

# GLOBAL HOUSING RETURNS, “BACKCASTING FINANCE,” AND THE EMERGENCE OF THE SAFE ASSET, 1465-2024

Paul Schmelzing

Discussion by

**Cameron LaPoint**

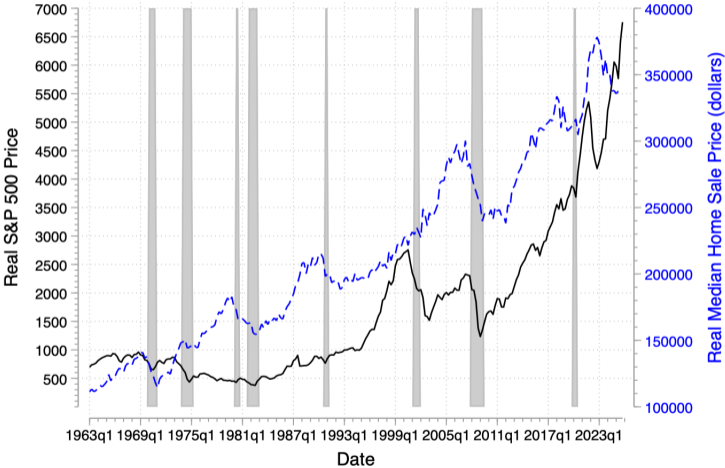
Yale SOM

Western Finance Association Meeting

Denver, CO

June 24th, 2026

# ONE INTERPRETATION OF THIS PAPER: WHEN, IF EVER, WAS THERE JUSTIFICATION FOR THE PHRASE: “AS SAFE AS HOUSES?”



Source: Median home sale prices from HUD and Real S&P 500 prices from Robert Shiller’s data

# THIS PAPER: FIVE CENTURIES (!!!) OF HOUSING RETURNS

- **Big archival contribution:** German repeat sales data back to 1465
  - ▶ 12 cities, 707 properties, 5,564 repeat sales pairs
  - ▶ 128,679 property-years in headline summary table
- **Methodological contribution:** use modern ML models to *backcast* house prices
- **Interpretation:** rising housing values reflect secularly falling discount rates
  - ▶ Recent decades are part of longer trends, not a one-off boom
  - ▶ Safe asset premium emerges in early modern Europe

Figure 2: German real house price index (RHPI, log), raw and sqm-basis, 1465-2025.



## THIS PAPER: ANSWER IS “BIG YES,” BUT TO WHAT QUESTION?

- **Most convincing:** long-run German urban property prices were more dynamic than existing hockey-stick narratives suggest
- **More ambitious:** the data identify global investable housing returns over 1465–2024
  - ▶ **Most (> 90%) of the total return to holding housing comes from rental yields**
- **Most ambitious:** the evidence pins down the secular decline in discount rates and the **emergence of sovereign safe assets**
- Roadmap for my discussion:
  1. Facts vs. measurement error, and how to benchmark
  2. Refining time series econometric tests
  3. What does this evidence tell us about the housing credit cycle hypothesis?
  4. When did a house stop becoming a home?

## COMMENT #1: FACTS VS. MEASUREMENT ERROR

- **Question:** are the new stylized facts raw empirical facts or constructed objects?
- The paper emphasizes 128,679 property-years, but index construction comes from much thinner independent variation
  - ▶ 707 properties; 5,564 repeat sales observations
  - ▶ Average continuous transaction history is roughly 201 years
  - ▶ Sample is still described as “small” relative to the German housing stock
- The underlying objects are typically in the “old” parts of cities
  - ▶ Good for long histories and consistent cadaster links
  - ▶ Potentially bad for representativeness and survival bias
  - ▶ **Are these cities defined in a geographically consistent way over time?**
    - ★ Ahlfeldt, Albers, Behrens (2026) “prime locations” shapefiles

# HOW SHOULD WE BENCHMARK THE MEASUREMENT?

- Repeat sales controls for **location/parcel fixed effects**, but not the whole housing bundle
  - ▶ Structures, density, floor space, infrastructure, legal rights, sanitation, and neighborhood amenities change
  - ▶ Paper acknowledges repeat sales indices do not fully strip out quality improvements
  - ▶ **Median German homeowner may have a 30–50% larger home over the long run**
    - ★ Long appendix discussion of this issue but no adjustment made
- **Suggestions:**
  1. Show raw transaction-level moments and city-by-city leave-one-out estimates
  2. Benchmark against hedonic indices where possible
  3. Separate land, structure, size, and amenity components
  4. **Systematically decompose role of weights, deflators, splices, and quality adjustments**

## WHAT HAPPENS IF WE PUT STANDARD ERRORS ON THESE SERIES?

- Standard repeat sales regression in modern data:

$$\log p_{i,t}^c = \delta_t^c + \eta_i^c + \epsilon_{i,t}^c$$

$$P_t^c = \exp(\delta_t^c)$$

- Generally don't report SEs on  $\delta_t$  given large transaction volume, but small  $N$  here
  - ▶ Appendix mentions “interpolating” points, presumably because some years do not have any transactions → also adjust for intra-year seasonality in cadastral records?
- **Is the U-shaped pattern in returns statistically significant?**

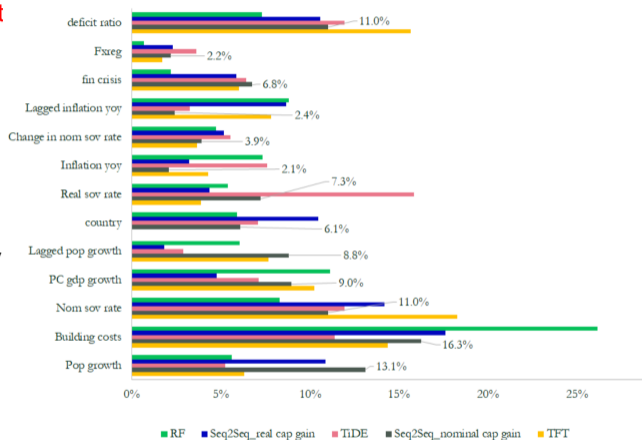
Table 3: Summary Statistics for Global data, GW (AW) basis, 1560-2020.

	nom TR	real TR	real capital gains	rental yd
1560-1660	7.3 (10.0)	5.9 (8.3)	-0.1 (-0.1)	6.0 (8.2)
1660-1760	8.0 (7.7)	7.6 (7.3)	0.7 (0.6)	6.9 (6.7)
1760-1860	6.3 (6.5)	4.7 (5.2)	-0.7 (-0.7)	5.4 (5.9)
1860-1960	9.5 (9.4)	6.9 (6.6)	0.1 (0.0)	6.8 (6.6)
1960-2020	10.6 (11.2)	6.8 (7.5)	1.7 (2.7)	5.1 (4.8)
Total				
1560-2020	8.1 (8.8)	6.3 (6.9)	0.2 (0.3)	6.1 (6.6)

# ADDITIONAL MEASUREMENT ERROR FROM ML BACKCASTING?

- **Global series combine two very different sources of variation**
  - ▶ Germany: archival repeat sales evidence
  - ▶ U.S./U.K./France/Netherlands: ML backcasts from modern training data
- **Concern:** common patterns may partly reflect model structure
- High modern-panel OOS fit is not evidence that the model recovers 16th century prices
  - ▶ Benchmark using within-country backcast?

Figure 4: Variable importance, by ML model.



- Database of Marfè & Pénasse (2024) might be useful to add more vars.

## BAYESIAN “BACKCASTING” AS A BENCHMARK

- Others in macro/finance have been using Bayesian approaches to recover series from sparsely populated + noisy data (e.g., Bouscasse, Nakamura, Steinsson 2025)
- **Canonical Bayesian state space setup:**

$$\underbrace{\Delta p_{i,t}^{obs}}_{\text{archival / repeat sales signal}} = \underbrace{\Delta p_{i,t}^*}_{\text{latent house price growth}} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim N(0, \sigma_{\text{source}(i,t)}^2)$$

$$\Delta p_{i,t}^* = \beta' X_{i,t} + f_t + u_i + \eta_{i,t}, \quad f_t = f_{t-1} + \nu_t$$

- **Let measurement error vary by source:** repeat sales, rents, CPI deflators, ML covariates
- Allow stochastic trends/breaks/time-varying volatility in  $f_t$
- Draw posterior paths for prices, yields, returns, and excess returns
- Compare ML point backcast to Bayesian posterior bands; re-run all stylized fact tests across posterior draws

## COMMENT #2: STATIONARITY DOES NOT EQUAL FORECASTABILITY

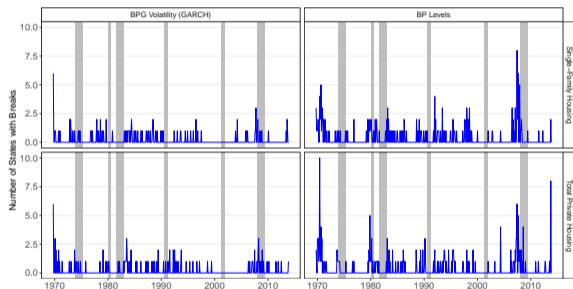
- Paper says: **“ADF-GLS tests ... confirm or reject forecastability”** and Table 4 “clearly confirm[s] forecastability”
- **But rejecting a unit root in returns is a weak diagnostic**
  - ▶ A stationary return series can be white noise
  - ▶ ADF tests generally underpowered, but perhaps less of an issue here with centuries of data
  - ▶ Forecastability in finance needs a time- $t$  information set, predictive regressions, OOS performance, and economic magnitude
- For annual asset returns, stationarity is usually the null hypothesis
- Core objects may instead be price-rent ratios, excess returns, rental yields, PV residuals, or expected return proxies

## STATIONARITY TESTS: WHAT WOULD CONVINC ME?

- Showing stationarity of rental yields is a core result of the paper if we care about under/overvaluation of housing
  - ▶ Gordon growth approaches (or discount to NAV for REITs) implicitly take this as given
- Main issue: many tested series are **generated, spliced, interpolated, and persistent**
  - ▶ Have already addressed measurement error in my previous comment
  - ▶ Short ADF-GLS lag choices in the appendix → **try business cycle lag order?**
- **How to move the tests out of the “old-school” 1990s literature:**
  1. Break-robust unit root tests with multiple endogenous breaks (Kim & Perron 2009)
  2. Block/bootstrap critical values that re-run the index construction
  3. Fractional integration estimates:  $d < 1$  but maybe not  $d = 0$
  4. Predictive regressions with rolling/blocked OOS splits
  5. Separate inference for raw archival series vs. ML-generated series

# DO WE CARE ABOUT TREND BREAKS OR VOLATILITY BREAKS?

## Bai-Perron Break Date Tests for State-Level Building Permits

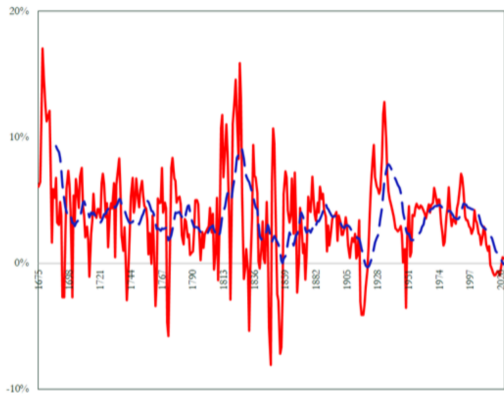


Source: Cortes & LaPoint (2024): "Housing *Is* the Financial Cycle: Evidence from 100 Years of Local Building Permits"

- Safety here is defined relative to sovereign debt interest rates
- Cortes & LaPoint (2024) focus on volatility of building permit growth over 100 years
- Volatility in house prices and quantities (building permits) seems to have declined over the long run
- Emergence of a safe asset seems more like a statement about a break in the **second moment** → can test this!

## COMMENT #3: WHAT ABOUT THE HOUSING CREDIT CYCLE?

- Paper recognizes credit supply explanations in the literature
  - ▶ Mian–Sufi style credit supply channel
  - ▶ Justiniano et al. (2019): mortgage rates capture loosening lending constraints
- New mortgage rate series (from the author's book) is a major contribution
- But the interpretation shifts toward **secular discount rate trends**
- **Credit cycles are about quantities and constraints, not rates alone**
  - ▶ Example: innovation of U.S. 30-year fixed-rate mortgage after Great Depression



German real mortgage rates, 1685-2022, from Schmelzing (2026)

# MORTGAGE RATES ARE NOT THE WHOLE CREDIT CHANNEL

- A long downward trend in mortgage rates may explain **valuation trends**
- Credit cycle hypotheses are about **amplification**:
  - ▶ LTVs, amortization, maturity, recourse, underwriting, collateral constraints
  - ▶ Lender balance sheets and local bank competition (e.g., Favara & Imbs 2015)
  - ▶ Borrower composition + speculation (Haughwout et al. 2011; Chinco & Mayer 2016)
- **Suggestions for isolating leverage channel:**
  1. Separate slow-moving discount rate trend from boom-bust credit quantities
  2. Interact mortgage rates with financial crises, bank deregulation, and collateral regimes
  3. Test whether high credit growth episodes predict stronger price run-ups and reversals
  4. Construct credit cycle diagnostics on top of discount rate diagnostics

## COMMENT #4: WHEN DID A HOUSE STOP BECOMING A HOME?

- **The concept of “housing” as an asset changes over five centuries**
  - ▶ Early stage: shelter and family wealth
  - ▶ Middle stage: rental asset and collateral
  - ▶ Modern stage: financial asset and speculative investment good
- Paper notes rental income was relevant throughout the horizon
  - ▶ Roughly 30% of taxable citizens in German urban areas rented in the late 15th century
  - ▶ Rises to roughly 60% by 1800
- **Owner-occupied vs. investment demand** changes risk, cash flows, holding periods, leverage, and the relevant discount rate

## TIME-VARYING QUALITIES LIKELY MORE CENTRAL TO THE STORY

- Repeat sales takes out the parcel fixed effect
- Concern: If modern housing bundles are better, safer, more regulated, more financeable, or more amenitized, then:
  - ▶ Part of “return” may be compensation for improved services
  - ▶ Part may be capitalized amenities rather than pure expected return → tax rate volatility is high in German historical context
  - ▶ Price appreciation per parcel is not necessarily an investable asset return
- **Suggestion: split housing into portfolios and check which long-run trends survive**
  - ▶ Owner-occupied primary residences, rental housing, second homes, speculative or investor-owned sales
- Probably not feasible to do all of these splits, given small  $N$  in German cadastral data
  - ▶ Data underlying other more modern series also rely on small  $N$  (e.g., Lyons et al. 2026 for U.S. house prices), so same issue applies

## WHERE DOES THIS LEAVE THE PAPER?

- **Strongest contribution:** new archival evidence that German house urban prices were dynamic before the 20th century
  - ▶ Rejection of the unique post-1960s “hockey stick” (Knoll et al. 2017)
- **Strong but needs error bands:** repeat sales index with quality, survival, and representativeness benchmarks
- **Most fragile:** global investable housing returns from ML backcasting + exact timing of safe asset emergence and the interpretation of yield spreads as safety premia

### My takeaway:

**Paper proves that pre-1800s housing prices moved around a lot. It has not fully proven that housing behaved like a stable global asset class.**



Yale SCHOOL OF MANAGEMENT

THANKS!

---