

1 Appendix: Net Yield Panel Data Instructions

In this document, we describe how to produce the net yield panel from publicly-available data.¹

1.1 Data files

The Census conducts the American Housing Survey, which is sponsored by the Department of Housing and Urban Development (HUD). The data files can be downloaded from their website at <http://www.census.gov/programs-surveys/> The survey has been conducted in every odd-numbered year since 1973. The codebook for 1997-2013 can be found at <http://www.census.gov/programs-surveys/ahs/data/2013> We use all of the odd-numbered years since 1985. Prior to 1985, the value of the home is a categorical variable, so we do not use those data.

Download the flat files and place them into folders to which you then need to point the code supplementing this data appendix, found in the file named “[TBD]”.

1.2 Data selection

We remove datapoints without an MSA identifier, as we will be building a panel by city. We remove units in housing projects, those with bars on the windows, and those that are rent stabilized. We remove datapoints missing data (such as tenure status, i.e., owner occupied or renter occupied).

We further clean the observations before the hedonic regression as follows

- Delete if the ratio of household income to house value is greater than 2 (This identifies data errors in the house value field)
- Delete if the ratio of household income to annual rent is greater than 100 (This identifies data errors in the annual rent field)

Throughout the rest of the analysis, we restrict the sample to the top 30 cities by datapoints (after cleaning) in 1985, the first year in the sample.

¹We do not supply the CoreLogic data used to compute house price appreciation (HPA) (and total returns) because of licensing agreements. Public-use HPA data from FHFA for the 100 largest metros are available at <http://www.fhfa.gov/DataTools/Downloads/Documents/HPI>

After cleaning, we still have over 5,000 houses in the sample for each year, as listed in Table 1. The sample is larger in some years because the Census chose to sample certain cities more intensely. For example, the sample size of Minneapolis in 2007 is 1,662.

Table 1: AHS sample size after cleaning

	N
1985	6,958
1987	6,051
1989	7,410
1991	6,027
1993	7,899
1995	8,578
1997	5,263
1999	7,692
2001	5,594
2003	7,968
2005	5,669
2007	12,443
2009	6,560
2011	27,715
2013	8,471
Total	130,298

1.3 Imputing rents with a hedonic model

We employ a hedonic model (as in Malpezzi 2002) to project the log rent of a home upon

- Metropolitan statistical area (MSA) fixed effect
- Year fixed effect
- Unit type fixed effect (i.e., detached, attached, condo, or apartment)
- Number of rooms
- Number of bathrooms
- Dummy for central air
- Year unit built categories (by decade to 1970, then every 5 years)

We do this to unit age because it is a categorical variable before 1995. We do not use square feet because of the large number of missing observations.

The hedonic model is estimated upon renter-occupied homes. It enables us to compute a rent for each owner-occupied home using its characteristics. We then have an estimate of a rent-to-price ratio for each owned home. The model estimates are presented in Table 2.

Table 2: Hedonic regression coefficients (N=34,960)

	coefficient	t value
Intercept	8.87	80.5
ROOMS	0.05	13.6
BEDS	0.02	4.7
BATHS	0.18	30.0
AIRSYS	0.15	19.9
Detached home	0.05	7.7

1.4 Aggregating rent-to-price ratios with nonparametric weights

We wish to find the median rent-to-price of *rental* homes, but our dataset has rent-to-price ratios computed on *owned* homes. To account for sample selection in our dataset of rent-to-price ratios, we weight the owned homes in a city to the distribution of rental homes. (We do not use the alternate methodology – to estimate the hedonic coefficients on owned homes, and then compute rent-to-price ratios on rental homes directly – because there are not enough rental homes for a meaningful sample in some city-year bins. Indeed, the very same cities that are less populated are the ones with a low ratio of rental homes.) This procedure is similar to the nonparametric approach used in Barsky et al (2002).

Sample selection is an issue because rent-to-price ratios are decreasing in house prices (and in house rents), as discussed in the section on house level data in the main text.

For each city in each year, we re-weight the owner-occupied houses as follows. First, line up all the houses by predicted rent. Then bin by percentile of predicted rent. Next determine the density of renter-occupied in the predicted rent space. Finally, compute the median rent-to-price ratio among owner-occupied, using the density of renter-occupied to take a weighted median.

Note that relative to an unweighted median, this nonparametric procedure reduces the weight on expensive homes, which are the same homes for which the hedonic model has the largest errors (because it is estimated upon rental homes, which are likely to be smaller

homes).

1.5 Vacancy data

To compute net yields from gross yields, we need to know the percentage of rental homes that sit vacant. We can get this information from the AHS as well. We use the same dataset (including removing units in housing projects, those with bars on the windows, those that are rent stabilized, and those missing data). We label a home as a vacant rental if the survey identifies it as for rent only, for rent or for sale, or rented but not yet occupied. The vacancy rate is the ratio of this number to this number plus the number of renter-occupied homes. For those city-year cells without enough datapoints, we use a projection from the rest of the dataset.

1.6 Tax rates

We also need a panel of tax rates to compute net yields. Our sources are Emrath (2002) for 1990 and 2000 tax rates from Census data, and the National Association of Home Builders (NAHB) for 2005 to 2012 tax rates from ACS data. The tax rate data are available by state.

1.7 Interpolating missing years

As the survey is biannual and the tax rates are from Census, we linearly interpolate the rent-to-price ratios, vacancy rates, and tax rates to even-numbered years and other missing years (in the case of the tax data).

1.8 Net yields

Starting from gross yields, we compute net yields using the follow costs, some of which are expressed as a percentage of rent and some of which are a percentage of home value. We use expense ratios from Morgan Stanley, “The New Age of Buy-To-Rent,” July 31, 2013. Similar, but less comprehensive, assumptions appear in Bernanke (2012) “The US Housing Market: Current conditions and policy considerations.” The assumptions underlying Core Logic’s Rental Trends, discussed below, are also broadly consistent with ours, however some of their cost estimates rely on direct proprietary data rather than ratios of rent or house price.

- Insurance: 0.375% of price
- Repairs: 0.6% of price
- Capex: 1.15% of price
- Property manager: 5.9% of rent
- Credit loss: 0.73% of rent
- Tax: on price
- Vacancy: on rent

We now have a panel dataset of net yields for $N=30$ cities for $T=29$ years.

1.9 Sanity checks

We employ alternate sources for several datapoints to check the quality of our data.

Gross yields: We have Zillow rent-to-price ratios from their own hedonic model applied to both rented and owned homes in their database. These data start in 2011.

Vacancy rates: We have national vacancy rates for rental homes from Census, who use the Current Population Survey (CPS) and Housing Vacancy Survey (HVS). We also have vacancy rates from CoreLogic, who use the USPS.

Net yields: We have CoreLogic cap rates (defined in the main text) starting in 2012.

2 Appendix: Other Data Sources

In addition to the net yield panel at the MSA level, we employ several other data sources at the MSA, zip, and house level to describe net yields, capital gains, and leverage.

2.1 House price appreciation from CoreLogic

To determine total returns at the MSA level, we pair an HPA value for each city-year cell with its net yield.

We have monthly HPI by core-based statistical area (CBSA) from CoreLogic for 1976-present. We use Tier 11 (all homes, including distressed) and also consider Tier 1 (sale price < 0.75 of the median). Because the AHS is conducted each year between May and

September, we set the HPI in each year to equal the HPI in June of that year. The HPA is then the June-to-June percentage increase in HPI.

We created our own translation table from MSA to CBSA with the help of the Missouri Census Data Center.

2.2 House price tiers

We match the HPI from CoreLogic in June 2014 with the Zillow Home Value Index from June 2014.² We can then determine the price level in each year from 1985-2014 by appropriately deflating with the house price index.

The house price tier assigned to each MSA and each zip code is dynamic, with transition matrix diagonals varying from 83% to 96%. The highest tier-to-same-tier transition rate is found at the highest-priced tier. At the MSA level, the probability of staying in the highest tier is 94%, and at the zip level, the probability is 96%.

2.3 Gross Yields from Zillow

We use Zillow's characteristic-adjusted rent-to-price ratios to sanity check our own calculations of gross yields in appendix subsection 1.9.

2.4 Net Yields from CoreLogic

At the MSA level, we use CoreLogic's net yields to sanity check our own calculations of net yields in appendix subsection 1.9. At the zip level, we use CoreLogic's net yields as our primary data source. Their RentalTrends database tracks median rents of 1, 2, 3, and 4 bedroom homes back to 2011 in 10,146 zip codes. The database calculates net yields, which are also referred to as capitalization rates, as depicted in Figure 1.

In the section on HPA at a zip code-level, we discuss observables correlated with HPA. One such observable is subprime origination. We get this data from CoreLogic's LoanPerformance dataset. It covers nonagency loans at origination and in subsequent performance. It enables us to study associations to subprime origination levels by zip code.

²Zillow data are publicly-available at <http://www.zillow.com/research/data/>

Figure 1: Net yield (or capitalization rate) calculation, source: CoreLogic

$$\text{Cap Rate} = \frac{\text{Revenue} - \text{Expense}}{\text{Cost}}$$

$$\text{ROI} = \frac{\text{Revenue} - \text{Expense} + (\text{Discount} - \text{Renovation} + \text{Cap Gain})_{\text{Annualized}}}{\text{Cost}}$$

Rent	Rent Amount Model
HOA fees	MLS data
Insurance	Formula: 0.35% of property value annually
Maintenance	Formula: 17.5% of rent
Management	Formula: 8% of rent
Property Tax	Property data
Resident Risk	SafeRent ScorePLUS
Vacancy Loss	Vacancy Rate Model
Property Value	AVMs, HPI Forecast

2.5 Zip-level covariates from the Census Bureau

When discussing observables correlated with HPA in the section on zip level, we look at age of housing stock by zip code. We also get this and other demographic data from the Census Bureau. They provide zip-level demographic data from the 1990 and 2000 Census. They also provide 5-year American Community Survey (“ACS”) estimates from 2011-2013.

2.6 AHS from Housing and Urban Development

Our primary data provides us with the cost of rent for rental units and an estimated home value for owner-occupied units. This home value is based upon the home owner estimating what they could sell the home for when surveyed. Although this variable is a subjective estimate, Kiel and Zabel (1999) show that it is actually quite reliable. Specifically, they find that the value is inflated by 5.1% on average and that, “the use of the owners’ valuations will result in accurate estimates of house price indexes and will provide reliable estimates of the prices of house and neighborhood characteristics.”

Using the sample described above, we divide the sample into owner occupied and renter subsamples. The two subsets of housing are significantly different. For example, the mean number of bedrooms in an owner occupied home is 3.2, while it is 2.0 in a renter occupied home. In appendix subsection 1.3, we use the AHS for a hedonic regression to calculate characteristic-adjusted rent-to-price ratios.

2.7 House level data from the bond markets

In the section on house-level data, we examine the houses operated by large SFR institutions. We have collected data from the bond markets on the fourteen SFR issuances as of January 2015. Each issuance comes with an appendix listing every property in the collateral, its acquisition prices, its net rent, and other characteristics. We also reference the reports on these issuances produced by Kroll Brothers Rating Agency.