

Give People Housing and They Will Have Kids. Housing Prices and Fertility across U.S. Counties, 2000–2023.

Filip Zaleski
University of Toronto

Fifth Undergraduate Tech EConference

May 7th 2025

Motivation & Research Questions

- **Motivation:**

- Aging population $\Rightarrow$ strain on pensions, healthcare, labor.
- Fertility decline threatens demographic sustainability.
- Housing costs $\Rightarrow$ may shape family-size decisions.

- **Research Questions:**

- **Primary:** How do county-level housing prices affect fertility rates in the U.S. (2000–2023)?
- **Sub-Question 1:** Is there variation along the urban–rural continuum?
- **Sub-Question 2:** Can online housing expectations predict housing prices — and thus inform mechanisms affecting fertility?

Key Literature:

- Ermisch (1999): housing costs delay leaving home and childbearing.
- Dettling & Kearney (2012): metro prices negatively correlate with births.
- Li (2024): global panel links affordability to fertility trends.

My Reflection: no study combines U.S. county-level variation, urban–rural stratification, measures of market expectations.

Features:

- Urban–rural gradient explored via density interactions.
- Short-run vs long-run: first-differences vs. levels models.
- ML supplement: regression trees & expectations-enhanced forests.
- Policy relevance: zoning reform, income supports, expectation monitoring.

Data: Overview

Structure:

- County-level panel (2000–2023) covering **between 62% and 76%** of the U.S. population across years.
- $N \approx 12,000$ observations (county-year pairs).

Variables:

- **Dependent:** General fertility rate (live births per 1,000 women 15–44).
- **Explanatory:** Zillow Home Value Index (ZHVI) $\Rightarrow$ 35th to 65th percentile of home value.
- **Controls:** unemployment, education mix, marriage rate, racial shares, density.
- **Market Expectations:** Reddit & Twitter (X) Posts, Google Search Patterns $\Rightarrow$ used to predict changes in house prices.

Sources:

- **Controls:** CDC, Zillow Research, U.S. Census Bureau/ Bureau of Labour Statistics, U.S. Department of Agriculture.
- **Market Expectations:** API Programs (Reddit API, Pytrends, Apify), FRED, U.S. Census Bureau, Redfin.

Data: Summary and Early Insights

Summary Statistics: All Variables

	Variable	Mean	Std	Min	Max	Median
1	Fertility Rate	61	10	21	112	61
2	ZHVI	231	146	44	1517	189
3	Population Density	902	2918	7	49536	351
4	Unemployment Rate	6	3	2	29	5
5	% White	82	13	19	98	85
6	% Black	12	12	0	74	8
7	% Indian American	1	3	0	47	1
8	% Asian/Pacific	5	6	0	71	3
9	% Bachelor+	30	10	10	77	29
10	% High School	28	7	7	50	28
11	% Some College	29	5	11	43	30
12	% Below High School	13	7	2	44	12
13	% Married	53	6	31	71	53

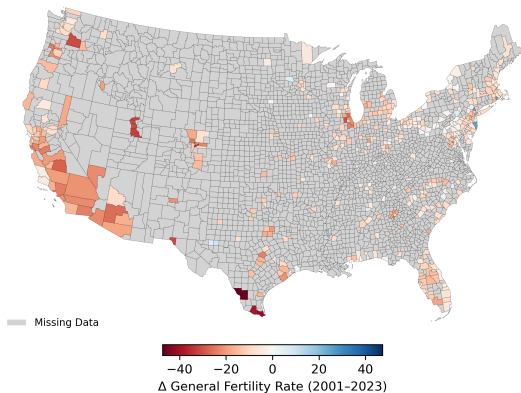
Notes: Fertility rate is per 1,000 women aged 15–44. ZHVI is reported in thousands of dollars. Population density is measured per square mile.

Insights:

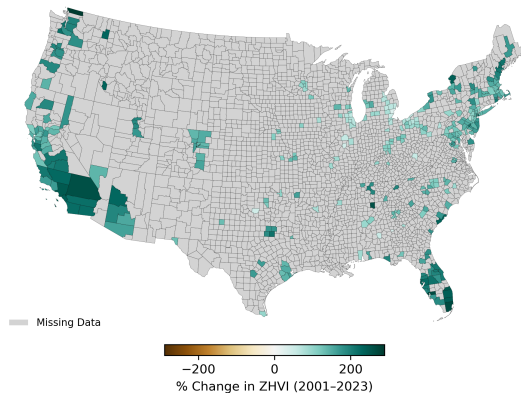
- High variance in fertility (21–112) & housing prices (ZHVI range 44–1517).
- Strong right-skew in population density (median 351 vs mean 902).
- Socioeconomic heterogeneity suggests fixed-effects essential.

Visual Motivation

Change in Fertility Rate Levels Over Time (2001-23)



Change in House Prices Over Time (Percentage, 2001-23)



- Urban counties (coastal CA, FL, IL) $\Rightarrow$ sharp fertility decline.
- Same counties show steepest price appreciation.
- Visual co-movement motivates econometric analysis.

Econometric Specifications

Fixed-Effects OLS

$$\text{Fertility}_{it} = \alpha + \beta \text{ZHVI}_{it} + \gamma \mathbf{X}_{it} + \mu_i + \delta_t + \varepsilon_{it}$$

Where:

- $\mathbf{X}_{it}$ matrix includes macro controls.
- μ_i = county fixed effect.
- δ_t = year fixed effect.

Extensions:

- Interactions with dummies (*Top 30% density*, density quintiles), or a continuous log (Density).
- Robust SEs clustered by county.

First-Differenced OLS

$$\Delta \text{Fertility}_{it} = \alpha + \beta \Delta \text{ZHVI}_{it} + \gamma \Delta \mathbf{X}_{it} + \varepsilon_{it}$$

- Captures **short-run** responses (wealth / expectations effects).

Main Results

Overall:

- The ZHVI coefficient is significant and negative across all specifications. Higher prices $\Rightarrow$ lower fertility after controlling for socioeconomic heterogeneity.

Preferred Specification Findings:

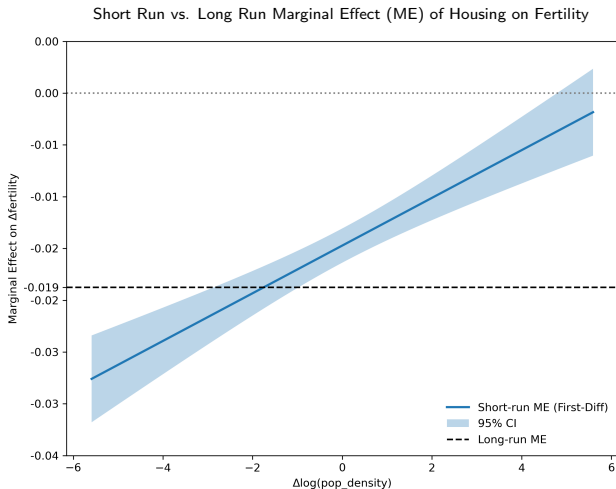
- **ZHVI coefficient** = **-0.019**. Each \$10k increase in house prices is associated with a 0.19-point reduction in the general fertility rate at mean density, holding other factors constant.
- **ZHVI $\times$ Density Coefficient** = **+0.001**. The negative price effect diminishes as density rises. But because house prices are much higher in dense urban counties, the overall affordability burden remains large.
- **Model Strength:** $R^2 = 0.84$ with full macro controls and fixed effects.

Short-run vs Long-run Effects

Idea: How do fixed-effects (long-run) models compare to differenced models (short run)?

Key insights

- **Long-run:** constant negative effect on fertility, $\beta = -0.019^{***}$.
- **Short-run:** effect varies with density change:
 - Shrinking counties ($\Delta \log \rho < 0$): stronger depressions (0.03).
 - Fast-growing counties ($\Delta \log \rho > 0$): attenuated—approaches zero (0.005).
- **Implication:** static wealth effects mask dynamic heterogeneity—policy must account for both level-based burdens and local growth shocks.



Regression Tree Insights

Why ML? These methods can capture non-linear thresholds and offer robustness against linear miss-specification.

Top Regression Tree Splits and Fertility Outcomes

Split	Variable	Threshold	Interpretation	Avg. Fertility
1	% Bachelor's	30.3%	Education $\Rightarrow$ delayed fertility	61.4 $\rightarrow$ 64.8
2	% under HS	22.1%	Less education $\Rightarrow$ higher fertility	64.8 $\rightarrow$ 75.6
3	% Married	48.5%	Lower marriage $\Rightarrow$ lower fertility	56.8 $\rightarrow$ 51.4
4	ZHVI	\$243k	Higher prices $\Rightarrow$ lower fertility	71.9 $\rightarrow$ 62.7
5	Top-30% Density	= 1	Urban dummy shows weaker outcome	51.4 $\rightarrow$ 46.3

Note: only top branches shown. Splits selected using variance reduction (CART); sample sizes omitted for readability.

- **Top splits:** education (% bachelor's, % under HS).
- **Secondary:** marriage, urban dummy, ZHVI.
- **Insight:** tree confirms fixed-effects pattern — affordability + education shape fertility.
- **Parsimony:** the most complex tree (18 vars) adds little predictive gain (4% MSE $\downarrow$).

Market Expectations: *Why Model Them?*

Behavioural lens. House–purchase and child-bearing plans react to *narratives* and price expectations, not just observed fundamentals. Laboratory and survey evidence shows subjective expectations can rival hard data in shaping economic choices (Roth & Wohlfart, 2020).

Empirical gap. Prior fertility–housing studies focus on prices or credit; few integrate *real-time* sentiment signals drawn from social media / search.

My approach

- 3 sources — *Twitter, Reddit, Google Trends* — complement depth, immediacy, and breadth.
- City-level panel (5 metros), 2010–2023, $\approx$ 70k tweets + 12k reddit posts, weekly GT indices.
- NLP scoring: VADER (Tweets) — FinBERT (Reddit) → annualised sentiment index.

Hypothesis

- Hypothesis: Negative expectations may reduce perceived affordability, which in turn could influence fertility decisions — though in this study, expectations are used to predict housing prices, not fertility directly.

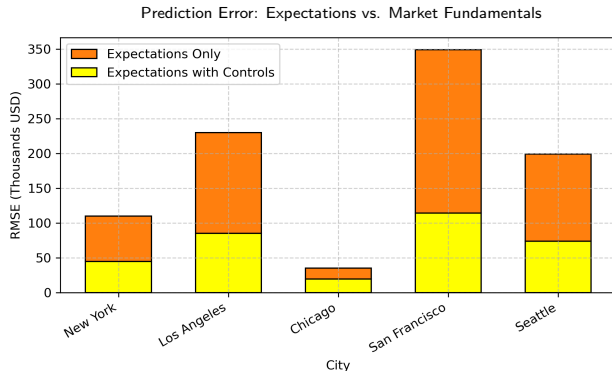
Market Expectations: Random Forest Insights

In the **Expectations-Only Random Forest**, Twitter dominates feature importance in all five cities – implying that:

- Either Twitter best captures real-time housing narratives, or
- VADER on Tweets outperforms FinBert on Reddit in sentiment scoring.

In the **Controlled Random Forest**:

- Adding fundamentals cuts RMSE by \$65k–\$235k per city (chart).
- *Median income* drives 64–94% of importance → purchasing power is key.
- *Mortgage rates* (2–11% importance) and *Unemployment* (1–40% importance) add marginal value.
- *Sentiment* still contributes 8% → narratives complement, but do not substitute, fundamentals.



Conclusion: Policy & Limitations

Policy Levers

- Expand housing supply in high-cost metros: zoning reform, streamlined approvals.
- Targeted income support: child tax credits, down-payment assistance.
- Geographic scaling: align family benefits with local housing costs.

Limitations & Future Work

- Expectations data coarse: small corpus & NLP context limits.
- Migration-fertility endogeneity: pursue IV & spatial RDD designs.

Happy to take questions. Thank You!

Contact: filip.zaleski@mail.utoronto.ca