Dynamic Panel Data Models

Jake Anderson

May 16, 2026

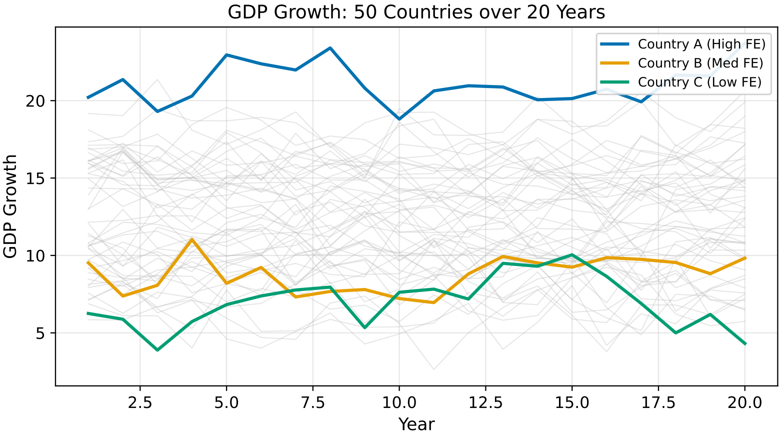
Outline

GDP Growth Across 50 Countries

Does this year's GDP growth depend on last year's? If so, how do we estimate that persistence?

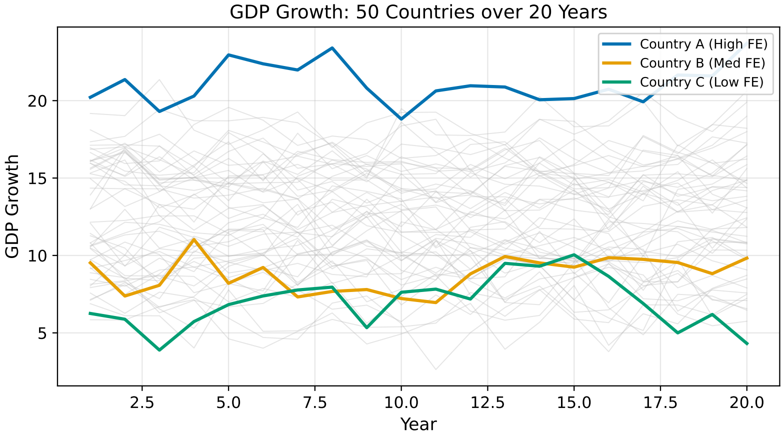
GDP Growth Across 50 Countries

Does this year's GDP growth depend on last year's? If so, how do we estimate that persistence?



GDP Growth Across 50 Countries

Does this year's GDP growth depend on last year's? If so, how do we estimate that persistence?



Some countries consistently grow faster than others. And growth seems **sticky**: a good year tends to follow a good year.

Persistence: The Pattern

Two features stand out in this panel:

Persistence: The Pattern

Two features stand out in this panel:

1. **Cross-country heterogeneity.** Country-specific factors (institutions, geography, culture) generate permanently different growth levels $\Longrightarrow$ fixed effects α_i.

Persistence: The Pattern

Two features stand out in this panel:

1. **Cross-country heterogeneity.** Country-specific factors (institutions, geography, culture) generate permanently different growth levels $\implies$ fixed effects α_i.
2. **Within-country persistence.** Even after accounting for country differences, growth this year is correlated with growth last year $\implies$ the lagged dependent variable $y_{i,t-1}$ belongs in the model.

Persistence: The Pattern

Two features stand out in this panel:

1. **Cross-country heterogeneity.** Country-specific factors (institutions, geography, culture) generate permanently different growth levels $\Longrightarrow$ fixed effects α_i.
2. **Within-country persistence.** Even after accounting for country differences, growth this year is correlated with growth last year $\Longrightarrow$ the lagged dependent variable $y_{i,t-1}$ belongs in the model.

$\Longrightarrow$ We need a model that includes **both** fixed effects and a lagged dependent variable.

The Dynamic Panel Model

$$y_{it} = \rho\, y_{i,t-1} + \beta\, \mathsf{invest}_{it} + \alpha_i + \varepsilon_{it}$$

The Dynamic Panel Model

$$y_{it} = \rho\, y_{i,t-1} + \beta\, \text{invest}_{it} + \alpha_i + \varepsilon_{it}$$

- y_{it}: GDP growth for country i in year t
- $y_{i,t-1}$: last year's growth (the **lagged dependent variable**)
- invest_{it}: investment rate (exogenous regressor)
- α_i: country fixed effect (unobserved, time-invariant)
- ε_{it}: idiosyncratic error ($\text{E}[\varepsilon_{it}] = 0$, serially uncorrelated)

The Dynamic Panel Model

$$y_{it} = \rho\, y_{i,t-1} + \beta\,\text{invest}_{it} + \alpha_i + \varepsilon_{it}$$

- y_{it}: GDP growth for country i in year t
- $y_{i,t-1}$: last year's growth (the **lagged dependent variable**)
- invest_{it}: investment rate (exogenous regressor)
- α_i: country fixed effect (unobserved, time-invariant)
- ε_{it}: idiosyncratic error ($\text{E}[\varepsilon_{it}] = 0$, serially uncorrelated)

The parameter ρ measures **persistence**: how much of last year's growth carries forward. If $|\rho| < 1$, the process is stationary.

The Dynamic Panel Model

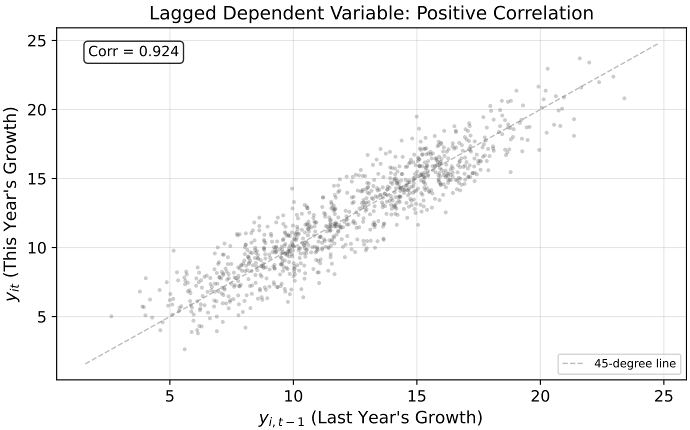
$$y_{it} = \rho\, y_{i,t-1} + \beta\, \text{invest}_{it} + \alpha_i + \varepsilon_{it}$$

- y_{it}: GDP growth for country i in year t
- $y_{i,t-1}$: last year's growth (the **lagged dependent variable**)
- invest_{it}: investment rate (exogenous regressor)
- α_i: country fixed effect (unobserved, time-invariant)
- ε_{it}: idiosyncratic error ($\text{E}[\varepsilon_{it}] = 0$, serially uncorrelated)

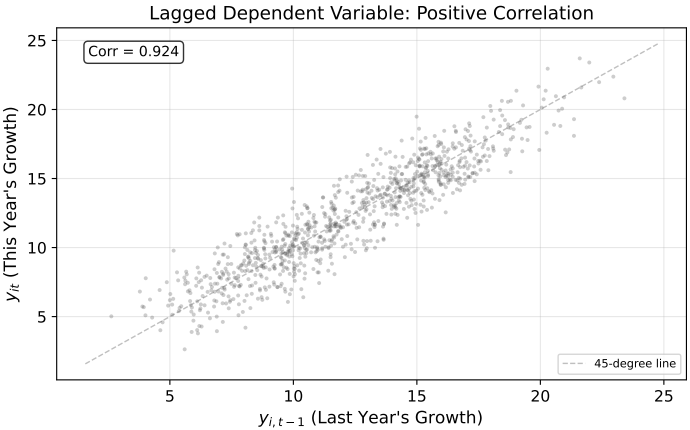
The parameter ρ measures **persistence**: how much of last year's growth carries forward. If $|\rho| < 1$, the process is stationary.

Goal: estimate ρ consistently. True value in our simulation: $\rho = 0.40$.

The Lagged Relationship

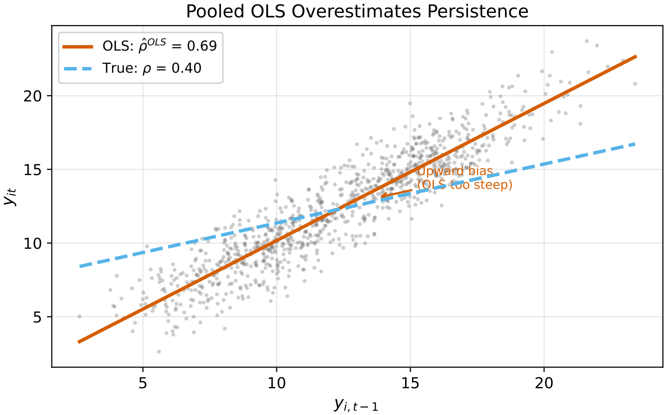


The Lagged Relationship

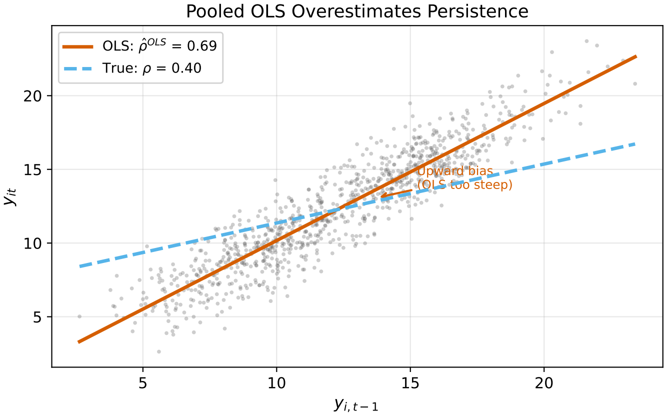


Strong positive correlation between y_{it} and $y_{i,t-1}$. But how much is true persistence, and how much is driven by the unobserved α_i?

Naive OLS: Just Regress y_{it} on $y_{i,t-1}$



Naive OLS: Just Regress y_{it} on $y_{i,t-1}$



OLS estimates $\hat{\rho}^{OLS} \approx 0.69$. The true ρ is 0.40. Why the overestimate?

Why OLS Fails: Omitted Variable Bias

Think of two countries:

Why OLS Fails: Omitted Variable Bias

Think of two countries:

- **Country A** has strong institutions (α_i high). Growth is *always* high. Both $y_{i,t-1}$ and y_{it} are large.

Why OLS Fails: Omitted Variable Bias

Think of two countries:

- **Country A** has strong institutions (α_i high). Growth is *always* high. Both $y_{i,t-1}$ and y_{it} are large.
- **Country C** has weak institutions (α_i low). Growth is *always* low. Both $y_{i,t-1}$ and y_{it} are small.

Why OLS Fails: Omitted Variable Bias

Think of two countries:

- **Country A** has strong institutions (α_i high). Growth is *always* high. Both $y_{i,t-1}$ and y_{it} are large.
- **Country C** has weak institutions (α_i low). Growth is *always* low. Both $y_{i,t-1}$ and y_{it} are small.

OLS sees: high $y_{i,t-1}$ followed by high y_{it}, low followed by low. It concludes the relationship is very strong. But the correlation is driven by **permanent country differences**, not persistence.

Why OLS Fails: Omitted Variable Bias

Think of two countries:

- **Country A** has strong institutions (α_i high). Growth is *always* high. Both $y_{i,t-1}$ and y_{it} are large.
- **Country C** has weak institutions (α_i low). Growth is *always* low. Both $y_{i,t-1}$ and y_{it} are small.

OLS sees: high $y_{i,t-1}$ followed by high y_{it}, low followed by low. It concludes the relationship is very strong. But the correlation is driven by **permanent country differences**, not persistence.

Formally, $y_{i,t-1}$ depends on α_i (since $y_{i,t-1} = \rho\, y_{i,t-2} + \beta\, \text{invest}_{i,t-1} + \alpha_i + \varepsilon_{i,t-1}$), so:

$$\text{bias} = \frac{\text{Cov}(y_{i,t-1}, \alpha_i)}{\text{Var}(y_{i,t-1})} > 0$$

Why OLS Fails: Omitted Variable Bias

Think of two countries:

- **Country A** has strong institutions (α_i high). Growth is *always* high. Both $y_{i,t-1}$ and y_{it} are large.
- **Country C** has weak institutions (α_i low). Growth is *always* low. Both $y_{i,t-1}$ and y_{it} are small.

OLS sees: high $y_{i,t-1}$ followed by high y_{it}, low followed by low. It concludes the relationship is very strong. But the correlation is driven by **permanent country differences**, not persistence.

Formally, $y_{i,t-1}$ depends on α_i (since $y_{i,t-1} = \rho\, y_{i,t-2} + \beta\, \text{invest}_{i,t-1} + \alpha_i + \varepsilon_{i,t-1}$), so:

$$\text{bias} = \frac{\text{Cov}(y_{i,t-1}, \alpha_i)}{\text{Var}(y_{i,t-1})} > 0$$

$\Longrightarrow$ OLS **overestimates** ρ because it confuses level differences with persistence.

Outline

Fixed Effects: The Natural Fix?

FE removes α_i by demeaning. The within-transformed model:

$$\underbrace{y_{it} - \bar{y}_i}_{\ddot{y}_{it}} = \rho \underbrace{(y_{i,t-1} - \bar{y}_i)}_{\ddot{y}_{i,t-1}} + \beta \, \ddot{\text{invest}}_{it} + \underbrace{(\varepsilon_{it} - \bar{\varepsilon}_i)}_{\ddot{\varepsilon}_{it}}$$

Fixed Effects: The Natural Fix?

FE removes α_i by demeaning. The within-transformed model:

$$\underbrace{y_{it} - \bar{y}_i}_{\ddot{y}_{it}} = \rho \underbrace{(y_{i,t-1} - \bar{y}_i)}_{\ddot{y}_{i,t-1}} + \beta\, \ddot{\text{invest}}_{it} + \underbrace{(\varepsilon_{it} - \bar{\varepsilon}_i)}_{\ddot{\varepsilon}_{it}}$$

No more α_i. Problem solved?

Fixed Effects: The Natural Fix?

FE removes α_i by demeaning. The within-transformed model:

$$\underbrace{y_{it} - \bar{y}_i}_{\ddot{y}_{it}} = \rho \underbrace{(y_{i,t-1} - \bar{y}_i)}_{\ddot{y}_{i,t-1}} + \beta \, \ddot{\text{invest}}_{it} + \underbrace{(\varepsilon_{it} - \bar{\varepsilon}_i)}_{\ddot{\varepsilon}_{it}}$$

No more α_i. Problem solved?

Not quite. There is a subtle problem with demeaning when you have a lagged dependent variable.

Nickell Bias: Demeaning Creates New Correlation

The demeaned lag $\ddot{y}_{i,t-1} = y_{i,t-1} - \bar{y}_i$ is correlated with the demeaned error $\ddot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$.

Nickell Bias: Demeaning Creates New Correlation

The demeaned lag $\ddot{y}_{i,t-1} = y_{i,t-1} - \bar{y}_i$ is correlated with the demeaned error $\ddot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$.

Concrete example. Suppose $\varepsilon_{i,5}$ is a large positive shock.

- $\bar{y}_i$ includes $y_{i,5}$, which is inflated by $\varepsilon_{i,5} \implies \bar{y}_i$ goes **up**.
- Look at year 4: $\ddot{y}_{i,4} = y_{i,4} - \bar{y}_i$ goes **down** (the mean was inflated by the year-5 shock, but $y_{i,4}$ was not).
- Meanwhile $\ddot{\varepsilon}_{i,5} = \varepsilon_{i,5} - \bar{\varepsilon}_i$ goes **up** ($\varepsilon_{i,5}$ is positive).

Nickell Bias: Demeaning Creates New Correlation

The demeaned lag $\ddot{y}_{i,t-1} = y_{i,t-1} - \bar{y}_i$ is correlated with the demeaned error $\ddot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$.

Concrete example. Suppose $\varepsilon_{i,5}$ is a large positive shock.

- $\bar{y}_i$ includes $y_{i,5}$, which is inflated by $\varepsilon_{i,5} \implies \bar{y}_i$ goes **up**.
- Look at year 4: $\ddot{y}_{i,4} = y_{i,4} - \bar{y}_i$ goes **down** (the mean was inflated by the year-5 shock, but $y_{i,4}$ was not).
- Meanwhile $\ddot{\varepsilon}_{i,5} = \varepsilon_{i,5} - \bar{\varepsilon}_i$ goes **up** ($\varepsilon_{i,5}$ is positive).

$\implies$ Positive demeaned error at t=5 is associated with a negative demeaned lag at t=4. The correlation is **negative**, so FE is biased **downward**.

Nickell Bias: Demeaning Creates New Correlation

The demeaned lag $\ddot{y}_{i,t-1} = y_{i,t-1} - \bar{y}_i$ is correlated with the demeaned error $\ddot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$.

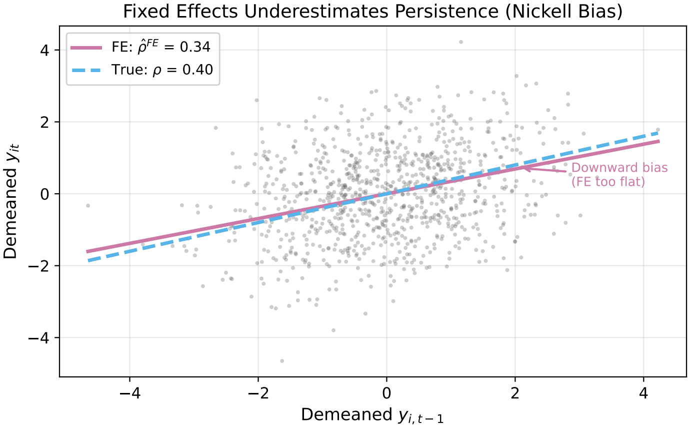
Concrete example. Suppose $\varepsilon_{i,5}$ is a large positive shock.

- $\bar{y}_i$ includes $y_{i,5}$, which is inflated by $\varepsilon_{i,5} \implies \bar{y}_i$ goes **up**.
- Look at year 4: $\ddot{y}_{i,4} = y_{i,4} - \bar{y}_i$ goes **down** (the mean was inflated by the year-5 shock, but $y_{i,4}$ was not).
- Meanwhile $\ddot{\varepsilon}_{i,5} = \varepsilon_{i,5} - \bar{\varepsilon}_i$ goes **up** ($\varepsilon_{i,5}$ is positive).

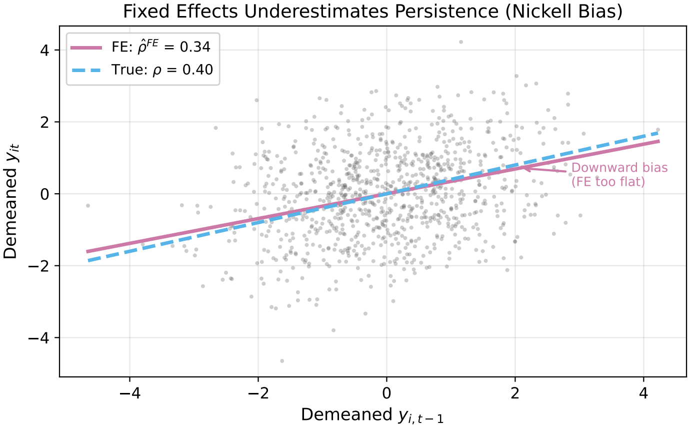
$\implies$ Positive demeaned error at t=5 is associated with a negative demeaned lag at t=4. The correlation is **negative**, so FE is biased **downward**.

This is the **Nickell bias** (Nickell, 1981). It arises purely from the mechanical relationship between demeaning and the lagged dependent variable.

FE Estimate: Biased Downward

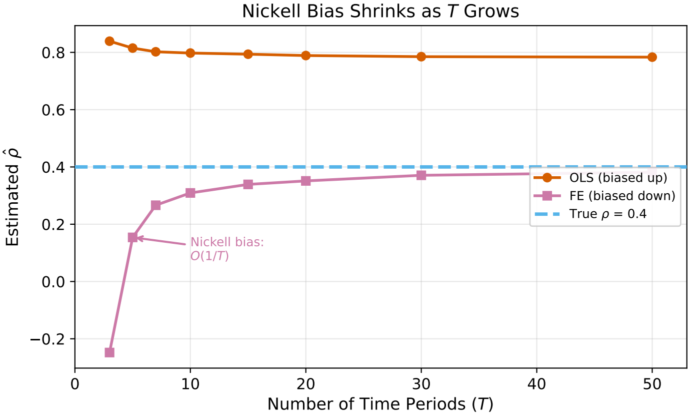


FE Estimate: Biased Downward

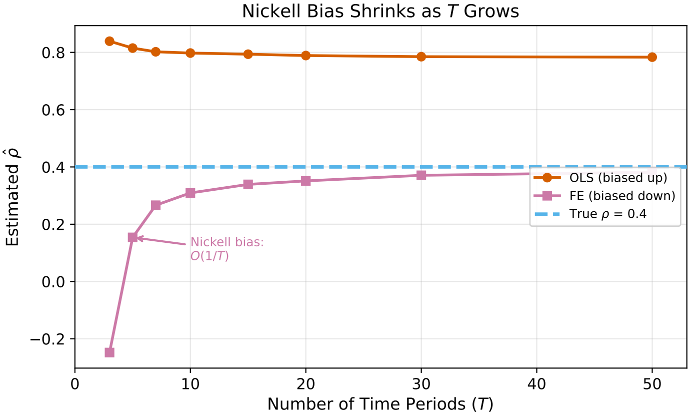


FE estimates $\hat{\rho}^{FE} \approx 0.34$, well below the true $\rho = 0.40$. FE overcorrects: it removes α_i but introduces a new bias in the opposite direction.

Nickell Bias Shrinks as T Grows



Nickell Bias Shrinks as T Grows



The FE bias is $O(1/T)$ (roughly proportional to $1/T$): it is severe for short panels ($T = 3, 5$) but shrinks as T grows. The OLS bias barely moves because it stems from cross-sectional heterogeneity, not from T.

The Bounds: OLS Up, FE Down

We now have two biased estimators that bracket the truth:

Estimator	$\hat{\rho}$	**Bias direction**
OLS	≈ 0.69	Upward (ignores α_i)
FE	≈ 0.34	Downward (Nickell bias)
True	**0.40**	

The Bounds: OLS Up, FE Down

We now have two biased estimators that bracket the truth:

Estimator	$\hat{\rho}$	**Bias direction**
OLS	≈ 0.69	Upward (ignores α_i)
FE	≈ 0.34	Downward (Nickell bias)
True	**0.40**	

In well-behaved dynamic panels, the true ρ lies **between** OLS and FE. This gives a useful sanity check for any estimator you apply.

The Bounds: OLS Up, FE Down

We now have two biased estimators that bracket the truth:

Estimator	$\hat{\rho}$	Bias direction
OLS	≈ 0.69	Upward (ignores α_i)
FE	≈ 0.34	Downward (Nickell bias)
True	**0.40**	

In well-behaved dynamic panels, the true ρ lies **between** OLS and FE. This gives a useful sanity check for any estimator you apply.

Both OLS and FE are biased. Is there a way out?

Two Problems, One Strategy

Where we stand:

- **Problem 1:** OLS ignores $\alpha_i \implies$ biased upward.
- **Problem 2:** FE removes α_i by demeaning, but demeaning creates Nickell bias $\implies$ biased downward.

Two Problems, One Strategy

Where we stand:

- **Problem 1:** OLS ignores $\alpha_i \implies$ biased upward.
- **Problem 2:** FE removes α_i by demeaning, but demeaning creates Nickell bias $\implies$ biased downward.

The root cause of Nickell bias is *demeaning*: it spreads every error across every time period. What if we eliminated α_i a different way?

Two Problems, One Strategy

Where we stand:

- **Problem 1:** OLS ignores $\alpha_i \implies$ biased upward.
- **Problem 2:** FE removes α_i by demeaning, but demeaning creates Nickell bias $\implies$ biased downward.

The root cause of Nickell bias is *demeaning*: it spreads every error across every time period. What if we eliminated α_i a different way?

The plan:

1. Eliminate α_i by **first-differencing** instead of demeaning.
2. First-differencing creates its own endogeneity problem, so find **valid instruments** for the resulting equation.

Two Problems, One Strategy

Where we stand:

- **Problem 1:** OLS ignores $\alpha_i \implies$ biased upward.
- **Problem 2:** FE removes α_i by demeaning, but demeaning creates Nickell bias $\implies$ biased downward.

The root cause of Nickell bias is *demeaning*: it spreads every error across every time period. What if we eliminated α_i a different way?

The plan:

1. Eliminate α_i by **first-differencing** instead of demeaning.
2. First-differencing creates its own endogeneity problem, so find **valid instruments** for the resulting equation.

$\implies$ This is the Arellano-Bond (1991) approach.

Outline

Step 1: First-Difference to Eliminate α_i

AB = Arellano-Bond. GMM = Generalized Method of Moments.

Instead of demeaning (which creates Nickell bias), **first-difference** the model:

$$\Delta y_{it} = \rho\,\Delta y_{i,t-1} + \beta\,\Delta \text{invest}_{it} + \Delta\varepsilon_{it}$$

Step 1: First-Difference to Eliminate α_i

AB = Arellano-Bond. GMM = Generalized Method of Moments.

Instead of demeaning (which creates Nickell bias), **first-difference** the model:

$$\Delta y_{it} = \rho\,\Delta y_{i,t-1} + \beta\,\Delta \text{invest}_{it} + \Delta\varepsilon_{it}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$.

Step 1: First-Difference to Eliminate α_i

AB = Arellano-Bond. GMM = Generalized Method of Moments.

Instead of demeaning (which creates Nickell bias), **first-difference** the model:

$$\Delta y_{it} = \rho \, \Delta y_{i,t-1} + \beta \, \Delta \text{invest}_{it} + \Delta \varepsilon_{it}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$.

The fixed effect α_i is gone (differenced out). But we still cannot run OLS on this equation.

Step 1: First-Difference to Eliminate α_i

AB = Arellano-Bond. GMM = Generalized Method of Moments.

Instead of demeaning (which creates Nickell bias), **first-difference** the model:

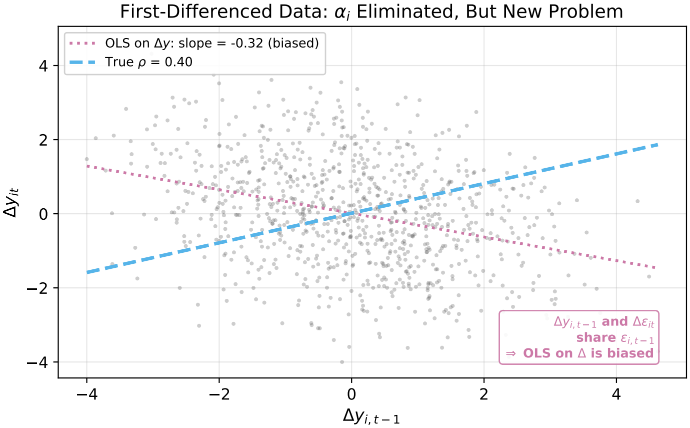
$$\Delta y_{it} = \rho\, \Delta y_{i,t-1} + \beta\, \Delta \text{invest}_{it} + \Delta\varepsilon_{it}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$.

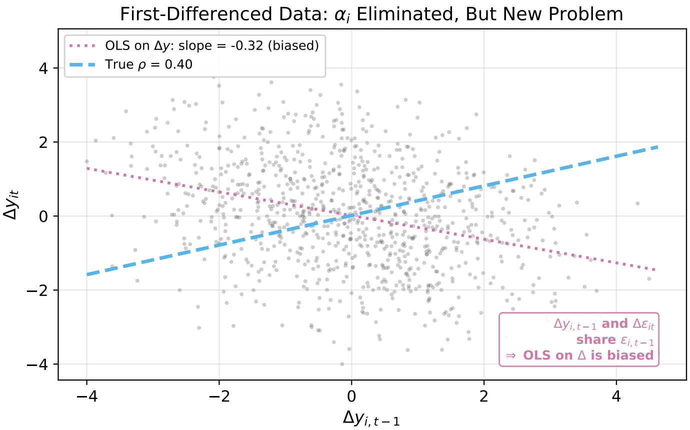
The fixed effect α_i is gone (differenced out). But we still cannot run OLS on this equation.

What goes wrong if we try?

OLS on First-Differenced Data: Still Wrong



OLS on First-Differenced Data: Still Wrong



OLS on the differenced data gives a slope of ≈ -0.32, wildly wrong (true $\rho = 0.40$). First-differencing removed α_i, but something else went wrong. What?

The Problem: $\Delta y_{i,t-1}$ and $\Delta \varepsilon_{it}$ Share a Term

Expand the terms:

$$\begin{aligned} \Delta y_{i,t-1} &= y_{i,t-1} - y_{i,t-2} \quad \text{contains } \varepsilon_{i,t-1} \\ \Delta \varepsilon_{it} &= \varepsilon_{it} - \varepsilon_{i,t-1} \quad \text{also contains } \varepsilon_{i,t-1} \end{aligned}$$

The Problem: $\Delta y_{i,t-1}$ and $\Delta\varepsilon_{it}$ Share a Term

Expand the terms:

$$\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2} \quad \text{contains } \varepsilon_{i,t-1}$$
$$\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1} \quad \text{also contains } \varepsilon_{i,t-1}$$

$\implies$ $\Delta y_{i,t-1}$ and $\Delta\varepsilon_{it}$ share $\varepsilon_{i,t-1}$, so:

$$\text{Cov}(\Delta y_{i,t-1}, \Delta\varepsilon_{it}) \neq 0$$

The Problem: $\Delta y_{i,t-1}$ and $\Delta \varepsilon_{it}$ Share a Term

Expand the terms:

$$\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2} \quad \text{contains } \varepsilon_{i,t-1}$$
$$\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1} \quad \text{also contains } \varepsilon_{i,t-1}$$

$\implies$ $\Delta y_{i,t-1}$ and $\Delta \varepsilon_{it}$ share $\varepsilon_{i,t-1}$, so:

$$\text{Cov}(\Delta y_{i,t-1}, \Delta \varepsilon_{it}) \neq 0$$

OLS on the first-differenced equation is **inconsistent**. We need an instrument for $\Delta y_{i,t-1}$.

The Problem: $\Delta y_{i,t-1}$ and $\Delta\varepsilon_{it}$ Share a Term

Expand the terms:

$$\begin{aligned} \Delta y_{i,t-1} &= y_{i,t-1} - y_{i,t-2} \quad \text{contains } \varepsilon_{i,t-1} \\ \Delta\varepsilon_{it} &= \varepsilon_{it} - \varepsilon_{i,t-1} \quad \text{also contains } \varepsilon_{i,t-1} \end{aligned}$$

$\implies$ $\Delta y_{i,t-1}$ and $\Delta\varepsilon_{it}$ share $\varepsilon_{i,t-1}$, so:

$$\text{Cov}(\Delta y_{i,t-1}, \Delta\varepsilon_{it}) \neq 0$$

OLS on the first-differenced equation is **inconsistent**. We need an instrument for $\Delta y_{i,t-1}$.

What variable is correlated with $\Delta y_{i,t-1}$ but uncorrelated with $\Delta\varepsilon_{it}$?

Step 2: Use $y_{i,t-2}$ as an Instrument

The **Arellano-Bond (1991)** insight: $y_{i,t-2}$ is a valid instrument for $\Delta y_{i,t-1}$.

Step 2: Use $y_{i,t-2}$ as an Instrument

The **Arellano-Bond (1991)** insight: $y_{i,t-2}$ is a valid instrument for $\Delta y_{i,t-1}$.

Relevance:

- $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$, and $y_{i,t-1}$ depends on $y_{i,t-2}$ through the dynamic model
- $\implies \text{Corr}(y_{i,t-2}, \Delta y_{i,t-1}) \neq 0$ ✓

Step 2: Use $y_{i,t-2}$ as an Instrument

The **Arellano-Bond (1991)** insight: $y_{i,t-2}$ is a valid instrument for $\Delta y_{i,t-1}$.

Relevance:

- $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$, and $y_{i,t-1}$ depends on $y_{i,t-2}$ through the dynamic model
- $\implies \text{Corr}(y_{i,t-2}, \Delta y_{i,t-1}) \neq 0$ ✓

Validity (exclusion restriction):

- $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$, which contains only period-t and period-$(t-1)$ errors
- $y_{i,t-2}$ depends on $\varepsilon_{i,t-2}$ and earlier, not on $\varepsilon_{i,t-1}$ or ε_{it}
- $\implies \text{Cov}(y_{i,t-2}, \Delta\varepsilon_{it}) = 0$ ✓ (assuming no serial correlation in ε)

Step 2: Use $y_{i,t-2}$ as an Instrument

The **Arellano-Bond (1991)** insight: $y_{i,t-2}$ is a valid instrument for $\Delta y_{i,t-1}$.

Relevance:

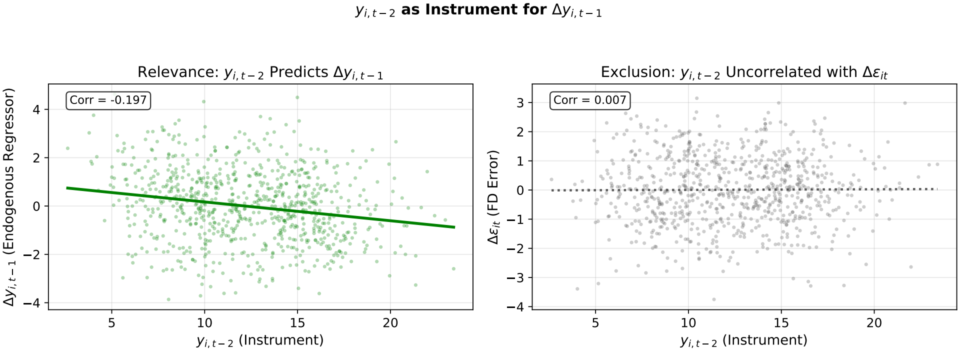
- $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$, and $y_{i,t-1}$ depends on $y_{i,t-2}$ through the dynamic model
- $\implies \text{Corr}(y_{i,t-2}, \Delta y_{i,t-1}) \neq 0$ ✓

Validity (exclusion restriction):

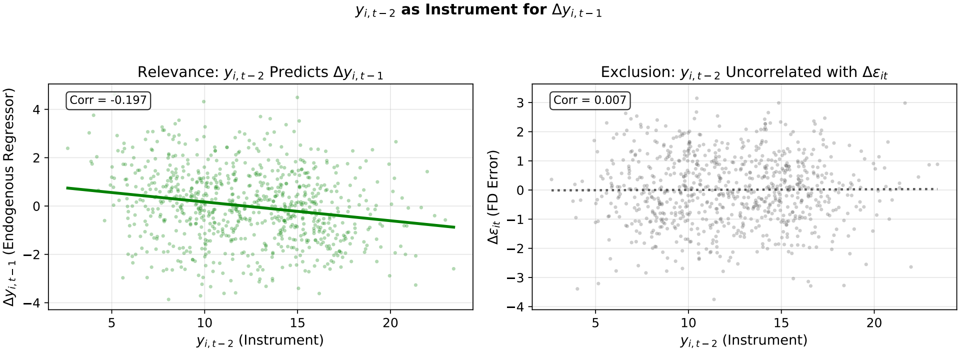
- $\Delta\varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$, which contains only period-t and period-$(t-1)$ errors
- $y_{i,t-2}$ depends on $\varepsilon_{i,t-2}$ and earlier, not on $\varepsilon_{i,t-1}$ or ε_{it}
- $\implies \text{Cov}(y_{i,t-2}, \Delta\varepsilon_{it}) = 0$ ✓ (assuming no serial correlation in ε)

This exclusion restriction is an **assumption**. It fails if the original errors ε_{it} are serially correlated, since then $y_{i,t-2}$ would correlate with $\varepsilon_{i,t-1}$ inside $\Delta\varepsilon_{it}$.

The Instrument in Action

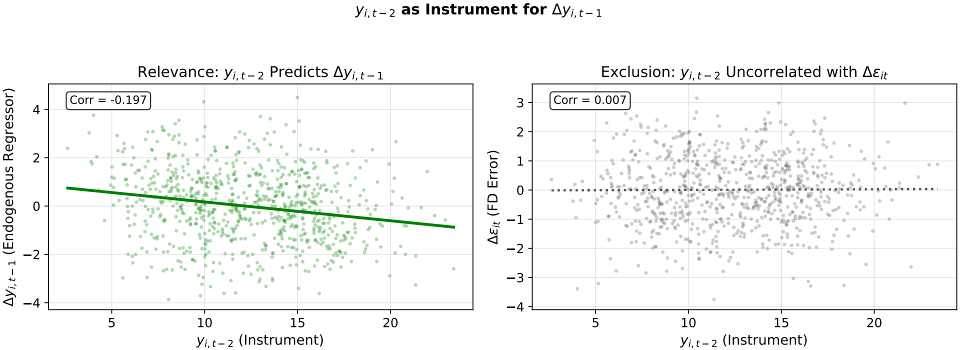


The Instrument in Action



Left: $y_{i,t-2}$ predicts $\Delta y_{i,t-1}$ (relevance). The correlation is negative because $y_{i,t-2}$ appears with a minus sign in $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$. Any nonzero correlation suffices for relevance.

The Instrument in Action



Left: $y_{i,t-2}$ predicts $\Delta y_{i,t-1}$ (relevance). The correlation is negative because $y_{i,t-2}$ appears with a minus sign in $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$. Any nonzero correlation suffices for relevance.

Right: $y_{i,t-2}$ is uncorrelated with $\Delta\varepsilon_{it}$ (exclusion). This single-period instrument looks weak, but AB uses many lags, which collectively provide strong identification.

Expanding the Instrument Set

The set of valid instruments **grows with** t:

Period	**Available instruments for** $\Delta y_{i,t-1}$
$t = 3$	$y_{i,1}$
$t = 4$	$y_{i,1},\ y_{i,2}$
$t = 5$	$y_{i,1},\ y_{i,2},\ y_{i,3}$
$\vdots$	$\vdots$
$t = T$	$y_{i,1},\ y_{i,2},\ \ldots,\ y_{i,T-2}$

Expanding the Instrument Set

The set of valid instruments **grows with** t:

Period	Available instruments for $\Delta y_{i,t-1}$
$t=3$	$y_{i,1}$
$t=4$	$y_{i,1},\ y_{i,2}$
$t=5$	$y_{i,1},\ y_{i,2},\ y_{i,3}$
$\vdots$	$\vdots$
$t=T$	$y_{i,1},\ y_{i,2},\ \ldots,\ y_{i,T-2}$

This creates many **moment conditions**. A moment condition is a restriction on population averages. Here, it says each instrument is uncorrelated with the error, the same logic as an IV exclusion restriction:

$$E[y_{i,s} \cdot \Delta\varepsilon_{it}] = 0 \quad \text{for all } s \leq t-2$$

Expanding the Instrument Set

The set of valid instruments **grows with** t:

Period	Available instruments for $\Delta y_{i,t-1}$
$t = 3$	$y_{i,1}$
$t = 4$	$y_{i,1},\ y_{i,2}$
$t = 5$	$y_{i,1},\ y_{i,2},\ y_{i,3}$
$\vdots$	$\vdots$
$t = T$	$y_{i,1},\ y_{i,2},\ \dots,\ y_{i,T-2}$

This creates many **moment conditions**. A moment condition is a restriction on population averages. Here, it says each instrument is uncorrelated with the error, the same logic as an IV exclusion restriction:

$$\mathrm{E}[y_{i,s} \cdot \Delta\varepsilon_{it}] = 0 \quad \text{for all } s \leq t-2$$

We need a way to combine all these moment conditions into a single estimate. That is what GMM does.

GMM: Combining Many Instruments

We have more instruments than parameters (the model is **overidentified**). Standard 2SLS can handle multiple instruments, but GMM provides optimal weighting when the number of moment conditions grows.

GMM: Combining Many Instruments

We have more instruments than parameters (the model is **overidentified**). Standard 2SLS can handle multiple instruments, but GMM provides optimal weighting when the number of moment conditions grows.

Intuition: GMM finds the $\hat{\rho}$ that makes all the instrument-error correlations as close to zero as possible.

GMM: Combining Many Instruments

We have more instruments than parameters (the model is **overidentified**). Standard 2SLS can handle multiple instruments, but GMM provides optimal weighting when the number of moment conditions grows.

Intuition: GMM finds the $\hat{\rho}$ that makes all the instrument-error correlations as close to zero as possible.

In Stata:

```
xtabond growth invest, lags(1) twostep
```

You specify the dependent variable, exogenous regressors, and the number of lags. The software constructs the full instrument matrix and solves the GMM problem.

GMM: Combining Many Instruments

We have more instruments than parameters (the model is **overidentified**). Standard 2SLS can handle multiple instruments, but GMM provides optimal weighting when the number of moment conditions grows.

Intuition: GMM finds the $\hat{\rho}$ that makes all the instrument-error correlations as close to zero as possible.

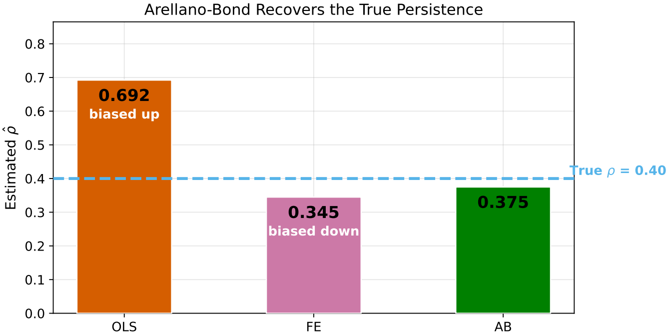
In Stata:

```
xtabond growth invest, lags(1) twostep
```

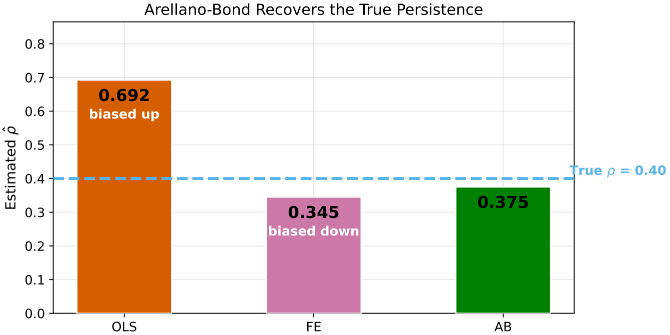
You specify the dependent variable, exogenous regressors, and the number of lags. The software constructs the full instrument matrix and solves the GMM problem.

Note: "One-step" GMM uses a preliminary weighting matrix; "two-step" re-estimates with an optimal weighting matrix from step-1 residuals. Two-step is more efficient.

The Result: AB Recovers the True ρ

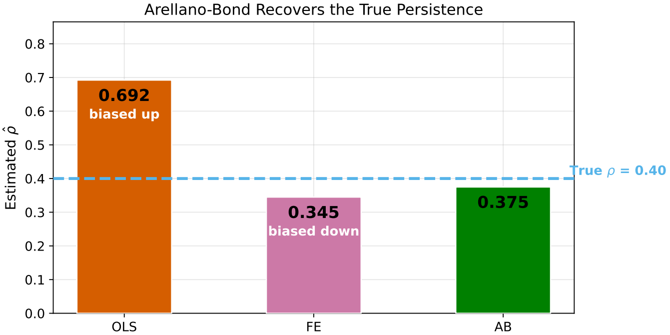


The Result: AB Recovers the True ρ



AB estimates $\hat{\rho}^{AB} \approx 0.37$, close to the true $\rho = 0.40$. It falls between the OLS upper bound (0.69) and the FE lower bound (0.34), as expected.

The Result: AB Recovers the True ρ



AB estimates $\hat{\rho}^{AB} \approx 0.37$, close to the true $\rho = 0.40$. It falls between the OLS upper bound (0.69) and the FE lower bound (0.34), as expected.

$\implies$ By first-differencing (to remove α_i) and instrumenting (to handle the $\Delta\varepsilon_{it}$ correlation), AB produces a consistent estimator.

Checking the AB Assumptions

AB relies on **no serial correlation** in ε_{it}. Two standard tests:

Checking the AB Assumptions

AB relies on **no serial correlation** in ε_{it}. Two standard tests:

1. AR(1) and AR(2) tests on $\Delta\varepsilon_{it}$:

- AR(1) in $\Delta\varepsilon$: we *expect* this (by construction, $\Delta\varepsilon_{it}$ and $\Delta\varepsilon_{i,t-1}$ share $\varepsilon_{i,t-1}$)
- AR(2) in $\Delta\varepsilon$: we should *not* find this. Why? $\Delta\varepsilon_{it}$ and $\Delta\varepsilon_{i,t-2}$ share no terms unless ε_{it} itself is autocorrelated. So AR(2) in the differenced errors is really a test for AR(1) in the original errors.

Checking the AB Assumptions

AB relies on **no serial correlation** in ε_{it}. Two standard tests:

1. AR(1) and AR(2) tests on $\Delta\varepsilon_{it}$:

- AR(1) in $\Delta\varepsilon$: we *expect* this (by construction, $\Delta\varepsilon_{it}$ and $\Delta\varepsilon_{i,t-1}$ share $\varepsilon_{i,t-1}$)
- AR(2) in $\Delta\varepsilon$: we should *not* find this. Why? $\Delta\varepsilon_{it}$ and $\Delta\varepsilon_{i,t-2}$ share no terms unless ε_{it} itself is autocorrelated. So AR(2) in the differenced errors is really a test for AR(1) in the original errors.

2. Sargan/Hansen test of overidentifying restrictions:

- We have more instruments than strictly needed. The Sargan test checks whether they all agree. If some instruments give very different answers, at least one is probably invalid.
- H_0: all instruments are valid. Rejection $\Longrightarrow$ at least one instrument is correlated with the error.

Checking the AB Assumptions

AB relies on **no serial correlation** in ε_{it}. Two standard tests:

1. AR(1) and AR(2) tests on $\Delta\varepsilon_{it}$:

- AR(1) in $\Delta\varepsilon$: we *expect* this (by construction, $\Delta\varepsilon_{it}$ and $\Delta\varepsilon_{i,t-1}$ share $\varepsilon_{i,t-1}$)
- AR(2) in $\Delta\varepsilon$: we should *not* find this. Why? $\Delta\varepsilon_{it}$ and $\Delta\varepsilon_{i,t-2}$ share no terms unless ε_{it} itself is autocorrelated. So AR(2) in the differenced errors is really a test for AR(1) in the original errors.

2. Sargan/Hansen test of overidentifying restrictions:

- We have more instruments than strictly needed. The Sargan test checks whether they all agree. If some instruments give very different answers, at least one is probably invalid.
- H_0: all instruments are valid. Rejection $\implies$ at least one instrument is correlated with the error.

$\implies$ Report both tests. If AR(2) is significant or Sargan rejects, the AB assumptions are in doubt.

Diagnostics: Our Simulation

In our simulated data (ε_{it} i.i.d. by construction), the tests should confirm AB is valid:

Diagnostics: Our Simulation

In our simulated data (ε_{it} i.i.d. by construction), the tests should confirm AB is valid:

Test	H_0	**Expected result**	**Conclusion**
AR(1) in $\Delta\varepsilon$	No AR(1)	Reject (by construction)	Normal
AR(2) in $\Delta\varepsilon$	No AR(2)	Fail to reject	Instruments valid
Sargan/Hansen	All IVs valid	Fail to reject	Instruments valid

Diagnostics: Our Simulation

In our simulated data (ε_{it} i.i.d. by construction), the tests should confirm AB is valid:

Test	H_0	**Expected result**	**Conclusion**
AR(1) in $\Delta\varepsilon$	No AR(1)	Reject (by construction)	Normal
AR(2) in $\Delta\varepsilon$	No AR(2)	Fail to reject	Instruments valid
Sargan/Hansen	All IVs valid	Fail to reject	Instruments valid

Since we generated the data with serially uncorrelated errors, we know the AB assumptions hold. In real applications, you must rely on these tests.

Diagnostics: Our Simulation

In our simulated data (ε_{it} i.i.d. by construction), the tests should confirm AB is valid:

Test	H_0	**Expected result**	**Conclusion**
AR(1) in $\Delta\varepsilon$	No AR(1)	Reject (by construction)	Normal
AR(2) in $\Delta\varepsilon$	No AR(2)	Fail to reject	Instruments valid
Sargan/Hansen	All IVs valid	Fail to reject	Instruments valid

Since we generated the data with serially uncorrelated errors, we know the AB assumptions hold. In real applications, you must rely on these tests.

$\Longrightarrow$ Always report AR(2) and Sargan/Hansen p-values alongside your AB estimates.

When Does AB Struggle?

AB uses only the **first-differenced equation**, instrumented with lagged levels. This works well when ρ is moderate.

When Does AB Struggle?

AB uses only the **first-differenced equation**, instrumented with lagged levels. This works well when ρ is moderate.

But when ρ is close to 1, AB runs into trouble. Why?

When Does AB Struggle?

AB uses only the **first-differenced equation**, instrumented with lagged levels. This works well when ρ is moderate.

But when ρ is close to 1, AB runs into trouble. Why?

- When $\rho \approx 1$, the series behaves like a random walk: levels barely change over time.
- Past levels carry almost no information about *changes* $\Longrightarrow$ the instruments (lagged levels) are **weak**.
- Weak instruments produce imprecise, unreliable estimates.

When Does AB Struggle?

AB uses only the **first-differenced equation**, instrumented with lagged levels. This works well when ρ is moderate.

But when ρ is close to 1, AB runs into trouble. Why?

- When $\rho \approx 1$, the series behaves like a random walk: levels barely change over time.
- Past levels carry almost no information about *changes* $\implies$ the instruments (lagged levels) are **weak**.
- Weak instruments produce imprecise, unreliable estimates.

$\implies$ We need additional instruments. Where do they come from?

System GMM: Blundell-Bond

Blundell and Bond (1998) proposed **System GMM**: stack two equations:

1. First-differenced equation (as in AB), instrumented with lagged *levels*
2. Levels equation, instrumented with lagged *differences*

System GMM: Blundell-Bond

Blundell and Bond (1998) proposed **System GMM**: stack two equations:

1. First-differenced equation (as in AB), instrumented with lagged *levels*
2. Levels equation, instrumented with lagged *differences*

The extra moment conditions improve efficiency, especially when:

- ρ is close to 1 (near unit root)
- T is small
- The variance of α_i is large relative to ε_{it}

System GMM: Blundell-Bond

Blundell and Bond (1998) proposed **System GMM**: stack two equations:

1. First-differenced equation (as in AB), instrumented with lagged *levels*
2. Levels equation, instrumented with lagged *differences*

The extra moment conditions improve efficiency, especially when:

- ρ is close to 1 (near unit root)
- T is small
- The variance of α_i is large relative to ε_{it}

In Stata: `xtabond2 growth L.growth invest, gmm(L.growth) iv(invest) twostep`

System GMM: Blundell-Bond

Blundell and Bond (1998) proposed **System GMM**: stack two equations:

1. First-differenced equation (as in AB), instrumented with lagged *levels*
2. Levels equation, instrumented with lagged *differences*

The extra moment conditions improve efficiency, especially when:

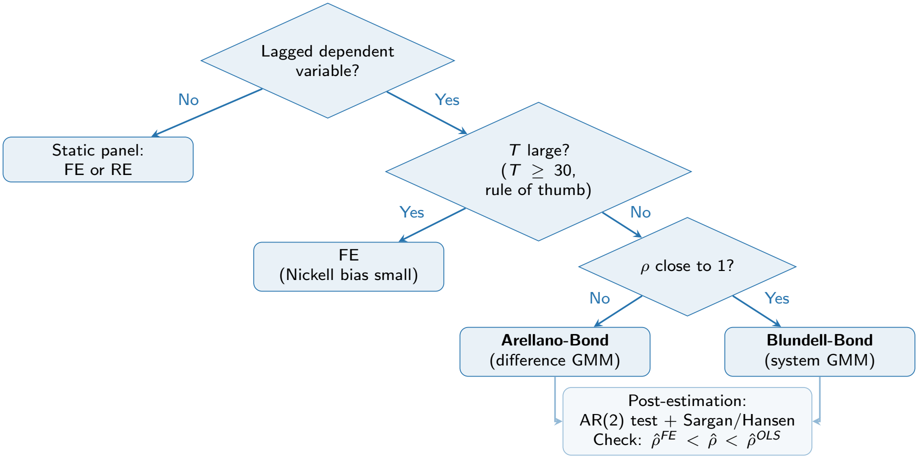
- ρ is close to 1 (near unit root)
- T is small
- The variance of α_i is large relative to ε_{it}

In Stata: `xtabond2 growth L.growth invest, gmm(L.growth) iv(invest) twostep`

$\implies$ System GMM adds an assumption: $\mathrm{E}[\Delta y_{i,t} \cdot \alpha_i] = 0$ (each country's growth has settled to its long-run average by the start of the sample). This is testable via the difference-in-Sargan test.

Outline

Decision Flowchart



Check whether $y_{i,t-1}$ is in the model; if T is small, choose AB or BB based on persistence; always run post-estimation diagnostics.

Outline

Summary

We started with a panel of 50 countries and wanted to estimate the persistence of GDP growth. Both OLS and FE are biased: OLS overestimates, FE underestimates.

Summary

We started with a panel of 50 countries and wanted to estimate the persistence of GDP growth. Both OLS and FE are biased: OLS overestimates, FE underestimates.

1. **OLS** ignores $\alpha_i \implies$ confuses level differences with persistence $\implies \hat{\rho}^{OLS}$ biased **upward**.

Summary

We started with a panel of 50 countries and wanted to estimate the persistence of GDP growth. Both OLS and FE are biased: OLS overestimates, FE underestimates.

1. **OLS** ignores $\alpha_i \implies$ confuses level differences with persistence $\implies \hat{\rho}^{OLS}$ biased **upward**.
2. **FE** removes α_i by demeaning, but demeaning creates a mechanical correlation between $\ddot{y}_{i,t-1}$ and $\ddot{\varepsilon}_{it} \implies \hat{\rho}^{FE}$ biased **downward** (Nickell bias, $O(1/T)$).

Summary

We started with a panel of 50 countries and wanted to estimate the persistence of GDP growth. Both OLS and FE are biased: OLS overestimates, FE underestimates.

1. **OLS** ignores α_i $\implies$ confuses level differences with persistence $\implies$ $\hat{\rho}^{OLS}$ biased **upward**.
2. **FE** removes α_i by demeaning, but demeaning creates a mechanical correlation between $\ddot{y}_{i,t-1}$ and $\ddot{\varepsilon}_{it}$ $\implies$ $\hat{\rho}^{FE}$ biased **downward** (Nickell bias, $O(1/T)$).
3. **Arellano-Bond** first-differences to remove α_i, then uses $y_{i,t-2}$ (and deeper lags) as instruments for $\Delta y_{i,t-1}$. Consistent when ε_{it} is serially uncorrelated.

Summary

We started with a panel of 50 countries and wanted to estimate the persistence of GDP growth. Both OLS and FE are biased: OLS overestimates, FE underestimates.

1. **OLS** ignores $\alpha_i \implies$ confuses level differences with persistence $\implies \hat{\rho}^{OLS}$ biased **upward**.
2. **FE** removes α_i by demeaning, but demeaning creates a mechanical correlation between $\ddot{y}_{i,t-1}$ and $\ddot{\varepsilon}_{it} \implies \hat{\rho}^{FE}$ biased **downward** (Nickell bias, $O(1/T)$).
3. **Arellano-Bond** first-differences to remove α_i, then uses $y_{i,t-2}$ (and deeper lags) as instruments for $\Delta y_{i,t-1}$. Consistent when ε_{it} is serially uncorrelated.
4. **Blundell-Bond** (System GMM) adds moment conditions from the levels equation. Preferable when ρ is close to 1 or T is very small.

Summary

We started with a panel of 50 countries and wanted to estimate the persistence of GDP growth. Both OLS and FE are biased: OLS overestimates, FE underestimates.

1. **OLS** ignores $\alpha_i \implies$ confuses level differences with persistence $\implies \hat{\rho}^{OLS}$ biased **upward**.
2. **FE** removes α_i by demeaning, but demeaning creates a mechanical correlation between $\ddot{y}_{i,t-1}$ and $\ddot{\varepsilon}_{it} \implies \hat{\rho}^{FE}$ biased **downward** (Nickell bias, $O(1/T)$).
3. **Arellano-Bond** first-differences to remove α_i, then uses $y_{i,t-2}$ (and deeper lags) as instruments for $\Delta y_{i,t-1}$. Consistent when ε_{it} is serially uncorrelated.
4. **Blundell-Bond** (System GMM) adds moment conditions from the levels equation. Preferable when ρ is close to 1 or T is very small.
5. **Sanity check:** a credible estimate should satisfy $\hat{\rho}^{FE} < \hat{\rho} < \hat{\rho}^{OLS}$.

Summary

We started with a panel of 50 countries and wanted to estimate the persistence of GDP growth. Both OLS and FE are biased: OLS overestimates, FE underestimates.

1. **OLS** ignores α_i $\implies$ confuses level differences with persistence $\implies$ $\hat{\rho}^{OLS}$ biased **upward**.
2. **FE** removes α_i by demeaning, but demeaning creates a mechanical correlation between $\ddot{y}_{i,t-1}$ and $\ddot{\varepsilon}_{it}$ $\implies$ $\hat{\rho}^{FE}$ biased **downward** (Nickell bias, $O(1/T)$).
3. **Arellano-Bond** first-differences to remove α_i, then uses $y_{i,t-2}$ (and deeper lags) as instruments for $\Delta y_{i,t-1}$. Consistent when ε_{it} is serially uncorrelated.
4. **Blundell-Bond** (System GMM) adds moment conditions from the levels equation. Preferable when ρ is close to 1 or T is very small.
5. **Sanity check:** a credible estimate should satisfy $\hat{\rho}^{FE} < \hat{\rho} < \hat{\rho}^{OLS}$.

$\implies$ When your panel model has a lagged dependent variable and small T, use AB or System GMM, not OLS or FE. First-difference to remove the fixed effect, then instrument to restore consistency.

Thank you!

jakeanderson@g.ucla.edu