DYNAMIC PANEL DATA MODELS IN PYTHON

FROM NICKELL BIAS TO SYSTEM GMM: A PRACTITIONER'S STUDY GUIDE

CORE FOCUS:

- Methodological foundations, diagnostic workflows, and Python implementation for estimating dynamic labor-demand equations.

DATASET CONTEXT:

Arellano and Bond (1991) UK Manufacturing Panel (N=140, T=7–9, 1,031 firm-years).

The Core Problem: Measuring Persistence

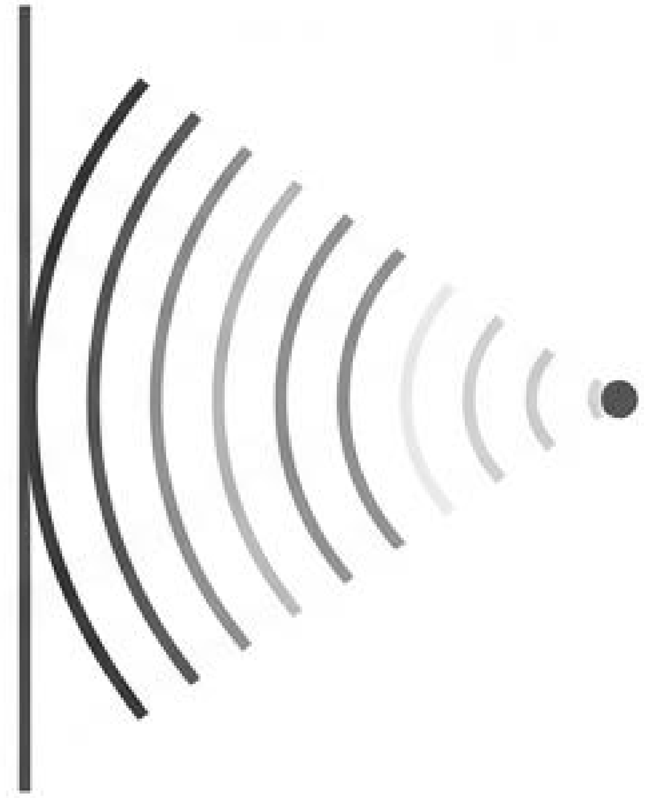
$$n_{it} = \rho n_{i,t-1} + \beta_1 w_{it} + \beta_2 w_{i,t-1} + \beta_3 k_{it} + \beta_4 k_{i,t-1} + \alpha_i + \delta_t + \varepsilon_{it}$$

Log employment, real wage, capital.

Unobserved permanent firm fixed effect.

The persistence parameter (The core focus of this deck).

Mental Model: **The Canyon Echo**



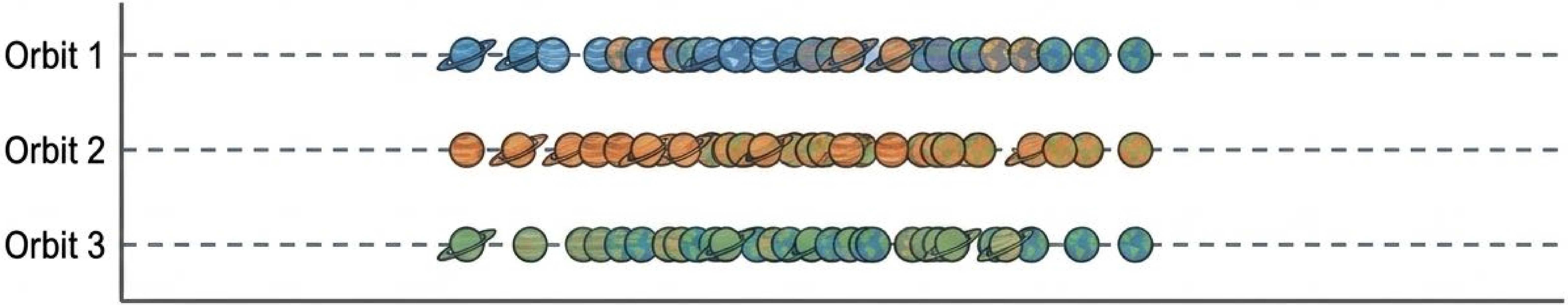
ρ is the echo strength of the labor market. If $\rho = 0.6$, shocks fade fast. If $\rho = 0.95$, the echo barely decays—what happens to a firm in 1980 is still audible in 1988.

The Endogeneity Trap (α_i)

- **The Nuisance**: The firm fixed effect (α_i) captures permanent traits like management quality or niche.
- **The Trap**: We put a lagged dependent variable ($n_{i,t-1}$) on the right-hand side, while α_i sits in the error term. $n_{i,t-1}$ is mechanically correlated with α_i because a firm with a high permanent level had high employment last year.

Data Reality: α_i dominates	
Between-firm SD:	1.339
Within-firm SD:	0.195
Firm differences are 7x larger than internal variance.	

Mental Model: Planets in Different Orbits



Regressing position on lagged position without removing α_i just measures which orbit the planet is in, mistaking orbit differences for persistence.

The Bias Bracket (Bond 2002)

Pooled OLS: $\hat{\rho} = 0.962$

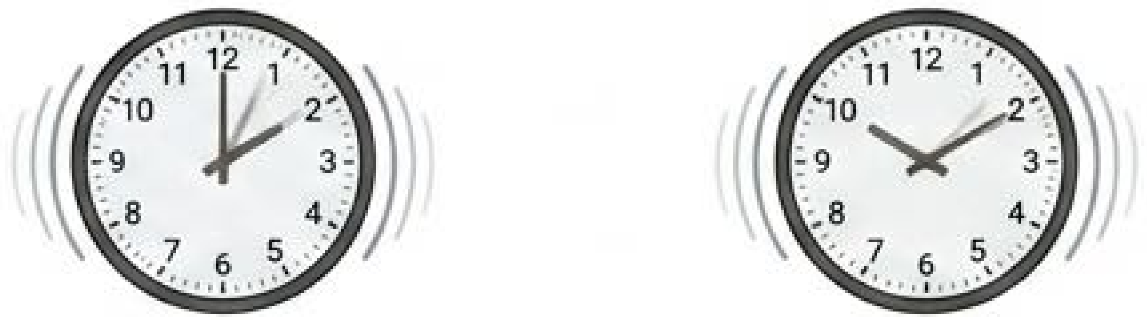
Biased UP (Ignores α_i: rewards the lag for work the firm effect is doing).

Yardstick

Fixed Effects: $\hat{\rho} = 0.626$

Biased DOWN (Nickell Bias: demeaning subtracts future shocks, inducing negative correlation of order 1/T).

Mental Model: Two Broken Clocks



Fast Clock

Slow Clock

Neither tells the time, but the true time is trapped between them. Any consistent estimator hugging the edges is highly suspicious.

The First Escape Route: Anderson-Hsiao IV

Step 1: Difference away α_i

$$\Delta n_{it} = \rho\, \Delta n_{i,t-1} + \cdots + \Delta\varepsilon_{it}$$

The constant α_i vanishes! But $\Delta n_{i,t-1}$ contains $\varepsilon_{i,t-1}$, which is also inside $\Delta\varepsilon_{it}$. It is still endogenous.

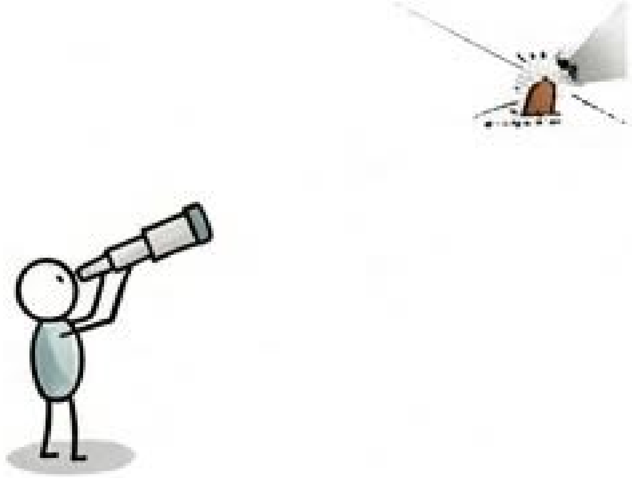
Step 2: Instrument

Use the level $n_{i,t-2}$. It predicts the change but predates the shock (Sequential Exogeneity).

The Result: $\hat{\rho} = 1.233$ (SE 0.478)

Verdict: Consistent but practically powerless. The CI width is 1.87 and spans [0.296, 2.170]. It is **explosive** and outside the bracket.

Mental Model: The Lonely Witness



The testimony is valid, but one witness seeing a persistent series from far away extracts far too little information to secure a it to secure a tight estimate.

Difference GMM (Arellano-Bond 1991)

The Insight:

If t-2 is valid, so is t-3, t-4, etc.

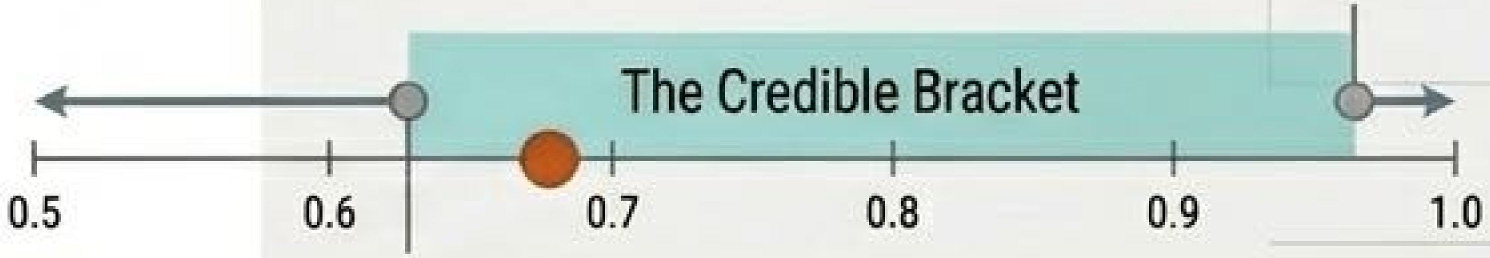
The Moment Condition:

$$E[n_{i,t-s}\,\Delta\varepsilon_{it}] = 0 \text{ for all } s \geq 2$$

GMM weights dozens of these zero-correlation conditions optimally.

$$\hat{\rho} = 0.679 \text{ (SE 0.089)}$$

Uses 91 instruments. AR(2) and Hansen J tests pass perfectly.



The Weak Instrument Trap: The estimate (0.679) sits just 0.053 above the Fixed Effects bound. Why? Because past levels barely predict future changes when ρ is near 1 (a near-random walk).

Mental Model:
A Courtroom of 40 Vague Witnesses

Adding 40 witnesses who only caught a vague glimpse adds degrees of freedom, but no real identifying power.

System GMM (Blundell-Bond 1998)

The System

- **Difference Equation** (Instrumented with lags)
- **Level Equation** (Instrumented with lagged differences)

$\hat{\rho} = 0.927$ (SE 0.079)

Uses 32 collapsed instruments.
AR(2) p=0.994, Hansen p=0.462

Comfortably inside the top half of the bracket.

- **The Fix:** If lagged levels weakly predict differences, use lagged differences to predict levels. Stack both equations to extract maximum information.
- **New Moment Condition:** $E[\Delta n_{i,t-1}(\alpha_i + \varepsilon_{it})] = 0$
- **Crucial Assumption:** Mean Stationarity. Firms' initial deviations from steady states are uncorrelated with their permanent levels α_i.

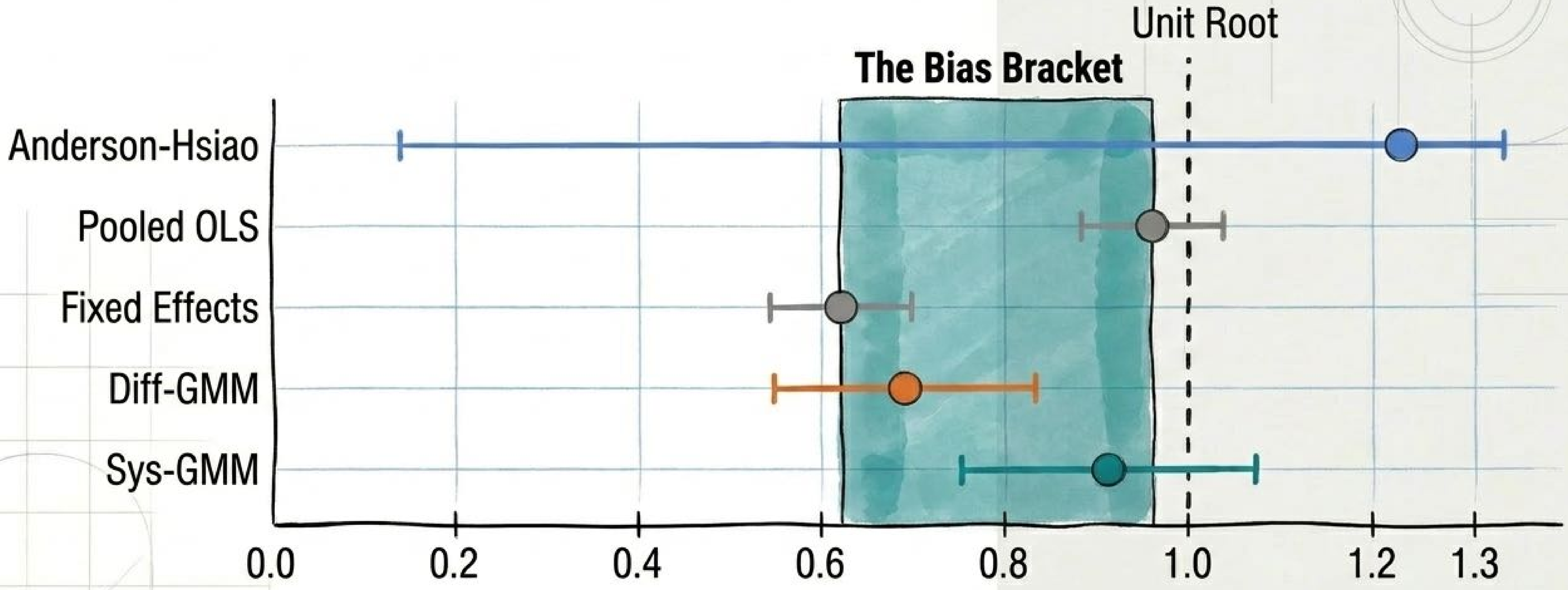
Mental Model: Ripples vs. Waterline

Learning a calm lake's depth from ripples alone is hard. Using the permanent waterline marks (levels) pins the answer down.

The Estimator Ladder (Synthesis Matrix)

Estimator	Treatment of α_i	Instrument Set	Result ($\hat{\rho}$)	Verdict
Pooled OLS	Ignored	None	0.962	Biased Up
Fixed Effects	Demeaned	None	0.626	Biased Down (Nickell, 1/T)
Anderson-Hsiao	Differenced	$n_{i,t-2}$	1.233	Powerless (Explosive CI)
Difference GMM	Differenced	All past levels	0.679	Weak Instruments (Hugs FE bound)
System GMM	Stacked	Levels + Differences	**0.927**	**Defensible Winner**

Visual Synthesis: The Forest Plot



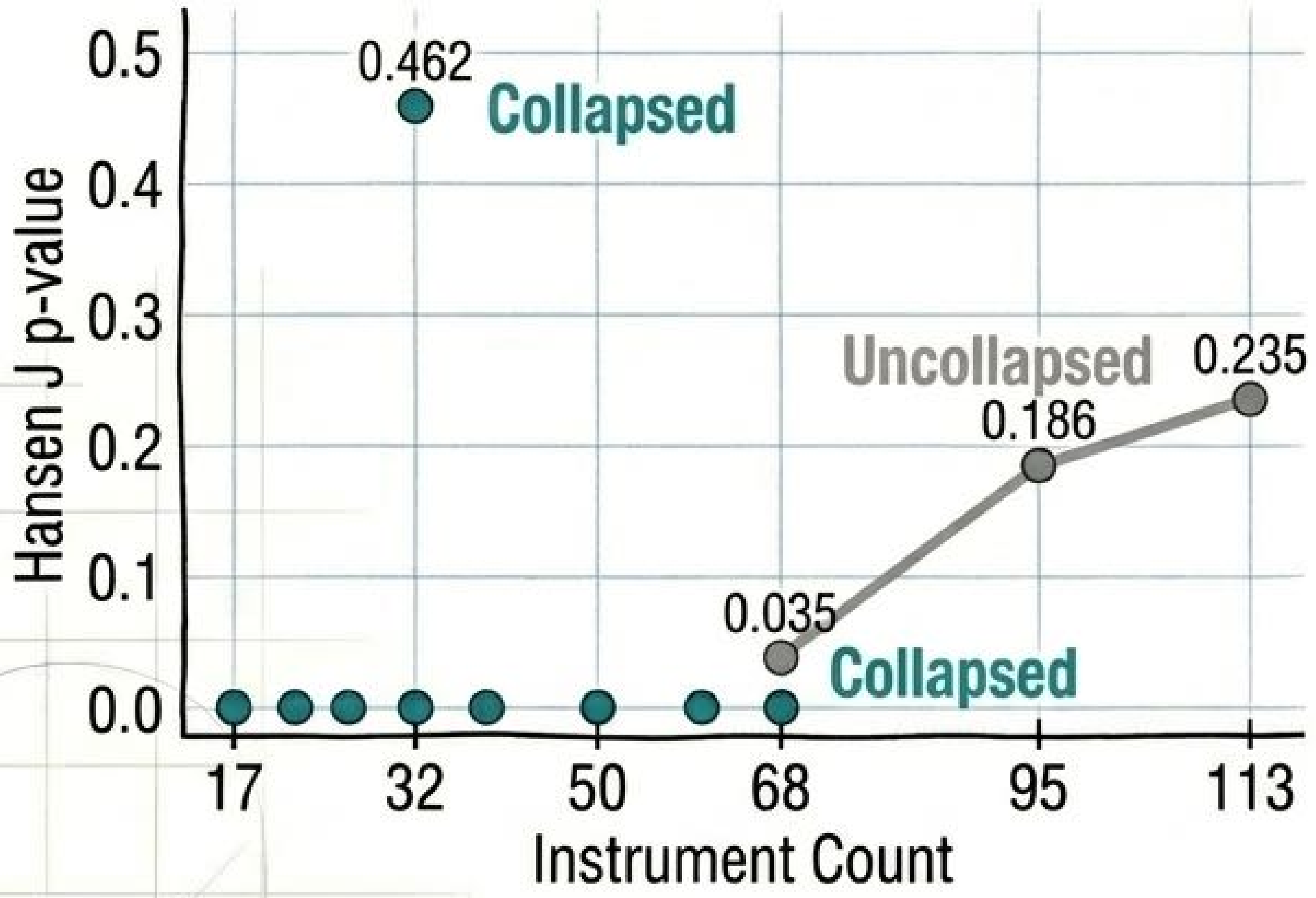
Core Insight: No single printed p-value separates the winner from the losers. Only the bracket logic and diagnostic workflow identify 0.927 as the true parameter.

Diagnostic Decoder Cheat Sheet

	AR(1) in Differences	• **Checks:** 1st-order serial correlation in $\Delta\varepsilon_{it}$. • **Rule:** MUST REJECT ($p < 0.05$). Adjacent errors mechanically overlap. Rejection is good news. • **Headline Result: z=-4.49, p=0.000 (Pass).**
STOP	**AR(2) in Differences**	• **Checks:** 2nd-order correlation (validates t-2 instruments). • **Rule:** MUST NOT REJECT. A rejection here means instruments are contaminated. • **Headline Result: z=-0.01, p=0.994 (Pass/Clean).**
	Hansen J Test	• **Checks:** Joint validity of overidentifying restrictions. • **Rule:** TWO-TAILED IN SPIRIT. $p < 0.05$ is invalid. But p approaching 1.0 means the test is overwhelmed by too many instruments. • **Headline Result: p=0.462 (Comfortable).**

The Danger of Instrument Proliferation

The Overfitting Drift



The Trap: “Use every lag” generates instruments quadratically (T^2). Too many overfit the endogenous variables and mathematically disarm the Hansen test, pushing the p-value falsely toward 1.0.

The Solution: Collapse the instrument matrix. Combines lags into one column per depth rather than one per time period.

Roodman’s Rule (2009): Keep total instrument count strictly below the number of groups (N=140). Our collapsed model uses 32.

Core Insight: High Hansen p-values often indicate overfitting from instrument proliferation. Collapsing the matrix is a critical robustness check to ensure the test’s validity.

Implementation Step 1: Data Prep & Baselines

Data Geometry: Short & Wide Panel (N=140, T=7-9).
Warning: Generating lags is expensive. 1 lag drops 13.6% of rows because there is no t=0 to look back to.

```
# Pooled OLS (Upper Bound)
feols("n ~ n_lag1 + w + w_lag1 + k + k_lag1 | year",
      data, vcov={"CRV1": "id"})

# Fixed Effects (Lower Bound)
feols("n ~ n_lag1 + w + w_lag1 + k + k_lag1 | id + year",
      data, vcov={"CRV1": "id"})
```

Cluster-robust standard errors are essential due to serial dependence witfirms.

Implementation Step 2: GMM Workflow in Python

```
# Tool: pydynpd (Syntax mimics Stata's xtabond2)
command = """n L1.n w L1.w k L1.k |
             gmm(n, 2:99, collapse)
             gmm(w, 2:99, collapse)
             gmm(k, 2:99, collapse) |
             timedumm"""

abond(command, df, ['id', 'year'])
```

Use **all available lags** starting from depth t-2 (Sequential Exogeneity).

Shrinks instrument matrix to solve the proliferation trap. Uses 32 instruments instead of 113.

Omitting `nolevel` forces the stacked System GMM instead of Difference GMM.

Note: The output always reports the 2-step Windmeijer finite-sample corrected standard errors.

The Practitioner's 7-Step Checklist

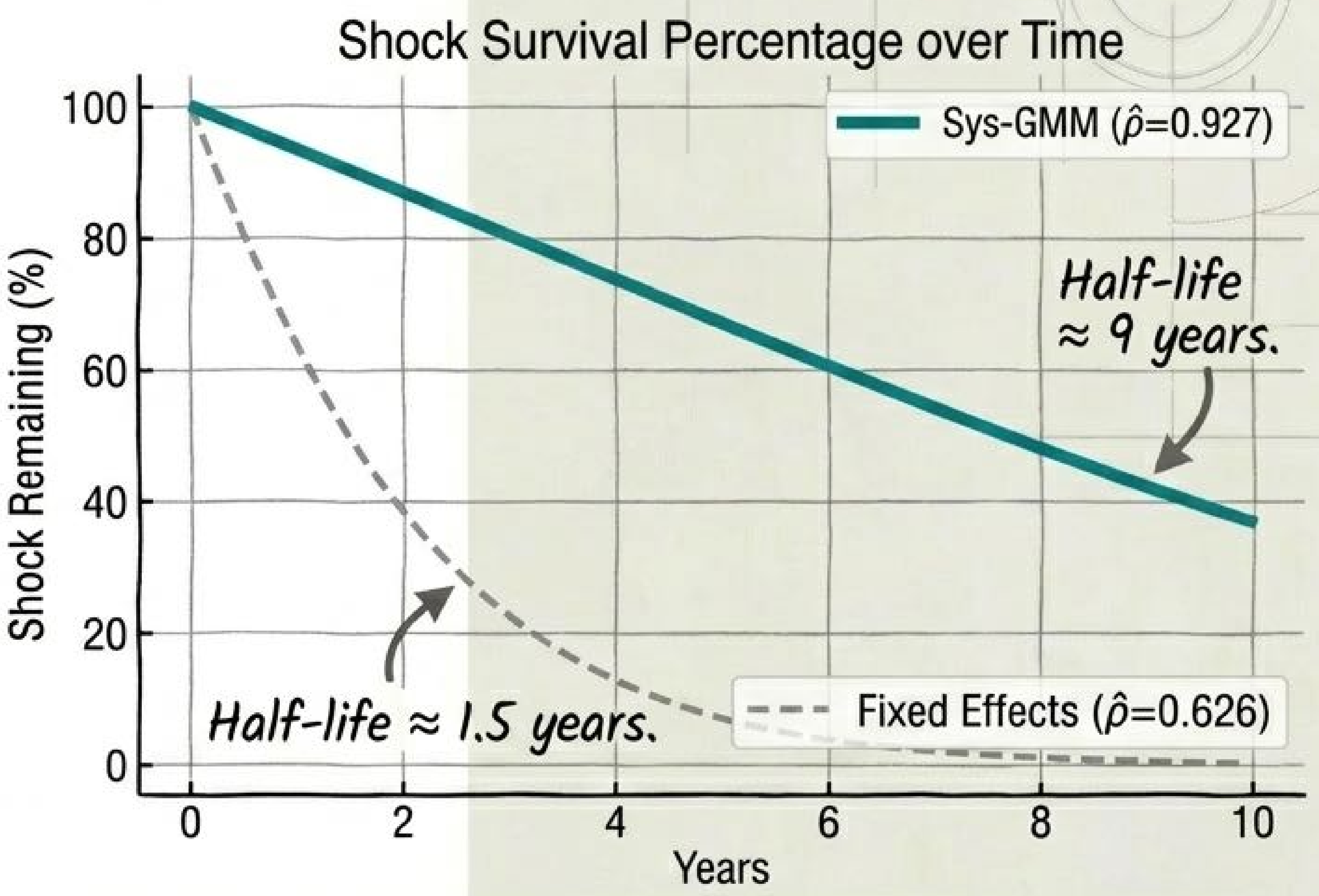
1. **Establish Bounds** — Run OLS (biased up) and FE (biased down) to define the bracket. ← *Crucial starting point!*
2. **Check Diff-GMM** — If it hugs the FE lower bound, suspect weak instruments.
3. **Upgrade to Sys-GMM** — Preferred for persistent series; verify the point estimate lands inside the bracket. ← *Standard practice!*
4. **Check AR Tests** — Ensure AR(1) mechanically rejects, and AR(2) does not reject.
5. **Check Hansen J** — Read two-tailed. Investigate values suspiciously approaching 1.0.
6. **Collapse Instruments** — Force instrument count < N. Accept larger standard errors as the price of mathematical honesty. ← *Tough but necessary!*
7. **Replicate Toolchains** — Always run a published benchmark dataset to verify software behavior.

Interpretation: What Does $\hat{\rho} = 0.927$ Mean?

Economic Reality vs. Bad Econometrics:

A policymaker using Fixed Effects assumes shocks evaporate in 1.5 years.

By using correct GMM, we see they linger for a decade—altering cost-benefit horizons by a factor of six.



Methodological Caveat: The 95% CI is [0.773, 1.081]. It includes the unit root. The defensible claim is the point estimate and lower bound, not strict stationarity.