working paper 15 11

Determinants of Differential Rent Changes: Mean Reversion versus the Usual Suspects

Randal Verbrugge, Alan Dorfman,
William Johnson, Fred Marsh III,
Robert Poole, and Owen Shoemaker

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 15-11 July 2015

Determinants of Differential Rent Changes: Mean Reversion versus the Usual Suspects

Randal Verbrugge, Alan Dorfman, William Johnson,
Fred Marsh III, Robert Poole, and Owen Shoemaker

We study 2001-2004 and 2004-2007 rent growth of 18,000 rental units, ending our study prior to the Great Recession. Which variables correlate with rent growth: Location? Age? Rent level? Occupancy duration? Structure type? The answers deepen understanding of the rental market, help statistical agencies make decisions about sample stratification and substitution, and expose coverage problems. We document significant rent stickiness. Initial relative rent level is the best predictor, though mainly due to mean reversion. "Location" comes in second, though often not statistically significantly: the relative value of location is persistent. Age and occupancy duration are also notable. Our findings are reassuring to statistical agencies.

JEL codes: E3, L8, R11, R21, R31

Keywords: location, rent stickiness, mean reversion, inflation measurement

Suggested citation: Verbrugge, Randal, Alan Dorfman, William Johnson, Fred Marsh III, Robert Poole, and Owen Shoemaker, 2015. "Determinants of Differential Rent Changes: Mean Reversion versus the Usual Suspects," Federal Reserve Bank of Cleveland, working paper no. 15-11.

Randal Verbrugge is a the Federal Reserve Bank of Cleveland (randal.verbrugge@clev.frb.org). Alan Dorfman is at the Bureau of Labor Statistics (Office of Survey Methods Research) William Johnson, Fred Marsh III, Robert Poole, and Owen Shoemaker are at the Bureau of Labor Statistics (Statistical Methods Division). They thank David Brickman, Jim Follain, Ronald Johnson, Gregory Kurtzon, Frank Ptacek, Christina Wang, Elliot Williams, Tony Yezer, and seminar participants at George Washington University and the 2009 AREUEA Mid-Year Meeting. The comments of two anonymous referees and the editor helped to substantially improve this paper. All views expressed in this paper are those of the authors and do not reflect the views or policies of the Bureau of Labor Statistics (BLS) or the views of other BLS staff members.

This paper was previously circulated under the title "Determinants of Rent Changes: Location, Location, Location?"

1. Introduction

Why do rents on some units rise faster than others? Which variables are correlated with subsequent rent movements? Many would like to know the answer: academic economists who want to understand (and model) the dynamics of the rental market; real estate participants and practitioners who want to understand and forecast rental prices; policy analysts who want to study changes in the cost of living or the impact of desegregation; and financial investors who undertake real estate investments. Statistical agencies also want to know the answer, since it will aid decisions about sample stratification and sample substitution, and could help address concerns of academics and the larger public about the adequacy of sample coverage.

We might expect location to be the predominant, or perhaps only, economically- and statistically-significant predictor of rent change, since "location-location-location" is the answer to many real estate questions. While the influence of location on prices is unmistakable, people do not have preferences over a particular latitude and longitude per se. Rather, they have preferences over the implicit and explicit services offered by homes at various locations.[5] Most data sets do not have adequate information on the bundle of characteristics associated with different locations. Thus, while study after study finds that "location, location, location" is a useful rule of thumb (see, e.g., Green and Malpezzi, 2003), data unavailability will often lead to concerns about omitted variable bias. Might there be a good proxy for locational characteristics?

One candidate – particularly in the context of predicting rent change – is initial relative rent level. Rent level should proxy for unobserved desirability; high-end apartments in the best part of town are more desirable (even after controlling for number of rooms, number of baths, and so on). As Goodman (2004) puts it, rent is "a single-dimensional summary of the market's valuation of all the physical, service, and locational attributes of an apartment." It is reasonable to expect that markets are differentiated or segmented, at least to some extent, by rent levels. For example, the upper end of the market might respond much more to changes in homeowner user costs, since the less affluent – who rent cheaper apartments – are probably not making a rent-

[5] Having said this, Goodman (2005) finds that the "zone" geographic identifiers in the American Housing Survey metro files explain more of the variance in rent than do a set of specific neighborhood attributes and assessments relating to safety, services, and amenities; moreover, adding these neighborhood attribute variables did not increase the explanatory power of specifications that already included the zone variables. Below we discuss how the variance in rent level is related to the variance in rent change.

versus-buy decision. Alternatively, urban development might concentrate on the upper end of the market. Hence, we might expect rent dynamics to differ by rent level.[6]

We study determinants of rent changes, investigating the practicality of rent level as a proxy for quality or segment of market. While the literature studying the determinants of rent *levels* is large (see, e.g., Green and Malpezzi, 2003), there is very little research studying differential rent *changes*, even though understanding rent change is at least equally interesting, particularly for market participants and for statistical agencies.[7] To conduct our analysis, we use microdata from the Bureau of Labor Statistics (BLS) Consumer Price Index (CPI) rental housing sample. We study 2001-2004 and 2004-2007 rent changes experienced by 18,000 rental units in 87 cities. When linked to Census neighborhood data, these data are expansive in both geographic and intertemporal detail – and unlike many other data sources, BLS data are sampled from the entire distribution of rental units, not just from larger professionally-managed properties. Our data set thus appears to be well-suited for such an inquiry.

[6] Three previous studies have uncovered a correlation between initial rent levels and subsequent rent inflation. In the context of determining the bias associated with different price index formulas, Shoemaker (1997) noted this correlation. McCarthy and Peach noted this correlation during their presentation of their work at the 2003 Brookings Conference on hard-to-measure services. Finally, Goodman (2005) constructed rent inflation indexes for the 1988-2000 period by market segment (bottom, middle, top) for various cities. Verbrugge (2008) constructed ex ante user costs, and compared these to overall rent inflation, and to rent inflation experienced by detached units; the rent measures were very similar, and both diverged markedly from user costs.

[7] This is not to say that the issue has never been discussed. The traditional monocentric “standard urban model” was developed in the 1960s. This influenced much of the early (1970s) work on hedonic house price indexes, which highlighted the role of distance from the central business district; urban economists offered many stories to explain why, for example, prices in suburbs might grow more or less rapidly than those in the central city. Theoretically, changes in the desirability of various locales – caused by trendiness, gentrification, the construction of new highways, crime, or other such factors – will cause differential price and rent movements; demographic shifts, such as reduced family sizes or aging, can alter the valuation of features such as additional bedrooms. Changes in regulation will affect the bottom of the market differently than the top (Green and Malpezzi, 2003), and a small ‘agent-based’ literature investigates complex local dynamics (see, e.g., Diappi and Bolchi, 2008). Still, the empirical literature has almost exclusively focused on differences in prices at a point in time, or looked at trends in aggregates. Shimizu, Nishimura and Watanabe (2010) and Gallin and Verbrugge (2014a) study rent stickiness and other dynamic features of rents, but do not focus on cross-sectional differences in rent inflation. In the one published study we were able to locate – Clayton (1998), a study of real estate market efficiency – rent inflation differences across regions of Vancouver are reported, but not discussed further. As noted previously, Goodman (2005) documents different rent inflation at different parts of the rent distribution. Regarding the paucity of research, Malpezzi, Pollakowski and Simmons-Mosley (2005, p.2) comment: “Interestingly, and perhaps surprisingly, little research has been done on what [the distribution of rent changes] looks like.” That study estimates the cross-section standard deviation of rent changes in 20 metropolitan areas (and makes reference to another unpublished paper, Follain, Kogut and Marshoun 2000); these volatility estimates are not the focus, but are rather used to parameterize a simulation study, so the salient point for our study is simply that they report non-zero estimates. Rent inflation divergence is perhaps implicit in, e.g., Saiz (2003). The empirical rent control literature is arguably related to this paper; Turner and Malpezzi (2003) is a survey of empirical research.

We focus on three-year rent changes. We do this for two reasons. First, we wish to abstract from short-run rent movements that may be transitory. One of our chief questions concerns bias in inflation measurement; this question requires a clear signal. Second, we wish to observe at least one rent change for each unit (although as we will see, even over this horizon a significant fraction of rental units do not experience a rent change – and this rent rigidity plays a role in our results).

We find that initial relative rent level is indeed a powerful predictor of subsequent rent changes. However, once one controls for a form of regression to the mean, the explanatory power of relative rent level falls significantly. Relative rent level is a "contaminated" proxy, either for segment of the market or for quality. (We discuss the endogenous relationship of rent level, location, quality, and market segment below.)

We find that location, age, and occupancy variables are the variables most reliably related to differential rent changes, that provision of utilities exerts a significant influence on rent change over this period, and that size (as proxied by the number of rooms) and structure type may also be related to rent change. Our findings regarding age, change in occupancy, and utilities support adjustments performed by statistical agencies, as explained below. Our findings also suggest that rental markets are segmented to some extent – by locality, rent level, and so on.

But our findings also suggest that the characteristics observable in our data – including location – have fairly modest power in explaining differential rent growth. Within a city, location has a statistically-significant relationship to rent growth in only about half of the cities investigated.[8] Other variables fare worse. In this sense, *none* of the variables included can really be considered a powerful explanatory variable, even though in any given city or time period, several variables might have had a statistically-significant relationship to rent growth.

We reach three chief conclusions. First, a seeming paradox: even though our location variable is often unrelated to rent change, our findings actually *support* the prevailing notion of "location, location, location". Why? As explained below, the results suggest that the relative desirability of location is generally quite persistent; a location which is currently desirable generally remains desirable three years later. If location had instead been strongly related to

[8] This finding initially appears to be somewhat at odds with the findings of Poole and Verbrugge (2008). We discuss the apparent contradictions below. Other studies, such as those of Valente et al. (2005) and of Hwang and Quigley (2009), emphasize the micro-spatial dimension in the *level* of real estate prices, but are not focused on studying changes in the cross-sectional distribution.

differential rent change (after controlling for other covariates), then this finding would suggest substantial (and rapid) *change* in the relative desirability of different locations. Second, our findings suggest caution in interpreting results when trying to use initial rent level as a proxy either for quality or for segment of the market. (This insight probably applies to other contexts as well.) Third, our findings are reassuring to statistical agencies that employ the rental-equivalence method (see Diewert 2008 for an exposition) and that follow current best practices, i.e., that primarily stratify their rental housing sample via location, that adjust for utilities provision, and that undertake aging and vacancy adjustments. As discussed below, there is little evidence that this set of methods is deficient and should be abandoned. Rent inflation in geographically-proximate units appears to be an adequate proxy for the change in the value of the flow of services received by homeowners, even when structure type, size, and rent level differ.[9]

Section 2 describes the data. Section 3 discusses the specification. Section 4 presents results: it identifies mean reversion, suggests a modification of the rent-level measure to address this problem, and presents results based upon this alternative measure. Section 5 concludes.

2. Data

2.1 Description

The CPI housing sample is the source of information on changes in the price of housing services for the two principal shelter indexes in the CPI, the residential rent index and the owners' equivalent rent (OER) index.[10] This sample is a stratified cluster sample that, at its onset, consisted of approximately 40,000 rental units.[11] The overall sample is divided into six panels, with rental units in a given panel surveyed every six months, resulting in two data collections per year per rental unit in the sample.

[9] Internal empirical research by the BLS has been uniformly supportive of this operational assumption.
[10] For more details about how the BLS produces these indexes, see Ptacek and Baskin (1996).
[11] Over time, the sample suffers a loss of units. Conversely, over this period the sample was augmented based on building permit data obtained from the Census Bureau and from canvassing of areas not requiring building permits.

When a unit is initiated into the sample, the BLS collects information on unit characteristics. For example, the BLS microdata contain the following information on each rental unit in its sample:

- location
- type of structure – mobile home; detached; semi-detached (duplex or townhome); multiple-unit dwelling with elevator; multiple-unit apartment dwelling without elevator; other.[12]
- year built
- duration of occupancy
- number of rooms
- utilities types, and contracted utilities provision

In subsequent measurement months, BLS field representatives gather data on the unit's structural changes, changes in the provision of utilities and amenities, duration in months of occupancy (if occupied), and rent. The data are of high quality. Odd-looking or abnormal reported rents are carefully investigated by BLS staff to determine if they are valid. Rent prices are adjusted if, for example, the tenant provides labor services to the landlord in exchange for reduced rent.

Our sample derives from a period during which the BLS used a sampling procedure that relied upon decennial Census data. This procedure explicitly involved geography. There were three steps: dividing each city into large regions, termed "strata;" selecting very small regions, termed "segments," from each strata; and selecting rental units from each segment.

In the first step, each Primary Sampling Unit (PSU)[13] was divided into six geographic regions, labeled "strata." The center-urban portion of the PSU was first divided into two strata, and then the remainder of the PSU was divided into four strata. Each stratum is geographically contiguous. In the next step, each stratum was divided into "segments," which correspond to Census blocks or block groups. Then segments were randomly selected from each stratum, using probability-proportional-to-size procedures. In the final step, the BLS would randomly select

[12] "Other" includes mixed commercial and residential structures, such as an apartment above a store front or professional office, and apartments in or above garages. It accounts for less than 1% of the sample. The sample excludes mobile homes that are not on permanent foundations or blocks, and houseboats which are not permanently moored.

[13] PSUs are urban areas spanning one or more cities; they vary in size, from huge to relatively small. For example, PSUs in the current BLS sample include: Chicago-Gary-Kenosha, Houston-Galveston-Brazoria, Anchorage, Buffalo-Niagara Falls, and Yuma, Arizona. Precise definitions are available in the BLS Handbook of Methods, Chapter 17, Appendix 5, at http://www.bls.gov/opub/hom/pdf/homch17.pdf.

housing units to visit, with the goal of finding at least five, and perhaps a few more, rental units within each segment. (For more details, see Ptacek and Baskin, 1996.)

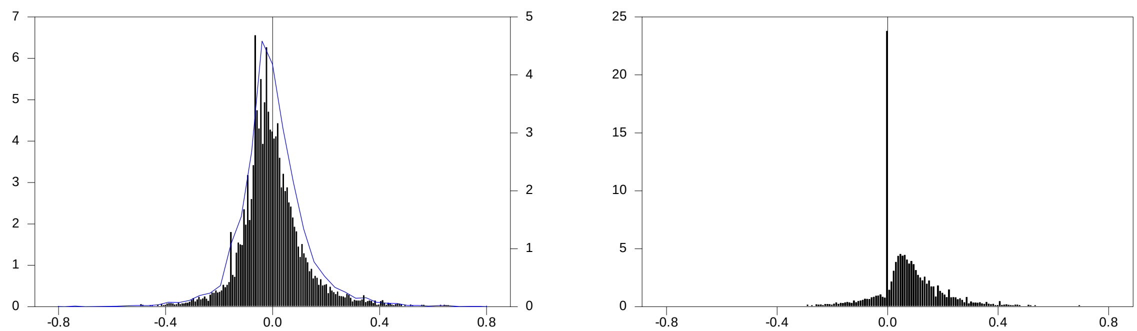
By merging the BLS data with publicly-available Census data, we are also able to determine associated (year 2000) demographic variables related to location, such as proportion of owner-occupied housing in the immediate neighborhood (Census block group), and the median income of the Census tract in which each unit is located.

Our sample was restricted so as to ensure that the data were suitable for this analysis. We retained in our data only those observations for which we have a reliable measure on each characteristic. Then, to be included in the 2001-2004 analysis, the unit had to have a valid rent in 2001 and 2004; and to be included in the 2004-2007 analysis, the unit had to have a valid rent in 2004 and in 2007. During both periods, our data consists of rental information pertaining to about 18,000 rental units, located in 87 PSUs.

2.2 Basic data characteristics

Over the 2001-2004 period, average rent growth in our sample is 7.6%. For each PSU, we also computed the standard deviation of rent growth, as well as its skewness and kurtosis. Across the 87 PSUs, the average standard deviation of cross-sectional rent growth is 12%, while average skewness is 0.84 and average kurtosis is 7.5. Over the 2004-2007 period, average rent growth in our sample is 7.8%. Across PSUs, the average standard deviation of rent growth is 12%, average skewness is 0.93, and average kurtosis is 8.6. In Table 1, we report some sample statistics related to unit characteristics which are taken from 2004, which is the midpoint of our sample. Figure 5 in the Appendix plots a histogram of the 87 rent growth standard deviations for each of the two time periods.

There is significant rent growth dispersion and significant rent stickiness. To illustrate this, we depict the distribution of the log-rent changes in the 2001-2004 data below in Figure 1. The left-hand figure presents the distribution of scaled rent changes; in particular, there we plot the distribution of $\left\{\Delta^{01-04}\ln r_i\left(t\right)-\Delta^{01-04}\ln \overline{r_k}\left(t\right)\right\}$, where $\Delta^{01-04}\ln r_i\left(t\right)$ is the 2001-2004 rent growth rate of unit *i*, and $\Delta^{01-04}\ln \overline{r_i}\left(t\right)$ is the 2001-2004 rent growth rate of the particular PSU *k* within which unit *i* is located. We plot both the estimated density and the histogram; the density is estimated using the Epanechnikov kernel function. For the right-hand figure, we instead plot the histogram for the raw data, i.e. $\Delta^{01-04}\ln r_i\left(t\right)$. Notice the large proportion of units that experience rent growth of zero. This finding, together with the shape of the left tail of the rent growth distribution, suggests downward price rigidity in rents, a point which has been emphasized by Genesove (2003); Gallin and Verbrugge (2014a) is a recent empirical study of rent stickiness (see also Hoffman and Kurz-Kim 2006), and Gallin and Verbrugge (2014b) provide a theoretical explanation of rent stickiness. We depict only those observations that lie between -0.8 and +0.8, which means that 40-odd outliers are trimmed in each case.



Panel A: Centered, with Estimated Density

Panel B: Raw Histogram

Figure 1: Distribution of 2001-2004 Rent Growth, Outliers Removed

3. Specification

This paper is primarily a hedonic regression analysis of rent changes (for recent book-length treatments of hedonic analysis in the real estate context, see Coulson (2008), or see Aizcorbe, forthcoming). In hedonic regressions, a measure of price – or price *change* (here, rent change) – is regressed upon supposed explanatory variables.[14] What explains rent? People rent a unit to obtain its varied services: both services specific to the *structure* (such as “a roof over their heads,” water, heat, a convection oven, etc.), and also services or amenities specific to that *location* (such as parking, local schools, a nice view, proximity to the highway, etc.). The equilibrium rental price obviously depends upon supply and demand. Supply of rental property in most given locations is relatively, but not *completely*, inelastic in the short run: housing

[14] “The job of hedonic analysis in real estate … is to investigate the relationship between the existence and amount of all of these characteristics–structural and locational–and the price that people are willing to pay for the unit.” (Coulson, 2008)

construction takes time, apartment buildings are not demolished overnight, and so on. Still, in many locations, there are various owned properties which could be converted to, or from, rental housing in response to price changes. Furthermore, renovation can change the structural characteristics of existing units. Demand for specific rental property ultimately depends upon demand for the various services provided by that location's properties.

Because price is determined by supply and demand, hedonic regression coefficients need not have a straightforward structural interpretation. This is related to a classic problem: to truly identify a demand curve, one must have truly exogenous variation in supply. In the context of this paper, the coefficients on age terms are a classic example of coefficients without a straightforward interpretation. These age coefficients represent an admixture of physical depreciation, demand effects such as desirability of vintage, and unobserved quality (see Clapp and Giacotto (1998) and Gallin and Verbrugge (2014c).) A lack of a structural interpretation is a serious problem in some contexts; for example, the focus of a particular study might be estimating physical depreciation, or understanding the chain of causation. Obtaining a structural interpretation will typically require a scientific theory, whether explicit or implicit.

In the present context, the lack of structural interpretation poses much less difficulty, since our more modest goal is to determine whether or not there exist covariates that are reliably associated with differential rent changes, with the chief purpose of determining whether existing methods used by statistical agencies are flawed. Further, because we do not necessarily need a structural interpretation of any given coefficient, we can even include endogenous regressors in our hedonic estimation. We include several of these variables both because they might be important correlates of differential rent changes, and because we wish to more clearly isolate the explanatory power of the remaining variables.[15]

[15] Rent level, location, quality and market segment are all probably jointly determined. This suggests high collinearity among the observables. Moreover, of these four variables, only rent level is measured without error. But even if we observed all four perfectly, we could only truly identify the causal influence of each variable with the use of a structural model, which is beyond the scope of this study. From a statistical standpoint, high collinearity may lead to imprecise coefficient estimates on individual regressors, and may lead one to falsely conclude that a variable is not important. (If the covariates are jointly powerful, we would see high regression R^2 despite low t-statistics.) Since – as in any hedonic study – we are not attempting to identify structural parameters nor make causal statements, such collinearity need not invalidate the investigation, although its presence would necessitate care in interpreting individual coefficient estimates and t-statistics. The Appendix 6.3 robustness check suggests that high collinearity is not a problem in this study. Having said that, our location variables are fairly coarse, and we suspect that we would find higher collinearity if our location variables were more refined.

To motivate the rent-change specification, assume that the underlying data-generating process can be adequately approximated by the model[16]

$$\ln r_i(t) = \alpha(t) + \tilde{X}_i \tilde{\beta}(t) + \left(U_i \delta(t) + v_i(t)\right),$$

where: $v_i(t)$ is an idiosyncratic influence on rent which is identically distributed, independent across units, and tentatively assumed to be independent across time; $\tilde{X}_i$ is a vector consisting of measured characteristics applicable to unit *i* (specified below); U_i is an unobserved characteristic ("quality") of unit *i* (assumed to be a mean-zero IID random variable that is observed by market participants, that is not present in our data, and that reflects the part of quality that is not reflected in the covariates in our data); and the coefficient vector $\tilde{\beta}(t)$ and the coefficient $\delta(t)$ are not assumed fixed over time.

Since the characteristic U_i is unobserved, in an empirical specification this variable is subsumed in the residual $\varepsilon_i(t)$, as in:

$$\ln r_i(t) = \alpha(t) + \tilde{X}_i \tilde{\beta}(t) + \varepsilon_i(t), \tag{1}$$

where

$$\varepsilon_i(t) \equiv U_i \delta(t) + v_i(t).$$

In this study, observed characteristics $\tilde{X}_i$ for each unit include of a measure of its age (namely, a dummy variable for decade built); a measure of its location;[17] a dummy variable for its structure type; a set of five rooms variables: number of bathrooms (and bathrooms^2), number of bedrooms (and bedrooms^2), and number of other rooms; and the length of occupancy of the renter at the beginning of the period. A second occupancy variable is also included, namely a "new tenant" dummy, which takes the value of one if a new tenant moved into the unit during the period, and zero otherwise. $\tilde{X}_i$ also includes four Census neighborhood variables pertaining to the unit: percent "white, non-Hispanic" in unit *i*'s Census block-group, percent of the aged-25-and-over population with at least some college in unit *i*'s Census block-group, percent renter in unit *i*'s Census block group, and median income in unit *i*'s Census tract. The characteristics vector also

[16] Again, we do not claim that the models in this paper have a structural interpretation.
[17] In a PSU-level regression, the location variable is a dummy variable indicating unit *i*'s stratum (defined in Section 2.1); in a pooled regression, there are two location dummy variables, one indicating the PSU that the unit is in, and the other indicating whether the unit is in the center-urban part of the PSU.

includes two dummy variables related to rent-control. The first, the "rent control" dummy, takes the value of one if the unit was under rent control during the entire period, and zero otherwise. The second, an "off-rent-control" dummy, takes the value of one if the unit came off rent control during the period, and zero otherwise.

Finally, $\tilde{X}_i$ also includes three additional measures that allow us to address concerns central to CPI construction. The first two measures are related to CPI aggregation weights for two different shelter inflation indexes, the Rent index and the OER index. Rent index growth derives from a (complicated) weighted average of changes in the market rents of individual rental units. And since the CPI in the US (and in many other countries) uses the rental equivalence method for estimating homeowner cost changes,[18] OER index growth also derives from a (complicated but different) weighted average of changes in the market rents of individual rental units. The rental unit sample is roughly the same for both indexes, but the two indexes employ different weights, reflecting the fact that (for example) a rental unit in a neighborhood that is predominately owner-occupied should have a greater influence on the OER index than on the Rent index. In keeping with this, the first additional measure in $\tilde{X}_i$, "relative Rent weight," reflects unit *i*'s relative importance (or relative aggregation weight) in the construction of the Rent index. Similarly, the other additional measure, "Relative OER weight," reflects unit *i*'s relative importance in the construction of the OER index. A larger weight indicates that, relative to other units in the PSU, unit *i*'s inflation has a larger impact on the overall PSU-wide inflation measure. We wish to investigate whether rent growth over this period has any relationship to these weights, holding other characteristics constant. (We discuss the third additional measure, namely relative rent level, below.) A finding of statistical significance would suggest that, all else equal, "owner-intensive" neighborhoods experience different rent growth than do "renter-intensive" neighborhoods.

When unit characteristics do not change over time, (1) leads to the following model for log rent-change:[19]

[18] The rental equivalence method assumes, not that the rent *levels* of rental units are good proxies for homeowner costs, but rather that the rent *changes* experienced by rental units are good proxies for homeowner cost *changes*. In other words, the assumption is that the implicit rent on an owned home rises at the same rate as do nearby market rents. For details about CPI shelter index construction, see Ptacek and Baskin 1996 and Poole and Verbrugge 2010.

[19] We explored other specifications, including using two-year log-rent changes (2001-2003) as the dependent variable, using rent relatives rather than log changes as the dependent variable, alternative means of describing the percent renter, etc.; none of these choices were consequential regarding inference.

$$\begin{aligned}\Delta \ln r_i(t) &= \Delta\alpha(t) + \tilde{X}_i \Delta\tilde{\beta}(t) + \Delta\varepsilon_i(t) \\ &\equiv \phi + \tilde{X}_i \tilde{\theta} + e_i(t)\end{aligned} \tag{2'}$$

where $e_i(t)$ is a mean-zero error term. We investigate three-year rent change, so that for the 2001-2004 period, $\Delta \ln r_i(t) = \ln r_i(2004) - \ln r_i(2001)$, and other variables are defined analogously. As this is an analysis of rent *change*, statistically-significant coefficient estimates will implicitly reflect *changes* in demand and supply conditions. If the relative equilibrium valuation of a particular characteristic (say number of rooms) does not change over the three-year period, then its coefficient estimate will not be statistically distinguishable from zero.

In fact, we actually use a different specification, namely

$$\Delta \ln r_i(t) = \phi + \tilde{X}_i \tilde{\theta} + \gamma \bullet dev_i + e_i(t) \tag{2}$$

The variable *dev* is the third additional measure. This variable refers to *relative rent level* of the unit. More specifically, if we are considering unit *i* in PSU *k*, then $dev_i \equiv \left(\ln r_i - \overline{\ln r_k}\right)$, which is the deviation of unit *i*'s log-rent from its PSU-average log-rent. This variable has two potential interpretations.

First, when used in the context of a regression like (2), this variable could be thought of as a proxy for unit quality: all else equal, a unit with a high *dev* value is a unit with high unobserved quality. If the coefficient estimate on this variable is not statistically significant, then we might conclude that unobserved quality is not an important issue. Second, this variable could be thought of as a proxy for segment of market: a high-*dev* unit is an expensive unit. Its inclusion in (2) could help answer the question: do different parts of the rental market distribution experience different inflation rates? For example, we might observe lower rent growth in the higher rent segment of the market. Such a rent change pattern might occur due to changes in supply, such as the construction of numerous high-end units. Or, such a rent change pattern might occur due to changes in demand, such as favorable conditions for buying rather than renting. If the coefficient estimate on this variable is not statistically significant, then we might conclude that the rental market is not appreciably segmented by rent level. In Sections 4.1 and 4.2, we investigate the usefulness of *dev* for purposes like these. In practice, the inclusion of *dev* led to surprising findings and additional insight into rent dynamics.

It may initially seem odd to include a variable like *dev* in a rent-change regression. But in this rent-change context, one can sensibly augment the list of independent variables in (1) with this variable.[20][21] In (2), we intentionally omit a date designation for this variable, for reasons which will be discussed in Section 4.1.

Model (2) was applied to data from each PSU (separately), and also to a pooled data set that includes data from all the PSUs. For the pooled-data regressions, we used specification (2), but used a different set of location dummy variables. In particular, we include PSU dummy variables in the pooled regressions, but the only other location variable is a dummy variable indicating city-center location.

Note again that the dependent variable in this specification is a rent *change*, rather than a rent *level*. Many studies of rental markets focus on how different characteristics influence the *level* of unit rent. We instead study the correlation between various characteristics and rent *changes*, an analysis that answers different questions. For example, as noted above, our findings may tell us about rental market segmentation. Furthermore, inflation measurement is obviously related to rent change rather than rent level. Note that low adjusted R^2s are common when modeling changes over time, and should not be interpreted as a symptom of poor model specification.

The null hypothesis is that all rental units, regardless of characteristics, share the same inflation rate; i.e., $H_0 : \tilde{\theta} = 0, \gamma = 0$. (In other words, though a unit's particular characteristics influence its rent level in 2001 and in 2004, these characteristics do not influence the *change* in rents between 2001 and 2004.) Alternatively, finding that $\theta_m \neq 0$ would imply that rent inflation is instead related to characteristic *m*, i.e. units whose characteristic *m* differed from its average would experience different-from-average rent inflation. As noted above, note that $\theta_m \neq 0$ would also imply that β_m in (1) has changed over time; thus, the estimated marginal impact of that characteristic on equilibrium rent has changed.

---

[20] One can think of (1) as providing an estimate of the function $r(X)$, rent as a function of characteristics X. Conversely, (2) can be thought of as providing an estimate of $r'(X)$, the derivative of $r(.)$ with respect to time. In the latter context, it is logically coherent to talk about how the derivative varies as a function of the level of the function.

[21] Why not simply use initial rent level, rather than a deviation? Consider the most relevant case, when dev_i refers to *initial* rent deviation. For a particular city taken in isolation, initial rent level would capture the same information as initial rent deviation; it would, for example, allow one to partition units into high-end, medium-end, and so on. However, in some regressions, we pool data from all cities. Using initial rent level would then no longer convey the notion of high-rent-vs-low-rent, since average rent levels vary so much across cities.

The simple "location-location-location" hypothesis would appear to suggest a non-zero estimate only for the "geography" coefficient. Perhaps surprisingly, the opposite might be more nearly true. This is because the relative value of location is persistent. For location to have an influence on rent *growth*, the value of location – such as the value of living in the Northeast region of the city – must *change* rapidly relative to the value of living elsewhere in the city over the three-year period.[22] Thus, "location-location-location" actually suggests that location should have little explanatory power for differential rent growth. (One must bear in mind, of course, that in the pooled regressions, the location dummy variables consist merely of city indicator variables and a dummy variable for "central city." This aggregate control for location does not adequately capture within-PSU differences across market segments. In the city-by-city regressions, location dummy variables consist of strata dummy variables.)

Since we care about inference, it is important to verify that the OLS assumptions are not rejected by the data. We found little evidence for heteroskedasticity in specification (2), but the residuals displayed significant kurtosis. For the purposes of inflation measurement, it is important to retain all valid data, but for the purposes of estimating regression coefficients, it is important to ensure that the results are not being driven by outliers. We report results based upon two rounds of trimming using studentized residuals, removing at each stage all observations whose studentized residuals were greater than 2.5. [23] This trimming resulted in the removal of about 7% of the sample on average, and reasonably satisfactory residual properties.

[22] In our cross-section regressions which pool data from the entire country, many PSUs have a statistically-significant coefficient estimate on the PSU-wide indicator variable. This implies that rent growth in that city differed significantly from average rent growth in the country, reflecting differing city-wide demand and supply conditions over the period. This finding obviously supports the conventional location hypothesis.

[23] Even after quite aggressive outlier treatment, the Shapiro-Wilks test indicated that normality of the estimated residuals was easily rejected in 38 PSUs. However, as outlier removal became more and more aggressive, inference remained quite stable. Heteroskedasticity did not appear to be a problem in these data; using White's test, it was rejected in 77 of the 87 PSUs. Results were insensitive to different outlier removal methods.

4. Mean Reversion, Rent Rigidity, and Results

4.1 Initial Versus Final Rent Deviation (in 2001-2004 Data)

In this subsection and the next, we study *dev*. We present selected results from 2001-2004 data. Recall that rent deviation at period *s* for unit *i* in PSU *k* is given by $dev_i(s) \equiv \ln r_i(s) - \overline{\ln r_k(s)}$, the deviation of unit *i*'s log-rent from its PSU-average log-rent at time *s*.

First, we investigated using initial rent deviation as the measure of relative rent level in (2). Loosely speaking, we thus asked whether high-end units experienced more or less inflation during the period. If "rent level" were irrelevant for rent inflation, we would expect no relationship between rent growth and rent level.

We next investigated using final rent deviation in (2). This variable raises concerns about interpretation, since final rent deviation is admittedly problematic from the endogeneity perspective: units experiencing above-average rent inflation will end up with a higher level of *dev*. However, over long periods of time, a high-end unit is a high-end unit, irrespective of whether we make this determination at the beginning or the end of the sample period. In any case, using final rent deviation in (2) is, at least, useful as a diagnostic device.

We focus on the statistical significance of the *dev* coefficient estimate. Figure 2 displays, for each of the 87 PSU regressions in the 2001-2004 period, the estimated regression coefficient on the $dev_i(2001)$ coefficient – graphed in order of increasing magnitude – along with ± one standard deviation bands. Initial rent level is found to be statistically significant at the 5% level in 60% the PSUs. The estimated impact is quantitatively very large: the average estimated coefficient is -0.09. Estimating (2) using pooled data from all PSUs yields a coefficient estimate on $dev_i(2001)$ of -0.046, with an associated t-statistic of -20. There is no doubt about the statistical significance of this variable. As to its economic significance, coefficient estimates imply that, after conditioning on all the other variables, the highest initial rent units in a PSU experienced 4-9% less inflation over the subsequent three years than the average. There are two apparent conclusions. First, initial rent level is a powerful predictor of future rent change. Second, high-end units experienced significantly lower inflation over the period.

These findings are, however, misleading. To illustrate the issue, consider using final rent deviation. Given the "a high-end unit is a high-end unit" principle, we might expect to confirm our findings above: high-end units would experience lower rent inflation. But using *final* rent deviation $dev_i(2004)$ as the measure of relative rent level suggests the *opposite* conclusion, namely that high-end units experienced significantly *higher* inflation. Figure 3 displays, for the 87 PSU regressions, the estimated *t*-statistics on the $dev_i(2004)$ coefficient. Final rent level is found to be statistically-significant at the 5% level in 64% the PSUs. The estimated impact is again very large: the average estimated coefficient is +0.09. In pooled data, the coefficient estimate was +0.05, with an associated t-statistic of 23.

These apparently contradictory results likely stem from rent rigidity and a form of mean reversion, as we explain next.

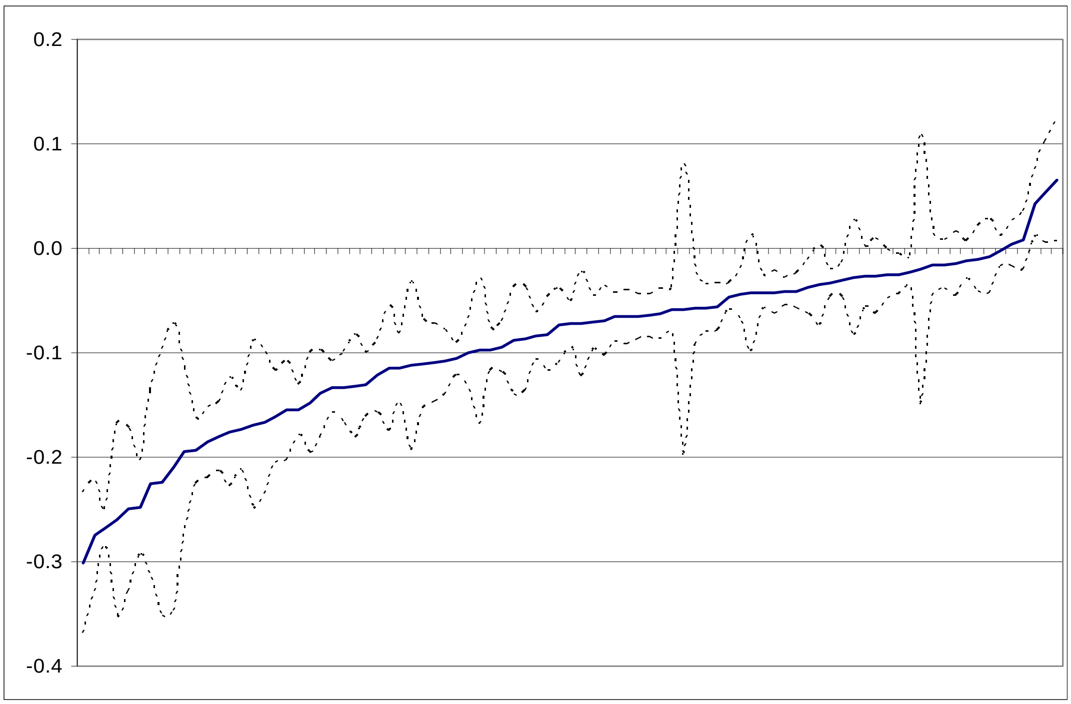


Figure 2: Ordered estimated $dev_i(2001)$ coefficients with ± 1 std. deviation bands

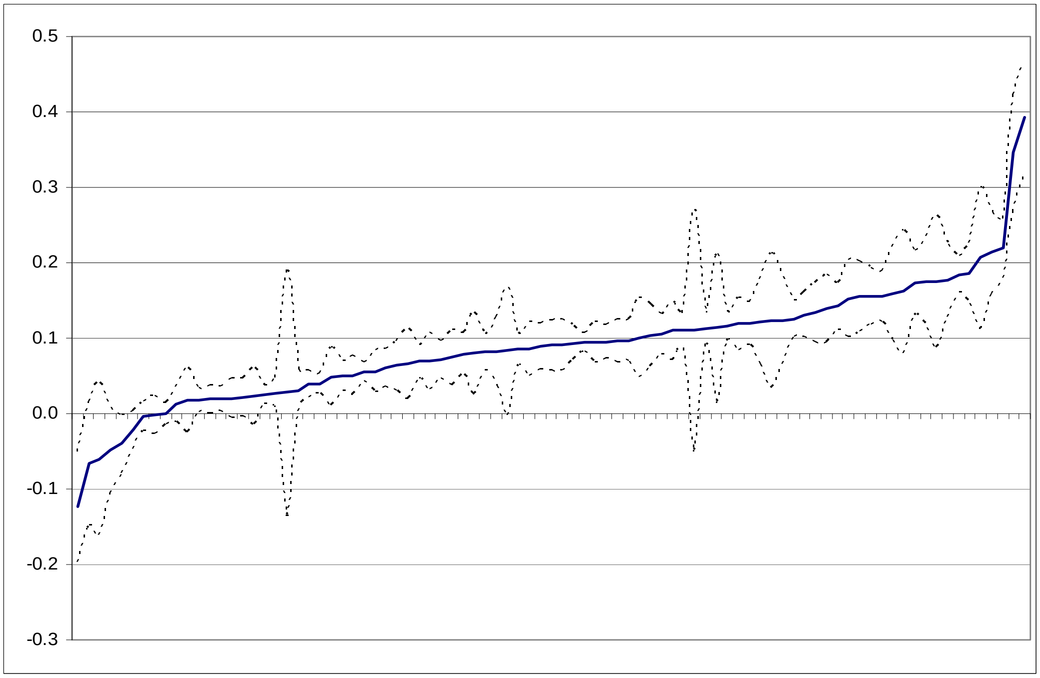


Figure 3: Ordered estimated $dev_i(2004)$ coefficients with ± 1 std. deviation bands

4.2 Mean Reversion

To explain these results, write out $dev_i(2001)$, using (1), as:

$$\begin{aligned} dev_i(2001) &= \left[\tilde{\alpha}(2001) + \tilde{X}_i\tilde{\beta}(2001) + \left(\delta(2001)U_i + v_i(2001)\right)\right] \\ &\quad - \left[\tilde{\alpha}(2001) + \bar{X}_i\tilde{\beta}(2001)\right] \\ &= \left(\tilde{X}_i - \bar{X}_i\right)\tilde{\beta}(2001) + \delta(2001)U_i + v_i(2001) \end{aligned}$$

where we have used the fact that both U_i and v_i are mean-zero random variables.

Now consider the error term in specification (2):

$$\begin{aligned} u_i(t) &\equiv \Delta e_i(t) \equiv e_i(2004) - e_i(2001) \\ &= U_i\Delta\delta + v_i(2004) - v_i(2001) \end{aligned} \tag{3}$$

Neither $dev_i(2001)$ nor $dev_i(2004)$ is independent of $u_i(t)$; in particular,

$$E\left(\left[dev_i(2001)\right]\left[u_i(t)\right]\right) = \delta(2001)\Delta\delta EU_i^2 - Ev_i^2(t) + Ev_i(t)v_i(t+3)$$

and

$$E\left(\left[dev_i(2004)\right]\left[u_i(t)\right]\right) = \delta(2004)\Delta\delta EU_i^2 + Ev_i^2(t) - Ev_i(t)v_i(t+3)$$

(If $v_i(t)$ is independent across time, the final term in each expression vanishes.)

These two expressions allow us to interpret the results in Section 4.1, and to assess the ability of *dev* to capture rental market segmentation or quality. Consider three key cases. First, suppose that $\Delta\beta$, $\Delta\delta$, and all second moments of $v_i(t)$ were all negligible. In that case, $dev_i(2001)$ and $dev_i(2004)$ would essentially coincide, and their troublesome correlation with the error term would essentially vanish, so that $\hat{\gamma}$ in (2) (using either *dev* measure) would suitably indicate how different rental market segments experienced different rental inflation rates. Second, if $v_i(t)$ were distributed IID with a large variance – or if $v_i(t)$ were strongly negatively correlated over time, as suggested in Section 4.4 below – then all else equal, this would simultaneously induce high negative correlation between $dev_i(2001)$ and $u_i(t)$, and high positive correlation between $dev_i(2004)$ and $u_i(t)$. Why? Consider unit *i*. Intuitively, if $v_i(2001) >> 0$, this likely implies $dev_i(2001) >> 0$. But by 2004, the influence of $v_i(2001)$ either goes away or is actually reversed. Hence, unit *i* will likely experience below-average inflation. The same argument holds in reverse for unit *j* with $v_j(2001) << 0$. Similarly, if unit *k* experiences $v_k(2004) >> 0$, this likely implies $dev_k(2004) >> 0$, and also likely implies that unit *k* experienced above-average inflation; and the same argument holds true in reverse if $v_k(2004) << 0$. This phenomenon is a form of mean-reversion, or regression to the mean.[24] Third, if $\Delta\delta$ were sufficiently large and positive (negative), this will induce positive (negative) correlation between the error and both *dev* measures even if the second moments of $v_i(t)$ were not insignificant.

It follows that the empirical results noted above are informative. Why? Since $\hat{\gamma} < 0$ and statistically significant when using $dev_i(2001)$, while $\hat{\gamma} > 0$ and statistically significant when using $dev_i(2004)$, we may reach three tentative conclusions. First, $\Delta\delta$ is probably not of overwhelming importance. Second, the second moments of $v_i(t)$ are probably sizable. The third conclusion follows: neither $dev_i(2001)$ and $dev_i(2004)$ can really identify the "segment of the market," nor can either one proxy for unobserved unit quality *U*.

[24] There is evidently no standard terminology to refer to this sort of phenomenon. Efron (2011) refers to a closely related phenomenon as "selection bias." In that context, units experiencing a large positive transitory shock were more likely to be "selected" into a "high" category …but he also states that it is "an example of regression to the mean." Liu, Moon and Schorfheide (2014) follow Efron in using the terminology "selection bias."

One could "remove" the part of the correlation between *dev* and the error term that is induced by the second moments of $v_i(t)$ by setting *dev* in (2) equal to the *average* of $dev_i(2001)$ and $dev_i(2004)$. However, this does not address the correlation stemming from the EU_i^2 term, so we do not pursue this approach here.[25] Instead, bearing in mind the caveats above regarding induced correlation, we use initial rent deviation, a predetermined variable, hereafter.

4.3 Determinants of Differential Rent Changes

Rather than reporting thousands of coefficient estimates and their standard errors, for brevity we confine ourselves to reporting some summary measures. We use specification (2), with the initial dev_i measures as discussed above.

Table 2 reports, for the 80-odd PSU regressions in the 2001-2004 and 2004-2007 periods,[26] the percentage of PSUs for which the exclusion test on the variable or variable group in question had an estimated *p*-value smaller than 5%. Many of the variables listed are actually groups of dummy variables; e.g., age, type of structure, and location fall into this category.[27] We remind the reader that statistical significance is not equivalent to economic significance. Some coefficient estimates are quite precise, but small in magnitude; others are estimated very imprecisely, but are much greater in magnitude.

[25] In a working paper version of the paper, we investigated other rent deviation measures, including the average noted here. (We also use an average deviation in one of our Appendix exercises.) Using any of the other measures, evidence for differential rent change by rent level was relatively weak: during either time period, relative rent level was not found to be statistically significant in two-thirds of the cities investigated. In the Appendix, we explore the use of year-2000 rent deviation for 2001-2004 rent changes, with similar results. Qualitative conclusions about other variables are mostly robust to using different measures of rent deviation, with an exception noted below.

[26] Two or three of the small cities did not contain enough sample to reliably estimate all the coefficients.

[27] The location variables merit further discussion. In PSU-by-PSU regressions, we use strata rather than segment (or Census tract) dummy variables for two reasons. First, many segments have only three or fewer units by 2004, so that segment dummy variables would be both imprecisely estimated and overwhelmingly costly in terms of degrees of freedom. Using Census tract dummy variables would not help, since in many cases there is only one sampled segment in each tract. For example, Chicago has over 850 Census tracts, but there are less than 500 rental units in the BLS sample, in 120 segments. Second, segments can be dominated by units from the same apartment complex. Usually a segment has about five rental units. It would not be interesting to discover that similar units within the same apartment complex share similar rent growth. (This motivates our "No Multiple Units" robustness check in the Appendix.) In *pooled* regressions, we use PSU dummy variables but must aggregate strata, so that strata 1 and 2 become a "city center" variable. As noted previously, we aggregate strata because "stratum 4"in Chicago has no relationship to "stratum 4" in Los Angeles, but we can still meaningfully ask whether city centers experienced slower rent growth. Chang and Coulson (2001), for example, find that central city employment dynamics can be quite different from those in the suburbs, and it is interesting to ask whether central cities are declining relative to their suburbs, or conversely being revitalized.

Aside from initial rent level, the three variables most often correlated with rent change (in order) are location (i.e., stratum), age (i.e., decade built), and the length-of-occupancy or tenure variables. The other two variables of note are the number of rooms[28] and structure type. However, for all five of these variables, the associated coefficient estimates are statistically significant at the 5% level in only about half of the PSUs. Of course, high collinearity among covariates might lead to imprecise coefficient estimates. But the results of the robustness exercise outlined in Appendix 6.3 suggest that collinearity is of modest importance. For example, even without any other covariates, location appeared to be statistically significant in less than 30% of the PSUs; and structure type did not appear to be of more importance when the set of other covariates is drastically reduced.

Regarding the sign of coefficient estimates, in general it is hard to discern a clear pattern across PSUs. The only variables whose influence appears to be broadly consistent across time periods and across PSUs are age and the two length-of-tenure variables.[29] In particular: Older units typically experience more inflation, a phenomenon we discuss below; units which turn over typically experience greater inflation; and conditional on not turning over, rent inflation is positively associated with length of tenure.[30]

Tables 3 and 4 present results from pooled regressions. In these tables we report selected coefficient estimates, along with estimated standard errors, *t*-statistics and *p*-values. The specification is not identical to that for Table 2, in that these pooled regressions include 86 PSU dummy variables, but – as noted above – reduce the number of within-PSU location variables from five to one. For brevity, these tables do not report coefficient estimates on the 10 individual decade-built dummy variables, the five rooms variables, or on the 86 individual PSU dummy variables.

There are three cautionary notes. First, these regressions are unweighted, which means that the implicit weight of a particular PSU will be proportional to its rent sample size, so bigger

[28] Recall that these estimates are conditional on initial relative rent level, which itself proxies for size.
[29] During the 2001-2004 period, three other variables exhibited a consistent pattern of influence across PSUs (percent renter, rent weight, and structure type detached), but during the 2004-2007 period they did not.
[30] Ideally, perhaps, we would follow Ambrose, Coulson and Yoshida (2013) and Ozimek (2014) and focus attention on newly-signed leases. This might provide a clearer picture of how valuations are changing. However, Gallin and Verbrugge (2014a) show that structure type is strongly correlated with length of time between lease renewals; results might end up being driven by rent change characteristics in large apartment complexes. Furthermore, using only newly-signed leases would drastically reduce the sample size.

cities are somewhat more influential.[31] Second, since the only within-city location variable is "city center" and "other," these pooled regressions do little to capture within-city differences across locations. Third, this pooling imposes a homogeneity assumption, namely that any particular variable has an identical influence on rent change across all PSUs. If the assumption is true, then this pooling will give rise to more accurate estimates of the influence of that particular variable. Conversely, if the assumption is false – if the influence of that particular variable across PSUs is actually heterogeneous – then this pooling will lead to misleading coefficient estimates. For example, consider the estimates on the rooms variables. These estimates will be misleading if several PSUs experienced a relative shortage of large apartments, while several other PSUs experienced a relative shortage of small apartments.

Results from pooled regressions are broadly consistent with the PSU-by-PSU patterns. In these all-US regressions, many of the variables contribute to explaining differences in rent inflation, in the sense that coefficient estimates are statistically significant, as conventionally defined. Most have muted economic significance: numerical magnitudes are generally modest compared to the ±7% inflation differences across PSUs. (Recall, though, that coefficient estimates represent the average influence of a particular variable across PSUs.)

There are a few variables whose impact is undeniably large during one period or the other. But only PSU (location), tenure variables, utilities provision, and rooms stand out as consistently strong influences.[32] Regarding the other location variable in these regressions, did the desirability of the central city versus the periphery change for many PSUs over one or both of these three-year periods? Perhaps; but if so, the sign of this change must have differed across PSUs, leading to the overall insignificant coefficient estimate. Even age does not appear to be a consistently strong influence, in that it does not appear to be important in the 2004-2007 period. (See Gallin and Verbrugge (2008, 2014c) for more evidence on the influence of age, which is not monotonic.) Neighborhood percent renter is a predictor with a large estimated impact in the 2001-2004 period, although the sign of the estimated coefficient flips during the 2004-2007 period. (As there are even fewer PSU-specific location variables in this pooled regression, it is possible that percent-renter is among a group of variables acting as location proxies.) Removal of

[31] The PSU sample size varies from a minimum of about 55 (~0.3% of the total) to a maximum of about 1020 (~5% of the total), although only sixteen PSUs have more than 300. The median is about 190 (~1% of the total).
[32] There is some reason to doubt the consistent influence of rooms, since in 2001-2004 data the result is not robust across alternative specifications of *dev* (see working paper version).

rent control has the expected large impact in the 2001-2004 period, but its effect is not discernible during the next period. Structure type appears to have a fairly modest influence overall influence, even though many structure-type variables have coefficient estimates which are statistically significant at conventional significance levels. Of the eight possible coefficient estimates (four structure types × two time periods), seven are negative, and four of these are statistically significant; this pattern suggests that garden-style apartment units (which comprise the majority of the omitted structure type) experienced modestly higher inflation during both periods. Conversely, mobile homes experienced significantly lower rent inflation during both periods. Detached units experienced slightly lower inflation than did other structure types during the initial period, but not during the second period. Finally, the provision of utilities (heat or electricity) added roughly 1-2% additional inflation.

4.4 Discussion

There are six chief findings. Rents are sticky. Initial relative rent level seems to matter and to lead to one conclusion, but final relative rent level seems to matter and yet to lead to the opposite conclusion. Both aging and a change in occupancy matter. Provision of utilities matters. Location matters, although perhaps less than might have been expected. Finally, differential rent changes are not fully explained using variables that are observable in our data set. What are the implications of these findings?

Several of the facts presented above, when combined with theory and evidence from other studies, have implications for rent dynamics in general, and for $v_i(t)$ in equation (1) in particular. We confirm that rents are sticky and frequently remain unchanged for long periods of time (see Figure 1); continuing tenants often enjoy below-average rent inflation, at least for a while.[33] We also confirm that rent growth experienced by newly-vacated units substantially exceeds average rent growth. (For more on these and other prominent patterns in the data, see Rivers and Sommers 1983, Genesove 2003, Crone, Nakamura and Voith 2009, and Gallin and Verbrugge 2014a.).

[33] However, we note that rent inflation is positively correlated with length of tenancy; there is evidently a distinction between shorter-tenure continuing tenants and longer-tenure continuing tenants. Eventually, even rather sticky rents will adjust.

But how do these facts, combined with our results regarding *dev*, relate to the properties of $v_i(t)$? When the rent level on a unit is reset, optimal behavior dictates that landlords take any expected rigidity into account. Suppose that the optimal rent on a unit is expected to grow over time. Then, when a landlord has a chance to reset the rent, she will set the new rent well *above* the current optimal level, since she is trying to keep the actual (rigid) rent as close to the optimal level as possible over the period during which the rent will be rigid (see Krsinich and Verbrugge (2015) for a theoretical treatment). But this large-rent-reset behavior implies that $v_i(t)$ has a large variance and has negative correlation over time. Why? Units which recently experienced a rent reset are more likely to have an above-average rent – a positive $v_i(t)$ – while units which have not experienced a rent reset for a long time are likely to have a below-average rent – a negative $v_i(t)$. Furthermore, units which experienced a rent reset just prior to the beginning of the period are, all else equal, somewhat less likely to experience a rent reset during the period. Conversely, units which did not experience a rent reset just prior to the period are, all else equal, somewhat more likely to experience a rent reset during the period.[34] Hence, we conclude that $v_i(t)$ is not IID with small variance, but instead has a large variance and negative correlation over time.

As discussed above, relative rent matters mainly because of a form of mean reversion. Hence, caution is required in interpreting results if one is trying to use relative rent level as a proxy either for quality or for segment of the market.

Three of the findings support standard statistical agency rent inflation adjustments. Regarding the importance of utilities, our findings underscore the necessity of utilities adjustments in estimating the value of shelter service flow (see Verbrugge 2012). Regarding change of occupancy, our results underscore the necessity of vacancy adjustments in order to avoid bias when a vacant unit drops out of the sample (see, e.g., Rivers and Sommers 1983 and Crone, Nakamura and Voith 2009, and see Krsinich and Verbrugge 2015 who derive results in a theoretical treatment of sticky rent dynamics).[35] Regarding aging, our results underscore the necessity of applying aging bias adjustments, adjustments which themselves must vary with age

[34] We are indebted to Christina Wang for this suggestion. Including the variable "length of occupancy" in the regression specification does not fully control for this phenomenon.

[35] Gabriel and Nothaft (2001) study rental vacancy incidence and duration using BLS data; Deng, Gabriel and Nothaft (2003) study duration of residence, using AHS and BLS data, and find hazard rates are time-dependent.

(see, e.g., Randolph 1988 and Gallin and Verbrugge 2008, 2014a). In keeping with other studies, we find that the influence of aging is relatively small.

How do we explain the apparently modest importance of location – and indeed, of most of the variables – and what is implied? As noted above, differential rent changes are not easily explained using variables that are observable in our data set. But these largely negative findings have some interesting and important implications, particularly for statistical agencies.

Regarding structure type, the statistical significance of *detached* housing is of key interest, since the BLS employs the rental-equivalence method for estimating the shelter cost inflation experienced by homeowners. Most owned housing is detached, but only 21% of the BLS rental sample is detached. Thus if detached housing experienced dramatically different rental inflation than did other units, critics might interpret this finding as a flaw in the rental equivalence method, or at least a flaw in the sampling procedure.[36] However, the detached-unit coefficient estimate is statistically significant at the 5% level for only about 20% of the PSUs. Moreover, in pooled data, while its influence is statistically-significant (though small) in the 2001-2004 period, its influence is undetectable in the 2004-2007 period. The robustness exercise in Appendix 6.3 likewise suggests that structure type is not strongly and systematically related to rent change. In sum, there is little evidence that detached units experience different rent inflation than do other rental units.

Regarding the number of rooms, the results are not entirely negative, and hint at the possibility that rent changes might vary with size. After all, controlling for the effects of relative rent level, structure type, and location, rooms variables are still jointly statistically-significant in about half of the cities investigated, and in the pooled regressions, during both periods. These findings suggest that further research may be warranted. However, evidence presented in the Appendix (Section 6.3) indicates that the rooms variables seem rather unimportant in terms of explaining rent change.

Our findings imply that the coefficients in equation (1) are somewhat stable over time. If a coefficient in (1) does not change much over time, then its importance in (2) will be diminished. This observation helps us interpret the fact that location does not appear to be a powerful predictor of relative rent change. This fact is consistent with a stable location

[36] This conclusion would not be immediate; for example, CPI aggregation weights might gracefully take this into account.

coefficient in (1) … and such stability is actually an implication of the "location, location, location" principle. To see why, note that the validity of this key principle would be called into question if the value of location were to change dramatically over a three-year period. Regarding coefficient stability in (1), the cross-section regression evidence provides far less support to this line of reasoning: in these regressions, coefficient estimates will be statistically significant only if coefficients change in a similar manner across many cities.

But there is a conundrum that we have not discussed yet. What accounts for the apparent discrepancy between the somewhat weak influence of location (and aggregation weight) on rent change, and the evidence in Poole and Verbrugge (2008) – a study covering approximately the same time period – that suggests that rent changes vary dramatically across various locales within a city? The findings of these studies are not contradictory. The apparent contradictions stem from the nature of the questions being asked, the measurement of location available in this study, and the tools being used to answer those questions. Poole and Verbrugge (2008) seek to understand why inflation rates in the two shelter price indexes in the CPI, OER and Rent, diverged markedly in the decade prior to the Great Recession. Within the context of producing alternative shelter price indexes, they explore the importance of aggregation weights and the like. However, they do not conduct any regression analysis, and thus cannot answer "all else equal" kinds of questions. For example, they find that, in the recent period, units with higher OER weight experienced dramatically less inflation than did units with higher Rent weight. Our results in Table 2 indicate that, *all else equal*, aggregation weight variables are only weakly related to rent inflation. Similarly, the results in Table 4 do not appear to strongly support their findings. But aggregation weights are strongly correlated with a number of other variables in our regressions, such as structure type, percent renter, and median tract income. When these variables are dropped from the regression, we obtain results that are consistent with those of Poole and Verbrugge (2008).[37]

[37] The analysis of Goodman (2005) appears to indicate a very clear relationship between rent inflation and rent level in each of the five cities investigated. That study is a decade-long study ending in 2002. Like Poole and Verbrugge (2008), despite being based upon hedonic regressions such as (1), that study does not attempt to speak to "all else equal" questions; it does not address the statistical significance of the relevant regression coefficients; and its focus is on the estimated increase in constant-quality rent by position in the local rent distribution.

5. Conclusions

While many studies attempt to determine which variables influence market rents, this study instead asks a different question: which variables influence rent change? We investigate the usual suspects – location, age, size, and so on – in addition to relative rent level, presumably a proxy for unobserved quality, or at least an indicator of segment of market.

We confirm several facts from previous studies, such as: rents are quite sticky; utilities provision matters for rent change; change in occupancy is associated with above-average rent growth; and aging is an important determinant of rent growth. These findings support conventional statistical agency adjustments for rent inflation.

We find that initial relative rent level is strongly associated with subsequent rent change – but this is mostly due to a form of mean reversion. This variable does not cleanly proxy for unobserved quality or segment of market, and caution is necessary when attempting to draw conclusions from the associated coefficient estimates. These mean-reversion insights are likely applicable to other situations in which an initial price is used as a predictor for subsequent price changes, or to situations in which markets are stratified by initial price.

But what is perhaps the most surprising finding is that there is not *any* single variable, including location, that has a statistically-significant correlation with 2001-2004 or 2004-2007 rent change across even two-thirds of the cities in the sample. Location itself is only statistically significant in about half of the cities (although as Brickman[38] points out, relative rent level may also be partly proxying for this variable). Even the unit's amenities, such as utilities-included and number of rooms, are uncorrelated with rent change in over half of the cities. How can this be explained?

Our results do not challenge the conventional view that location has a first-order influence on rents, or the findings of a large body of research that studies the various influences on the level of market rents at a point in time. Our results do, however, suggest that these variables are much more strongly correlated to rent *level* than to rent *change*. Putting this differently, the value of a particular location or amenity tends to be fairly persistent. Indeed, such

[38] Discussion of this paper, 2009 AREUEA Mid-Year Meeting. It is possible that the treatment of location is too crude – in that our location measures cover fairly large regions of a city, and other neighborhood-level variables only partly proxy for location – and that other approaches such as a spatial modeling approach (e.g., Valente et al., 2005) might provide more support for the importance of location on rent change.

persistence in value is actually implied by the standard “location-location-location” principle. If a given location is desirable, it is usually still desirable three years later. After controlling for other covariates, a strong relationship between location and differential rent change would suggest rapid change in the relative desirability of various locations.

Our results are generally reassuring to statistical agencies. Such agencies go to a great deal of trouble and expense to obtain a representative sample, and face difficult decisions regarding how to appropriately address sample attrition. Currently, the BLS uses geographically-based sampling: the implicit assumption is that, after conducting utilities-, vacancy-, and aging-adjustments, rent inflation in geographically-proximate rental units is an adequate proxy for the change in the value of the flow of services received by homeowners, even when structure type, size, and rent level differ. Our findings are generally supportive of the conventional set of statistical agency best practices. Why? While it is challenging to find variables correlated with subsequent rent change, the best-performing variables (in this sense) are all used in the appropriate ways. Location, the best variable, is used as the basis of sample stratification. Utilities-, vacancy-, and aging-adjustments are clearly all necessary. Variables not used in sampling, such as structure type, size, and rent level, do not appear to be strongly and systematically related to differential rent changes. We did not locate evidence for problematic collinearity between rent level, location, quality and market segment. But if high collinearity were present, while this would certainly make inference about their individual effects challenging, it would not rule out sampling based only upon geography – since such collinearity would imply that location adequately captures the influence of the other variables.

6. Appendix: Robustness checks and other figures

6.1 Using $dev_i(01,04)$ and $dev_i(02,03)$.

In a previous working-paper version of this paper, we made use of two different *dev* measures, $dev_i(01,04)$ and $dev_i(04,07)$, where $dev_i(\mathrm{t}_1,\mathrm{t}_2) \equiv \frac{1}{2}\left(dev_i(\mathrm{t}_1) + dev_i(\mathrm{t}_2)\right)$. Results were qualitatively similar, although there was somewhat less evidence for the importance of *any* of the variables, and results were not terribly supportive of the hypothesis of rental market segmentation by rent level. In particular, even when we used the variable to partition the sample into "high-end," "low-end," and "other," the set of dummy variables was statistically-significant in only about one-third of the cities investigated for either time period.

Using 2001-2004 data, we also compared $dev_i(01,04)$ with $dev_i(02,03)$. Figure 4 is a depiction of the two histograms of the respective t-statistics pertaining to this coefficient; the shaded bar at 0, for example, indicates that the *t*-statistic on the $dev_i(02,03)$ coefficient was in the interval [-0.5,+0.5) for 27 of the PSUs. The similarity of the histograms indicates that the results are broadly consistent with each other. Over this time period, only about a quarter of the PSUs have *dev* coefficient estimates that are statistically significant at the 5% level. The range of estimates depicted in the histogram indicates the cross-PSU variability in influence that was noted previously. In pooled 2001-2004 data, the coefficient estimate on *dev* is +0.01, with a t-statistic in the neighborhood of 4.5, irrespective of whether $dev_i(01,04)$ or $dev_i(02,03)$ is used.[39]

[39] An additional round of trimming outliers, based upon this linear specification, reduces the p-values even further.

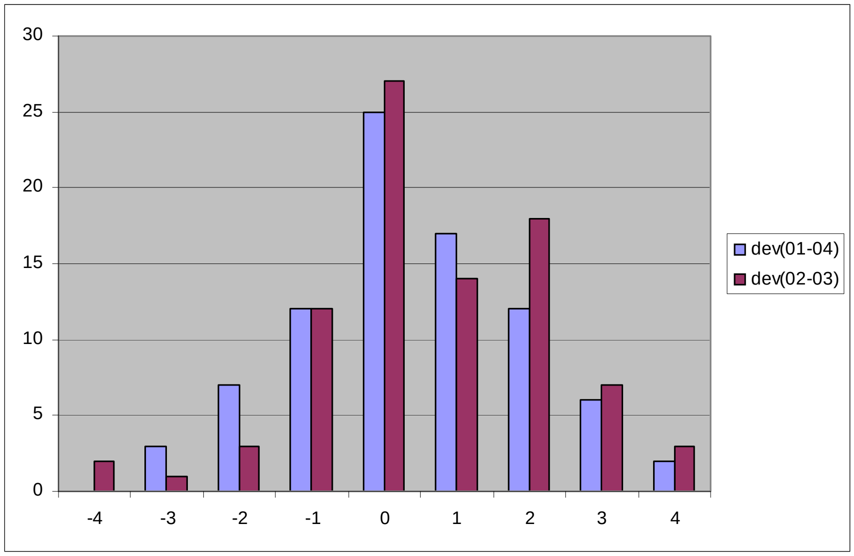


Figure 4: Histograms of *t*-statistics on $dev_i(01,04)$ and $dev_i(02,03)$ coefficients

6.2 No-multiple-units (7 PSUs)

For seven large PSUs[40] we examined the 2001-2004 microdata and manually removed (to the extent possible) multiple units from the same apartment complex. The rationale for doing so is that the rent inflation experienced by units within the same complex is likely highly correlated – as the same landlord sets these rents – and that this "apartment complex" variable might well be proxied by other variables. We present selected results in Table 5. Our major qualitative results are unchanged: age, location, and occupancy are significant predictors; percent renter in these 2001-2004 data is a significant predictor; and relative rent level is not a terribly convincing predictor for future rent change. The structure type "detached" is never estimated to be statistically significant at the 5% level (although at the 10% level, it is statistically significant in two PSUs, Chicago and Dallas-Fort Worth).

[40] These seven PSU's were Chicago-Gary-Kenosha, Dallas-Fort Worth, Detroit-Ann Arbor-Flint, Houston-Galveston-Brazoria, Los Angeles-Long Beach, Los Angeles Suburbs, and Philadelphia-Wilmington-Atlantic City.

6.3 Restricted Models: Model Selection via AIC-C, and Segmentation by Rooms

It might be argued that the main specification (2) suffers from two drawbacks. First, as there is collinearity amongst the variables, it is possible that the statistical significance of many covariates is being masked by the presence of other covariates. Second, the standard urban model suggests that the predominant segmentation of markets should be on the basis of a "quantity" measure, such as square feet or number of rooms. Furthermore, owing to the relative importance of the rooms variables (even given the presence of a relative rent level measure), one might worry that "size" should in fact be considered as a stratification variable for statistical agencies. To address these concerns, we performed a model-selection exercise on 2001-2004 data, as follows.[41] The approach is to proceed, in a stepwise fashion, to build up to a final specification which is (almost) identical across PSUs. The approach is to include, at each step, only variable groups (if any) which satisfy two criteria: 1), the group most improves model fit; and 2), the group improves the fit enough to reduce (improve) the AIC-C criterion. Note that a major collinearity problem will imply a very small model.

Given our model-selection and stratification focus in this exercise, our treatment of several variables changed. We reduced the number of age dummies, as well as the number of separate heating-and-cooling dummies, and removed the indicator for change-in-occupancy. Furthermore, our treatment of rooms changed significantly: instead of using number of bathrooms and so on as five separate covariates, we instead partitioned the universe of rented units into six *size* groups based upon bedrooms and bathrooms as follows:

- one bedroom and one bathroom
- two bedroom
- three bedroom and one bathroom
- other three bedroom
- four or more bedrooms
- other

[41] We are indebted to Tony Yezer for suggesting this exercise.

Five dummy variables are then necessitated. We considered including or excluding the entire group of five dummy variables. This treatment of size allows one to more readily assess whether segmentation of markets by size is important.

As noted above, we considered the inclusion or exclusion of groups of variables, such as three age bins, based upon the AIC-C information criterion. As it would be impossible to exhaustively test every combination, we instead proceeded in several steps.

In step one, after outlier treatment based upon the full model, we began with a “base-model” specification which included only the constant term. On a PSU-by-PSU basis, we then considered specifications which added *one* additional group of variables, group by group, and computed the AIC-C in each case.

The “location” group won the implicit horse race in 29% of the PSUs, and our restricted group of heating-and-air-conditioning (“HVAC”) variables was second with 17% of the PSUs. Age came in third, winning in 14% of the PSUs. Two usual suspects fared quite poorly. Structure type won the race in only 10% of the PSUs, while coming in dead last for 19% of the PSUs. Only the size group fared worse than structure type, but it was much worse, coming in last for nearly half (47%) of the PSUs. (It is likely that the influence of size on rent is highly persistent. In that case, as discussed in Section 4.4, size may well fail to have significant correlation with differential rent change.) Collinearity is clearly not masking the importance of either of these variables. Given our results, both the location and HVAC groups were included in the new “base model” for step two.

In step two, for each PSU we began with a constant, the location dummies, and the HVAC dummies. Then, for each PSU, we added each group of variables in turn, to see which would result in the largest information criterion gain. Type of structure won in 31% of the PSUs, while “none” won in only 10%, so the structure dummy variables were added to the base model for the next step.

In step three, we repeated the exercise once again, using our new base model. In this step, “none,” the age dummies, the variable tenancy duration, and three dummy variables corresponding to neighborhood-percent-renter each won about 20% of the time. (We note that the size (rooms) variables continued to remain relatively unimportant.)

In the final step, with a base model that added these new three sets of variables, we examined whether adding either size or the relative rent level variable would be selected on the basis of AIC-C. In this final step, relative rent level outperformed “no additional variables” and size in the large Northeast cities, and size outperformed “no additional variables” and relative rent level in the large Southern cities; in all other geographic groups, “no additional variables” was typically selected by AIC-C.

At this point, using these restricted models (including relative rent level in the large Northeast cities, and the rooms variables in the large South cities), we repeated the PSU-by-PSU regressions and compared the results to those in Table 2. Surprisingly, using these restricted models, evidence for the influence of any group of variables was generally *weaker*. Neighborhood-percent-renter – which we had divided into three category variables – was the exception: it was statistically significant at the 5% level in 40% of the PSUs. Location, rent control, HVAC, and structure type remained statistically significant in about the same proportion of PSUs as previously. Age and occupancy were statistically significant in a smaller proportion.

While these model-selection findings are certainly not definitive – our size measure does not exactly correspond to square feet, for example – neither do these results suggest a serious multicollinearity problem (as the resulting models are fairly large), nor do they provide much support to the view that statistical agencies should include size and structure type as stratification variables.

6.4 Using $dev_i(2000)$

Operationally, a statistical agency is able to use rental units already in the sample to proxy for missing sample. But in the context of drawing new sample, it is usually impossible to use contemporaneous rent data as a sampling frame. Instead, rent level data would only be available with a lag. We investigated whether year-2000 rents would still have a statistically-significant relationship to 2001-2004 rent change. In this specification we used $dev_i(2000)$ and $dev_i(2000)^2$ as regressors. This variable set was found to be statistically-significant at the 5% level in 43% of the PSUs. Still, this finding is of questionable use to statistical agencies.

6.5 Other specifications

We investigated a modest number of other specifications, such as a more conventional age, age^2, and age^3 specification. Our qualitative conclusions are unchanged.

6.6 Standard deviations of cross-sectional rent growth

These estimates may be helpful for the calibration of rental market models.

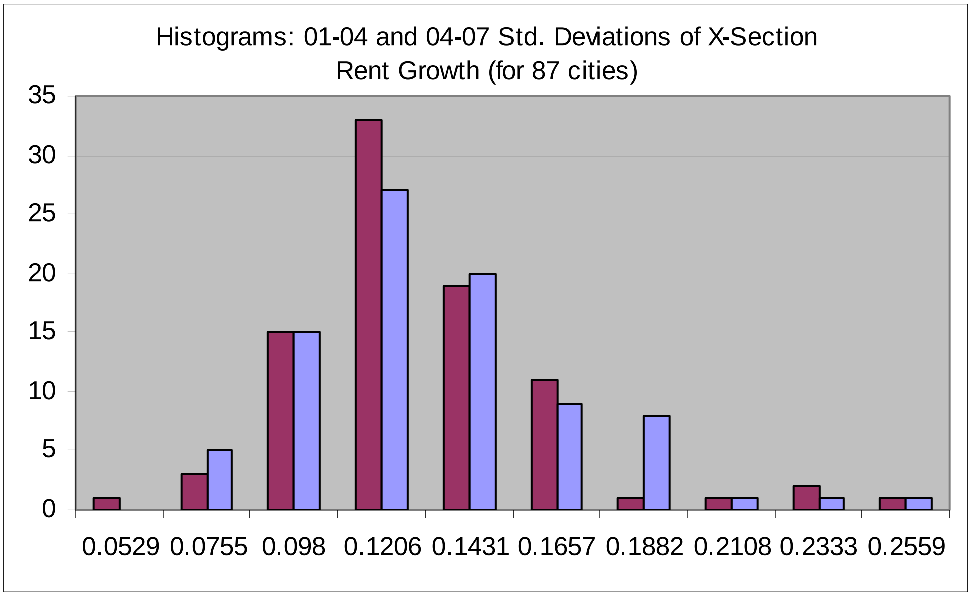


Figure 5: Histograms of standard deviations of cross-sectional rent growth

Note that these estimates are based upon the raw data, before trimming. (Obviously, outliers will strongly influence the estimated standard deviation.)

7. References

Aizcorbe, Ana (forthcoming) *A Practical Guide to Price Index and Hedonic Techniques* Oxford University Press.

Ambrose, Brent W., N. Edward Coulson, and Jiro Yoshida (2013) "The Repeat Rent Index." Manuscript, Pennsylvania State University.

Clapp, John M. and Carmelo Giaccotto (1998) "Residential Hedonic Models: A Rational Expectations Approach to Age Effects." *Journal of Urban Economics* 44.3, 415-37.

Clayton, Jim. (1998) "Further Evidence on Real Estate Market Efficiency." Journal of Real Estate Research, 15.1-2, 41-57.

Coulson, N. Edward (2008) *Hedonic Methods and Residential Valuation*. Working draft for AREUEA monograph.

Crone, Theodore M., Leonard I. Nakamura, and Richard Voith (2010) "Rents Have Been Rising, Not Falling, in the Postwar Period" *The Review of Economics and Statistics* 92.3, 628-642.

Chang, Sheng-Wen, and N. Edward Coulson (2001) "Sources of Sectoral Employment Fluctuations in Central Cities and Suburbs: Evidence from Four Eastern U.S. Cities." *Journal of Urban Economics* 49, 199-218.

Deng, Yongheng, Stuart A. Gabriel, and Frank E. Nothaft (2003) "Duration of Residence in the Rental Housing Market." *Journal of Real Estate Finance and Economics* 26.2-3, 267-285.

Diappi, Lidia, and Paolo Bolchi (2008) "Smith's rent-gap theory and local real estate dynamics: A multi-agent model." *Computers, Environment and Urban Systems* 32.1, 6-18.

Diewert, W. Erwin, "The Treatment of Owner Occupied Housing and Other Durables in a Consumer Price Index," in W.E. Diewert, J. Greenlees and C. Hulten (eds), *Price Index Concepts and Measurement*, NBER Studies in Income and Wealth, University of Chicago Press, forthcoming 2008.

Efron, Bradley (2011) "Tweedie's Formula and Selection Bias." Manuscript, Stanford University.

Follain, James R., David Kogut, and Michael Marschoun (2000) "Analysis of the Time-Series Behavior of Rents of Multifamily Units." Paper presented to the American Real Estate and Urban Economics Association, Boston, January 2000.

Gabriel, Stuart A., and Frank E. Nothaft (2001) "Rental Housing Markets, the Incidence and Duration of Vacancy, and the Natural Vacancy Rate." *Journal of Urban Economics* 49.1, 121-149.

Gallin, Joshua, and Randal Verbrugge (2008) "Improving the CPI's Age-Bias Adjustment: Leverage, Disaggregation and Model-Averaging." Mimeo, Bureau of Labor Statistics.

Gallin, Joshua, and Randal Verbrugge (2014a) "Sticky Rents." Mimeo, Federal Reserve Bank of Cleveland.

Gallin, Joshua, and Randal Verbrugge (2014b) "A Theory of Sticky Rents." Mimeo, Federal Reserve Bank of Cleveland.

Gallin, Joshua, and Randal Verbrugge (2014c) "Panel Data Estimates of Age-Rent Profiles for Rental Housing." Mimeo, Board of Governors of the Federal Reserve System.

Goodman, Jack (2004) "Determinants of Operating Costs of Multifamily Rental Housing." Manuscript, Harvard University.

Goodman, Jack (2005) "Constant Quality Rent Indexes for Affordable Housing." Working Paper W05-4, Joint Center for Housing Studies, Harvard University.

Genesove, David (2003) "Nominal Rigidity in Apartment Rents." *The Review of Economics and Statistics* 85.4, 844-853.

Green, Richard, and Stephen Malpezzi (2003) *A Primer on U.S. Housing Markets and Housing Policy.* Washington DC: The Urban Institute Press.

Hoffman, Johannes, and Jeong-Ryeol Kurz-Kim (2006) "Consumer Price Adjustment Under a Microscope: Germany in a Period of Low Inflation." European Central Bank Working Paper Series No. 652.

Hwang, Min, and John M. Quigley (2009) "Housing Price Dynamics in Time and Space: Predictability, Liquidity and Investor Returns." *The Journal of Real Estate Finance and Economics* 41.1, 3-23.

Krsinich, Frances, and Randal Verbrugge (2015) "Housing Exit and Entry Bias in the CPI." Mimeo, Federal Reserve Bank of Cleveland.

Liu, Laura, Hyunsik Roger Moon, and Frank Schorfheide (2014) "Forecasting with Dynamic Panel Data Models." Manuscript in preparation, University of Pennsylvania.

Malpezzi, Stephen, Henry O. Pollakowski, and Tammie X. Simmons-Mosley (2005) "The Distribution of Rent Changes Within a Market, and Implications for Second-Generation Rent Control." Mimeo, University of Wisconsin.

McCarthy, Jonathan, and Richard W. Peach (2004) "Housing Trends in the 1990s: The Effects on Rent Inflation and Its Measurement in the CPI." Mimeo, Federal Reserve Bank of New York.

Ozimek, Adam (2014) "Sticky Rents and the CPI for Owner-Occupied Housing." Manuscript, Econsult Solutions.

Poole, Robert (2006). Internal BLS Memo.

Poole, Robert, and Randal Verbrugge (2010) "Explaining the Rent-OER Inflation Divergence, 1999-2007." *Real Estate Economics* 38.4, 2010, 633-657.

Ptacek, Frank, and Robert M. Baskin (1996) "Revision of the CPI Housing Sample and Estimators," *Monthly Labor Review* (December 1996), 31-9.

Randolph, William C. (1988) "Housing Depreciation and Aging Bias in the Consumer Price Index." *Journal of Business and Economic Statistics*, 6.3, 359-371.

Rivers, Joseph D., and John P. Sommers (1983), "Vacancy Imputation Methodology for Rents in the CPI," *Proceedings of the ASA Economics and Business Section*, 201-205.

Saiz, Albert (2003) "Room in the Kitchen for the Melting Pot: Immigration and Rental Prices." *The Review of Economics and Statistics* 85.3, 502-521.

Shimizu, Chihiro, Kiyohiro G. Nishimura, and Tsutomu Watanabe (2010) "Residential rents and price rigidity: Micro structure and macro consequences." *Journal of the Japanese and International Economies* 24.2, 282-299.

Shoemaker, Owen (1997) "Performance Comparisons of Laspeyres Indexes with Geometric Mean Indexes in the U.S. Consumer Price Index." *ASA Proceedings*, December.

Turner, Bengt, and Stephen Malpezzi (2003) "A Review of Empirical Evidence on the Costs and Benefits of Rent Control." Manuscript, University of Wisconsin, Madison.

Valente, James, ShanShan Wu, Alan Gelfanc, and C.F. Sirmans (2005) "Apartment Rent Prediction Using Spatial Modeling." *The Journal of Real Estate Research*, Jan-March.

Verbrugge, Randal (2008) "The puzzling divergence of rents and user costs, 1980-2004." *Review of Income and Wealth* 54.4, 671-699.

Verbrugge, Randal (2012) "Do the CPI's Utilities Adjustments for OER Distort Inflation Measurement?" *Journal of Business and Economic Statistics* 30.1, 143-148.

Table 1: 2004 Descriptive Statistics

Variable	Rent Sample Distribution Statistics 1%	25%	50%	75%	99%
Log rent	\$221	\$493	\$665	\$898	\$2208
Age	4	23	35	54	128
Bedrooms	0	1	2	2	4
Bathrooms	1	1	1	2	3
Other rooms	1	2	2	3	4
Length of occupancy (months)	1	10	26	69	319
Number of sampled rental units in PSU	59	130	190	263	918

Percent of Rental Sample Featuring:

Variable	%	Variable	%	Variable	%
Single family detached	21	Electric heat	41	Central A/C	44
Duplex/townhouse	18	Gas heat	50	Window A/C	15
Multi-unit w/ elevator	9	Other heating fuel	1	Other A/C	11
Multi-unit w/o elevator	50	Heat included	18	Rent-controlled	3
Mobile home	2	Electricity included	7		

Table 2: PSU-by-PSU Summary of Statistical Significance

Specification: $\left[\ln\left(rent_{i,t}\right)-\ln\left(rent_{i,t-3}\right)\right]=\phi+\tilde{X}_{i,t*}\tilde{\theta}+u_{i,t}$ [42]

Variable	**2001-2004** % of PSUs[1] p-value $< 5\%$	**2004-2007** % of PSUs[1] p-value $< 5\%$
Location (6 Strata)[2]	53	67
Decade built[2]	45	60
Occupancy variables[2]	45	55
Room variables[2]	41	54
Structure type[2]	52	44
Detached unit	22	17
Initial relative rent level (*dev*)	60	60
Heat or Electricity Included[2]	39	46
Heating fuel type[2]	14	28
Aggregation weight variables[2]	27	37
Rent control variables[2]	33	43
All Census variables (joint)[2]	51	33
Poverty rate	23	23
% renter	24	13
Median tract income	22	18
% white, non-Hispanic	19	19
% over-25 some college	22	18
Median Adjusted R^2	0.35	0.30
Median N	187	181

[1] Percentages are reported as a percent of the PSUs for which this effect is estimable; rent control, for example, is rare and occurred in only 8 PSUs.
[2] This is a group of variables, as explained in the text.

[42] The notation $t*$ indicates that two of the variables, namely one occupancy variable and one rent control variable, are measures which depend upon values across the entire time period, and not just the initial period.

Table 3: 2001-2004, Pooled PSUs

Specification: $\left[\ln\left(rent_{i,t}\right)-\ln\left(rent_{i,t-3}\right)\right]=\phi+\tilde{X}_{i,t*}\tilde{\theta}+u_{i,t}$

Variable	Coefficient (std. error)	|t-statistic|	*p*-value
Intercept	–0.043 (0.0610)	–0.71	0.478
PSU[1]			0.000
Decade built (age)[1]			0.000
Central city	+0.001 (0.0015)	0.45	0.650
Initial length of occupancy	–0.000 (0.0000)	0.37	0.713
Change in tenant	+0.031 (0.0015)	19.86	0.000
Rooms[2]			0.002
Structure type: Mobile home	–0.024 (0.0058)	4.08	0.000
Detached	–0.006 (0.0023)	4.08	0.000
Semi-detached	–0.002 (0.0019)	0.93	0.350
Multi-unit w/elevator	+0.011 (0.0027)	3.91	0.000
Initial relative rent level ($dev_i(2001)$)	–0.046 (0.0023)	20.01	0.000
Heat included	+0.007 (0.0039)	1.76	0.078
Electricity included	+0.019 (0.0032)	5.86	0.000
Heating fuel type[2]			0.511
Rent weight (% deviation)	+0.000 (0.0009)	0.44	0.658
OER weight (% deviation)	–0.003 (0.0009)	3.07	0.002
Rent-controlled	+0.004 (0.0045)	0.85	0.393
Came off rent-control	+0.057 (0.0104)	5.56	0.000
% poverty	+0.021 (0.0081)	2.61	0.009
% renter	–0.036 (0.0042)	8.60	0.000
Median tract income	–0.000 (0.0000)	4.07	0.000
% white, non-Hispanic	–0.004 (0.0039)	0.91	0.361
% over-25 some college	–0.006 (0.0054)	1.03	0.302
Adjusted R^2	0.22		

[1] Only the joint significance level for the set of dummy variables reported.
[2] Only the joint significance level for the set of variables reported

Table 4: 2004-2007, Pooled PSUs

Specification: $\left[\ln\left(rent_{i,t}\right)-\ln\left(rent_{i,t-3}\right)\right]=\phi+\tilde{X}_{i,t*}\tilde{\theta}+u_{i,t}$

Variable	Coefficient (std. error)	|t-statistic|	*p*-value
Intercept	+0.013 (0.0167)	0.789	0.430
PSU[1]			0.000
Decade built (age)[1]			0.068
Central city	–0.001 (0.0014)	0.98	0.323
Initial length of occupancy	+0.000 (0.0000)	3.14	0.002
Change in tenant	+0.030 (0.0015)	20.19	0.000
Rooms[2]			0.000
Structure type: Mobile home	–0.017 (0.0061)	2.88	0.004
Detached	–0.001 (0.0023)	0.47	0.632
Semi-detached	–0.002 (0.0018)	0.83	0.405
Multi-unit w/elevator	–0.003 (0.0025)	1.02	0.307
Initial relative rent level ($dev_i(2004)$)	–0.032 (0.0022)	14.61	0.000
Heat included	+0.020 (0.0037)	5.53	0.000
Electricity included	+0.010 (0.0030)	3.18	0.001
Heating fuel type[2]			0.424
Rent weight (% deviation)	–0.001 (0.0008)	1.62	0.104
OER weight (% deviation)	–0.002 (0.0009)	2.23	0.026
Rent-controlled	–0.004 (0.0042)	0.89	0.369
Came off rent-control	–0.011 (0.0114)	1.00	0.318
% poverty	–0.002 (0.0078)	0.28	0.774
% renter	+0.009 (0.0040)	2.40	0.016
Median tract income	–0.000 (0.0000)	1.70	0.089
% white, non-Hispanic	+0.014 (0.0038)	3.84	0.000
% over-25 some college	+0.004 (0.0051)	0.86	0.387
Adjusted R^2	0.24		

[1] Only the joint significance level for the set of dummy variables reported.
[2] Only the joint significance level for the set of variables reported

Table 5: Seven PSU's, No Multiple Units

Specification: $\left[\ln\left(rent_{i,t}\right)-\ln\left(rent_{i,t-3}\right)\right]=\phi+\tilde{X}_{i,t^*}\tilde{\theta}+u_{i,t}$

	# (of 7) PSUs significant at the	
Variable	5%	10%
Location (6 Strata)	3	4
Decade built (age)	2	3
Occupancy variables	3	4
Structure type	2	3
Rent level ($dev_i^{02,03}$)	1	1
Number of rooms	3	4
% renter	3	3
Median tract income	0	1
% white, non-Hispanic	3	4
Average Adjusted R^2	0.21	