ONLINE APPENDIX

Disentangling Rent Index Differences: Data, Methods, and Scope

by Brian Adams, Lara Loewenstein, Hugh Montag, and Randal Verbrugge

A RENT DATA SOURCES AND INDICES

A.1 The BLS Housing Survey and the CPI Rent Index

The BLS Housing Survey uses a multistage sampling design meant to draw a sample representative of rental expenditure.[1] The first stage selects large geographic areas called "primary sampling units" (PSUs). PSU definitions now match metropolitan and micropolitan statistical areas. Before the BLS redesigned its geographic sample in 2018, PSUs had been modified metropolitan statistical areas and groups of counties with smaller towns (Paben et al. 2016). Each PSU is subdivided into segments, which become the fundamental units for sampling and weighting. Segments are often Census block groups. Segments are selected using a probability-proportional-to-size (PPS) method, where "size" is an estimate of total shelter expenditure within the segment. Finally, the BLS randomly samples enough rental units to yield at least five responding units per segment.

The BLS selected a new sample in 1999. Subsequently, the survey lost units to demolition, to conversion to other uses, or to respondent non-cooperation.[2] The survey periodically added new units sampled from construction permit data. More recently, the BLS implemented a rolling sample replacement design, with a new sample drawn starting in 2012. Since 2016, units remain in the sample for only six years; one-sixth of the sample is replaced annually.

CPI rent is calculated using the average six-month change in economic rent in that month's sample, which is converted into a monthly change by taking its sixth root. Let $\text{rent}_i^*(t)$ denote economic rent. Then the rent index at time t for a particular index area is constructed as

$$I^R(t) = \left(\frac{\Sigma_i w_i \text{rent}_i^*(t)}{\Sigma_i w_i e^{F_{i,t}} \text{rent}_i^*(t-6)} \right)^{1/6} I^R(t-1) \tag{A.1}$$

where w_i is the weight for unit i,[3] and $F_{i,t}$ is an age-bias factor that lowers the rent level in period $t-6$ to account for the fact that the observed change in rent will understate the constant-quality change in rent.[4]

---

[1] For more details on the design of the Housing Survey sample see Ptacek (2013).

[2] Gallin and Verbrugge (2016) suggest that sample attrition was concentrated in higher-quality units; such attrition influences aging bias estimates, among other things.

[3] A unit's weight in the rent index depends on the estimated aggregate rent payments from its segment and the response rate for the segment.

[4] For more details on the construction of CPI rent see Verbrugge and Poole (2010) or the BLS Handbook

Indices are calculated for each index area, which is either a large PSU or the set of PSUs representing the smaller cities in a Census division. The national index derived from the average of changes in the index area indices weighted by rent expenditure in that index area. Until January 2023 the aggregation weights were updated every two years, so that the indices in year t are aggregated using expenditures from $t-1$ or $t-2$. Starting in January 2023, the aggregation weights are updated annually.

A.2 CoreLogic SFRI

The CoreLogic SFRI employs an arithmetic repeat-rent methodology using rental listings of single-family properties in the Multiple Listing Service (MLS). CoreLogic collects these data from participating Realtor boards. By 2020, CoreLogic had on average 10 years of history for these boards, and it had more than 20 years of data in some markets. CoreLogic creates rent indices for CBSAs for which it has sufficient data. The national SFRI is then a weighted average of the available CBSAs, where the weight is based on the value of the rental housing stock in each CBSA (Boesel et al. 2021).

A.3 Zillow Observed Rent Index

ZORI is a repeat-rent index that begins in 2014. It is based on Zillow's proprietary rental data from rental listings on its website and from MLS listing data. Its estimation methodology proceeds in three stages. First, Zillow estimates equation 1 in the main text unweighted. In the second stage, Zillow regresses the squared residuals from the first stage on weights created by comparing the distribution of structure type and age of rental properties in Zillow's data to that in the American Community Survey (ACS) in each respective year. The predicted values from this second stage are used in a weighted least squares regression of equation 1; this index forms the ZORI. Once constructed, the index is smoothed using a three-month exponentially weighted moving average.

Both SFRI and ZORI are based solely on the rents paid by new tenants, not tenants renewing a lease. The MLS data set underlying both is not representative of the general rental market. The Census's 2018 Rental Housing Finance Survey estimates that only 11 percent of single-unit rental properties are listed using a real estate agent (and thus listed in the MLS).[5] On average, rental listings in the MLS are more expensive, larger, and newer than newly occupied rental units in the AHS (see Table 1).

---

of Methods.

[5]See Choi and Young (2020) for the differential advertising strategies of landlords.

A.4 The ACY Marginal Rent Index

The ACY MRI of Ambrose et al. (2022) is based on the product of two series for large multifamily properties from Real Capital Analytics (RCA): the commercial property price index (CPPI), which is a repeat-transaction index, and the monthly average multifamily capitalization rate for transacting properties (or income yield). The product of these two series produces a baseline net rent index that is then re-scaled to match a former index created by the same authors from Experian RentBureau data, called the repeat-rent index (RRI),[6] which was discontinued in 2010.

The CPPI and the multifamily capitalization rate are based on RCA's database of commercial properties. The capitalization rate is based on the last month's net operating income for each property (not a historical average) and is therefore forward looking. However, while the database maintained by RCA is comprehensive, it is limited to properties worth at least $2.5 million or more. The data underlying the ACY MRI are therefore very different from those underlying either the SFRI or the CPI rent index.

B REPRESENTATIVITY: FURTHER DETAILS

In the main body, we briefly discussed the representativity of data underlying the SFRI, the ZORI, the MRI, and indices based on BLS data. In this section, we discuss the sample representativity of two other data sources, as well as other information pertinent to comparison studies like this one.

Why is sample representativity important? A non-representative sample is, effectively, a sample that has been conditioned on a variable, such as geography or structure type. (Equivalently, non-response bias is a chief concern in many contexts.) "Location-location-location" has been an aphorism in real estate since at least the 1920s, and rent growth can vary significantly within and across cities (Verbrugge and Poole 2010). Real estate markets are segmented by location, but also by structure type (Adams and Verbrugge 2021). Thus, rental market dynamics vary not only by location, but also by structure type (within a location). A data source that is restricted along one of these two dimensions will feature rent movements that may differ from the average.

[6]See Ambrose et al. (2015).

B.1 The American Housing Survey

In Table 1 we compare the BLS Housing Survey to the American Housing Survey (AHS). We discuss the construction of the AHS here.

The AHS is a longitudinal housing unit survey conducted biennially by the US Department of Housing and Urban Development in odd-numbered years and designed (after weighting) to represent the US housing stock (and not US housing expenditure). Based upon 1980 Census data, the national sample underwent a redesign in 1985, with a base sample size of approximately 47,000 housing units (owned and rented). However, few homes remained in the panel over its entire length; over the 1985-2013 period, 100,000 different homes were included.[7] In 2005, the national sample was improved in two ways: first, mobile home coverage was adjusted by replacing the units currently in the sample with mobile homes selected from Census 2000, and, second, assisted living housing units selected from Census 2000 were introduced into the sample. A new representative national sample of approximately 85,000 housing units was drawn for the 2015 AHS using the master address file (MAF) as the sampling frame, with additional oversampling of selected metropolitan areas and HUD-assisted housing units. The total sample size beginning in 2015 is about 115,000 housing units.

The AHS collects information about units' physical characteristics (including the physical condition of homes), information on neighborhoods, information on the characteristics of people who live in the homes, vacancies, home improvements, and housing costs. In Table 1, we use the national sample in 2015; the AHS sample was redrawn at this date, so the sample is discontinuous there.[8] AHS data do not identify whether utilities are included in the contract rent.

B.2 Geographic representativity of MLS and BLS Housing Survey

BLS data are representative of expenditures across urban areas in the US, and AHS data (and ACS data, to a somewhat lesser extent) are representative of housing units across the entire US. To convey a sense of the coverage of MLS data versus BLS data, Figure B.1 maps what locations are most sampled in Los Angeles, where both data sources have many observations. The BLS sample is concentrated in its selected segments, but these segments are spread throughout the metropolitan area.

[7] An interesting aspect of the AHS is that a housing unit's transitions between owner-occupied and renter-occupied are observable; see, for example, Foote et al. (2021). Conversely, in the BLS data, transitions from owner-occupied to renter-occupied occur outside of the sample, and a transition from renter-occupied to owner-occupied will typically imply that the unit drops from the sample.

[8] We do not have access to Zillow microdata or to RCA CPPI microdata, so their corresponding summary statistics are not included in the table.

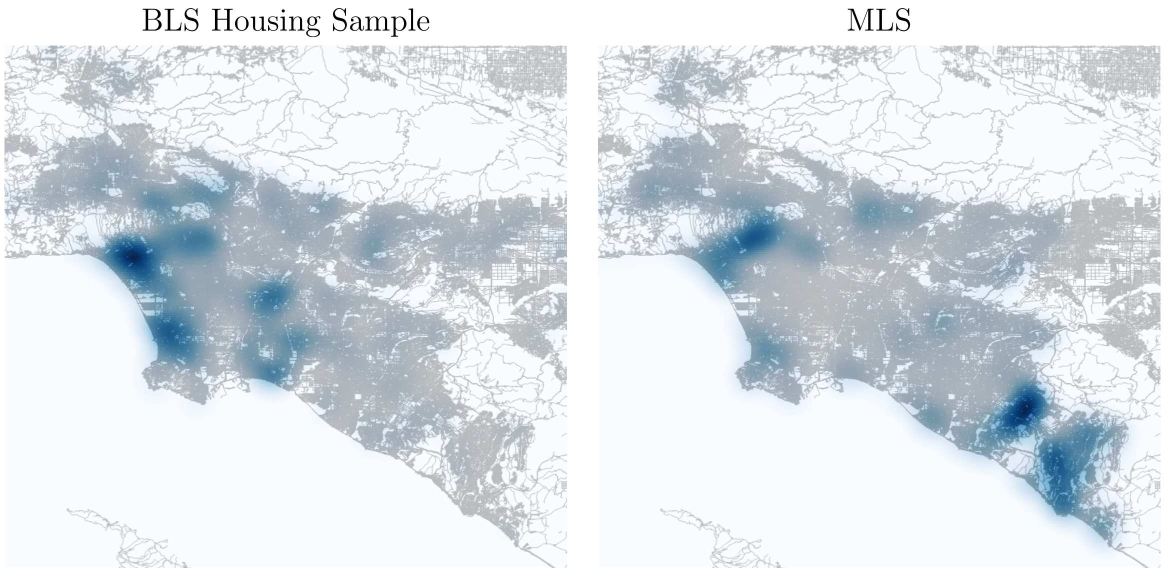

Figure B.1. Heatmap of Sample Locations in Los Angeles. *Note:* The left panel shows the geographic distribution of the sample in the BLS Housing Survey. The right panel shows the geographic distribution in the MLS data. *Source:* BLS Housing Survey and CoreLogic.

C MLS-BASED REPEAT-RENT INDEX

The CoreLogic SFRI is based on housing units listed for rent in the MLS. Because CoreLogic also provides access to the underlying data, these data are often used by researchers. While the rental rates and geographic dispersion of rental units in the MLS data are not representative (see Table 1 and Section B.2 above), our results suggest that over our sample period, the rent growth of properties listed in the MLS is representative. Most researchers use these data for a specific area. We therefore created a series of rent indices using the MLS data for areas that match the PSUs in the BLS Housing Survey. Our methodology is identical to that described in Section II, including that we remove any properties that the listing indicates were recently renovated or remodeled. We then compared the resulting MLS and CPI-data-based indices, and found that they consistently gave similar results — although the CPI-based indices are more volatile, reflecting their smaller sample size. Figure C.2 contains two examples. Our findings should provide some confidence to researchers who wish to use the MLS data to measure local rent growth.

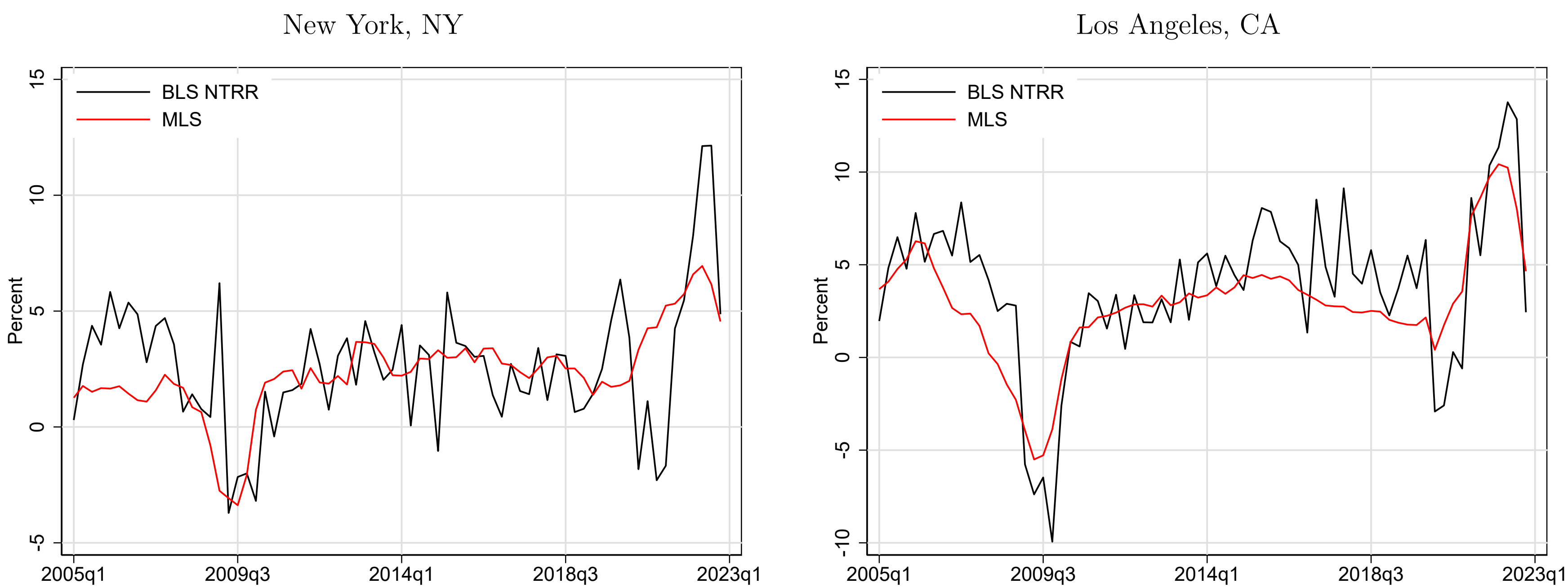

Figure C.2. PSU-LEVEL INDICES USING BLS NEW-TENANT DATA AND MLS RENTAL LISTINGS. *Note:* Areas are defined as the respective CBSA. *Source:* BLS Housing Survey and CoreLogic Multiple Listing Service data.

D CONSTRUCTING VARIANCE ESTIMATES FOR REPEAT-RENT INDICES

The BLS microdata are derived from a multistage sampling design. These data are then used to create a repeat-rent index, whose four-quarter growth rate is then computed. This is a nonlinear function of the data. In such cases, variances are unknown. To determine whether these indices are statistically indistinguishable, we estimate variances of the quarterly estimates using a bootstrap analysis, following Wolter (2007). These methods are even applicable to estimators deriving from complex sample survey designs.

The basic idea involves forming random groups by resampling housing units at random with replacement within each PSU in the BLS rental sample. We create k groups of housing units for each PSU, and then use the groups to create k PSU-specific repeat-rent indices. Next, we use the upper-level weights to aggregate the resampled PSU quarter-on-quarter changes to create k national repeat-rent quarter-on-quarter changes. The estimate of the variance $v\left(\hat{\theta}\right)$ in any given month is given by:

$$v\left(\hat{\theta}\right) = \frac{1}{k-1}\sum_{j=1}^{k}\left(\hat{\theta}_j - \hat{\theta}\right)^2, \tag{D.2}$$

where $\hat{\theta}$ is the average estimate across the k groups, and $\hat{\theta}_j$ is the estimate from group j, and we have suppressed the time subscript. Since k is not large in our applications, the confidence interval takes the form

$$\hat{\theta} \pm t_{k-1,\alpha/2}\sqrt{v\left(\hat{\theta}\right)}, \tag{D.3}$$

where $t_{k-1,\alpha/2}$ is the upper $\alpha/2$ percentage point of the t distribution.

E DYNAMIC RELATIONSHIPS AND CPI RENT FORECASTING IMPLICATIONS

To explore the dynamic relationships between the various rent index inflation rates, as well as to assess potential forecast gains for CPI rent using the SFRI, we estimate vector error-correction models (VECM) on pairwise sets of series. These highlight both the long-term relationship and their shorter-run dynamics. The VECM are specified as

$$\Delta y_t = \alpha(\gamma + \beta' y_{t-1}) + \upsilon + \sum_{i=1}^{3}\Lambda\Delta y_{t-i} + \epsilon_t \tag{E.4}$$

where $y_t = (y_{1,t}, y_{2,t})'$ is a vector of two rent indices (e.g., $y_{1,t} = \ln(\text{CPI rent}_t)$, $y_{2,t} = \ln(\text{SFRI}_t)$); $\Delta y_t = y_t - y_{t-1}$; γ and $\upsilon = (\upsilon_1, \upsilon_2)'$ are constants; Λ is a matrix of coefficients on lag terms; the parenthetical expression $(\gamma + \beta' y_{t-1})$, which is the (stationary) error-correction term, describes the long-term cointegration relationship between y_1 and y_2; and $\alpha = (\alpha_1, \alpha_2)'$ determines the speed at which each variable adjusts back toward this cointegrating relationship. We normalize β_1 to 1. We estimate these relationships on pre-pandemic data, to avoid overfitting based on one extreme episode. Table E.1 reports the values of α and β along with some standard errors.

Except in the ATRR-CoreLogic dynamic relationship, error-correction terms appear to play a notable role. In particular, in the other four pairwise relationships, the statistical significance of at least one α coefficient estimate suggests that error-correction terms will contain predictive content.

We explore the predictive content of the SFRI for CPI rent using the Bayesian information criterion (BIC). A stepwise model-selection search found the optimal model for quarterly CPI rent growth included lags 1 to 4 of quarterly CPI rent growth, lags 2 to 4 of quarterly CoreLogic growth, and the error-correction term. The BIC for this model is 1.610. Dropping the error-correction term from the model increases the BIC to 1.761. Finally, dropping all CoreLogic terms from the model yields a BIC of 2.218. These results indicate that the SFRI has strong predictive content for CPI rent, and that the error-correction term is likely a useful predictor.

Table E.1. Pairwise Vector Error Correction Results

	vecout				
	NTRR-CoreLogic	ATRR-CoreLogic	CPI Rent-NTRR	CPI Rent-CoreLogic	NTRR-ATRR
Cointegrating Equation					
γ	-0.780	-0.787	1.201	0.375	0.168
(Std Error)	(0.045)	(0.013)	(0.166)	(0.184)	(0.080)
β_1	1.000	1.000	1.000	1.000	1.000
β_2	-0.856	-1.156	-1.227	-1.074	-1.032
(Std Error)	(0.008)	(0.073)	(0.030)	(0.033)	(0.014)
Speed of Adjustment					
α_1	-33.224	-3.552	-13.483	-12.175	-101.688
(Std Error)	(17.078)	(5.186)	(2.424)	(3.488)	(34.030)
α_2	2.586	4.817	32.885	3.655	-0.388
(Std Error)	(11.398)	(8.684)	(20.386)	(10.938)	(10.976)

F ADDITIONAL FIGURES REFERENCED IN TEXT

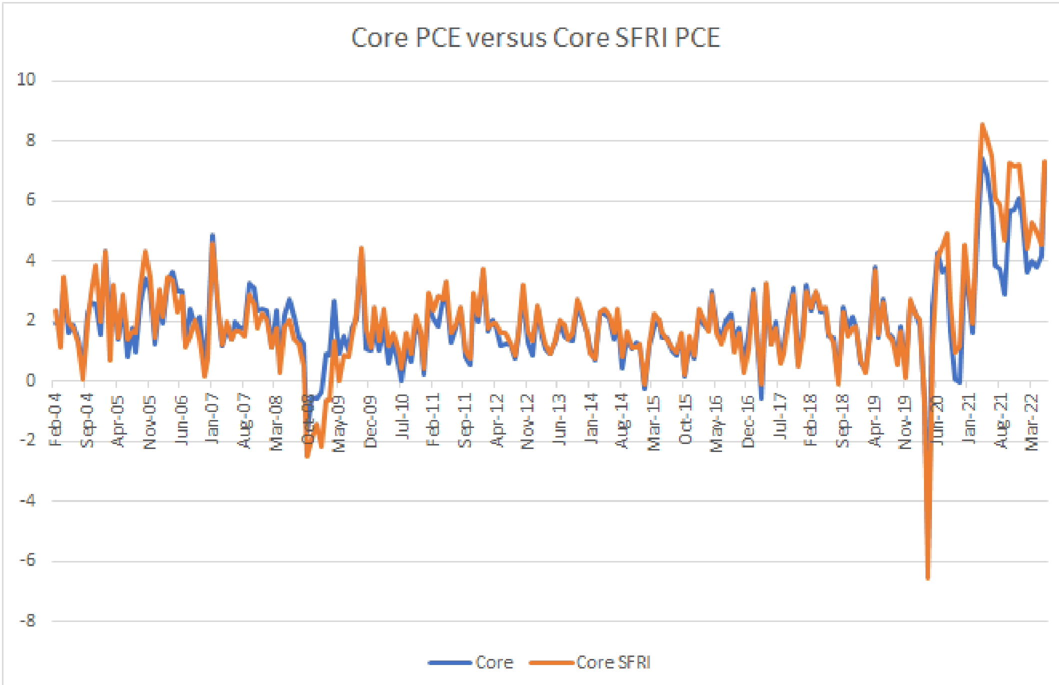

Figure F.3. Monthly Annualized Core PCE inflation versus Core-SFRI PCE inflation. *Note:* Core is standard core PCE inflation, which includes CPI rent inflation as a component. Core SFRI is an alternative inflation series that replaces CPI rent with rent inflation based on the SFRI . *Source:* Bureau of Economic Analysis, CoreLogic, and authors' calculations.

References

Adams, Brian and Randal Verbrugge (2021). "Location, location, structure type: Rent divergence within neighborhoods." Working Paper 21-03, Federal Reserve Bank of Cleveland. doi:10.26509/frbc-wp-202103.

Ambrose, Brent W, N Edward Coulson, and Jiro Yoshida (2015). "The repeat rent index." *The Review of Economics and Statistics*, 97(5), pp. 939–950. doi:10.1162/REST_a_00500. Publisher: MIT Press.

Ambrose, Brent W, N Edward Coulson, and Jiro Yoshida (2022). "Housing rents and inflation rates." *Journal of Money, Credit and Banking*. doi:10.1111/jmcb.12971.

Boesel, Molly, Shu Chen, and Frank E. Nothaft (2021). "Housing preferences during the pandemic: effect on home price, rent, and inflation measurement." *Business Economics*, 56, pp. 200 – 211. doi:10.1057/s11369-021-00241-4.

Choi, Jung Hyun and Caitlin Young (2020). "Owners and renters of 6.2 million units in small buildings are particularly vulnerable during the pandemic." *Urban Wire*. URL https://www.urban.org/urban-wire/owners-and-renters-62-million-units-small-buildings-are-particularly-vulnerable-during-pandemic.

Foote, Christopher L, Lara Loewenstein, and Paul S Willen (2021). "Cross-sectional patterns of mortgage debt during the housing boom: Evidence and implications." *The Review of Economic Studies*, 88(1), pp. 229–259. doi:10.1093/restud/rdaa034.

Gallin, Joshua and Randal J Verbrugge (2016). "Panel data estimates of age-rent profiles for rental housing." Working Paper 16-30, Federal Reserve Bank of Cleveland. doi:10.26509/frbc-wp-201630. URL https://www.clevelandfed.org/en/newsroom-and-events/publications/working-papers/2016-working-papers/wp-1630-age-rent-profiles.aspx.

Paben, Steven P., William H. Johnson, and John F. Schilp (2016). "The 2018 revision of the Consumer Price Index geographic sample." *Monthly Labor Review*. doi:10.21916/mlr.2016.47. URL https://www.bls.gov/opub/mlr/2016/article/the-2018-revision-of-the-CPI-geographic-sample.htm.

Ptacek, Frank (2013). "Updating the rent sample for the CPI Housing Survey." *Monthly Labor Review*. doi:10.21916/mlr.2013.25. URL https://www.bls.gov/opub/mlr/2013/article/updating-the-rent-sample-for-the-cpi-housing-survey.htm, institution: US Bureau of Labor Statistics.

Verbrugge, Randal and Robert Poole (2010). "Explaining the rent–OER Inflation divergence, 1999–2007." *Real Estate Economics*, 38(4), pp. 633–657. doi:10.1111/j.1540-6229.2010.00278.x.

Wolter, Kirk M. (2007). *Introduction to Variance Estimation*. Springer, New York, 2 edition. doi:10.1007/978-0-387-35099-8.