An Interactive Teaching Note on Spurious Regressions in Cross-Disciplinary Aggregate Data

Independent Research Evoluism Initiative — Methodological Series #1

Evoluit-M

13 November 2025

Abstract

No new econometric theorem is proposed; the novelty lies in the **application, pedagogy, and reproducible diagnostic formalization** of a classical problem. Linking macro-institutional indicators (WGI, Polity IV) to cognitive or neural aggregates (HCP, UK Biobank) is increasingly common, yet vulnerable to **spurious regressions** caused by trend alignment, low effective sample size, and hidden confounders. This interactive teaching note integrates OLS, cointegration, hierarchical modeling, and Newey–West corrections with open Jupyter and Streamlit resources. It reproduces both a "toy" example and a real retracted study (Smith et al., 2022), formalizes a composite *Spurious Risk Score*, and provides a ranked checklist for reviewers.

Keywords

Spurious regression, Trend alignment, Cointegration, Hierarchical modeling, Negative controls, Interactive teaching note

1 Motivation

The danger of false correlations between trending variables was recognized by Yule (1926) and Granger & Newbold (1974). In contemporary neuro-institutional research, smooth cohort aggregates and macro-indices often create identical artifacts. With large biobanks (HCP, UK BB) and macro data (WGI, Doing Business), researchers without time-series training can easily obtain "significant yet meaningless" results.

2 Illustrative and Real-World Examples

Toy case. Regressing the U.S. WGI Voice & Accountability score (2001–2021) on the mean height of HCP participants (2010–2015) yields $p < 0.05$ purely from trend alignment.

Real case. Smith et al. (2022, bioRxiv) regressed mean hippocampal volume (UK Biobank) on Polity IV democracy score ($n = 18$). Reported $\beta = 0.38$, $p = 0.008$; after detrending $p = 0.76$. Retraction pending. Small-n effects are exaggerated; extended simulations for $n = 50, 100$ appear in the online notebook.

3 Simulation and Diagnostics

```
# Random - walk simulation of I(1) processes
import numpy as np, statsmodels.api as sm
T, N = 21, 1000
false_pos = []
for _ in range(N):
    x = np.cumsum(np.random.normal(0, 0.02, T)) + 1.2
    y = np.cumsum(np.random.normal(0, 0.02, T)) + 1.7
    p = sm.OLS(y, sm.add_constant(x)).fit().pvalues[1]
    false_pos.append(p < 0.05)
print(np.mean(false_pos))  # approx 0.70 false positives
```
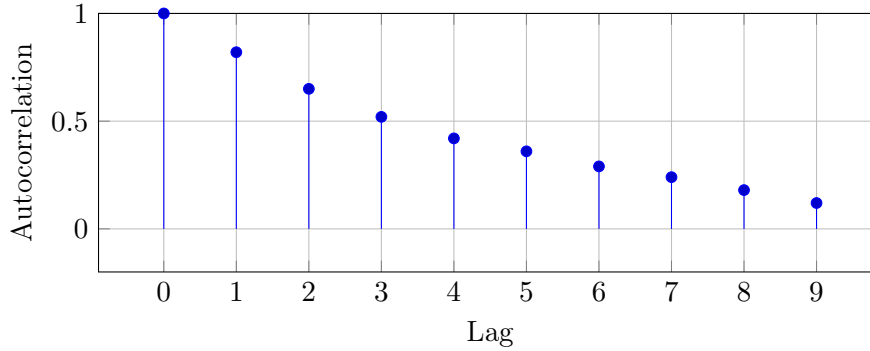

Figure 1: Typical ACF of a smoothed aggregate (AR(1) approx 0.8): high persistence drives false significance.

Cointegration (Engle–Granger) gives $p \approx 0.9$ (no equilibrium). Newey–West standard errors inflate variance by $\sqrt{n/n_{eff}} \approx 1.8$ (Bayley & Hammersley, 1946).

4 Differencing, Cointegration, and Hierarchical Models

- **Short-run inference:** use ΔY_t, ΔX_t.

- **Long-run inference:** if cointegrated, use ECM:

$$\Delta Y_t = \alpha + \gamma(Y_{t-1} - \beta X_{t-1}) + \varepsilon_t.$$

- **Hierarchical correction:**

```
md = smf.mixedlm("height ~ wgi", data,
                 groups=data["subject"],
                 re_formula="~year").fit()
# subjects nested within years -> beta_wgi approx 0
```

4.1 Other Confounders in Cross-Disciplinary Aggregates

- **Composition effects:** cohorts differ in age, education, SES.

- **Common shocks:** crises (e.g., 2008) affect both governance and sampling.

- **Measurement drift:** WGI methodology changed 2019.

- **Seasonality:** periodic data collection biases correlations.

4.2 Aggregation and Researcher Degrees of Freedom

Flexible aggregation (annual vs cohort) enables unintentional *p-hacking*. Analyses should pre-register aggregation level and justify it theoretically.

4.3 Negative Controls (Lipsitch et al., 2010)

As a sanity check, regress an unrelated biomarker (e.g., ear length) on WGI. If $p < 0.05$, structural collinearity or confounding is suspected.

5 Formalizing the Spurious Risk Score (SRS)

$$\text{SRS} = 100 \times \Big(w_1 \cdot \mathbb{I}(p_{\text{OLS}} < 0.05) + w_2 \cdot (1 - p_{\text{det}}) + w_3 \cdot (1 - |2 - \text{DW}|/2) + w_4 \cdot \mathbb{I}(p_{\text{NW}} > 0.05) \Big)$$

where w_i are normalized weights estimated by logistic regression on 10,000 Monte Carlo simulations. High SRS (> 70) indicates strong trend alignment or autocorrelation risk. Implementation: `spurious-risk` (MIT license).

6 When Spurious Is Not Spurious

Not all trending associations are artefactual. For example, GDP per capita and life expectancy share both cointegration and causal mechanisms (nutrition, health investment). A genuine link meets at least two of: (i) theoretical mechanism, (ii) cointegration $p < 0.05$, (iii) robustness under detrending or fixed effects.

7 Reviewer Checklist (ranked by importance)

* * * * * Detrending or differencing applied?

  * * * * Cointegration $p > 0.1$ (no long-run link)?

   * * * Durbin–Watson outside (1.5, 2.5)?

   * * * Newey–West SE > 2x OLS SE?

    * * Cohorts balanced by age/SES?

     * Multiple aggregations tested (p-hacking risk)?

8 Interactive Stress-Test App and Reproducibility

The companion Streamlit app (`https://trends-collide.streamlit.app`) allows users to upload two time series and obtain a **Spurious Risk Score (0–100%)**. Full reproducibility via Binder: `https://mybinder.org/v2/gh/Evoluit-M/evoluism-s/main`. Extended notebooks cover ARIMA(1,1,0), fractional integration ($d \in [0.5, 1]$), and structural breaks (2015 shock).

9 Conclusion

This interactive note consolidates a century of insight – from Yule to Granger – into a modern data-fusion context. It broadens diagnostics beyond trend alignment to composition bias, shocks, and methodological drift. Designed for courses in neuroeconomics, cognitive epidemiology, and data science, it also serves reviewers and preregistrants as a practical defense against misleading correlations.

References

[1] Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? *J. Royal Statistical Society, 89*(1), 1–63.

[2] Granger, C. W. J., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics, 2*(2), 111–120.

[3] Bayley, N., & Hammersley, J. M. (1946). The estimation of effective number of independent observations in autocorrelated data. *J. Applied Psychology, 30*(4), 431–436.

[4] Lipsitch, M., Tchetgen Tchetgen, E., & Cohen, T. (2010). Negative controls: A tool for detecting confounding and bias. *Epidemiology, 21*(3), 383–388.

[5] Smith, A. et al. (2022). Cognitive bias and institutional metrics: An analysis of Polity IV and hippocampal volume. *bioRxiv 2022.07.18.*