## Dynamic Panel Data Model

- $y_{it} = \theta y_{i,t-1} + \alpha_i + u_{it}, i = 1, \ldots, N, t = 1, \ldots, T$
  - $E(\alpha_i) = 0$, $E(u_{it}) = 0$.
  - $u_{it}$ iid.
  - $\alpha_i$ iid.
  - $E(\alpha_i u_{it}) = 0$

# Role of Initial Conditions

$y_{it} = \theta y_{i,t-1} + \alpha_i + u_{it}$.

The initial conditions are not specified in the base model. Different assumptions about the initial conditions affect the interpretation of how $y_{it}$ is generated (Anderson and Hsiao, 1981, "Estimation of Dynamic Models with Error Components, *JASA*; 1982, "Formulation and Estimation of Dynamic Models Using Panel Data, *J. Econ.*).

1. $y_{i0} = \tau_i, \;\; \tau_i$ and $\alpha_i$ are independent. A cross-sectional unit starts from a random initial position and gradually drifts towards its equilibrium level.

2. $y_{i0} = \alpha_i$. The individual effect represents the initial endowment and the effect of the initial endowment cumulates over time.

3. $y_{i0} = \frac{\alpha_i}{1-\theta} + v_{i0}, Var(v_{i0}) = \sigma_v^2$. Each individual's time series is stationary.

## Dependence of MLE on Initial Conditions

- $y_{it} = \theta y_{i,t-1} + \alpha_i + u_{it}$.
- Suppose $u_{it}$ and $\alpha_i$ are normally distributed.
- Anderson and Hsiao (1981): "The properties of the MLE depend crucially on the assumption of the initial conditions. Different assumptions about the initial conditions call for different methods to obtain the MLE. Mistaking one case for the other will generally not lead to asymptotically equivalent formulas. Consequently, the misused estimator may be inconsistent. Unfortunately, we usually have little information to rely on in making a correct choice of the initial conditions."

# Instrumental Variables Estimators

- $y_{it} = \theta y_{i,t-1} + \alpha_i + u_{it}, i = 1, \ldots, N, t = 1, \ldots, T, u_{it}$ iid.
- $y_{it} - y_{i,t-1} = \theta(y_{i,t-1} - y_{i,t-2}) + u_{it} - u_{i,t-1}$. $\Delta y_{it} = \theta \Delta y_{i,t-1} + \Delta u_{it}$
- Anderson and Hsiao proposed two instrumental variables estimators of $\theta$ that are consistent regardless of the initial conditions.
  1. $E[y_{i,t-2}\{\Delta y_{it} - \theta \Delta y_{i,t-1}\}] = E[y_{i,t-2} \Delta u_{it}] = 0$.
     $\hat{\theta}_{IV,1} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta y_{it}) y_{i,t-2}}{\sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta y_{i,t-1}) y_{i,t-2}}$.
  2. $E[\Delta y_{i,t-2}\{\Delta y_{it} - \theta(\Delta y_{i,t-1})\}] = E[(\Delta y_{i,t-2})(\Delta u_{it})] = 0$.
     $\hat{\theta}_{IV,2} = \frac{\sum_{i=1}^{N} \sum_{t=3}^{T} (\Delta y_{it})(\Delta y_{i,t-2})}{\sum_{i=1}^{N} \sum_{t=3}^{T} (\Delta y_{i,t-1})(\Delta y_{i,t-2})}$
- Anderson and Hsiao compared the asymptotic variances of $\hat{\theta}_{IV,1}$ and $\hat{\theta}_{IV,2}$ and gave guidelines for how to choose between the estimators.

# Moment Conditions for Dynamic Panel Data Model

- A semiparametric version of dynamic panel data model:
  - $y_{it} = \theta y_{i,t-1} + \alpha_i + u_{it}, t = 1, \ldots, T$ (1)
  - $E^*(u_{it} | y_{i0}, \ldots, y_{i,t-1}) = 0$ (2)
- (1) is equivalent to system of equations

  $y_{i1} = \theta y_{i0} + \alpha_i + u_{it}$ (3)

  $\Delta y_{it} = \theta \Delta y_{i,t-1} + \Delta u_{it}$
- (3) is not informative about $\theta$ in absence of restrictions on joint distribution of $(\alpha_i, y_{i0})$.
- Set of distributions $F$ for i.i.d. random vectors $(y_{i0}, \ldots, y_{iT})$ satisfying (1) and (2) for $\theta = \theta_0$ is equivalent to the set of distributions $F$ satisfying moment restrictions

  $E[y_{i,t-s}(\Delta y_{it} - \theta_0 \Delta y_{i,t-1})] = 0,$

  $s = 0, \ldots, t-2, t = 2, \ldots, T$ (4)
- All the moment conditions in (4) cannot be used directly in the method of moments because it leads to an overdetermined system of estimating equations.

## Generalized Method of Moments

- Let $\mathbf{z}_i = (y_{i0}, \ldots, y_{iT}, \mathbf{x}_{i0}, \ldots, \mathbf{x}_{iT})'$,
  $g(\mathbf{z}, \theta) = y_{i,t-s}(\Delta y_{it} - \theta \Delta y_{i,t-1})] = 0$,
  $s = 0, \ldots, t-2, t = 2, \ldots, T$.
- Note that $E_F[g(\mathbf{z}, \theta)] = 0$ for the true $F$, $\theta$.
- In analogy with the method of moments, we can obtain a consistent estimate of $\theta$ from minimizing a weighted sum of the sample moments:

  $$\hat{\theta} = \text{arg min}_{\theta} g(\mathbf{z}, \theta)' W g(\mathbf{z}, \theta)$$

  Equivalent to solving the estimating equation

  $$(\frac{\partial g(\mathbf{z}, \theta)}{\partial \theta})' W g(\mathbf{z}, \theta) = 0$$

## Properties of GMM

- First order asymptotic theory of GMM
  - Hansen (1982,*Econ.*): The asymptotically optimal weight matrix $W$ is $S = [E(gg')]^{-1}$. The following two step procedure is asymptotically as efficient as using the weight matrix $S$: (i) Use a consistent estimate of $\theta$ to estimate $S$; (ii) Estimate $\theta$ by using the weight matrix $\hat{S}$.
  - Chamberlain (1987,*J.Econ.*): The two-step GMM procedure is semiparametrically efficient.
- The first order asymptotic theory does not reflect the sampling fluctuations in estimating the weight matrix. This is particularly problematic when
  - The number of estimating equations is large.
  - Some of the estimating equations are “weak” in that they provide little information about $\theta$.

## Estimation Using Additional Moment Conditions

- We can construct a more efficient GMM estimator than that based on $E[y_{i,t-s}(\Delta y_{it} - \theta_0 \Delta y_{i,t-1})] = 0, s = 0, \ldots, t-2, t = 2, \ldots, T$ if we are willing to make additional assumptions.
- Example: The stationarity condition.
  $E[\alpha_i y_{i1}] = E[\alpha_i y_{i0}] \Rightarrow$
  $E[(\sum_{t=2}^{T} y_{it} - \theta y_{i,t-1})\Delta y_{i1}] = 0$

## Gains and Losses from Additional Assumptions

$T = 5$, $n = 250$, normal disturbances, 100 iterations.

Setting I: $\theta = .5$, initial conditions stationary.

Setting II: $\theta = .5$, $y_{i,-2} = \alpha_i$.

Setting III: $\theta = .9$, initial conditions stationary.

Setting IV: $\theta = .9$, $y_{i,-2} = \alpha_i$.

Setting V: $\theta = .5$, initial conditions stationary, double exponential disturbances.

RMSE of different estimators

| Estimator | I | II | III | IV | V |
|---|---|---|---|---|---|
| BM | .09 | .10 | .39 | .04 | .49 |
| S | .35 | .23 | .17 | .19 | .58 |
| BM+S | .09 | .15 | .25 | .17 | .56 |

BM=basic model, S=stationarity, BM+S=basic model + stationary

# Moment Condition Selection

## Description of problem

How to select which moment conditions to use? Two considerations:

1. Some moment conditions may be based on assumptions about which the researcher is unsure but the moment conditions would provide substantial information about $\theta$ if the assumptions are approximately true.

2. Using some true moment conditions may degrade estimation if the moment conditions are only weakly informative about $\theta$.

# Review of Literature

- **Pretesting**: The estimator is chosen by stepwise hypothesis tests. Overidentification tests for selecting moment restrictions. Two problems with this approach:
  1. The selection of significance levels of the tests is subjective and their interpretation is unclear.
  2. The tests are based on an asymptotic theory which is unreliable for typical sample sizes being considered.
- Andrews and Lu (2001, *J.Econ.*): Variable selection and moment selection is based on penalizing the GMM objective function for the number of parameters and number of moments. For example, Andrews' "BIC" criterion for a model with $b$ variables and $c$ moment conditions is

$$J_n(b,c) - (c-b)\ln n$$
$$J_n = g(\mathbf{z},\hat{\theta})'\hat{S}g(\mathbf{z},\hat{\theta})$$

- Hong, Preston and Shum (2003) developed an approach based on empirical likelihood that is similar in spirit to Andrews and Lu's.

## Motivation for New Approach

- Problem with Andrews and Lu' approach: Seeks to choose all possible correct moment restrictions. Does not work well in problems where a subset of the moment restrictions have weak information and it is better not to use them.
- Our two stage approach:
  1. Eliminate all inconsistent estimators using empirical likelihood test statistic.
  2. Choose estimator with best variance for parameter(s) of interest using the bootstrap.

# Empirical Likelihood

- Let $\mathbf{z}_i = (y_{i0}, \ldots, y_{iT}, \mathbf{x}_{i0}, \ldots, \mathbf{x}_{iT})'$, $g(\mathbf{z}, \theta) = y_{i,t-s}(\Delta y_{it} - \theta \Delta y_{i,t-1})] = 0$, $s = 0, \ldots, t-2, t = 2, \ldots, T$.
- Note that $E_F[g(\mathbf{z}, \theta)] = 0$ for the true $F$, $\theta$.
- The empirical likelihood $L_E(\theta)$ of $\theta$ is the following:

$$\sup\{\prod_{i=1}^{n} p_i | p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g(\mathbf{z}_i, \theta) = \mathbf{0}\}$$

- Empirical likelihood and two step GMM are asymptotically equivalent to the first order.
- Empirical likelihood tests of moment restrictions are asymptotically optimal in the large deviations sense of Bahadur and Hoeffding (Kitamura, *Econ.*, 2000).

## Setup for Two Stage Approach

- We observe i.i.d. $d$ dimensional random vectors $\mathbf{z}_i$ from $F_0 \in \mathbf{M}$ where $\mathbf{M}$ is the space of probability measures on $(R^d, B^d)$.
- We are interested in a $p$-dimensional parameter $\theta \in \Theta$.
- Basic model: We know that $F_0$ belongs to a family of probability measures $\mathcal{P}$,
  $\mathcal{P}(\theta) = \{F \in \mathbf{M} : E_F[g_{\mathcal{P}}(\mathbf{z}, \theta)] = 0\}$,
  e.g., $\mathcal{P} = \bigcup_{\theta \in \Theta} \mathcal{P}(\theta)$
- Let $\mathcal{Q}_m = \bigcup_{\theta \in \Theta_m} \mathcal{Q}_m(\theta)$ be subsets of $\mathcal{P}$, where the members of $\mathcal{Q}_m$ have an $r_m$ dimensional vector of moment restrictions $E[g_m(\mathbf{z}, \theta)] = 0$.
- $g(\mathbf{z}, \theta) = y_{i,t-s}(\Delta y_{it} - \theta \Delta y_{i,t-1})] = 0$,
  $s = 0, \dots, t-2, t = 2, \dots, T$. $\mathcal{Q}_1$ = $\mathcal{P}$, $\mathcal{Q}_2$:
  $E[\{g(\mathbf{z}, \theta), E[(\sum_{t=2}^{T} y_{it} - \theta y_{i,t-1})\Delta y_{i1}]\}]$.

## First Stage: Eliminate Wrong Models

- Let $\hat{\theta}_m$ be the maximum empirical likelihood estimator associated with $\mathcal{Q}_m$.
- Compute the empirical log-likelihood ratio $l_m = 2\{n \log n - \log L_{\mathcal{Q}_m}(\hat{\theta}_m)\}$ where $L_{\mathcal{Q}_m}(\theta)$ is the empirical likelihood of $\theta$ under $\mathcal{Q}_m$.
- Eliminate $\mathcal{Q}_m$ from consideration if

$$l_m > \max\{(r_m - p + q_m)n\epsilon_n, \chi^2_{r_m-p+q_m;1-\alpha}\}$$

  where $n\epsilon_n \to \infty$ and $\epsilon_n \to 0$.
- This ensures that wrong models will consistently eliminated and correct models will consistently not be eliminated.
- Rationale is related to both asymptotic $\chi^2$-distribution and the theory of moderate deviations for empirical likelihood ratio statistics.
- A typical choice of $\epsilon_n$ is $\frac{\log n}{2n}$ in analogy with BIC.

## Second Stage

- Among the models which have not been eliminated from consideration in Step 1, choose the associated estimator $\hat{\theta}_m$ which minimizes an estimate of a measure of the size of $Cov(\hat{\theta})$.
- Measuring the "size" of $Cov(\hat{\theta}_m)$:
  - The size of $Cov(\hat{\theta}_i)$ is measured in a way that is suited to the goals of the estimation problem. For example, if we are interested in one particular component of $\theta$, then we use the variance of the estimator of that component.
  - In the presence of moment conditions which only weakly identify the parameters of interest, the bootstrap provides far better estimates of $Cov(\hat{\theta}_m)$ than the asymptotic variance.

# Estimates of Variance

$y_{it} = \theta y_{i,t-1} + \alpha_i + u_{it}$

Each time series is stationary, unit normal disturbances, $\theta = .9$.

| | Actual | Bootstrap | Asymptotic |
|---|---|---|---|
| BM | 0.17 | 0.14 | 0.02 |
| BM+S | 0.08 | 0.09 | 0.01 |
| S | 0.02 | 0.02 | 0.01 |

BM=base model, S=stationarity moment restriction,

BM+S=base model plus stationarity moment restriction

## Simulation Study

$$y_{it} = \theta y_{i,t-1} + \alpha_i + u_{it}$$
$$(\alpha_i, u_{it}) \overset{iid}{\sim} N(0, I)$$

- Three models: basic model, basic model plus stationarity moment restriction and stationarity moment restriction alone.

- Three designs:
  - (I) The individual time series are stationary and $\theta = .9$ so that each individual time series is highly persistent.
  - (II) The individual time series are stationary and $\theta = .5$.
  - (III) The individual time series are nonstationary and $\theta = .9$. Each individual time series began two periods before the initial observation, $y_{i,-2} = \alpha_i$.

## Results of Simulation Study

Table 1: Design I: Stationary, highly persistent time series

| | RMSE | | |
|---|---|---|---|
| | $N = 100$ | $N = 250$ | $N = 500$ |
| A& L BIC | 0.530 | 0.501 | 0.369 |
| Two stage | 0.389 | 0.273 | 0.209 |

Table 2: Design II: Stationary, moderately persistent time series

| | RMSE | | |
|---|---|---|---|
| | $N = 100$ | $N = 250$ | $N = 500$ |
| A& L BIC | 0.158 | 0.096 | 0.062 |
| Two stage | 0.176 | 0.107 | 0.065 |

Table 3: Design III: Nonstationary, persistent time series

| | RMSE | | |
|---|---|---|---|
| | $N = 100$ | $N = 250$ | $N = 500$ |
| A& L BIC | 0.257 | 0.059 | 0.029 |
| Two stage | 0.254 | 0.078 | 0.034 |

# Conclusion

- Summary
  - Our methodology addresses problem of how to decide whether to use additional plausible, but uncertain, assumptions in parameter estimation for dynamic panel data model.
  - Our two stage approach provides a general methodology for selecting moment conditions when some proposed moment conditions may not be true and other proposed moment conditions may be true but too weak to be useful.