Corporate Finance Data
&
The Role of Dynamic Panels

Mark Flannery, University of Florida
Kristine W. Hankins, University of Kentucky

Panel Data

- Fixed Effects Matter
- Growing Focus on Methodology
  - Peterson, RFS 2009
  - Wintoki, Linck, & Netter, JFE 2012
  - Gormley & Matsa, Working paper

Dynamic Panel Prevalence

- Payout
- Capital Structure
- Cash Flow & Investments
- Corporate Governance/Ownership
- Banking & Financial Development

  - More complicated econometrics

Resources

- This paper
  - Flannery and Hankins, "Estimating Dynamic Panel Models in Corporate Finance"
    - *Journal of Corporate Finance*, forthcoming
    - Also on SSRN
- Slides & Further References:
  - http://gatton.uky.edu/faculty/hankins/
    - Google: Kristine Hankins, U of Kentucky

Dynamic Panel Issues

$$y_{it} = y_{it-1} + X_{it} + F_i + \varepsilon_{it}$$

Lag

Fixed Effect

$$y_{it} = fn(F_i) \longrightarrow y_{it-1} = fn(F_i)$$

Correlation between lag and FE creates a bias

Ignoring FE creates omitted variable problem

Dynamic Panel Issues

Demean

$$y_{it} - \overline{y_i} = (y_{it-1} - \overline{y_{i-1}}) + (X_{it} - \overline{X_i}) + (\varepsilon_{it} - \overline{\varepsilon_i})$$

- Short panel bias

$$\overline{\varepsilon_i} = \frac{1}{T}\sum \varepsilon_{it} \quad (including \ \ \varepsilon_{it-1})$$

First difference

$$y_{it} - y_{it-1} = (y_{it-1} - y_{it-2}) + (X_{it} - X_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1})$$

Correlation still exists

Creates downward bias in lag coef. estimate

Example of Short Panel Bias

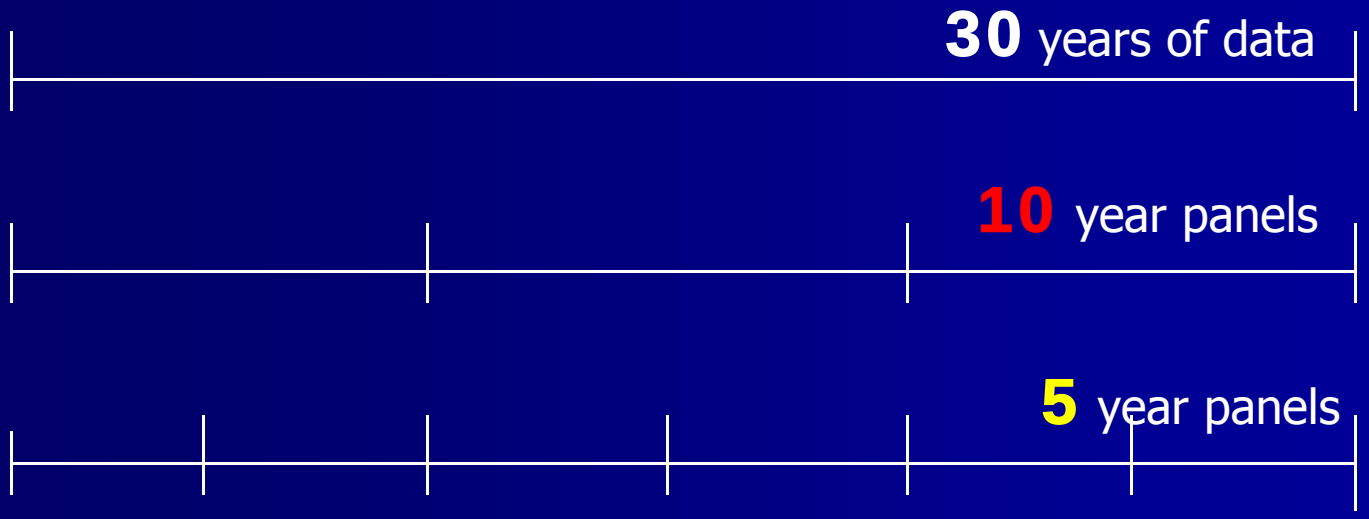
- Partial adjustment toward target leverage

$$MDR_{i,t+1} = (1-\lambda)MDR_{i,t} + (\lambda\beta)X_{i,t} + \lambda F_i + \delta_{i,t+1}$$

MDR	market debt ratio
λ	adjustment speed
X	firm controls
F	fixed effect

Degree of Bias?

- How much does panel length matter?
- Can't compare short and long panel firms
- Same data estimated over different horizons



Same Data, Varied Panel Length

Adjustment Speed

	OLS	FE	BB
30 Year Panels	13%	25%	15%
10 Year Panels	13%	44%	18%
5 Year Panels	13%	66%	18%

Possible Solutions

- Ignore short panel bias
  - FE
- Instrument
  - Traditional
  - GMM
    - Difference: Arellano and Bond ’91
    - System: Blundell and Bond ’98
    - Long difference: Hahn, Hausman, and Kuersteiner ’07
      Huang and Ritter ‘09

Instrument for Endogeneity

$$y_{i4} - y_{i3} = (y_{i3} - y_{i2}) + (\varepsilon_{i4} - \varepsilon_{i3})$$

- Arellano Bond ("Differences" GMM):
  - Lagged levels (y_{i2}, y_{i1}) are valid instruments for first differenced variables
- Blundell Bond ("System" GMM):
  - Lagged differences (y_{i2}- y_{i1}) also valid for levels eq.
- Long Difference:
  - LD_{ALL}: $(y_{it}- y_{i1}) = \delta (y_{it-1}- y_{i0}) + (v_{it} - v_{i1})$
  - LD_4: $(y_{it}- y_{it-4}) = \delta (y_{it-1}- y_{it-5}) + (v_{it} - v_{it-4})$

Possible Solutions

- Ignore short panel bias
  - FE **Bias depends on T**
- Instrument
  - Traditional **Hard to find**
  - GMM
    - Diff (Arellano Bond); System (Blundell Bond) **No 2nd Order Serial Corr.**
    - Long diff **Untested in unbalanced panels**
- Correct for bias
  - Bias-corrected LSDV (LSDVC) **Assumes exogeneity**
    - Kiviet '95, Bruno '05

Best in Corp Fin?

- Existing research
  - OLS < FE < AB < BB < LSDVC or LD
  - Econometrics Lit: Simple models
    - IID errors, 0 or 1 RHS variable
- Corporate Finance: More complicated
  - Multiple independent variables
  - Correlated with one another
  - Slow-changing
  - Endogenous

Methodology

- Generate data
  - Parameters + randomly generated errors
- Estimate
  - OLS, FE, GMM_{AB}, GMM_{BB}, LD_4, LD_{ALL}, LSDVC
- Save estimates, repeat 500 times
- Compare estimates and true values
  - RMSE

Simulate Data

- Specify 'true' model and parameters

$$y_{it} = \gamma y_{it-1} + \Sigma \beta_j x_{ijt} + \eta_i + \varepsilon_{it}$$

Lag **Multiple Xs** **FE** **Error**

$$x_{ijt} = \rho x_{ijt-1} + \alpha_1 y_{it-1} + \alpha_2 \eta_i + \xi_{ijt}$$

Endogeneous

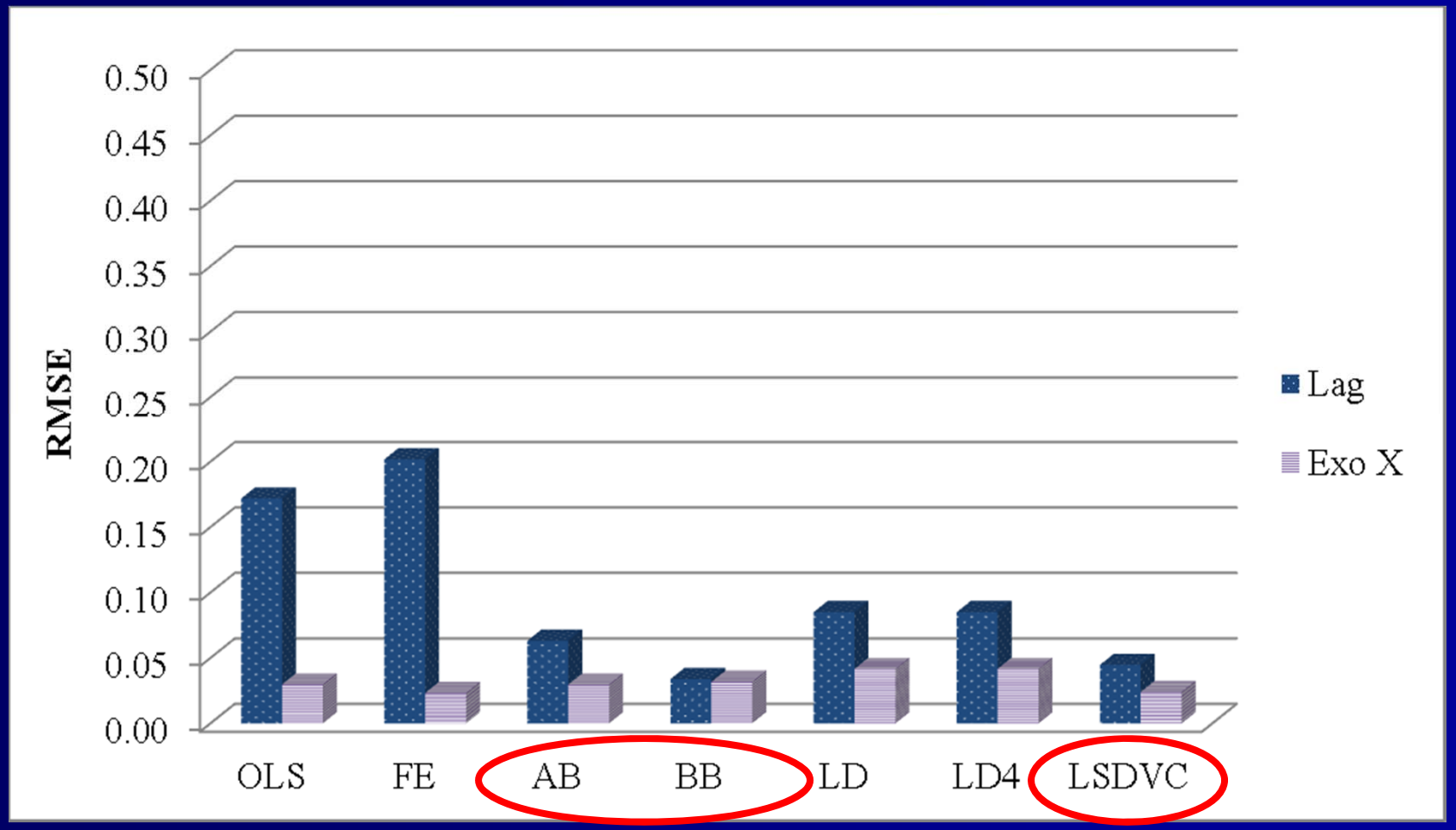
$$\varepsilon_{it} = \delta_1 \varepsilon_{it-1} + \delta_2 \varepsilon_{it-2} + \omega_{it}$$

Serial Correlation

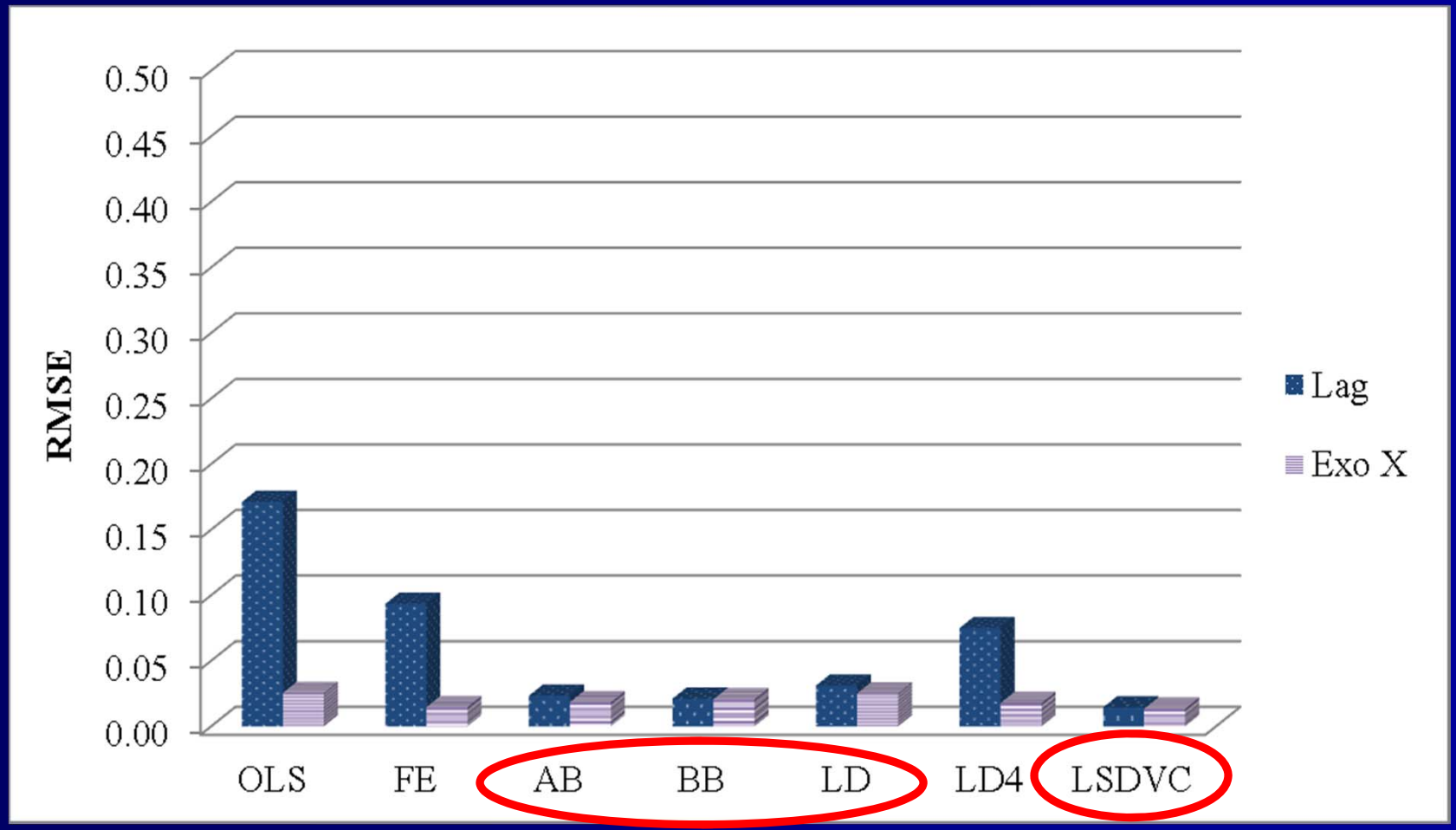
Overview of Simulations

- Corporate Panel Structure
  - Panel length
  - Persistence of lag
  - Exogenous variable structure (Xs covary)
- Common Limitations
  - Unbalanced panels
  - Missing observations
  - Dependent variable censoring or clustering
- Endogeneity, 2^{nd} Order Serial Correlation

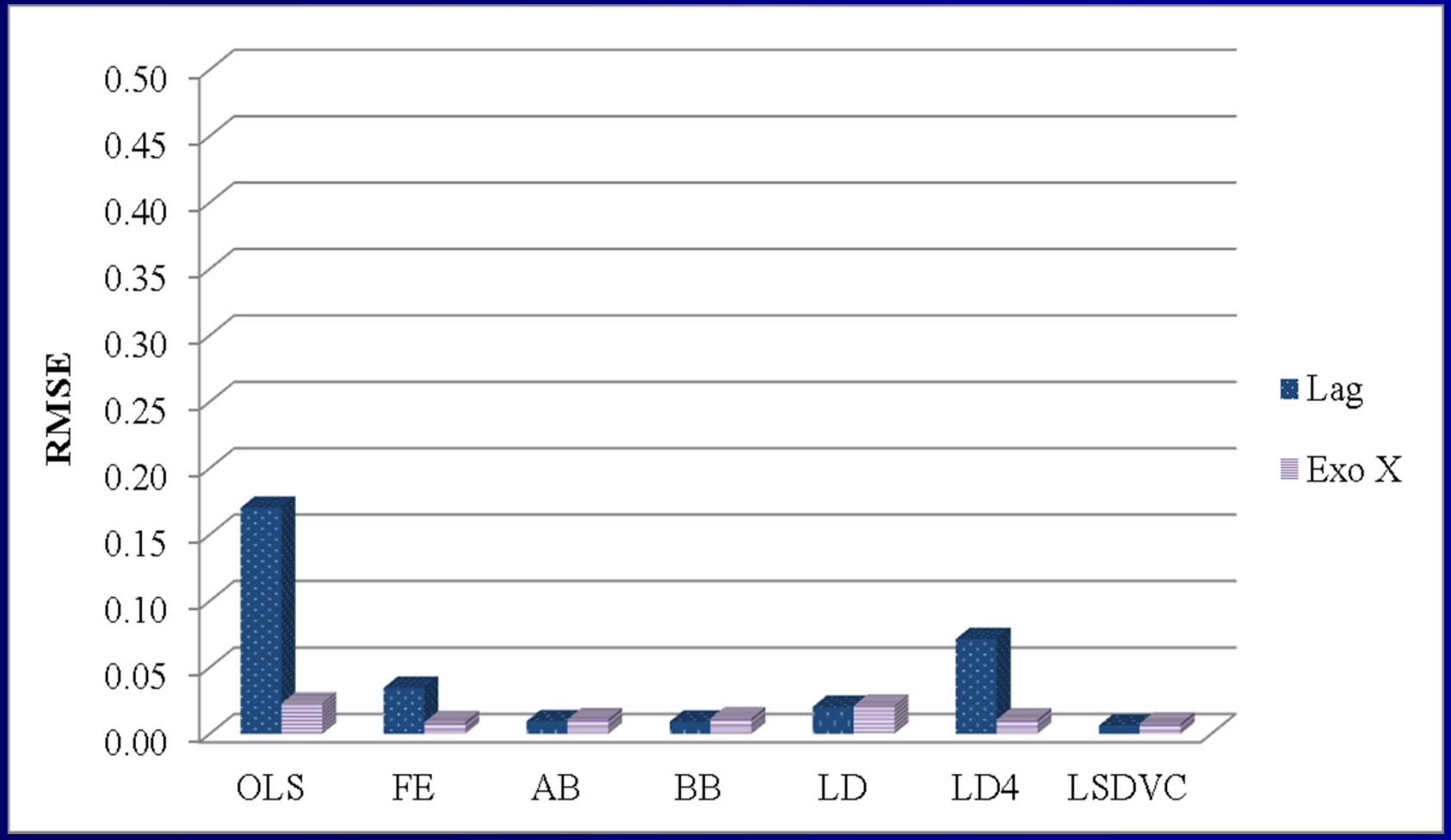
IID, T=6


RMSE
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
OLS
FE
AB
BB
LD
LD4
LSDVC
Lag
Exo X

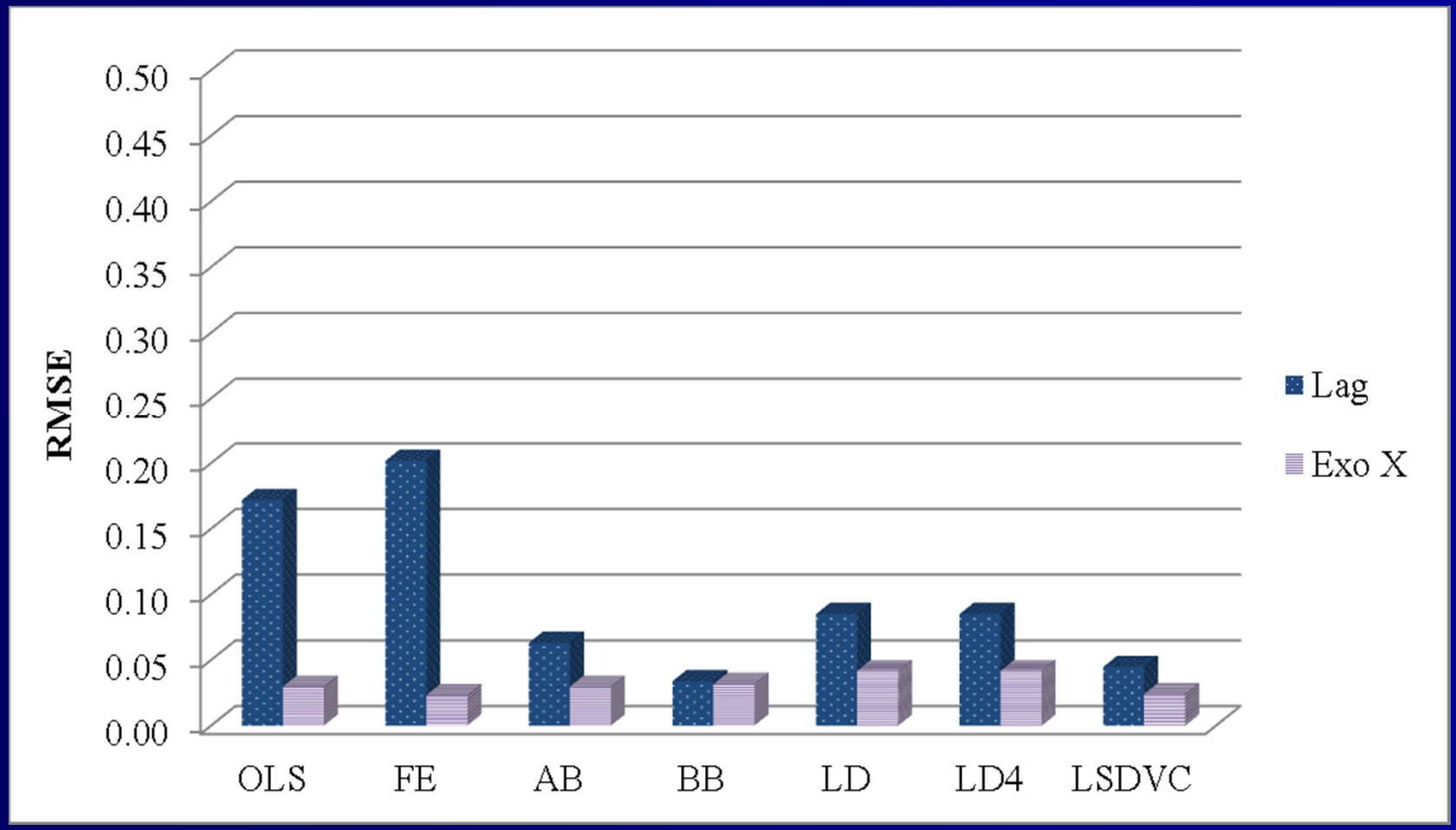
IID, T=12


RMSE
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
OLS
FE
AB
BB
LD
LD4
LSDVC
Lag
Exo X

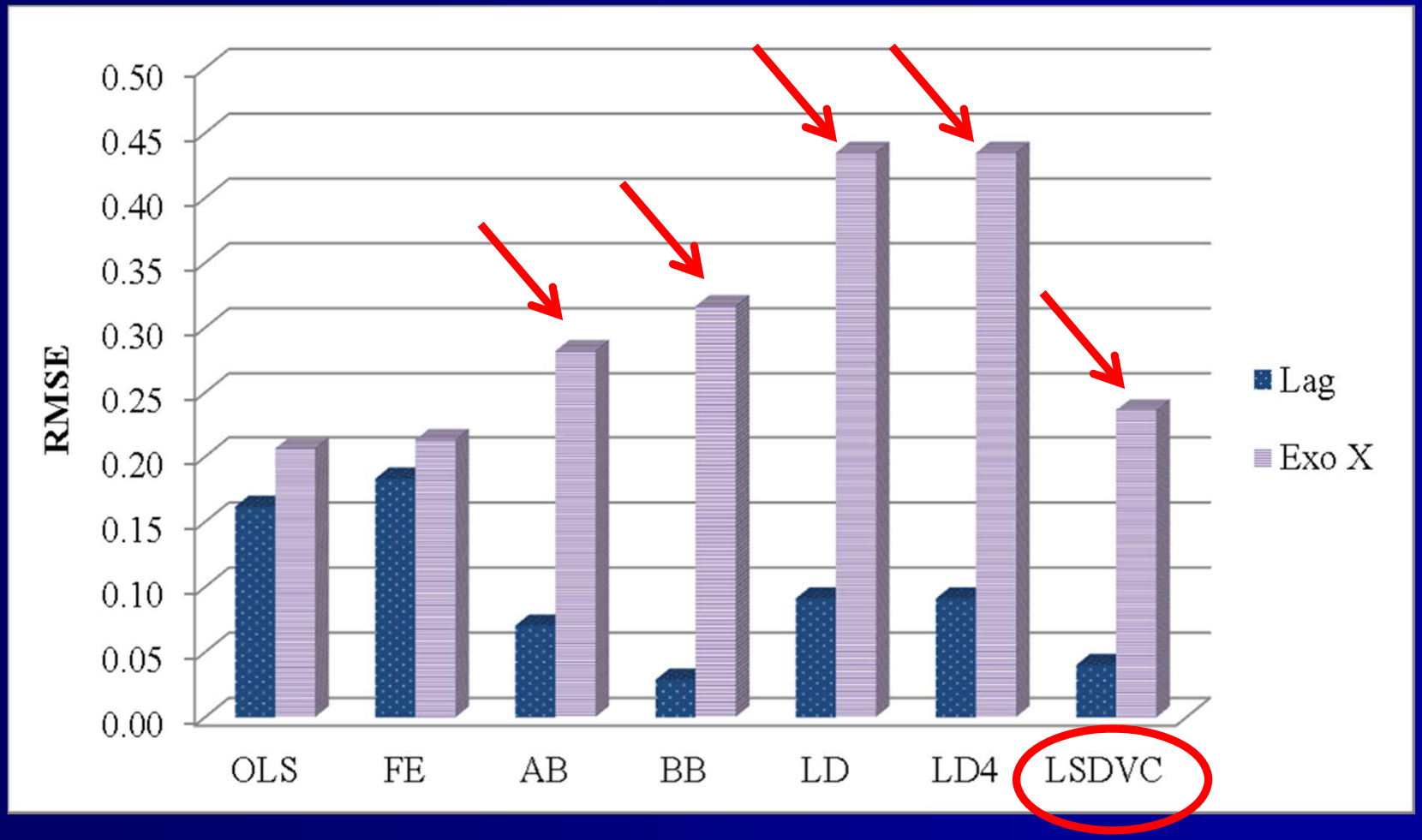
IID, T=30


RMSE
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
OLS
FE
AB
BB
LD
LD4
LSDVC
Lag
Exo X

IID, T=6


RMSE
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
OLS
FE
AB
BB
LD
LD4
LSDVC
Lag
Exo X

Compustat Innovations, T=6



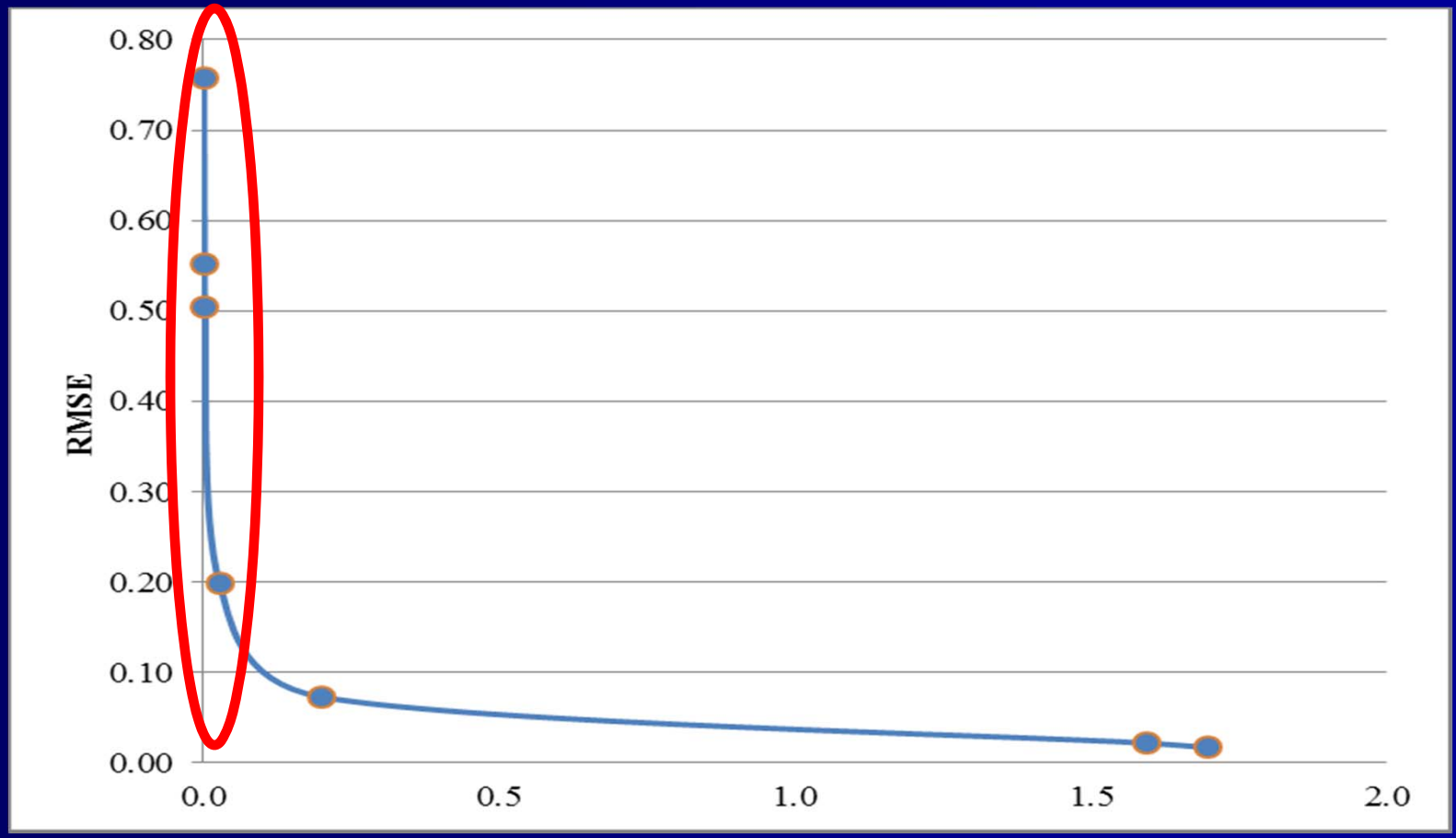
What Changed??

- Errors no longer IID for Xs

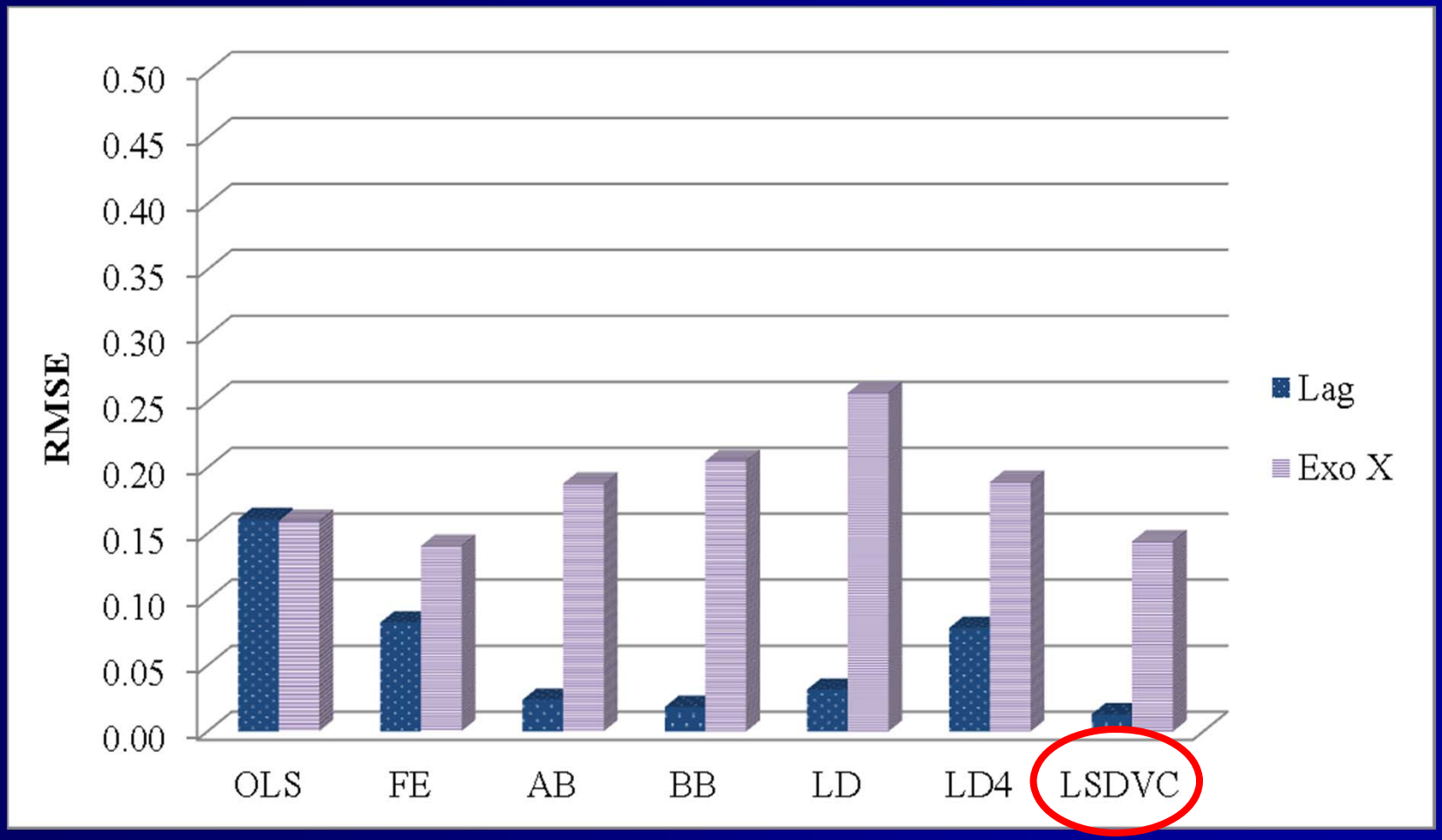
$$x_{ijt} = \rho_j x_{ijt-1} + \xi_{ijt}$$

- Drawn from joint normal distribution
  - Compustat var-cov matrix
  - Size of error term varies
  - Xs covary

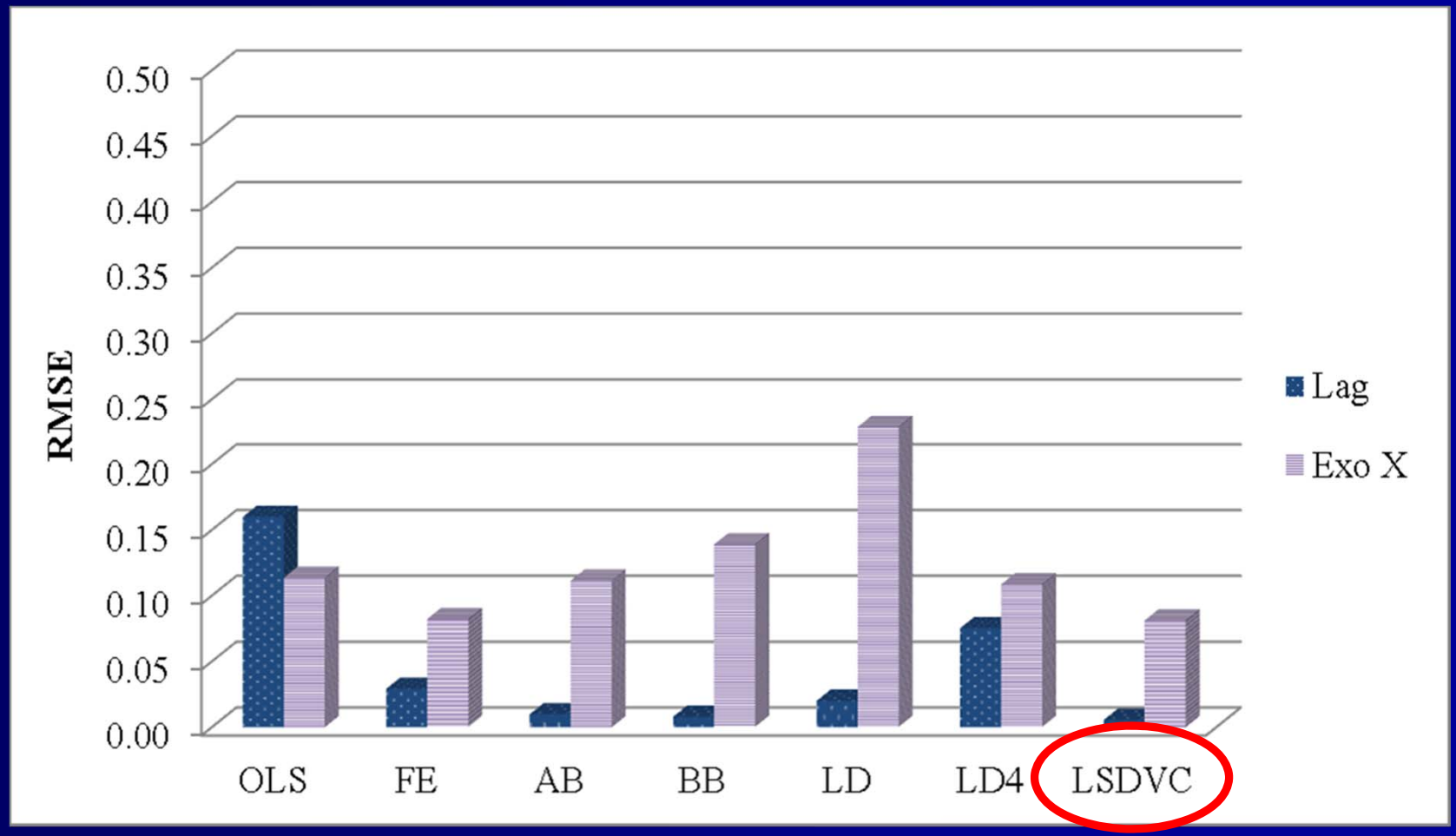
Large Errors with Sluggish Variables



Compustat Innovations, T=12



Compustat Innovations, T=30



Common Characteristics of CRSP-Compustat Data

- Unbalanced panels
- Missing observations

} LSDVC preferred

- Dependent variable censoring
- Dependent variable clustering

} BB better for Lag Dep Var

- LSDVC assumes exogeneity
- BB is invalidated by 2nd order serial corr

Endogeneity

- Wintoki, Linck, & Netter JFE 2012

$$x_{ijt} = \rho x_{ijt-1} + \alpha_1 y_{it-1} + \alpha_2 \eta_i + \xi_{ijt}$$

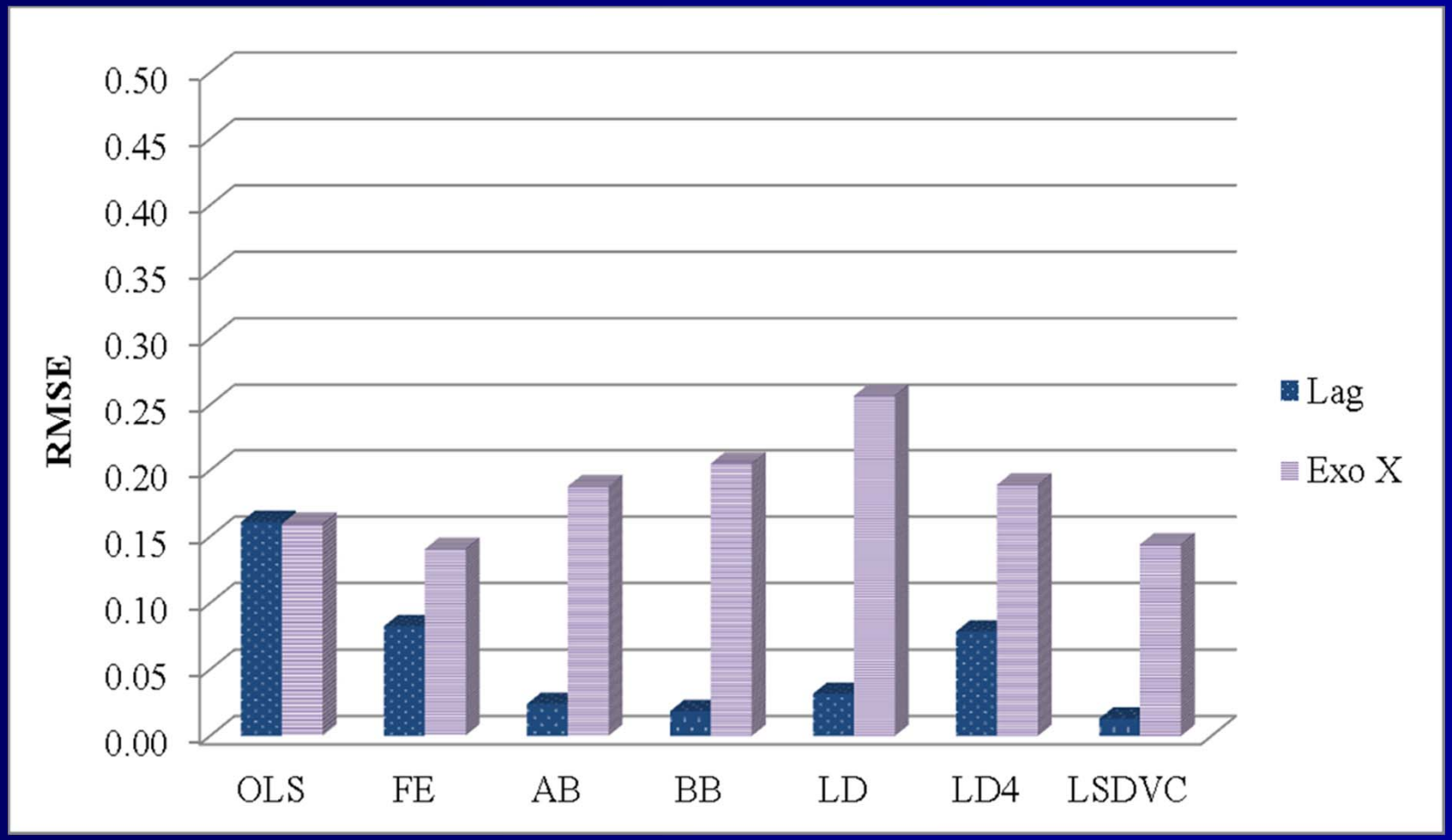
Lag, Fixed Effect

  - "Low" Endogeneity $\alpha_1 = \alpha_2 = 0.01$
  - "High" Endogeneity $\alpha_1 = \alpha_2 = 0.05$

- Wooldridge test for exogeneity

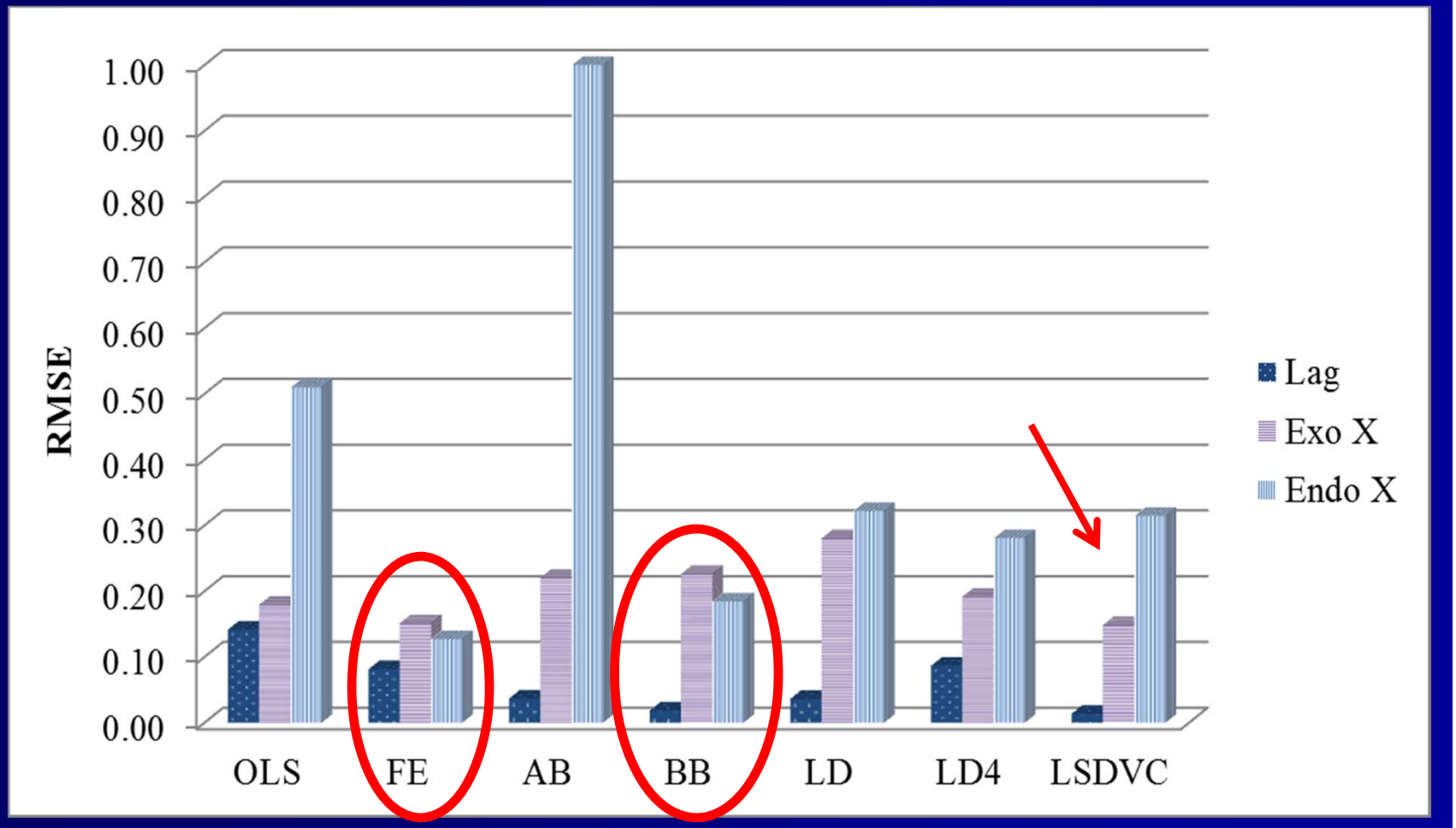
No Endo

T=12, Compustat Innovations



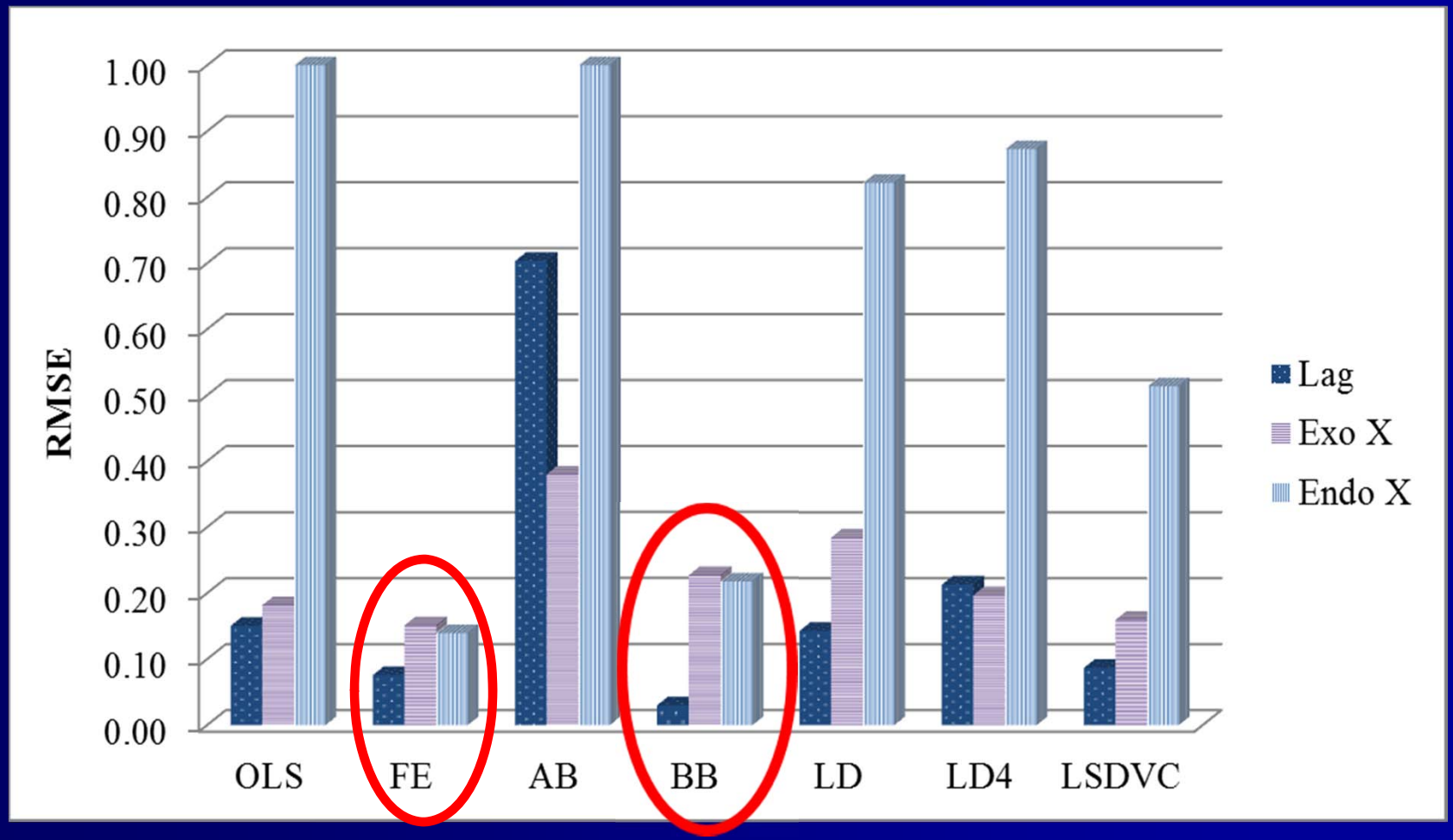
Low Endo

T=12, Compustat Innovations



High Endo

T=12, Compustat Innovations



Endo + Corp Fin Issues

- Unbalanced
  - Endo vars coefficients difficult to estimate
    - No methodology is remotely accurate
  - BB, LD, LSDVC best
    - But only reliable for low lag persistence (γ=0.2)

Endo + Corp Fin Issues

- Missing
  - BB best, FE only with low persistence
- Censoring
  - BB best, FE only if research interest is Xs
- Clustering
  - FE outperforms BB
  - But only tested on balanced panel

2nd Order Serial Corr

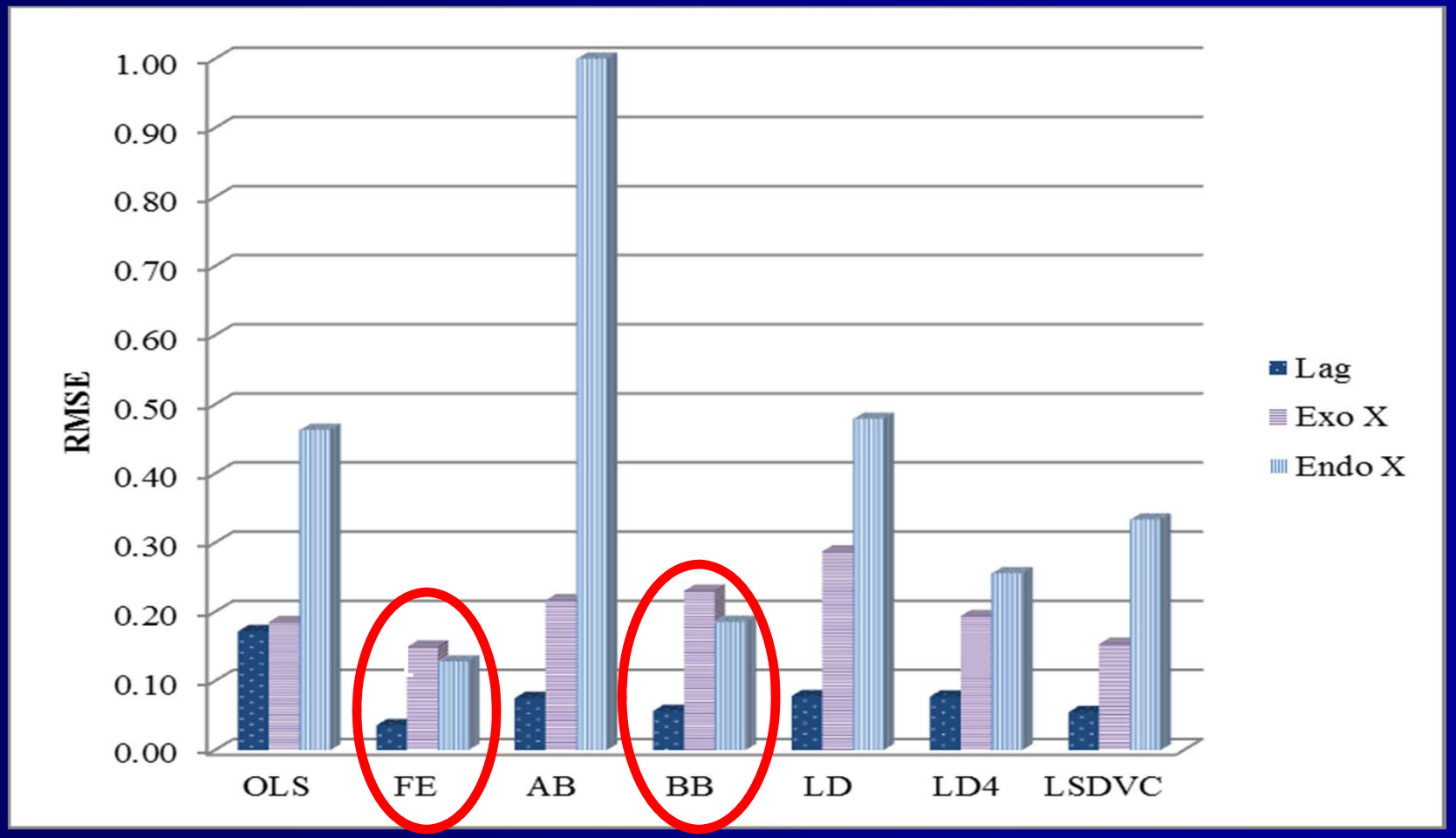
- Although Blundell Bond (BB) dominates with endogeneity
  - IVs are invalid with 2nd order serial corr
  - How important?
- Modify basic error term ($\delta_1 = 0.10$, $\delta_2 = 0.05$)

$$\varepsilon_{it} = \delta_1 \varepsilon_{it-1} + \delta_2 \varepsilon_{it-2} + \omega_{it}$$

- Evaluate with varying endogeneity
  - No, Low, and High

2nd Order Serial Corr

Low endogeneity



2nd Order Serial Corr

- LSDVC and FE best without endo
- BB slightly less accurate on lags
  - Not a large shift
- BB > LD (designed for serial corr!)
  - At least at T=12

Conclusions

- Econometrics matter
- Caution!!
  - Short panel bias exists
  - Sluggish variables difficult to estimate
  - AB problematic with endogeneity
    - Wooldridge endogeneity test
- Best Choices
  - No endo: LSDVC, BB
  - With endo: BB, FE