The Price of Housing in the United States, 1890-2006

Ronan C. Lyons (Trinity College Dublin)
Allison Shertzer (University of Pittsburgh and NBER)
Rowena Gray (UC Merced)
David Agorastos (University of Pittsburgh)

May 17, 2023

Motivation

- Housing is both a key service and an important asset
  - Housing is the single largest component of consumer spending
  - Real estate is typically the largest component of household portfolios
  - 30% of wealth is real estate owned by households, 25% is real estate owned by other sectors

- Despite the centrality of housing to the American economy, reliable series on the sales price of housing is limited prior to 1975, especially at the city level

- Rental data is limited to the BLS (1914-), AHS (1973-), and census (1930-) for most of the 20th century

Why do we need better U.S. data?

▶ Debate on performance of housing as an asset (Gyourko et al. 2013, Leamer 2015, Piketty 2014, Rognlie 2014, Jorda et al. 2019)

▶ Do existing price indices accurately reflect business cycle fluctuations? (Fishback and Kollman 2014, Ozimek 2013)

▶ Accurate information on growth of housing costs important for understanding living standards over the long run (Margo 1992, Officer and Wilson 2006)

▶ Entire classes of economic models cannot be implemented without more spatially disaggregated housing price data, for instance hedonic and equilibrium sorting models (Epple and Sieg, 1999; Black 1999; Bayer et al. 2007)

Historical Prices in Housing Project (HiPHoP) Dataset

▶ We digitized 2.6 million sale and rental listings from historical newspapers to create new series for owned and rented housing

▶ Series covers 30 cities, chosen to obtain a mix by economic trajectory and region

▶ Each observation includes the listed price, and measures of location, size (rooms/bedrooms) and type (apartment/house) – plus rental frequency for rental listings

▶ Most comprehensive and consistently collected data on market prices of housing on an annual or quarterly basis to date for U.S. cities spanning the entire twentieth century

Today's Talk

- Existing sources of housing price data for the U.S.
- HiPHoP dataset
- Benchmarking
- Findings:
  1. Much more pronounced interwar housing cycle, including boom in 1920s: housing prices increased before the 1990s
  2. Real rents have a flat trend and do not closely follow business cycle – but with far more inflation that BLS measure, implications for CPI
  3. Capital gains vary by city but are close to zero 1890-1945
  4. More housing price growth where land is scarce before WWII, while restrictive zoning associated with more price growth after 1980

Limitations of existing data on sale prices

Pioneering national index of Shiller starts in 1890:

- ▶ 1890-1934: based on a **survey of owner valuations taken in 1934** for 22 cities using recollection of transaction price (Grebler et al. 1956)
- ▶ 1934-1953: simple average over five cities (Chicago, LA, NOLA, NYC, DC) of **median home price** using N=30 listings from newspapers
- ▶ 1953-1975: based on a truncated sample of **government-backed mortgages**, with price ceiling changing over time
- ▶ 1975-1992: CSW (S&P500) **repeat sales, including appraisals** from OFHEO data
- ▶ 1992-present: Case-Shiller-Weiss (now owned by CoreLogic) **repeat sales index** based on single-family homes that were sold at least twice

Limitations of existing data on sale prices, II

▶ Shiller index shows a very muted interwar housing cycle, no 1920s boom (Fishback and Kollman 2014)

▶ From 1950-2000, census microdata suggest much greater increases in sales prices relative to Shiller, esp. in 1970s (Davis and Heathcote 2007)

▶ Greenlees (1982) argues statutory mortgage size limits causes GSE part (1953-1975) of Shiller index to grow too slowly

▶ Not possible to follow individual cities annually before 1975

Limitations of existing data on price of rented housing

▶ Rees and Jacobs (1961) provide a simple average rent price collected from listings of six city newspapers from 1890-1914

▶ There is no market rental series for U.S. cities available before 2000 (Experian)

▶ BLS "rent of primary residence" collected in surveys covering 29 cities starting in 1914; used for CPI

▶ Housing Survey Program aims to capture price changes for same house; "periodic sample revisions" do not reflect particularly rapid changes to housing stock

▶ Scholars have raised concerns about downward bias in shelter components of CPI from depreciation, new technologies, and tenant non-response (Gordon and Van Goethem 2003 and Crone et al. 2010)

HipHop Dataset

▶ We utilize online repositories of digitized newspapers, particularly newspapers.com, to create a consistently collected annual series

▶ Real estate sections are sampled to obtain approx. 250 sales and 150 rentals per year (or four times a year before covid)

▶ Criteria for valid listing: price, location, size, and type

▶ We digitized every attribute in listing

▶ Sample of 30 cities aims to achieve regional representation and diversity of economic trajectories

HiPHoP Dataset Cities and Newspapers

City	Newspaper	Start Year	
		Rent	Sales
Atlanta	Constitution	1890	1890
Baltimore	Sun	1890	1908
Boston	Boston Globe	1890	1890
Charleston	Post-Courier	1894	1911
Chicago	Tribune	1890	1890
Cincinnati	Enquirer	1890	1890
Cleveland	Plain Dealer	1894	1890
Dallas	Morning News	1890	1890
Detroit	Free Press	1890	1890
Houston	Post / Chronicle	1896	1901
Las Vegas	Review Journal	1948	1943
Los Angeles	Times	1890	1890
Louisville	Courier-Journal	1890	1890
Memphis	Commercial Appeal	1891	1890
Miami	Herald / News	1914	1912
Minneapolis	Star Tribune	1890	1890
Nashville	Tennessean	1890	1890
New Orleans	Times-Picayune	1890	1893
New York	Times	1890	1890
Philadelphia	Inquirer	1891	1891
Phoenix	Arizona Republican	1912	1910
Pittsburgh	Post Gazette	1892	1890
Portland	Oregonian	1898	1898
Salt Lake City	Tribune	1891	1890
San Diego	Union	1907	1890
San Francisco	Chronicle / Examiner	1890	1890
Seattle	(Daily) Times	1890	1897
St. Louis	Post Dispatch	1890	1890
Tampa	Tribune / Bay Times	1915	1905
Washington D.C.	Post	1890	1890

HipHop Sampling Approach
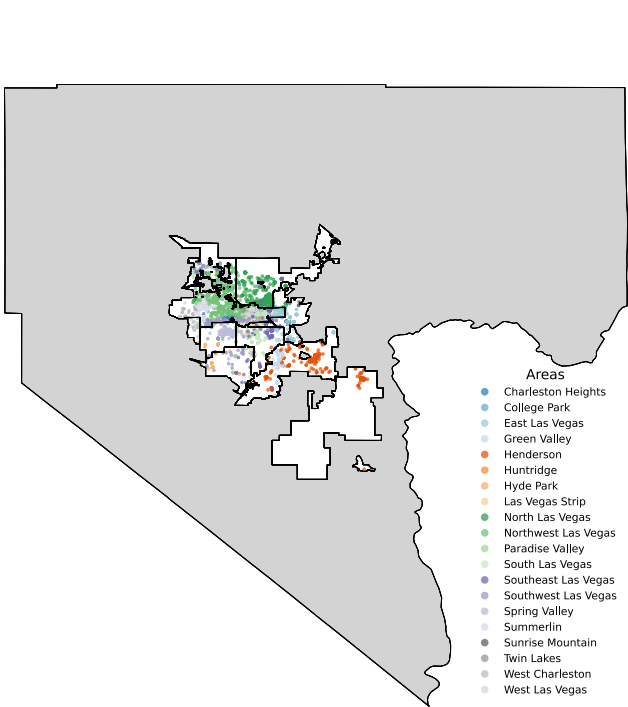
HipHop Sampling Approach

HipHop Sampling Approach

HipHop Data Dimensions

1. Segment: sale or rental
2. Price: as stated, use algorithm to determine frequency of rental payment
3. Size: number of rooms common earlier, then bedrooms in postwar period
4. Location: ML approach turns address, intersection, or "area" into a consistent set of geographic areas
5. Rentals: payment frequency (with rents converted to monthly; imputation for missing frequency, based on known weekly/monthly and monthly/annual ratios)
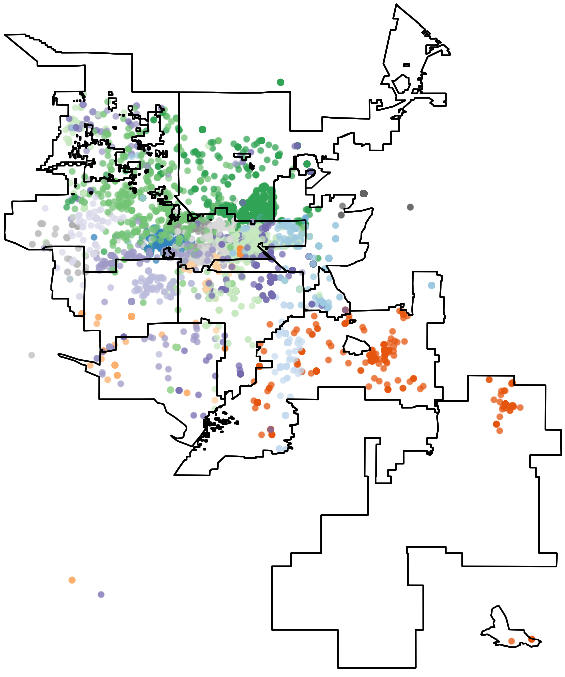
HiPHoP Area Construction

▶ We manually clean neighborhoods from newspapers and geocode addresses.

▶ Simple ML approach to assign listings to a set of 20 neighborhoods per city using prices, latitude and longitude of addresses, and neighborhood.

HiPHoP Area Example: Las Vegas

Areas
- Charleston Heights
- College Park
- East Las Vegas
- Green Valley
- Henderson
- Huntridge
- Hyde Park
- Las Vegas Strip
- North Las Vegas
- Northwest Las Vegas
- Paradise Valley
- South Las Vegas
- Southeast Las Vegas
- Southwest Las Vegas
- Spring Valley
- Summerlin
- Sunrise Mountain
- Twin Lakes
- West Charleston
- West Las Vegas

(a) Clark County

(b) Las Vegas-Henderson-Paradise

Benchmarking Analysis

▶ To understand our data better, we benchmark to several existing sources:

1. Census microdata versus HiPHoP data.

2. HiPHoP rent index vs. Rees + BLS/adjusted BLS series.

3. HiPHoP sales index vs. each Shiller component.

Census Benchmarking

1. HiPHoP matches census well in most years
2. Some evidence of sticky rents/lack of homeowner awareness of inflation in some city-year pairs
3. Some evidence of positive selection in some city-year pairs, well addressed by controlling for dwelling size
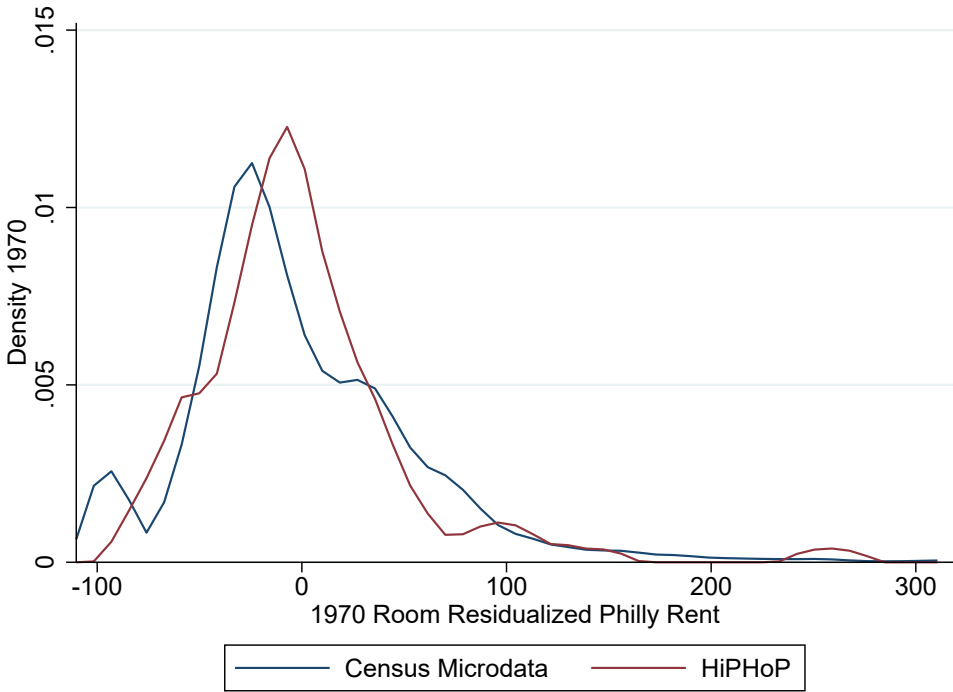4. Biggest divergence is due to rent control and census topcoding, not the business cycle

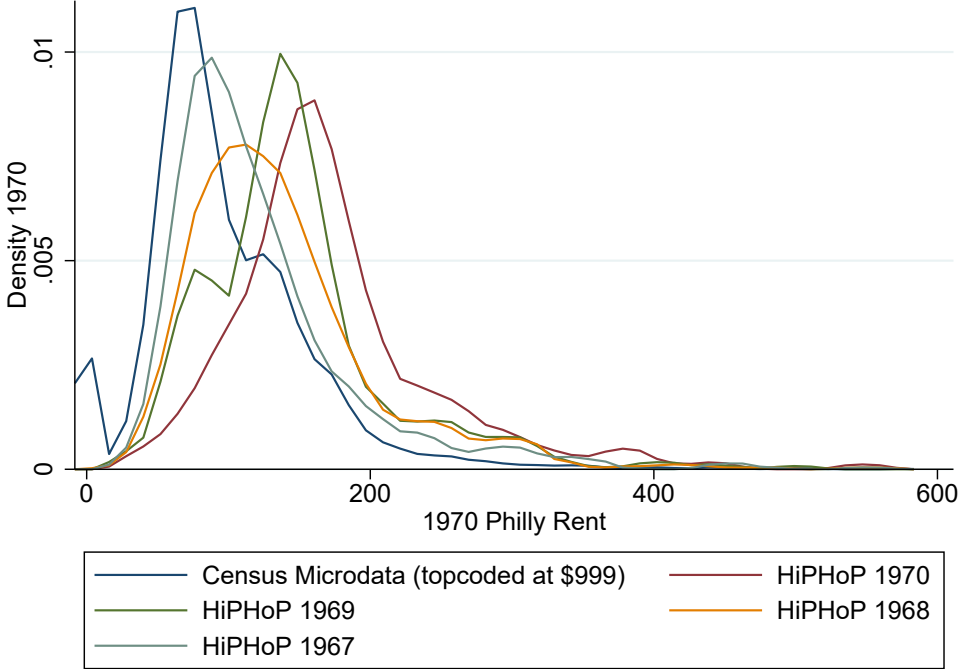# Census benchmarking: Philadelphia Rents (2000)
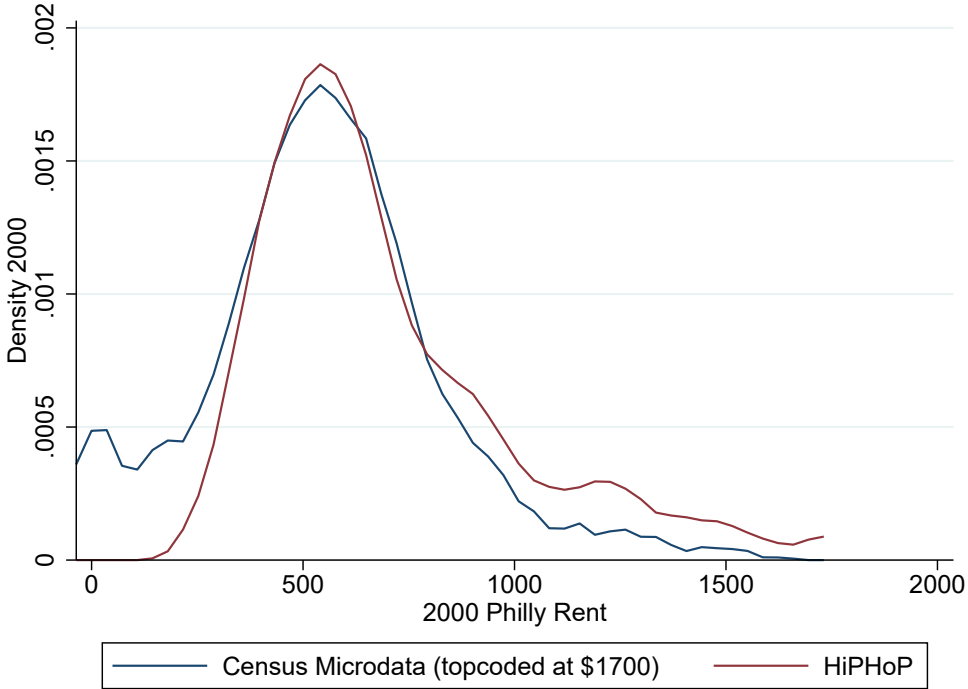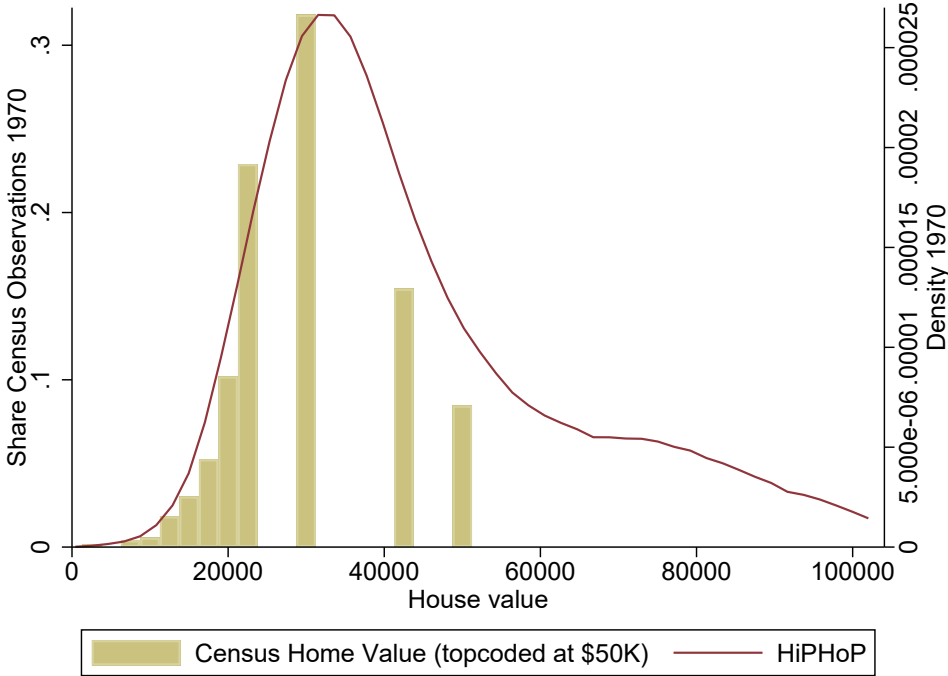
Census benchmarking: Philadelphia Rents (1970)
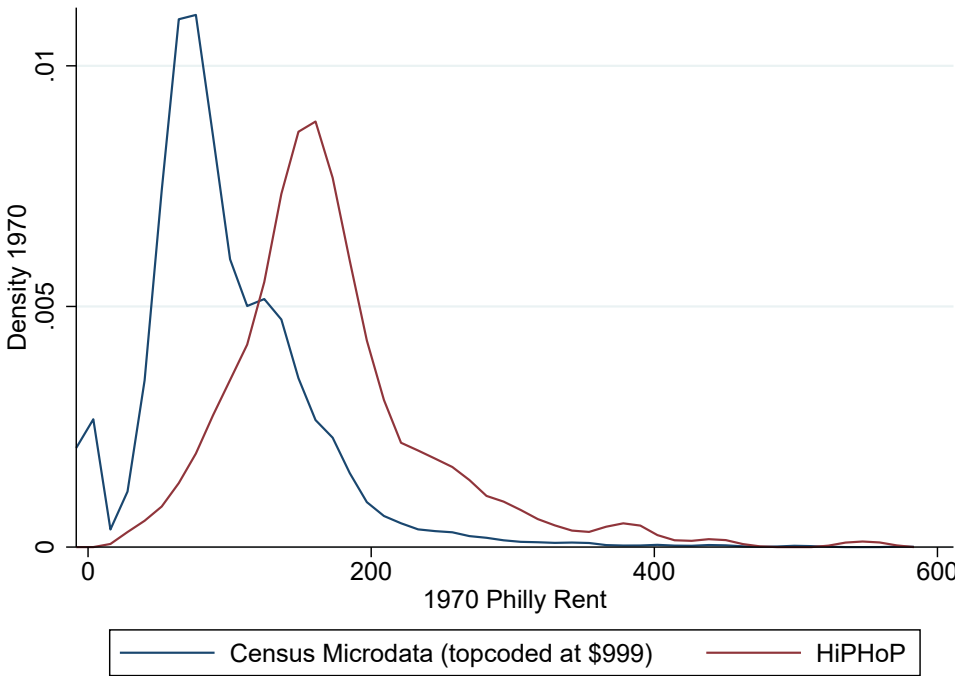
Census benchmarking: Residualized Philadelphia Rents (1970)

Census benchmarking: Lagged Philadelphia Rents (1970)

Census Topcoding: San Francisco Sales (1970)

Benchmarking Analysis

▶ To understand our data better, we benchmark to several existing sources:

1. Census microdata versus HiPHoP data.

2. HiPHoP rent index vs. Rees + BLS/adjusted BLS series.

3. HiPHoP sales index vs. each Shiller component.

Hedonic model

▶ Repeat sales are used in the U.S. currently but constant-quality assumption can bias these (Nowak and Smith 2020) and they are based on small ($\approx$7%) parts of the housing market in CS (Nagaraja et al. 2014)

▶ However, hedonic methods can also exhibit upward bias due to unobserved quality changes (Ross and Shen 2021)

▶ We control for type, neighborhood, size, rent payment frequency, and include various "rolling windows" to allow unobserved characteristics to have different baseline impacts (3 year baseline)

▶ Will show reassuring alignment with repeat-sales indices like CSW after 1980, despite differences in methodology

Hedonic Pricing Model

▶ Regression equation, running one city at a time and aggregating by population:

$$\ln\left(\text{Price}\right)_{i,t} = \alpha + \underbrace{\sum_{\min(Y),y\neq BY}^{\max(Y)} \beta_y \; \mathbb{1}_{(y=t)}}_{\text{Coefficients of Interest}} + \underbrace{\mathbf{X\Gamma}}_{\text{Controls}} + \varepsilon_{i,t}$$

▶ Controls: location, size, payment frequency, and dwelling type
▶ Can show 2, 3, and 5 year rolling window
▶ Transformation and rebasing of estimates:

$$\iota_t = \iota_{t-1} \exp\left(\beta_t\right) : t \in \{1891, 2006\}$$

RPI

▶ Current series from BLS suggest that real rents have fallen in the postwar era.

▶ Many adjustments have been proposed to address downward bias in shelter components of CPI from depreciation, new technologies, and tenant non-response (Gordon and Van Goethem 2003 and Crone et al. 2010), particularly before 1996.

▶ Our series matches adjusted series - real market rents have a remarkably flat trend over the whole twentieth century, fluctuating with a 15% band around 1890 levels.

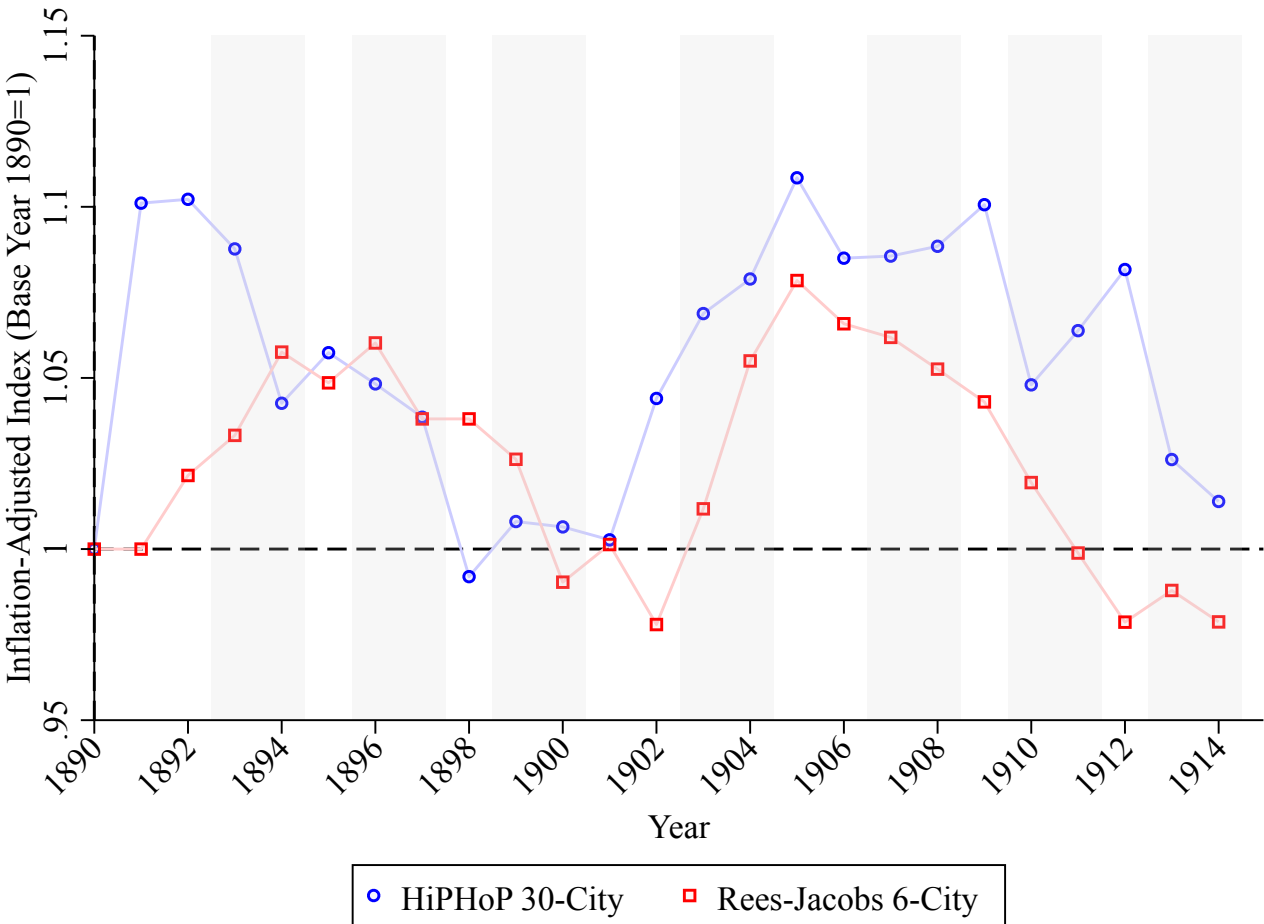▶ Also tracks Rees and Jacobs (6-city index based on newspaper) well.

BLS vs. "Adjusted" BLS vs. HiPHoP Robust RW Robust Geo
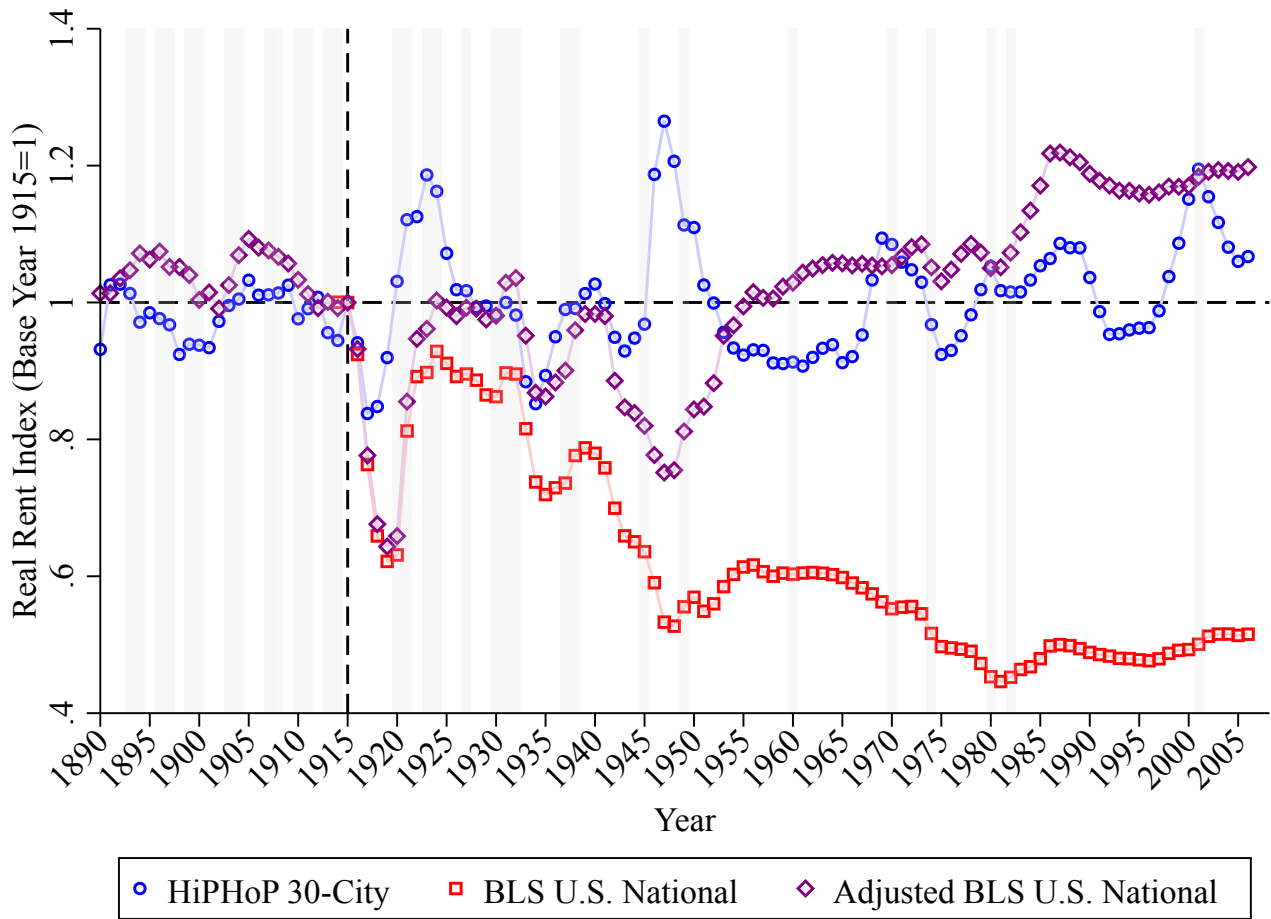
Rees-Jacobs vs. HiPHoP

HPI Results

▶ Let's begin with the pioneering work of Shiller, who constructed the first national, long-run housing price series for the United States.

▶ Probably impossible to start before 1890 or so except for New York, so we can't do multiple centuries as in Korevaar et al. (2022) for European cities.

▶ "Irrational Exuberance" finding is that real housing prices have only increased twice since 1890, first after WWII and second since 1997 or so.

▶ Also that housing prices didn't really boom and bust between the world wars.
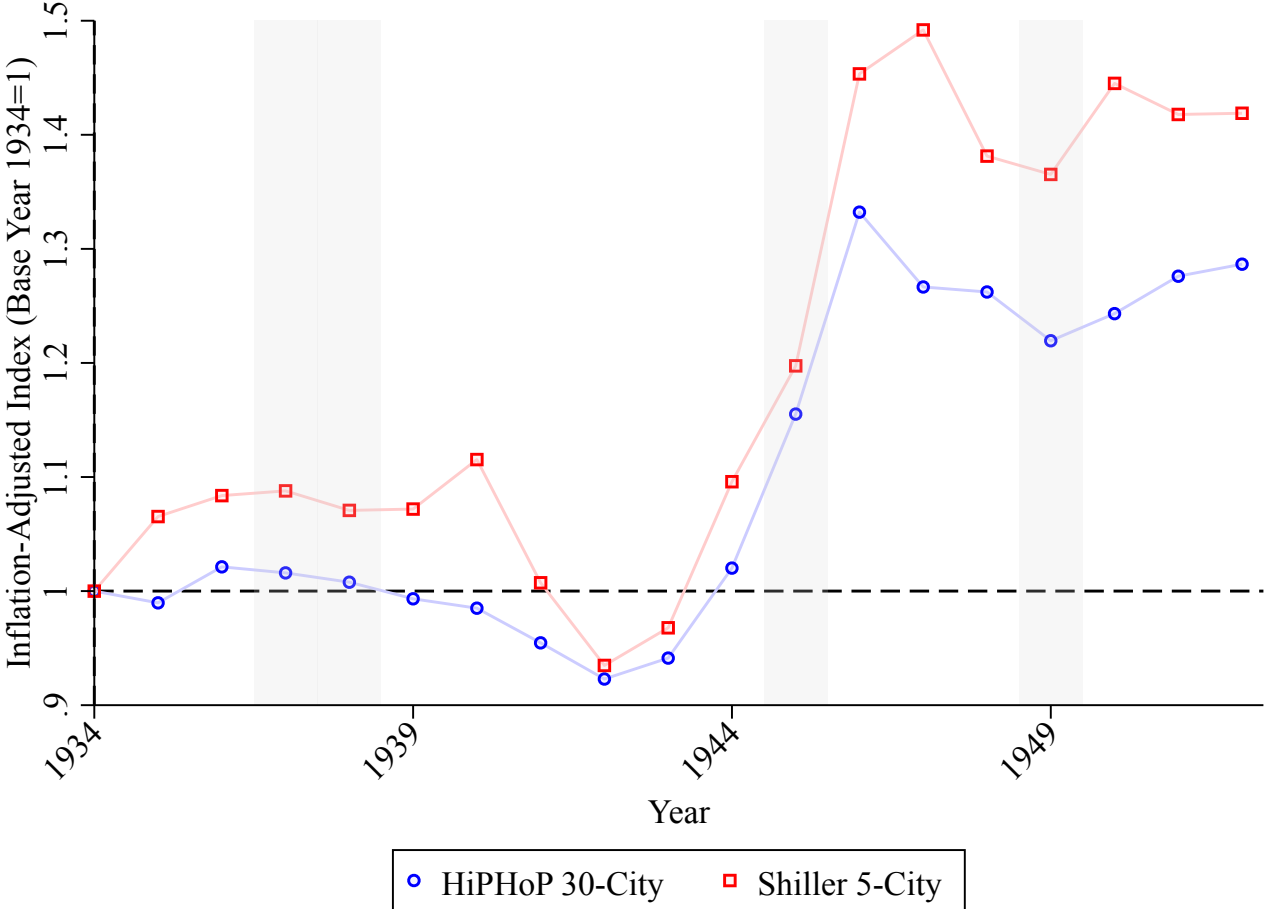
HiPHoP vs. Shiller: 1890-1933
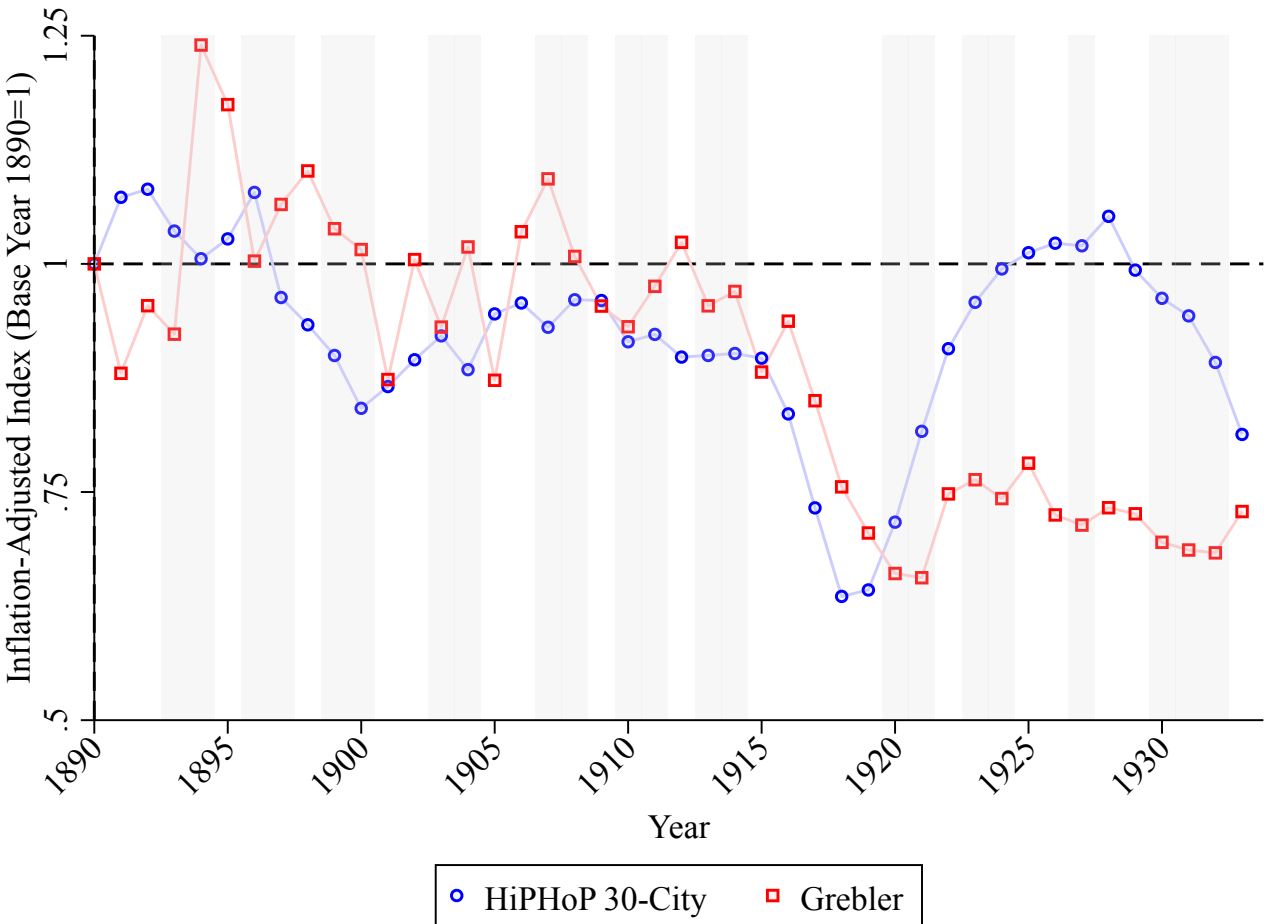
HPI Benchmarking: 1890-1933

- ▶ (Lack of) interwar housing cycle in Grebler et al. (1956) survey noted by others, including Fishback and Kollman (2014)
- ▶ We find that real prices are lower in 1940 than in 1930, consistent with the census, New Deal data, and Nicholas and Scherbina's (2013) study of NYC housing transactions
- ▶ Prices rose by 18% between 1920 and 1929 and then fell by 39% by 1935. Prices did not recover to 1928 peak until 1946
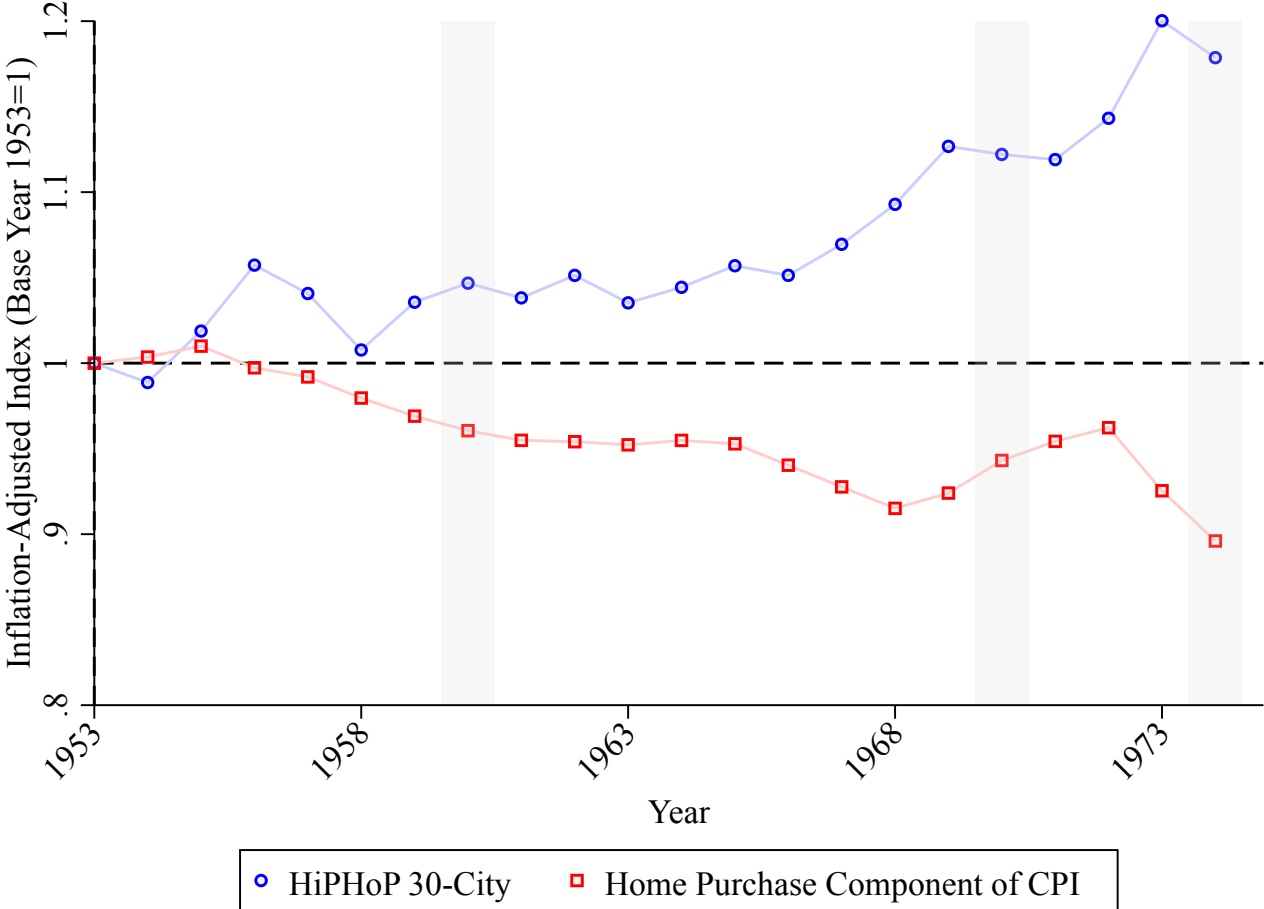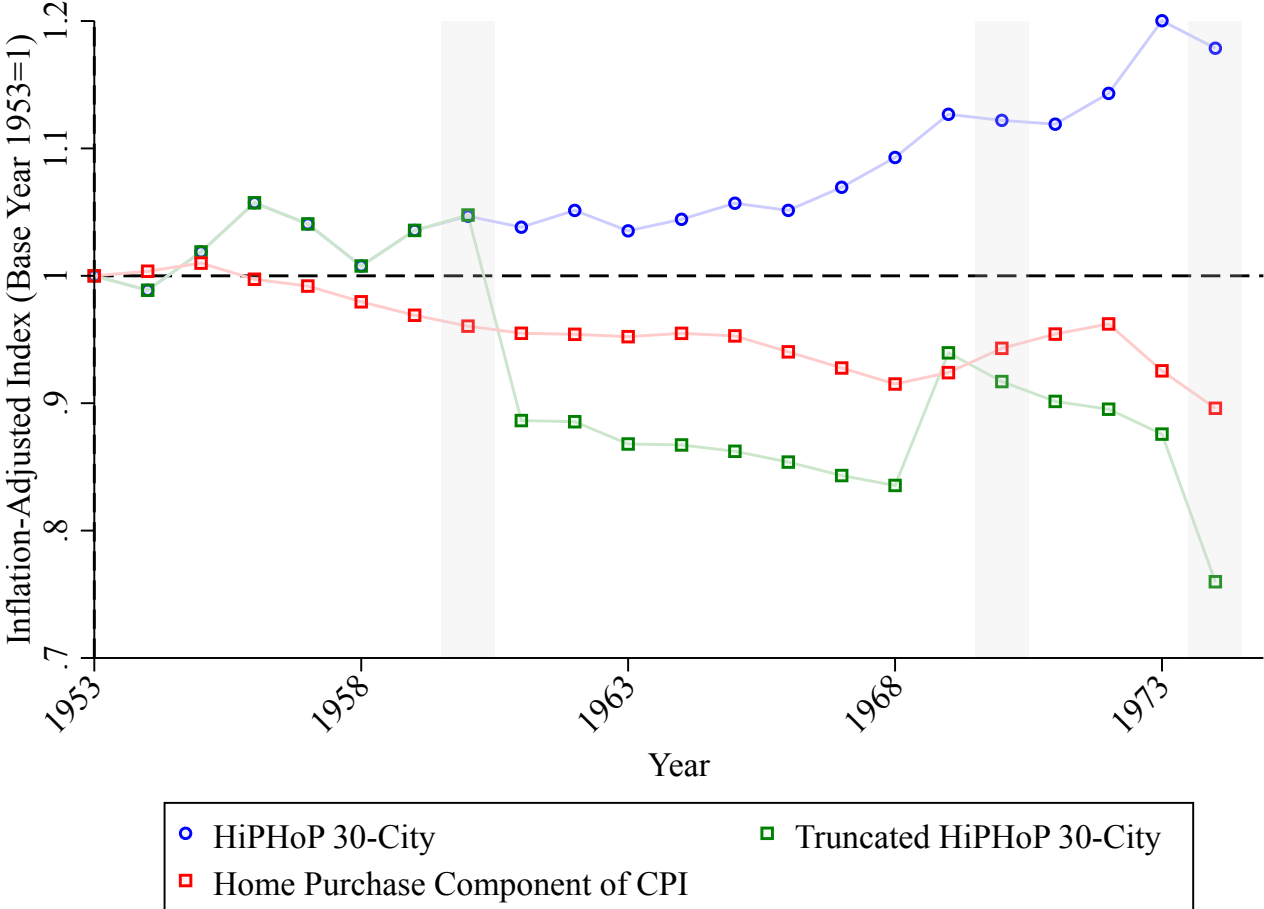
HiPHoP vs. Shiller: 1934-1952

HPI Benchmarking: 1934-1952

- ▶ Difference arises because we are making many adjustments (size, location, rolling windows) and different cities (30 vs. 5).

- ▶ Hedonic adjustments reduce price inflation.

- ▶ We find 28% vs. Shiller's 42%.

HiPHoP vs. Shiller: 1953-1974

HiPHoP (Truncated like FHA) vs. Shiller: 1953-1974

HPI Benchmarking: 1953-1974

▶ We find evidence in favor of Greenlees' critique: real housing prices increased by 18% between 1953-1974 rather than falling

▶ Statutory limit was $30,000 in 1973, $45,000 in 1974, and $60,000 in 1977 (Vandell, 1995): approximately half our 1970s dataset would be excluded

▶ Our evidence consistent with notion that many houses were excluded in major cities, limiting the appearance of housing price growth

▶ Real housing price growth thus starts much earlier, consistent with the census (Davis and Heathcote, 2007)

HiPHoP vs. Shiller: 1975-2006
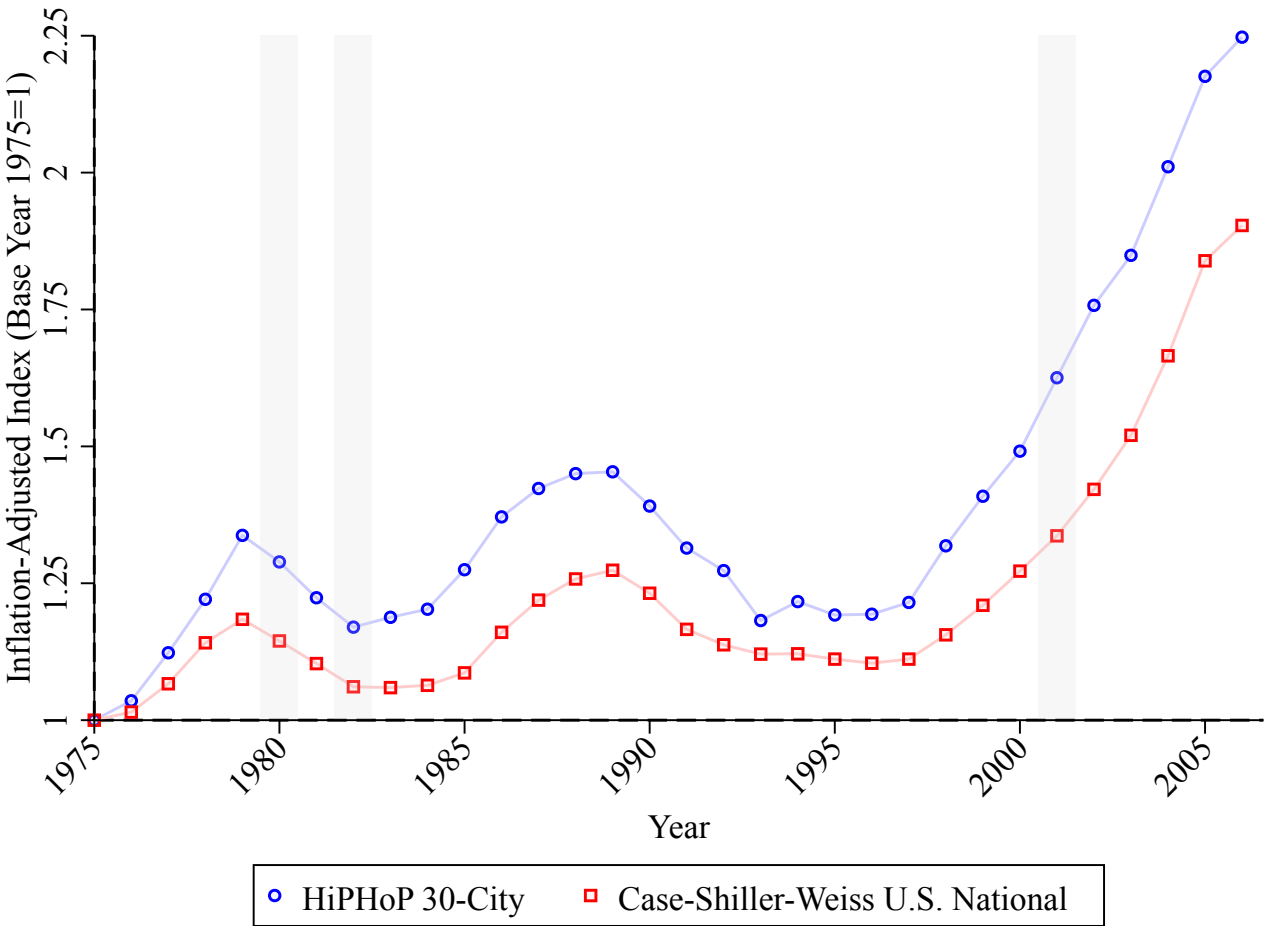
HPI Benchmarking: 1975-2006

▶ We find real price growth of 125% over the 1975-2006 period compared to the 90% found in the CSW series

▶ Factor of 1.95 for Shiller versus 2.54 for HiPHoP for the entire 1954-2006 period, driven by differences between before 1980

▶ CSW use OFEO appraisals before 1992, bias on this not clear

▶ SFH versus entire owned housing market

▶ Different city coverage

▶ We can use city-level indices to investigate further
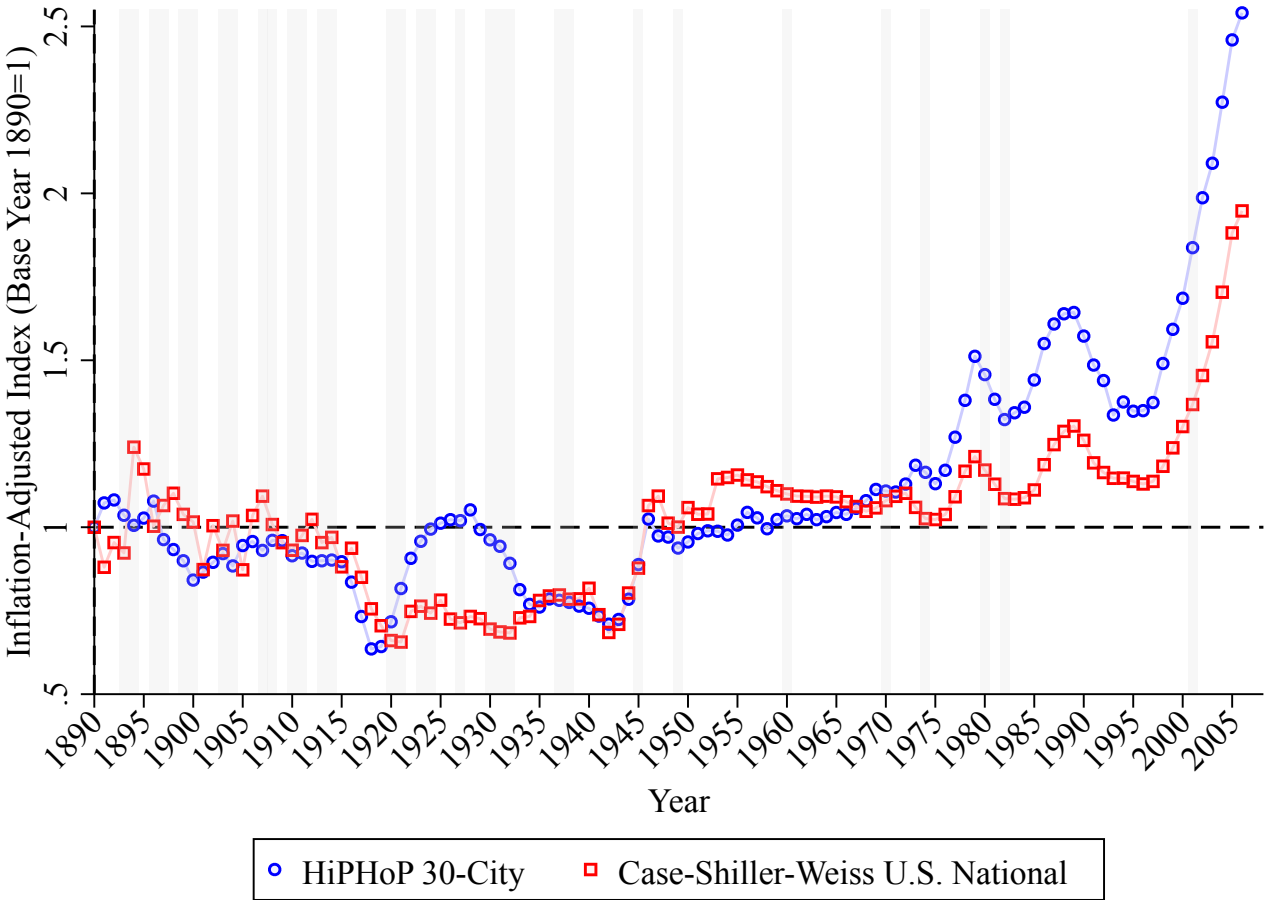
City HiPHoP vs. Shiller: 1987-2006
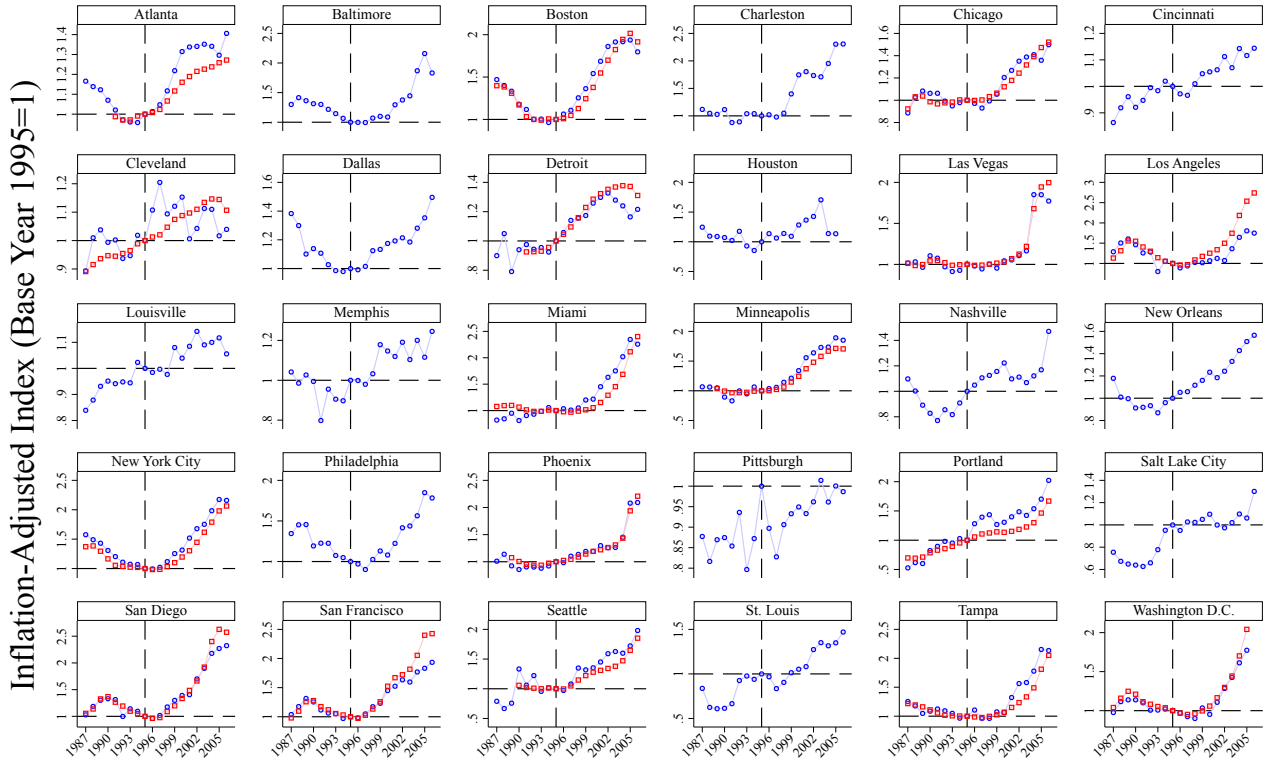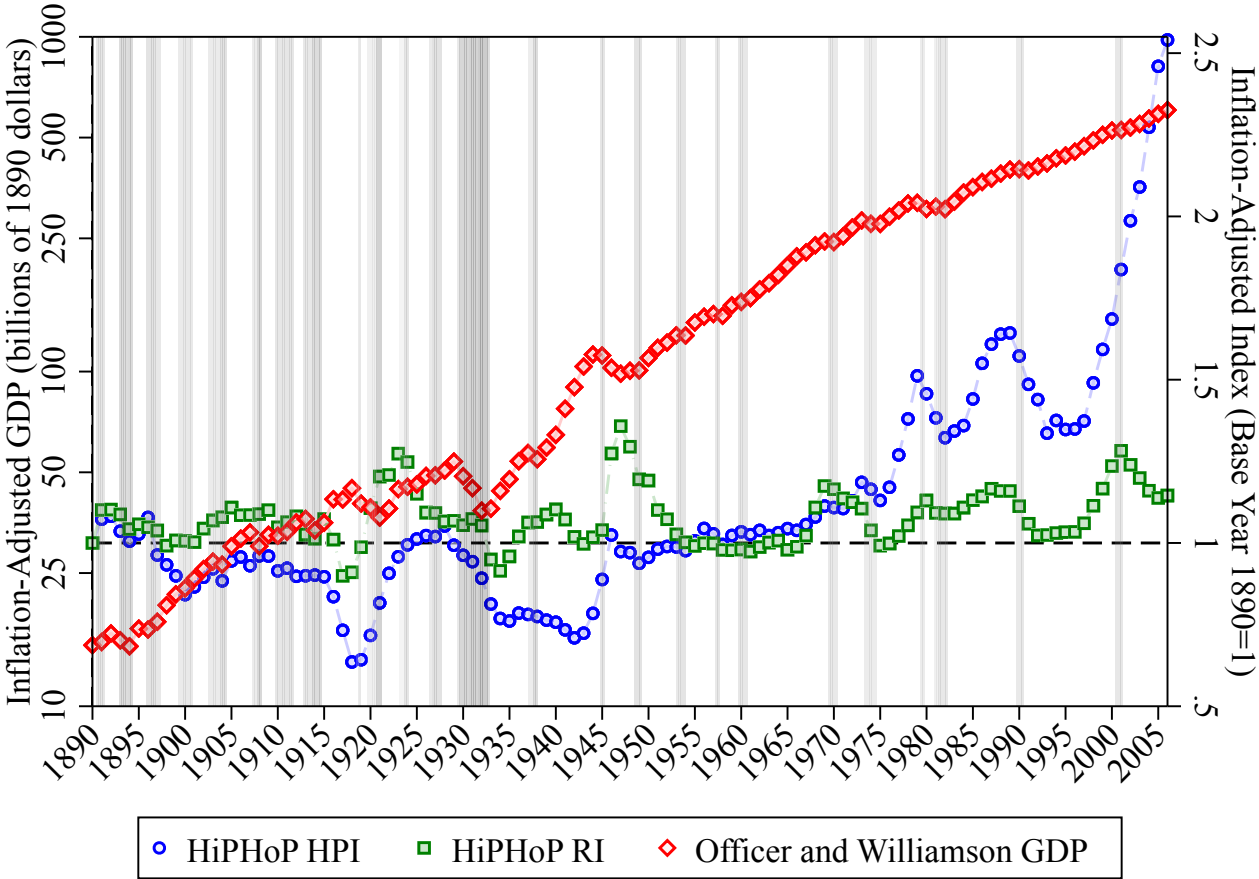
HiPHoP vs. Shiller Indices Robust RW Robust Geo

Implications for U.S. Housing Markets

1. Housing prices and the business cycle

2. Nominal rental growth and the CPI

3. Capital gain by city and time period

4. Why has housing appreciated in some cities more than others?

Housing Cycle and the Business Cycle

Note: Intensity of recession indicated by gradient of shading from light (minor) to dark (major).

Housing Cycle and the Business Cycle

▶ Housing prices and GDP are strongly linked, moving together in 76 out of 116 for sales prices and 65 for rents. In other years they fall soon after an economic downturn (after 1896 recession and after 1980).

▶ Only after both World War I and II do we see economic contractions accompanied by surging housing prices.

▶ Two cycles – Great Depression and late 1980s/early 1990s – in which rents fall, then prices, then GDP

▶ A similar pattern at the end of our dataset: rents fall from 2002, while prices and GDP will fall after 2006

▶ After 1970s, rents keep their overall flat trend while housing takes off: suggests joint role of financialization and housing supply.

Nominal Rental Growth and the CPI
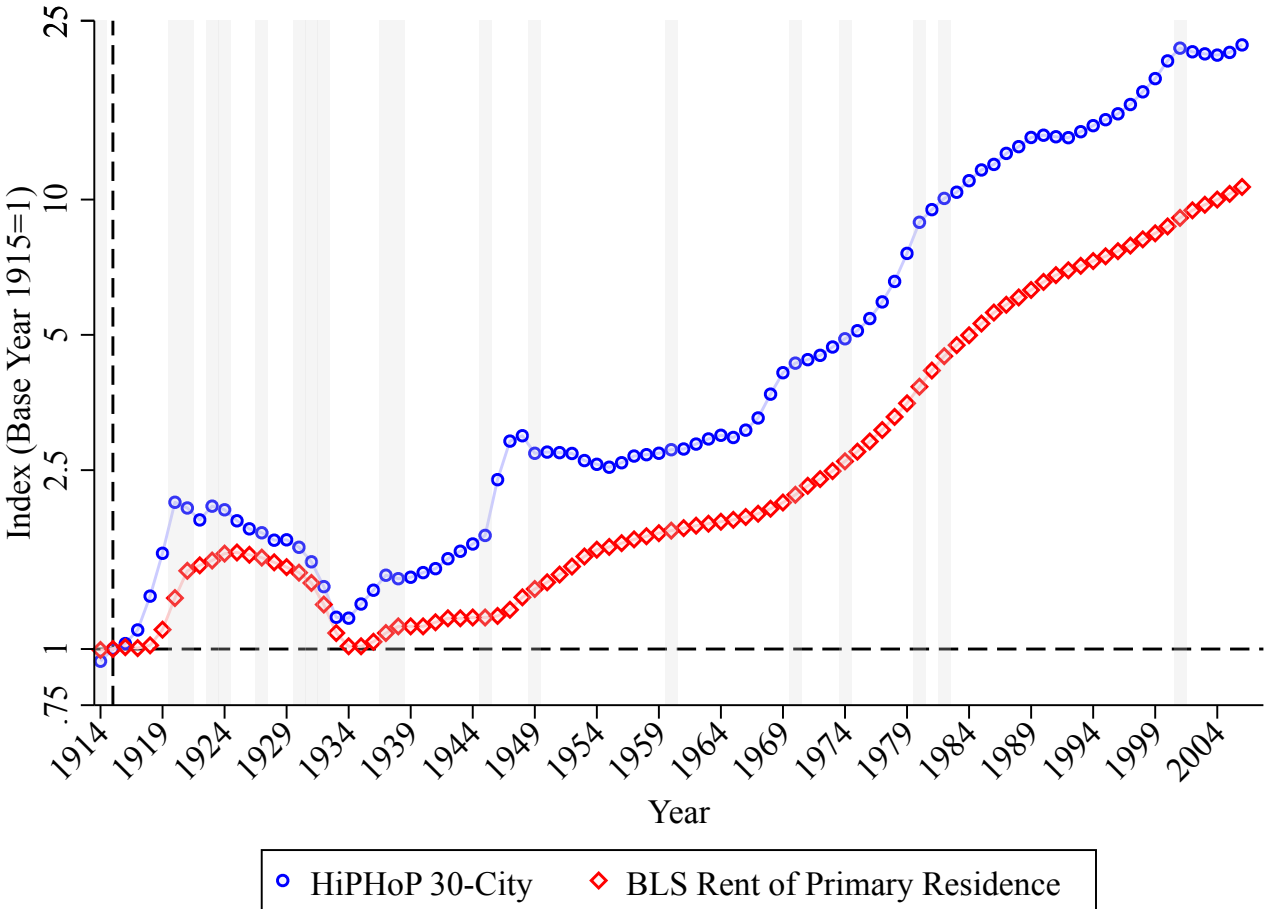
Nominal Rental Growth and the CPI, 1914-2006

▶ Our index suggests nominal rents grew 23.5x (3.5% p/yr) – while the BLS RoPR grew 10.7x (2.6% p/yr)

▶ Key question: with ex-shelter CPI at 3.3% p/yr 1914-2006, did rents rise by substantially less than other prices?

▶ Context: market rents used for owner-occupier implicit rents (since 1980s and ideally before that) – what is impact on CPI of switching to market rents measure?

▶ A CPI with two components, ex-shelter (75%) and housing (25%), would grow by 3.3% rather than 3.1% per year, 1914-2006

▶ Difference driven by two war/postwar periods (1914-1920, 105% vs 65%, and 1940-1950, 85% vs 25%) and period of higher inflation (1965-1985, 300% vs 175%)

City-level Capital Gains

Capital gain:

$$H_{c,t} = \frac{\mathsf{HPI}_{c,t} - \mathsf{HPI}_{c,t-1}}{\mathsf{HPI}_{c,t-1}}$$

Real capital gain:

$$h_{c,t} = \frac{1 + H_{c,t}}{1 + \pi_t} - 1$$

such that $\pi_t = \frac{\mathsf{CPI}_t - \mathsf{CPI}_{t-1}}{\mathsf{CPI}_{t-1}}$.

Average annual real capital gain:

$$\bar{h}_c = \frac{1}{|T|} \sum_{t=1890}^{2006} h_{c,t}$$

such that $t \in T = \{1890, 2006\}$.

City-level Capital Gain

City	Period				Full
	1890-1929	1930-1945	1946-1980	1981-2006	Time-Series
Atlanta	0.02	0.36	0.87	0.90	0.52
Baltimore[a]	1.01	0.92	1.73	1.84	1.47
Boston	0.26	-0.28	1.51	3.44	1.28
Charleston[b]	–	4.39	1.14	3.94	2.76
Chicago	1.29	0.07	0.43	2.25	1.08
Cincinnati	0.73	-0.97	0.66	0.55	0.43
Cleveland	1.08	0.10	1.07	-0.01	0.70
Dallas	-0.14	0.98	0.06	-0.65	-0.04
Washington DC[c]	-0.21	1.10	1.45	2.61	1.09
Detroit	1.61	-0.91	-0.25	-0.01	0.34
Houston[d]	0.42	1.60	1.39	-0.03	0.81
Los Angeles	0.88	3.13	1.97	1.12	1.57
Louisville	0.11	-0.23	0.53	0.58	0.29
Las Vegas[e]	–	–	1.75	1.83	1.78
Memphis	-0.47	1.41	0.23	1.56	0.45
HH Nat'l	0.16	-0.56	1.50	2.28	0.94

Values: arithmetic mean over annual real capital gain. HH Nat'l: population weighted.
Note: a) begins 1908, b) begins 1911, c) begins 1890, d) ends 2005.

City-level Capital Gain (Continued)

City	Period				Full Time-Series
	1890-1929	1930-1945	1946-1980	1981-2006	
Miami[a]	2.93	4.12	1.19	3.12	2.55
Minneapolis	0.13	1.03	1.65	2.02	1.14
New Orleans[b]	-0.70	-0.19	1.56	1.00	0.43
Nashville	-0.65	-0.25	1.22	2.23	0.61
New York	0.51	-1.50	2.07	3.96	1.48
Pittsburgh	-0.14	-1.36	0.90	-0.01	0.03
Philadelphia	0.18	-2.10	1.31	2.13	0.64
Phoenix[c]	-1.56	3.21	1.61	2.50	1.49
Portland, OR[i]	1.92	0.80	2.23	3.85	2.32
San Diego	2.46	3.88	2.62	3.37	2.91
Seattle[j]	2.36	3.63	4.11	3.26	3.32
San Francisco	-0.27	1.73	3.24	2.79	1.75
Salt Lake City	-0.44	2.76	1.85	1.43	1.11
St Louis	0.16	-2.48	1.40	3.34	0.88
Tampa Bay[k]	-0.49	3.23	1.66	2.22	1.54
HH Nat'l	0.16	-0.56	1.50	2.28	0.94

Values: arithmetic mean over annual real capital gain. HH Nat'l: population weighted. Note: e) begins 1908, f) begins 1911, g) begins 1900/ends 2005, h) begins 1943, i) begins 1898, j) begins 1897, k) begins 1905.

Capital gains

▶ The capital gain associated with owning housing was close to zero nationally between 1890 and 1945, before considering depreciation and maintenance. Shiller was right!

▶ Substantial heterogeneity across cities, with some West Coast cities having an average real capital gain over 2% in every period.

▶ Detroit's gain is negative after 1929.

Why did housing prices increase more in some cities relative to others?

▶ We ask how land availability and zoning stringency as measured in 2000 affected housing price growth at the city level using measures from Saiz (2010).

▶ Zoning was introduced in the 1920s. WRLURI survey done around 2006. Fundamental endogeneity problem, but still interesting.

▶ Land availability price gradient steepest in 1930-1945 period.

▶ Zoning price gradient is always positive but becomes steeper every period. By 1980-2006 it is triple the 1890-1922 period. Fischel (2001) is right.

Land availability
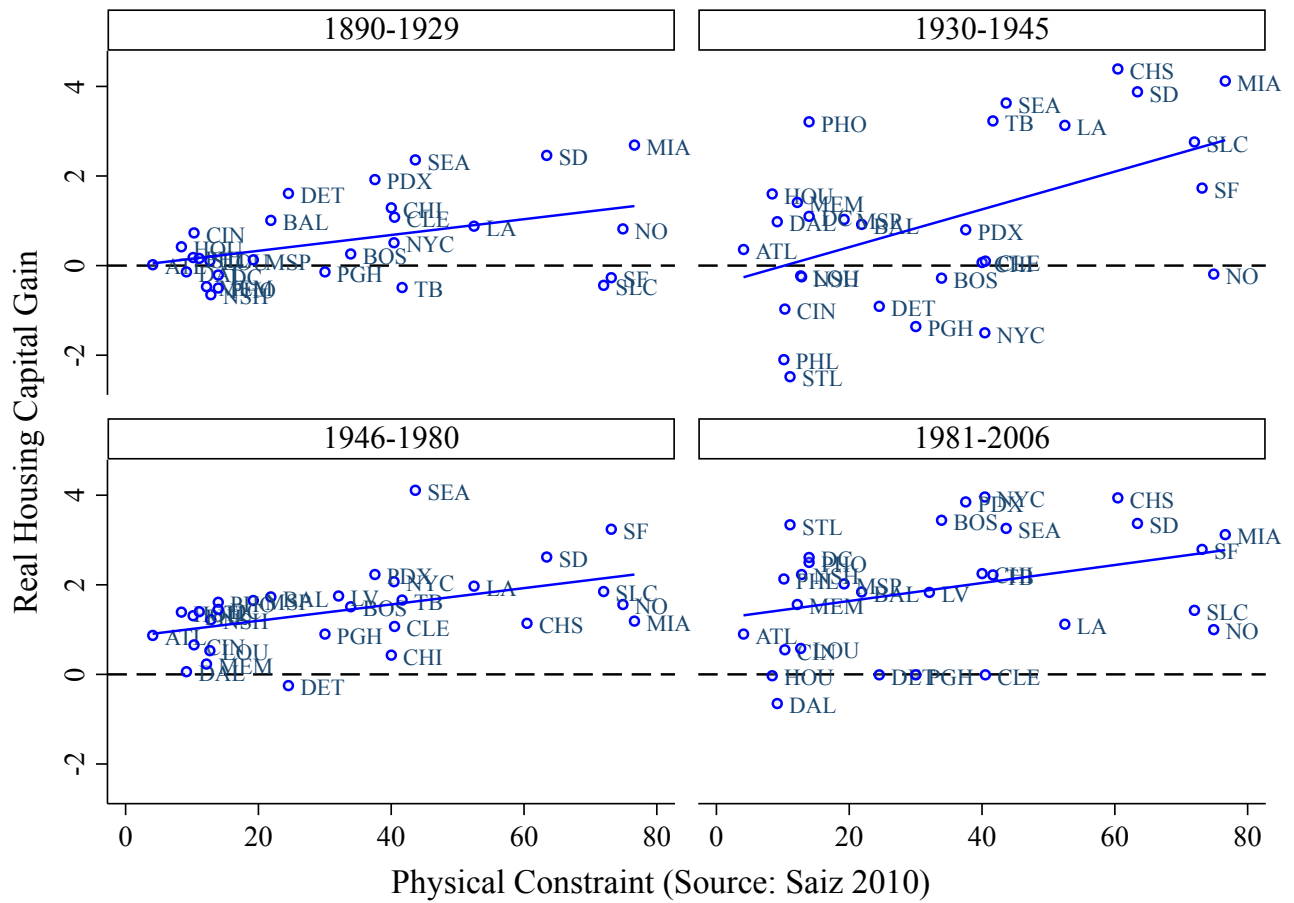
Zoning Stringency in 2006
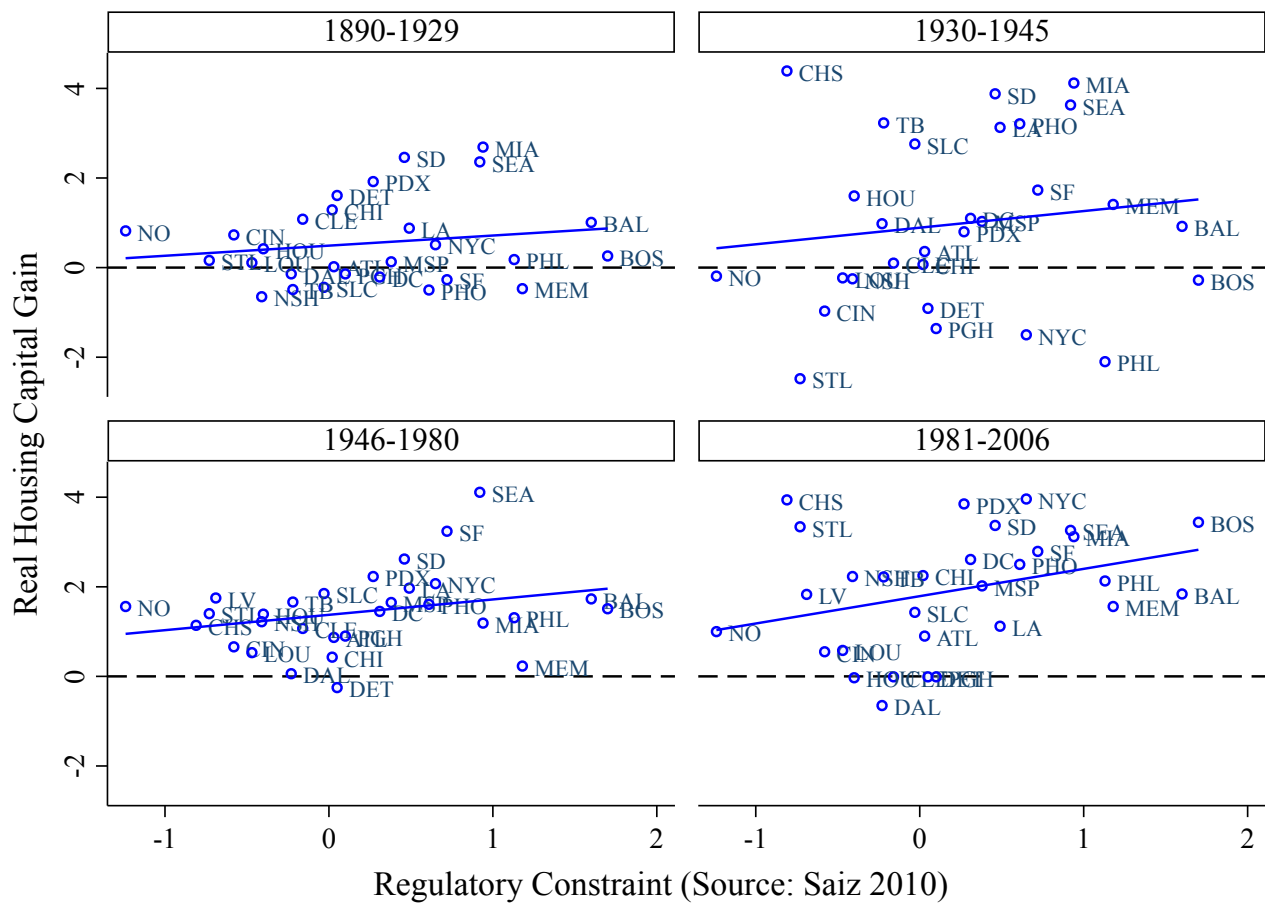
Conclusion

- ▶ HiPHoP has rent and price indices for 30 cities for 1890-2006.
- ▶ Confirms many limitations of existing series, plus provides consistently-collected, city-level series for the first time for both owned and rented housing.
- ▶ Provides new insight into the history of U.S. housing markets and many opportunities for future research.

HiPHoP HPI, Varying Rolling-Window Size ◂ Back
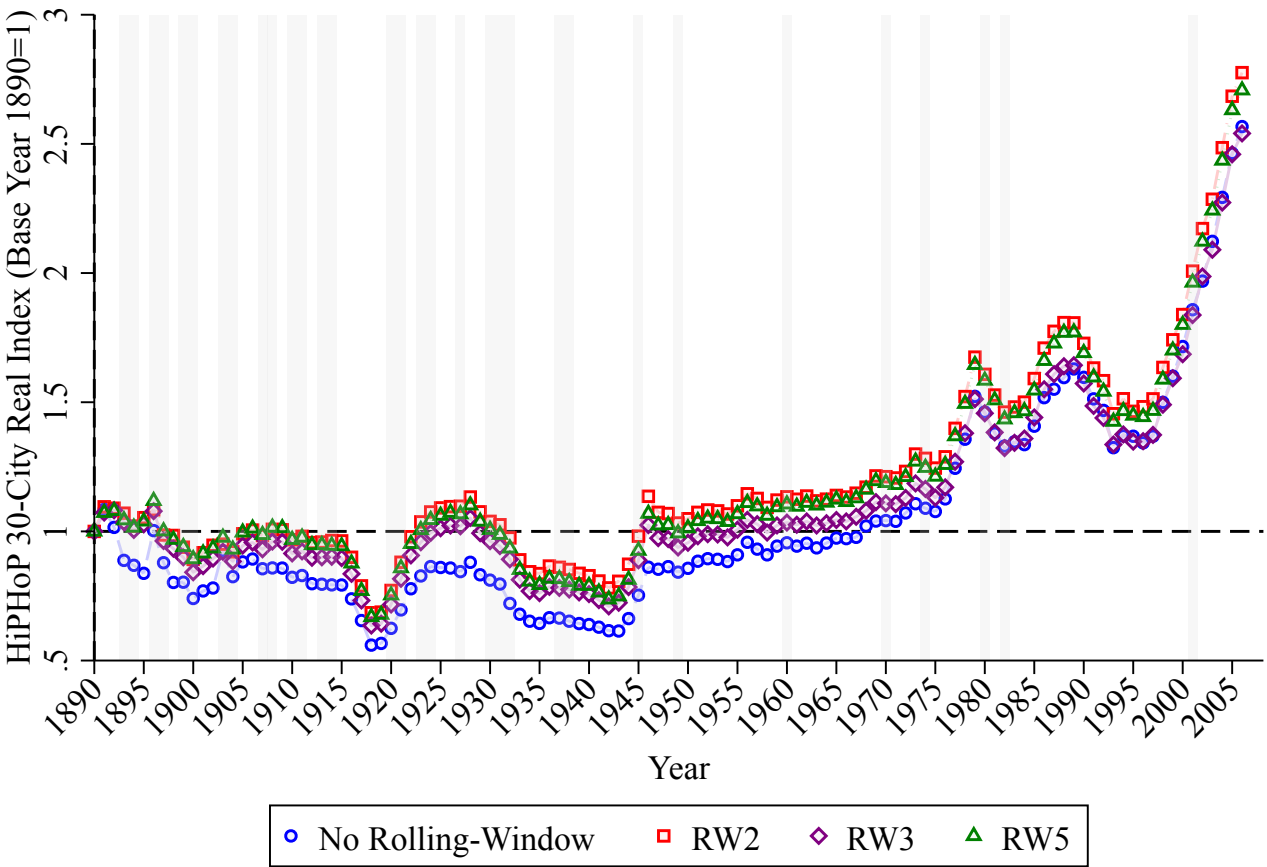
Note: RW# means Rolling-Window of Window Size # and Step Size 1.

HiPHoP RPI, Varying Rolling-Window Size ◂Back
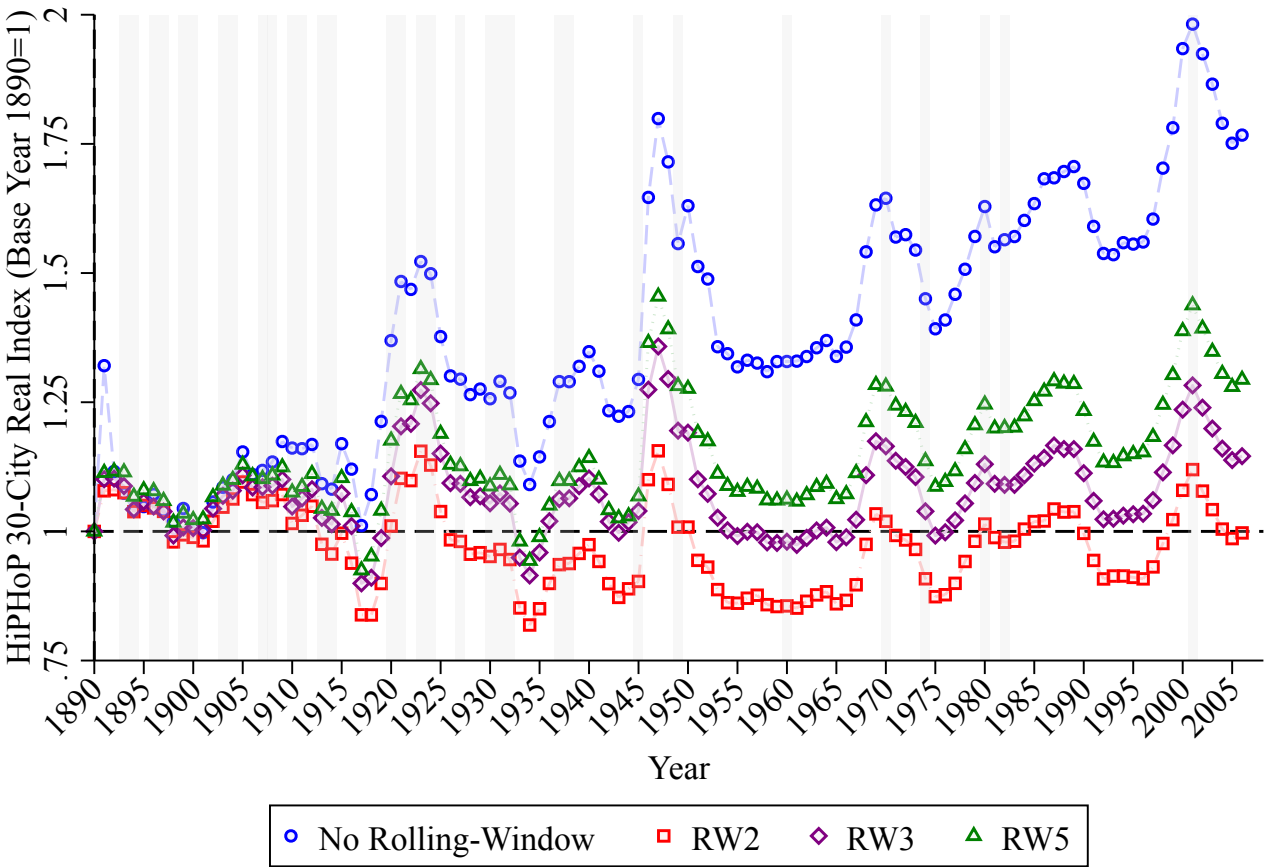
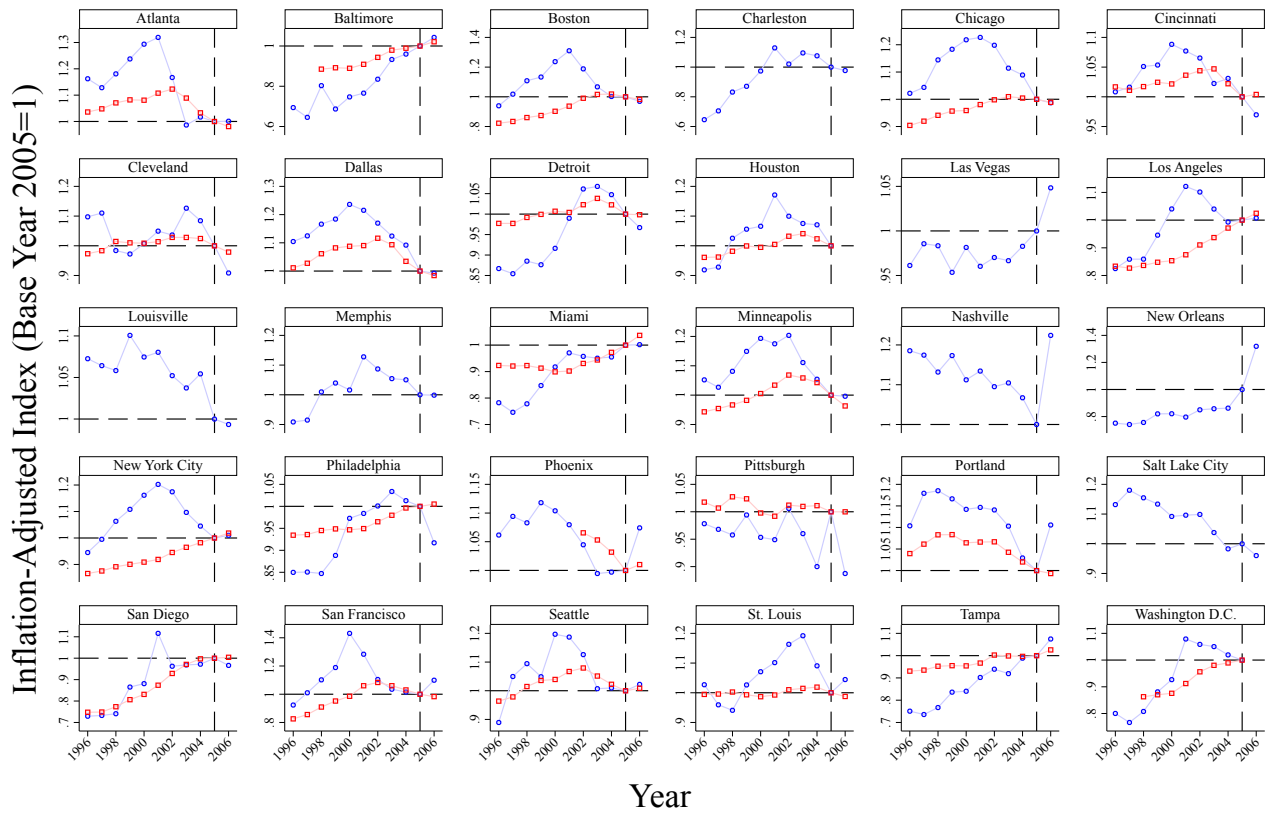Note: RW# means Rolling-Window of Window Size # and Step Size 1.

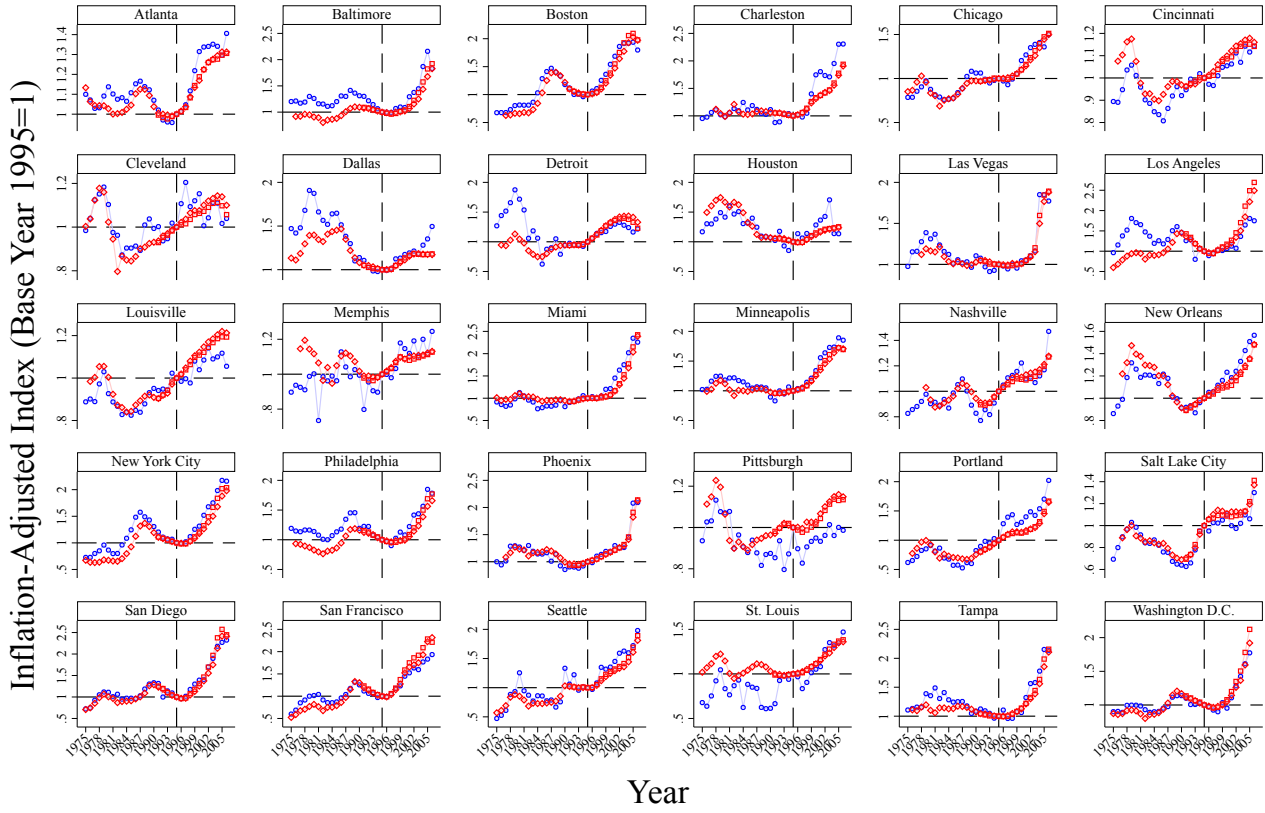# HiPHoP HPI, Geography ◂ Back
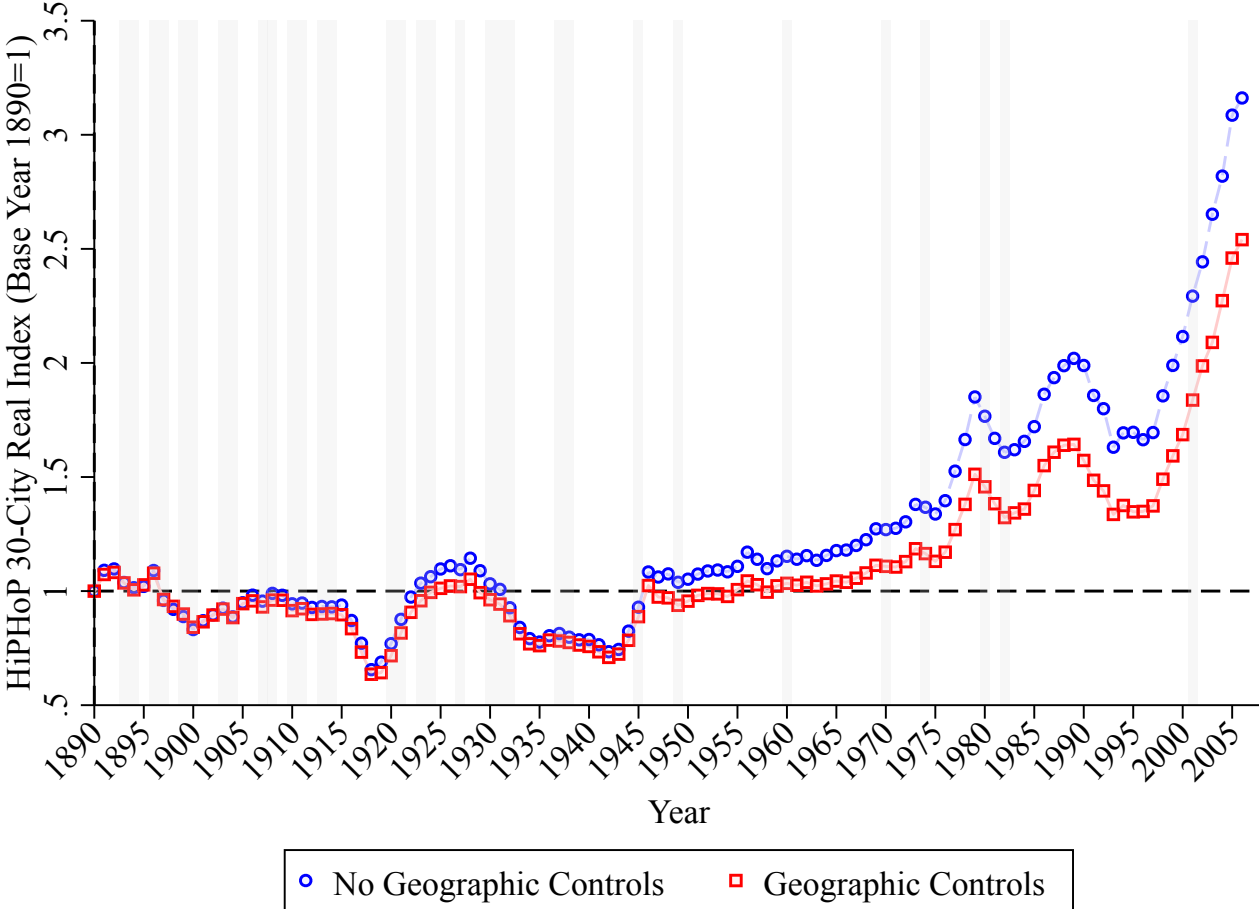
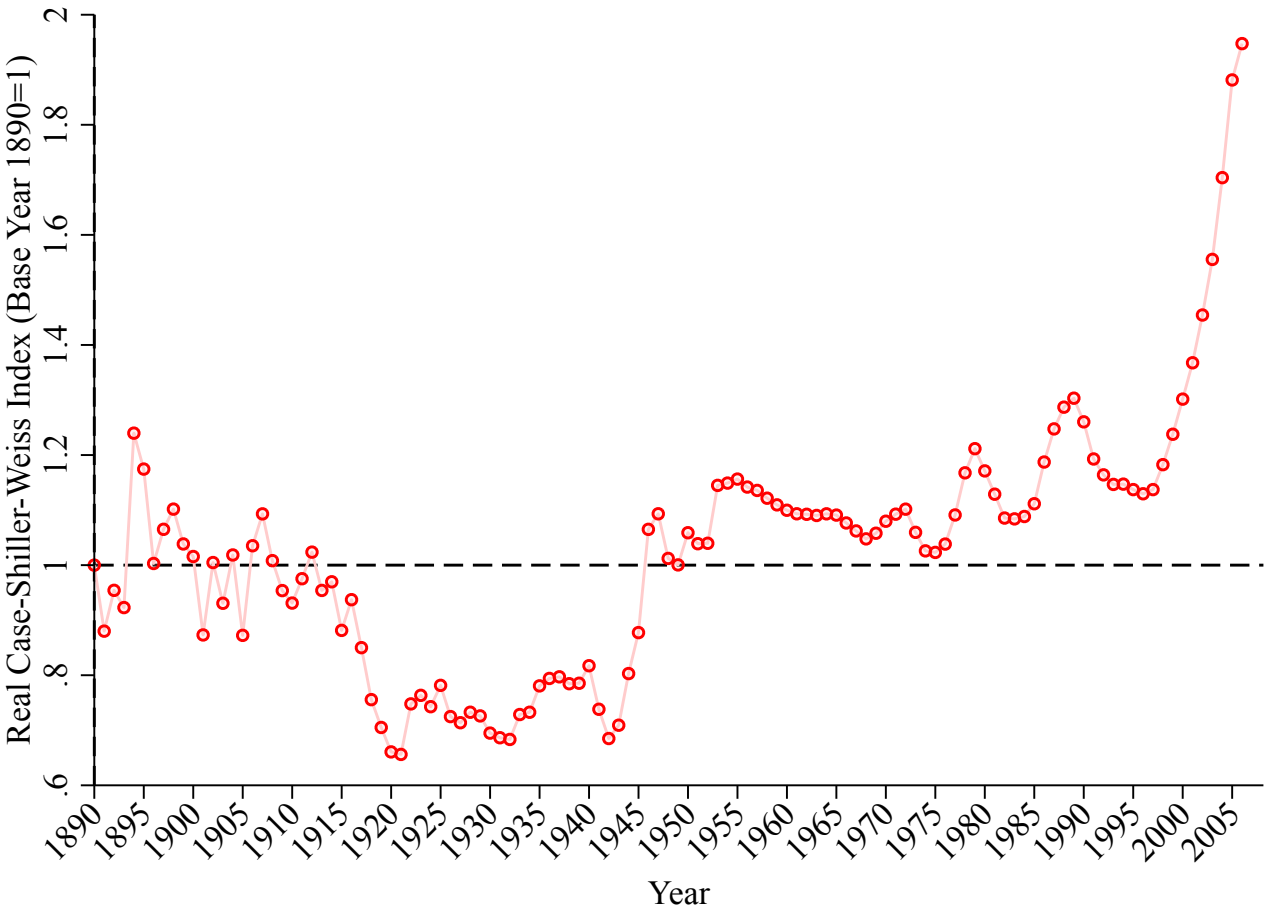# HiPHoP RPI, Geography ◂ Back
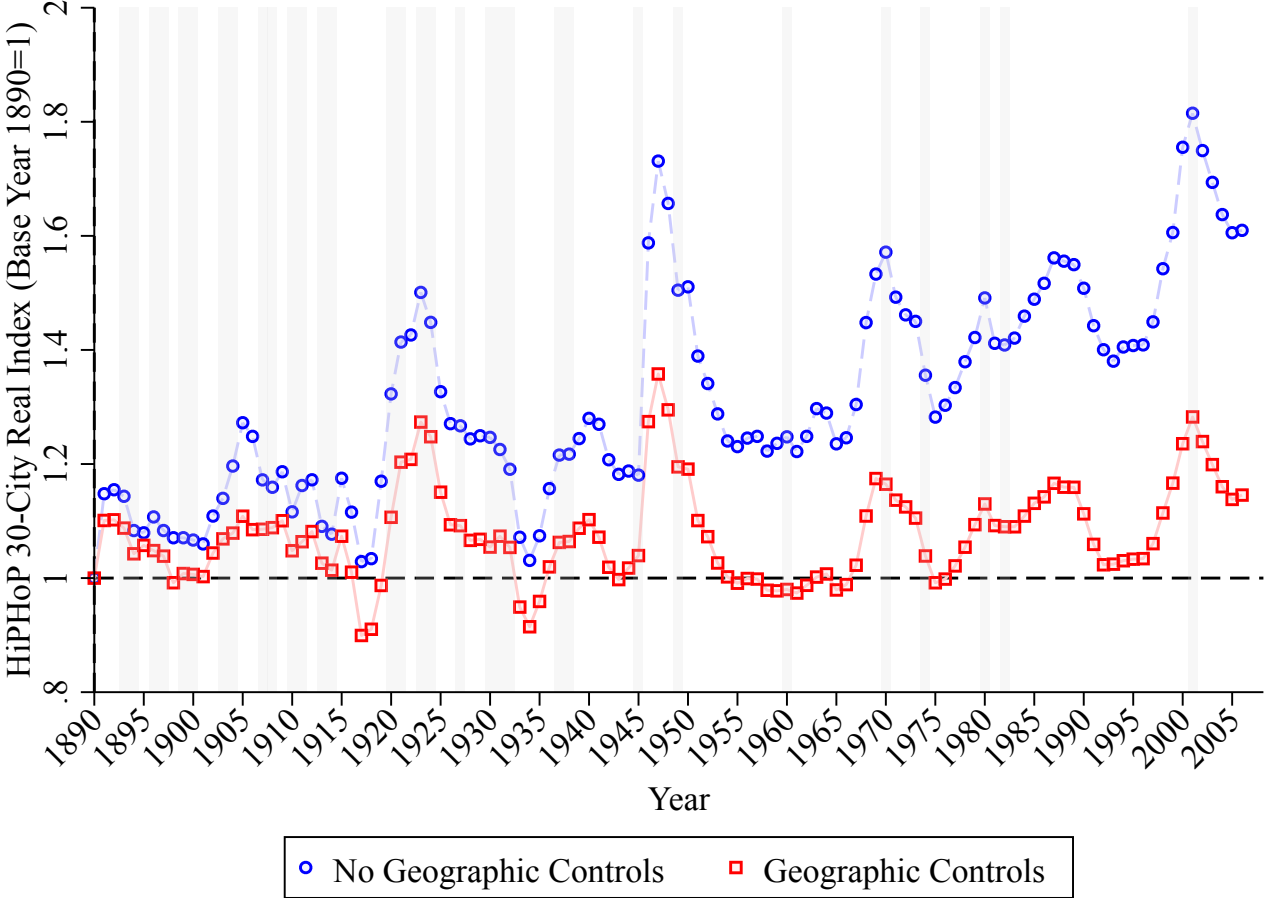
City-level comparison to BLS

City HiPHoP vs. FHFA: 1975-2006

Shiller Index (now CSW)

City-level HPI

City-level RPI

Real HiPHoP 30-City Index (Base Year 1948=1)

Year