ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Gray, Rowena

Working Paper

Selection bias in historical housing data

QUCEH Working Paper Series, No. 2018-01

Provided in Cooperation with:
Queen's University Centre for Economic History (QUCEH), Queen's University Belfast

Suggested Citation: Gray, Rowena (2018) : Selection bias in historical housing data, QUCEH Working Paper Series, No. 2018-01, Queen's University Centre for Economic History (QUCEH), Belfast

This Version is available at:
https://hdl.handle.net/10419/175181

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

QUCEH WORKING PAPER SERIES
http://www.quceh.org.uk/working-papers

SELECTION BIAS IN HISTORICAL HOUSING DATA

Rowena Gray (University of California, Merced)

Working Paper 2018-01

QUEEN'S UNIVERSITY CENTRE FOR ECONOMIC HISTORY
Queen's University Belfast
185 Stranmillis Road
Belfast BT9 5EE

January 2018

Selection Bias in Historical Housing Data*

Rowena Gray**

Abstract

A new sample containing rental price and characteristic data for over 15,000 New York City units was collected from historical newspapers for the period 1880 to 1910. These units were geolocated to the historical map of Manhattan Island to explore their geographic coverage, using Geographic Information System (GIS) software. This paper presents the new sample and discusses its representativeness of the New York City housing market during the sample period, with reference to the (limited) previous measures available in the literature and an analysis of the summary statistics of various subsamples of the data which can highlight selection biases. Finally, an analysis of the social status and ethnic composition of individuals located in the sample units in Census year 1880 is presented. Understanding the biases that might be present in this new sample will inform its usefulness in uncovering the workings of historical housing markets and in contributing to the scarce available information on historical housing costs.

Keywords: real estate markets, historical house rental price data, selection bias, US economic history, New York City, historical GIS.

JEL Classification: J15, R31, N91.

** University of California, Merced. Email: rgray6@ucmerced.edu.

* Gray acknowledges support from an Economic History Association Cole Grant which facilitated the collection of rental data, and a UC Merced Committee on Research Grant which facilitated geocoding of the data. Excellent research assistance was provided by Jordan Ahid, Frank Chou, Cecilia Garcia, Roman Guerrero, Justin Holtermann, Sewon Kim, Jason Lee, Alexandra Marr, Charles Martin, Chris Medina, Jacob Miller, Michael Mishkanian, Vanhien Nguyen, Jose Ulloa, Joselyn Ulloa, Daisy Urbina and Victoria Whitford. I acknowledge Carlos Villarreal who built the historical GIS maps of Manhattan Island and geocoded the first round of observations. Richard Wright provided excellent GIS assistance.

A. Introduction

This paper presents a new dataset of housing rents for New York City for the years 1880 to 1910, which was gathered from information on rental prices and housing characteristics gleaned from five different, popular newspapers circulating during that period. It then discusses the selection problems that arise from using this approach, employing a variety of methods to test the extent of these problems. Firstly, the secondary literature was scanned for rental price data and any corresponding locations or characteristics, for a holistic comparison to the newspaper sample. Then, I present comparisons of summary statistics of location, rental price and unit characteristics from the following subsamples of the data: by newspaper (which allows us to infer the bias that would have existed had only one or two newspaper sources been used); according to whether price data was present in the advertisement or not (to assess the representativeness of the sample units); and according to the month in which the advertisement appeared (to explore biases introduced by the sampling procedure). Then, the paper goes further to explore the social status of individuals living in the sample buildings during the Census year 1880. The characteristics of these "sample" individuals are compared to ward averages from the 100% count Population Census data for New York City.

B. Existing Literature

Few papers explore historical housing markets, due to data limitations.[1] The most commonly used sources are institutional and tax records and newspaper advertisements. For England and Wales, Clark (2002) collected data on 11,188 rental contracts over the very long run (1640-1909) from the records of charity groups. He used the observations on which there were multiple transactions to compute a quality-adjusted rental index and used this to construct estimates of trends in housing quality, showing that quality increased little during the Industrial Revolution even while rents increased. Devaney (2010) looked at deed and rent roll data from the City of London, to map out trends in commercial office rents over almost a century up to 1959. It similarly finds no great increase in rents over that period, with more marked short run fluctuations. Samy (2015) used auction yearbooks and building society mortgage data to construct price and rental indices for London residential property from 1895 to 1922. Drelichman and Gonzalez Aguda (2014) constructed a dataset of rents for Toledo, Spain, over the years 1489-1600 based on the records of the Cathedral Chapter of the city. Eichholtz et al (2012) use annual data from institutions such as orphanages and hospitals for Amsterdam for the years 1550-1850 to construct a constant-quality rental index using the repeated measures approach. Most recently, Eicholtz et al (2017) uses a patchwork of sources to construct rental series for several European cities.

[1] I discuss here only the existing literature relating to rents. There is also a growing body of work on sales prices, most recently Knoll et al (2017).

The use of newspaper advertisements as a source of rental information goes back at least to Rees (1961). He used advertisements from 6 cities as a source of housing rental information, although he made no attempt to control for unit characteristics. For New York City, he chose the New York World because of its working class target audience, and this paper also makes use of that publication, as outlined below. Margo (1996) constructed a quality-adjusted index of rents for New York City and its surroundings for the period 1830-1860, using newspaper advertisements for this period. His sample contained fewer than 1,000 observations, but was comprehensive enough to estimate the capitalized value of various unit-level characteristics as well as the distance to City Hall. It also used 3 different newspapers, but they did not generally overlap in the years consulted so that the sample at any given point in time might be skewed towards the clientele of the particular paper.

A recent study by Kholodilin (2016) presents monthly data on asking rents in Berlin gathered from advertisements found in the city's two most popular advertising newspapers, from 1909 to 1917, and explores the determinants of trends in the time series during and after World War I and its consequent population and building fluctuations. It is, to our knowledge, the only other study that has geocoded rental information from historical newspapers and thus made use of detailed locational as well as unit characteristics.

More generally, advertisements containing asking prices have been used before. Examples include Schulz et al (2014) for labor markets and Raff and Trajtenberg

(1997) for automobiles. Whether using actual rental data derived from institutional records or asking rents drawn from newspaper advertisements, existing historical research on housing has paid little attention to the potential selection biases, given the dearth of available sources. Samy (2015) does discuss the need to draw his sample of mortgage records across the calendar year to avoid seasonal bias. The advantage of institutional records is that they usually provide actual rents, but institutions may have held a peculiar portfolio of properties relative to the average living dwelling. Institutions might have had particular rules about who could rent the properties and how often they would be rented and to what types of individuals. The advantage of newspaper-created samples is, then, that they are comprised of units presented to the open market for rent, which are more likely to be representative of housing market activity at each snapshot of time.

C. New Housing Data

Information on rental prices and characteristics at the unit-level was collected from newspaper advertisements. This data was geocoded to historical maps of the island of Manhattan using the program GIS. The main challenge in completing the geocoding task was in locations where public housing projects (or private developments such as at Stuyvesant Town) or other major new buildings (Penn Station, for example) have been constructed since 1880 which may have changed the street map substantially.

Data was collected mostly from the months of March through May for each year from the following newspapers: New York Herald (NYH), New York Sun (NYS), New York Times (NYT), New York World (NYW) and the Brooklyn Daily Eagle (BDE, the last being used in only a few cases). Each year is represented by data from at least two of the above newspapers. Advertisements were coded where information was available on (at a minimum) the asking rental value and the exact street address of the unit. Various additional information was included in the typical advert, such as whether or not the unit was furnished; some measure of unit size such as the number of rooms or number of stories; whether the unit was located close to public transport; and whether an array of amenities such as electricity, steam heating or fixtures and fittings were included. Figure 1 shows the distribution of observations according to newspaper source.

Figure 1 about here

Some idiosyncrasies of this period suggest that selection bias is not a substantial issue. The city was mostly a rental market. Only 9.63% of household heads reported real estate holdings in 1870[2], which decreased to 1.7% owner-occupied dwellings by 1940.[3] Thus, movements in rental prices were relevant for the vast majority of New Yorkers. It was a very active market, with families moving frequently based on their

[2] 1% sample of the Population Census from IPUMS.
[3] 1940 Housing Census figures. Home ownership across outer boroughs was greater and households were moving out of Manhattan to achieve home ownership throughout the latter part of the sample period considered here.

economic fortunes of the particular year. An 1875 city directory described how "Of all the civilized people on the face of the earth the inhabitants of New York appear to be the most inclined to move about" (Sherzer 1992: 19). The market was fairly organized so that, even though written leases were very uncommon before World War I, the oral terms were quite standard and movement followed the following pattern. Landlords would announce to tenants early in the calendar year what the rental price would be in the subsequent year and tenants would decide whether to stay or move, with moving usually taking place on or around May 1st. Any attempt to change the rent mid-year could be resisted by tenants (Fogelson 2013: 32). This motivated the data collection strategy of this paper and we do indeed find clustering of rental observations in the months of March-April-May. A secondary moving date of October 1st is mentioned by, for example, Fogelson (2013: 21) and motivated collection of September rents when April was missing. Rees (1961) also looked at April and September rents, for these reasons.

Data from Gentzkow et al (2014), deposited with ICPSR, provides information on the circulation of the newspapers used for this paper, which range from 34000 for the NYT in 1880 to almost half a million for the NYW in 1896. These were all popular newspapers and were main sources of information on available rentals, and advertised themselves as such. Day (1999: 38) describes how immigrant banks advertised frequently in storefronts and papers. This suggests that even immigrant groups used newspapers as a source of information for their business life.

I collected data on newspaper costs and advertising prices from the sample newspapers. The price of buying a daily copy of each paper declined somewhat over our sample period. The NYT was the most expensive and attracted a higher-class clientele but its price (in 2017 dollars) fell from 96 cents in 1880 to 26 cents in 1910. The NYS cost 48 cents in 1880 and 52 cents in 1910. The Conference on Research in Income and Wealth (1975) describes how newspapers and periodicals were a regular part of the household budget for most families. The price of a line of advertising also declined, most notably in the NYT, where it went from 16 2017 dollars per line in 1880 to 3.86 2017 dollars per line in 1910. The NYH in 1910 charged a bit more, up to $7.72 per line. This price has implications about the type of units and rooms that were likely to be advertised. The cheapest rooms for rent or rent and board would likely only be advertised when multiple rooms were available, as otherwise an advertisement could not be justified. Our strategy of building the sample from April advertisements, during the busiest moving period, should make it more likely that we capture some of the lowest end of the market. However, it is likely that the sample does not fully capture the lower tail. We do not observe very many advertisements mentioning subletting, for example, even though that was a common practice. A unit that was listed at a particular price and size may in fact have been occupied by two families, thus reducing rents and housing quality. Further, fashionable and expensive units might not have listed a price (we do observe many advertisements that simply mention a fair or negotiable rent) or been advertised in newspapers for reasons of privacy or exclusivity, so we may miss the extremes of the market but should have good coverage of the bulk in between. Below, I discuss a comparison of unit

characteristics for sample observations and units advertised without a price, which sheds some light on this question.

A variety of different types of apartments and homes are advertised for rent in the various newspapers. The NYT does appear to have targeted a wealthier clientele, with a higher proportion of units being located next to Central Park and the more prominent avenues running North-South from the Park to downtown. Large townhouses are advertised alongside smaller apartments or parts of houses to rent. Some advertisements offer commercial premises, sometimes with residential rooms attached (792 observations in total, which are omitted from the analysis below). All of the publications carried "rooms for let" or "boarders wanted" type of adverts, targeting those who simply need 1 room, furnished or unfurnished, in a large house. In the data collection process, we sampled all of these different types of advertisements. Examples can be found in the appendix.

Figures 2 and 3 respectively show the full Manhattan sample for all 31 years, geocoded to the historical map and a summary of the spread of observations across neighborhoods. These neighborhood definitions were constructed using a modern neighborhood boundary shapefile, with some aggregations for neighborhoods that were new and small historically. Figure 2 shows that there is good overall coverage of the island, with the exception of the Lower East Side. Figure 3 confirms this and also shows that Midtown, Harlem and the Upper West Side were areas that showed a lot of rental activity during this period, with relatively little activity on the southern

tip of the island, which was already commercial, and the northern extremes, which were only becoming developed.

Figures 2 & 3 about here

D. Comparison to Existing Literature

This section presents information on actual rental prices in New York City gleaned from the secondary literature on housing.[4] It is noted that, while these are somewhat informative in validating the new dataset, a main aim of gathering the new 15000-plus observation sample was to increase sample sizes, and control more rigorously for quality differences across units by gathering data on the full set of location and unit-specific characteristics that comes from using the wealth of data from the newspaper advertisements and geocoding that data to historical maps of Manhattan. However, since advertisements provide only asking rents and not actual rents paid, it is useful to make the comparison to what has been documented in previous work. The use of asking rents in this dataset may give a more accurate picture of the housing market (where, again, renting predominated), as it does not limit the sample to units that have already been rented for an agreed price. This sample may be more representative for those interested in knowing general facts about the historical New York City housing market. This contrasts with the more modern literature on housing

[4] All figures given in this section are in nominal, contemporary dollars.

which relies mostly on sales transactions, limiting the sample to houses that are in fact sold, which is also quite a rare event.

The Charity Organization Society of the City of New York's tenth volume (1900: 3) provides some rental data and description of apartments where textile manufacturing was taking place (these were typically in or near the Lower East Side and may have been tenement buildings). On East 12th Street they list an apartment of 4 rooms, inhabited by 2 Austrian families for a total of 7 people, renting for $14. On Elizabeth Street 2 Italian families rented 3 rooms for $6, while another 2 Italian families rented 3 rooms for twice that price. In my sample are single, furnished rooms listed in 1895 and 1896 on East 12th Street for between $2.25 and $4.25 per week. Another single, furnished room was listed in 1900 for $3. Scaling up to four rooms, the asking prices in my sample are in line with the actual prices given for East 12th Street at similar points in time.[5]

The Evening World newspaper, (September 23, 1908), listed some non-location specific rental and budget information in their editorial section, discussing the housing market prevailing in New York City at that time. One letter to the editor mentions that the writer was then paying $30 per month for a 4-room heated flat, while others mention yearly rents of $288-312. The new rental dataset lists 4-room

[5] There are no observations on Elizabeth Street in the newspaper rental sample. Gabaccia (1984: 74) provided more rents in her in depth study of Elizabeth Street, around 1905. Two room apartments cost $9.50 monthly in a dumbbell tenement. 50% of apartments had 3 rooms, costing $8-$15 per month while the 20% with 4 rooms charged about $20 in a new law tenement.

apartments in 1908 for $19-65 per month in various locations within the city, averaging to $39.50. The new dataset therefore may portray rents as higher than actual rents being paid, which could be accurate for 1908 since there may have been some price reduction following the 1907 recession which asking rents might not have fully adjusted to. Without location information in the secondary literature, it is difficult to know which data is biased.

Jacob Riis (1997) cited going rents around 1900 to be $6 for a rear tenement, $17 for 4 front rooms in more modern tenements. Again, these do not specify exact locations but probably refer to the Lower East Side tenement district. The new dataset shows single, furnished, front rooms in 1900 had asking rents of $1.50-4 per week, suggesting that $6-16 for four rooms was reasonable.

Dora Costa's dataset on women giving birth in the Lying In Hospital in New York City lists actual monthly rents paid by those households.[6] I looked at 28 observations from 1910 and compared them to my sample of asking rents. A number of observations were close enough geographically for direct comparison. For example, 1 unit from my sample was 733 feet from 1 of Costa's observations on West 81st Street. The nominal monthly rent from my sample was $12 and $8 for the Costa observation. Another observation was 976 feet away on East 11th Street and again the rent listed in my

[6] The dataset was downloaded from Dora L. Costa's website on 8/6/2015: http://www.econ.ucla.edu/costa/data.html. It contains exact addresses.

dataset was $12 and in Costa's $8. These are most convincing because they contain actual addresses and can be compared to co-located observations.

Anbinder (2001) describes an overcrowded Bayard Street apartment in 1885 where lodgers paid 5 cents per spot to sleep, while a bed at a lodging house at 508 Pearl Street in 1882 cost 12 cents per night and 10 cents for a basement room. The new dataset does list a couple of unfurnished rooms to rent in Pearl Street in 1883 and 1884, and those cost $0.83-1.42 per night. This is substantially more than Anbinder finds, suggesting that the newspapers do not advertise places such as 508 Pearl Street, which is an extremely cheap boarding house.

Chapin's (1909) study of living standards in New York City cites a 4-room tenement on Essex Street with bathroom for $18 per month in 1905. The new dataset has few residential listings in this area but it does show an apartment on Bowery Street offered for $15 in 1901 and nightly hotel room rates on Bowery of 25 cents in 1903. The per room rate for this hotel is greater than for Chapin's tenement, but assuming that it comes furnished and with some hotel amenities, this is not out of line.

Much of the secondary literature provides rental information for those living in tenements, who made up a large share of New York City's population. The new, newspaper-derived, sample may reflect more the changing nature of the New York City housing market and may be more representative of classes above tenement-dwellers. These kinds of comparisons are fairly crude, and as King's *Handbook of New*

York City (1892) outlined, apartment rents showed huge variation at this time, depending on a variety of factors such as location, size etc. The direct comparisons suggest that our data is in line with what previous researchers have found, but the newspaper-derived sample may represent a different basket of properties, targeting a wealthier renter, compared to those identified in previous work. We move next to more quantifiable comparisons, in order to better assess the size and direction of bias in this rental data.

E. Subsamples of Data

This section looks at some summary statistics of various relevant subsamples of the dataset. The full sample includes 15068 data points, of which 15063 could be geocoded. The sample is further reduced to 12783 once some oversampled months are dropped from 1890 and 1891. The 788 commercial observations were dropped, as were those not located on Manhattan, leaving a sample of 11,995.

Observations without a rental price

There are 285 observations that did not list rental prices. These were collected to help judge the representativeness of the main sample of 11710 observations which contain rental data. Table 1 presents summary statistics of this full sample, plus the sample containing no rental data. This last sample is more closely compared to a subset of the full data which is limited to the years in which there is at least one data point without rental information, so the final column reports differences between

those 2 samples. The samples mainly differ according to composition in terms of unit mix. Advertisements where no rental price is listed are statistically significantly more likely to be houses or boarding opportunities while the main sample is composed predominantly of apartments. This unit mix also accounts for the fact that the listings without rental prices are more likely to be furnished, since boarding houses were almost always furnished. The average location of the observations without rental prices is 20 blocks South of the limited years sample with prices. This is also likely driven by the unit mix, as boarding houses were located closer to downtown. If we restrict to only apartments, where board is not included, then the samples become more similar. The number of rooms advertised does become significantly higher in the non-price sample, but they are statistically similar based on decoration, elegant, furnished and heating and more similar in terms of all improvements. Taken together, this evidence suggests that there is a systematic difference in what types of units are advertised with and without prices in these historical newspapers, but that their characteristics are not significantly different when restricted to similar unit types.

Table 1: Summary Statistics by Rental Price Availability

	Sample with Rental Prices	**Limited Years with Prices**	**Sample without Prices**	**Difference**
Rooms	5.1	5.3	5.7	-.37
Decorated	10.1	11.1	9.8	1.3
Elegant	8.6	9.5	12.3	-2.8
Improvements	23	23.8	16.8	7***
Furnished	23	19.4	33.7	-14.3***
Heating	15.2	17.6	13.7	3.9*
Transport	9.7	9.6	3.2	6.5***
House	14.5	13.1	19.6	-6.6***
Board	7.8	5.8	18.6	-12.8***
Apartment	58.5	64.4	43.5	20.9***
Obs	11710	7466	285	

These averages are for residential units. Most statistics are % of apartments reporting a particular amenity. The last column depicts the difference between columns 3 and 4.

1890-1891

More observations were collected for the years of 1890 and 1891 than any other year (1837 and 1421 respectively). This is because all months of the year were considered, not just the main sample months of March-May and the secondary months of August-October. So, we can use this oversample to explore the representativeness of the usual

sample months that were used for this paper. The NYS was consulted for the full 24 months, so we restrict to that sample, which was all coded by the same person, and focus on residential apartments that did not include board. For 1891, April was indeed the modal month and for 1890 it was March that produced the most usable observations. Table 2 displays some summary statistics for sample months (March-May and August-October) and non-sample months (the remaining months).

Table 2: Summary Statistics by Period

	Sample Months	Non-Sample Months	Difference
Monthly Rent	$30.78	$26.33	4.45***
Rooms	5.66	5.55	.12
Decorated	28.4	30.84	-2.5
Elegant	27.3	19.3	8***
Heating	21.5	20.2	1.3
Furnished	3.4	8.7	-5.3***
Transport	5.2	10.6	-5.3***
Observations	553	321	

Notes: These averages are for 1890 and 1891, for the sample drawn from the NYS, of residential apartments that do not offer board and for which a monthly rent could be calculated. Most statistics are % of apartments reporting a particular amenity.

Table 2 shows that rents were significantly higher in the sample months, despite the number of rooms in these units being similar. The difference in rents is almost 17%. Sample month units were less likely to be furnished but were more likely to highlight the elegance of the listing. They were also less likely to report being close to any public transport option, although the average sample locations are very similar (within 7 blocks). Figure 4 plots the observations for 1890 and 1891, where black circles are from non-sample months and white circles are sample months. This map suggests little difference in location of listings. Though more investigation is needed, this analysis suggests that units advertised in the sample period may have been listed at higher prices than others, perhaps because of actual higher quality or because many people were moving at that time and the available units were those where landlords were looking to raise rents the most (so that the current tenants preferred to move out). Sample and non-sample month units are broadly similar along most dimensions, though.

Figure 4 about here

Observations by newspaper

The 4 main newspapers used in this paper advertised different types of units. One quarter of units advertised in the NYT were houses, while only 5% of the NYW sample comprised houses. The share of apartments was fairly similar across all 4 samples, at about 60%, but was lowest for the NYS which had only 48%. The NYS gives us the largest proportion of rooms with board that we observe in the sample. We look now

at summary statistics for each of the 4 newspapers, but limited to apartments within those samples, and compare each sample to the full sample averages.

Table 3: Differences by Newspaper Source

	NYH	**NYS**	NYW	NYT	**NYT, <1900**
Monthly Rent	51.06***	**30.19*** **	27.28***	50.48***	**53.15*** **
Rooms	5.9***	**5.2*** **	4.7***	5.8***	**6.5*** **
Decorated	.08***	**.16*** **	.20***	.11***	**.13** **
Elegant	.12	**.12**	.14***	.09***	**.11**
Heating	.14***	**.23**	.17***	.31***	**.40*** **
Furnished	.20***	**.05*** **	.06***	.15***	**.11*** **
Transport	.20***	**.11**	.07***	.13	**.07** **
Year	1902	**1893**	1896	1899	**1891**
Observations	1235	**1056**	1321	2209	**1027**

Notes: Summary statistics are given for residential units in apartment listings found in the 4 newspapers. Data in columns 1-4 were compared for statistically significant differences to the overall sample average for apartments, while column 5 (NYT sample, restricted to 1880-1899, was compared to the NYS sample, covering the same dates).

Table 3 shows that, even if we restrict to comparable housing units, the listings in the various newspapers are systematically different from the sample average. The

newspapers were sampled in slightly different years, driven by online availability of each source, and Table 3 shows the average time period for each paper, which compares to that of the overall sample (1897). The NYT was available in almost all of the 31 years (29 in fact), while the NYS was available until 1900 (20 years in total). To investigate whether the date of observations is the key drivers of differences across papers here, we can restrict further to comparing apartments listed in the NYS and NYT, up to 1900. This is shown by the columns in bold in Table 3. The NYS and NYT had similar numbers of observations of apartments up to 1900 but the NYT apartments were clearly superior in terms of size (proxied by number of rooms), the provision of heating and with a higher likelihood of being furnished. These units had much higher asking rents which seem warranted, given these advantages. Figure 5 plots the NYT and NYS pre-1900 data side by side on the Manhattan map. NYT data is plotted as green dots, the NYS as red dots. The maps show that there is not a drastic difference in the geographic location of these advertisements, although the NYS concentrates more on both the lower part of the island and on the upper portions which were very newly developing in this period. The NYT is slightly more concentrated in more fashionable areas around Central Park.

Figure 5 about here

Overall, this section shows that, if the goal is to build a picture of the overall housing market for New York City in this historical period, consulting more than one or even

two sources is crucial. Newspapers advertised to distinctive clientele and with distinctive units.

F. Building-Level Census Data

I searched for the new sample addresses found in newspapers in the 1880 Census and identified individuals living at those addresses. We cannot say for sure that they paid the asking rent listed in the advertisements or that they lived in an identical unit, but it is likely that their units were very close to what was advertised.

I looked up 270 NYH, NYS and NYT observations for 1880 in the full count data for that Census, and located 180 addresses. Misspellings and missing address information is the likely reason for the low percentage of addresses found. Addresses may have been misread in the original newspapers too, so that we were looking for an address that didn't exist. I have data on 1849 individuals from the 1880 Census who lived at the sample addresses.

Firstly, I constructed summary statistics of characteristics of households that were in enumeration districts found in the new rental sample and for those enumeration districts on which I have no rental price information. This revealed that there is a systematic difference between rental sample districts and non-rental sample districts

along almost all dimensions that I measured.[7] The former were statistically significantly more likely to contain units occupied by native-born individuals whose parents were also native-born (you might call this group "native-squared"), whereas the non-rental sample districts contained a greater share foreign-born and second generation immigrants. This again suggests that newspapers may have been advertising predominantly to the more stable, existing populations than to immigrant groups. Perhaps because of these demographic differences, the rental sample districts were also less likely to report women working and they were engaged in higher status jobs, whether using the occupation scores provided by the Census or the HISCO classification.[8]

Focusing only on enumeration districts that contained rental observations, I then compared means for observations that were in the rental sample with those that were not. The means for these two groups were much more similar than I found in the above exercise. The main significant differences are found in occscores, which are significantly higher for the rental data group, even though the HISCO score does not significantly differ, and the number of people per household, which is significantly smaller for the rental data group. The nativity variables are well balanced across the two groups, while the share non-white is negligible in the rental data group compared

[7] I do not report the full table of summary statistics here but the results are available upon request.

[8] I only matched to HISCO for the occupations that were reported by rental sample observations so the HISCO means are for a subset of total observations. HISCO codes were obtained from this website, and have been used as a measure of occupational status in many other studies: http://historyofwork.iisg.nl.

to the non-rental data group. This last effect is interesting again as it may indicate that newspapers did not target the non-white market in its advertisements, and indeed, only in a small fraction of cases did an advertisement specify that it wanted to attract "colored" tenants. It may be that this group mainly rented as sub-letters, or found accommodation through some other medium.

In general, this dip into the Census data highlights that subletting must have been a much bigger phenomenon than is identified through the newspaper rental advertisements. In only a handful of cases was subletting explicitly mentioned by the adverts, but there were 6.8-6.9 people living in these households on average, with only .11 boarder per family reported. It seems likely that either multiple families lived in a household (sometimes including a boarder), or the definition of a family extended to grandparents and other relatives, probably adult ones given that the average age observed was about 29. This needs to be born in mind when thinking about an advert as advertising a family unit.

This gives a rough picture of the types of people found in the sample houses. More could be done to investigate further the importance of housing as a component of living standards for this period, and possibly to investigate bias in housing markets during this period.

E. Conclusion

Rents have received relatively little attention in the historical literature on living standards. This paper presents a new sample that can be used to build a picture of historical housing markets and to construct new estimates of quality-adjusted housing rents that might update existing cost of living measures. The results of analyzing the current sample for selection biases suggest that additional sources be consulted to address the relative lack of observations in the Lower East Side area of Manhattan in particular, which was densely populated with immigrants at this time. This might include Yiddish-language newspapers or records of immigrant societies. In general, a broader range of sources might be consulted, such as rental records from institutional landholders on the island. This last approach mirrors that taken by G. Clark 2002, Devaney 2010, and Drelichman and Gonzalez Aguda 2014. There is an archive for the Trinity Church properties that may be fruitful and which this author has just visited and collected rental data from 1900-1920 for and lease data from 1868 to 1900.

Finally, it should be acknowledged that concern about selection bias in this sample will depend on the use to which the data is put. Where the data is used to, say, construct a quality-adjusted rental price index for the island of Manhattan then the use of hedonic regression methodology and the fact that the final dataset proved rich in unit characteristics and is matched at a very fine level to nearby amenities and disamenities will allow the user to control for composition bias in the sample. This would not be possible with a less rich dataset or if the repeated sales/letting

methodology were preferred. Similarly, to construct quality-adjusted indexes for different social strata or different geographic areas, we would want to be sure that the sample had good coverage in all of these dimensions. Were we forced to use a repeated transactions approach then these composition issues might be of greater concern.

References

Anbinder, Robert C. (2001) Five Points. New York: Free Press.

Chapin, Robert C. (1909) The Standard of Living Among Workingmen's Families in New York City. New York: Russell Sage Foundation.

Charity Organization Society of the City of New York (1900) The Charities Review, Volume X. New York.

Clark, Gregory (2002) "Shelter from the Storm: Housing and the Industrial Revolution, 1550-1909." *Journal of Economic History* 62 (2): 489-511.

Conference on Research in Income and Wealth (1975) Trends in the American Economy in the Nineteenth Century. New York: Arno.

Day, Jared N. (1999) Urban Castles: Tenement Housing and Landlord Activism in New York City 1890-1943. New York: Columbia University Press.

Devaney, Steven (2010) "Trends in Office rents in the City of London: 1867-1959." *Explorations in Economic History* 47: 198-212.

Drelichman, Mauricio, and David Gonzalez Aguda (2014) "What Price a Roof? Housing and the Cost of Living in Sixteenth-Century Toledo?" *Explorations in Economic History* 54 (1): 27-47.

Eicholtz, Piet, Matthijs Korevaar, and Thies Lindenthal (2017) "500 Years of Urban Rents, Housing Quality and Affordability in Europe" Working Paper.

Eicholtz, Piet, Stefan Straefmans, and Marcel Theebe (2012) "The Amsterdam Rent Index: The Housing Market and the Economy, 1550-1850." *Journal of Housing Economics* (21): 269-282.

Fogelson, Robert M. (2013) The Great Rent Wars: New York, 1917-1929. New Haven: Yale University Press.

Gabaccia, Donna R. (1984) From Sicily to Elizabeth Street: Housing and Social Change Among Italian Immigrants, 1880-1930. Albany: State University of New York Press.

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson (2014) "Competition and Ideological Diversity: Historical Evidence from US Newspapers." *American Economic Review* 104 (10): 3073-3114.

Kholodolin, Konstantin A. (2016) "War, Housing Rents, and Free Market: Berlin's Rental Housing Market during World War I." *European Review of Economic History* (20): 322-344.

Kholodolin, Konstantin A. (2012) "Noise Expectation and House Prices. The Reaction of Property Prices to an Airport Expansion." DIW Discussion Paper 1212.

King, Moses (1892) King's Handbook of New York City: An Outline History and Description of the American Metropolis. Boston: Moses King.

Knoll, Katharina, Moritz Schularick, and Thomas Steger (2017) "No Price Like Home: Global House Prices, 1870-2012." *American Economic Review* 107 (2): 331-353.

Margo, Robert A. (1996) "The Rental Price of Housing in New York City, 1830-1860." *Journal of Economic History* 56 (3): 605-625.

Raff, Daniel M.G., and Manuel Trajtenberg (1997) "Quality-Adjusted Prices for the American Automobile Industry: 1906-1940," in Timothy F. Bresnahan and

Robert J. Gordon (eds.) The Economics of New Goods, NBER Volume 58. Chicago: University of Chicago Press: 71-108.

Rees, Albert (1961) Real Wages in Manufacturing, 1890-1914. Princeton: Princeton University Press.

Riis, Jacob (1997) How the Other Half Lives. New York: Penguin. First edition New York: Scribner.

Ruggles, Steven J., Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek (2010) Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Samy, Luke (2015) "Indices of House Prices and Rent Prices of Residential Property in London, 1895-1939." University of Oxford Discussion Papers in Economic and Social History 134.

Scherzer, Kenneth A. (1992) The Unbounded Community: Neighborhood Life and Social Structure in New York City, 1830-1875. Durham: Duke University Press.

Schulz, Wiebke, Ineke Maas, and Marco H.D. van Leeuwen (2014) "Employer's Choice-- Selection through Job Advertisements in the Nineteenth and Twentieth Centuries." *Research in Social Stratification and Mobility* (36): 49-68.

Van Leeuwen, Marco H.D., Ineke Maas, and Andrew Miles (2002) HISCO: Historical International Standard Classification of Occupations. Leuven: Leuven University Press.

Figures

Figure 1: Rents by Newspaper Source

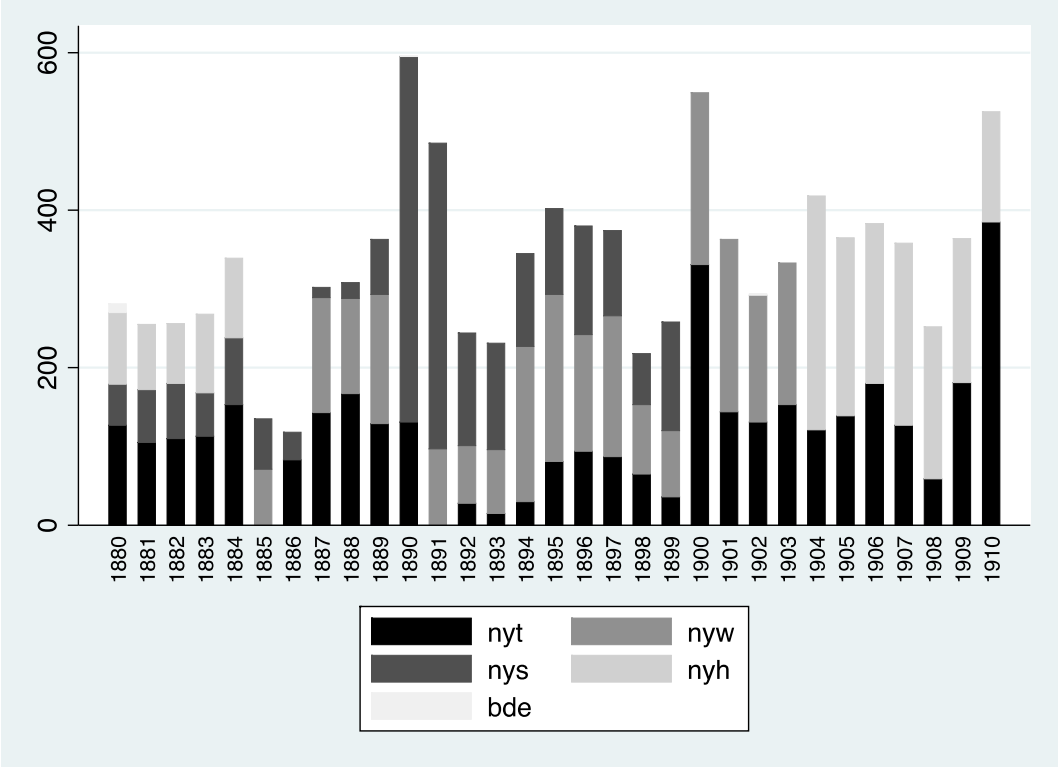


Figure 2: Geographic Coverage

Notes: The observations off the Manhattan map are the few observations coded that lie in the Bronx, Brooklyn or other outlying areas.

Figure 3: Rents by Neighborhood

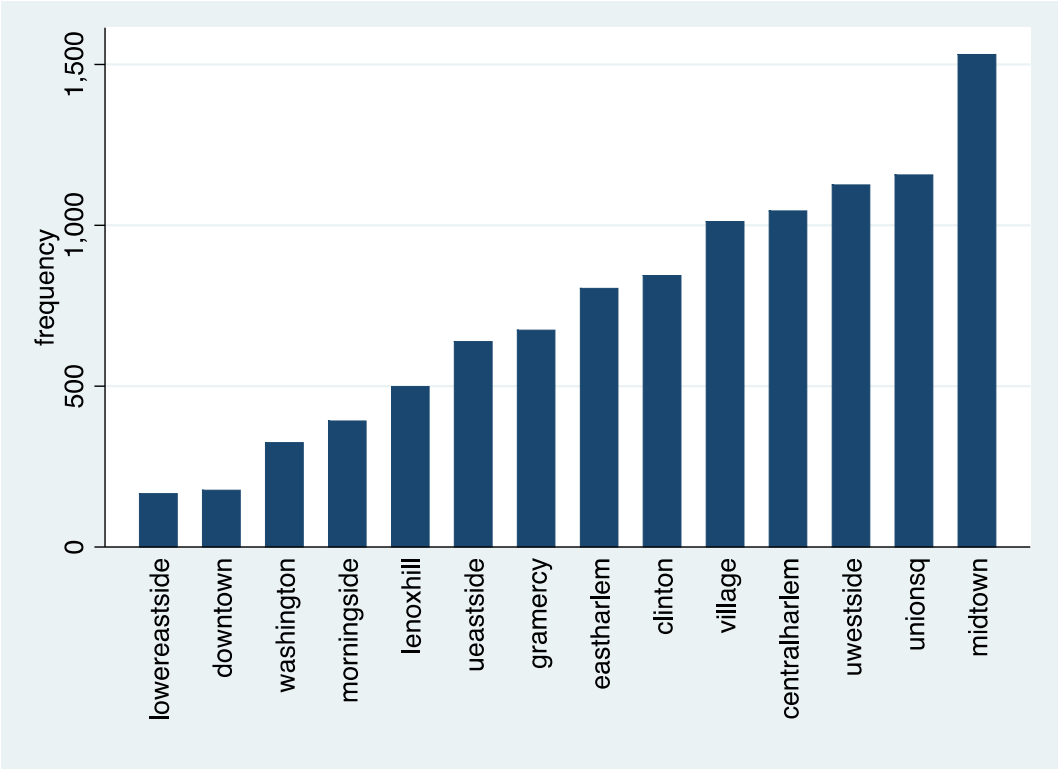


Figure 4: 1890-1891 Rental Data, by Sample Status

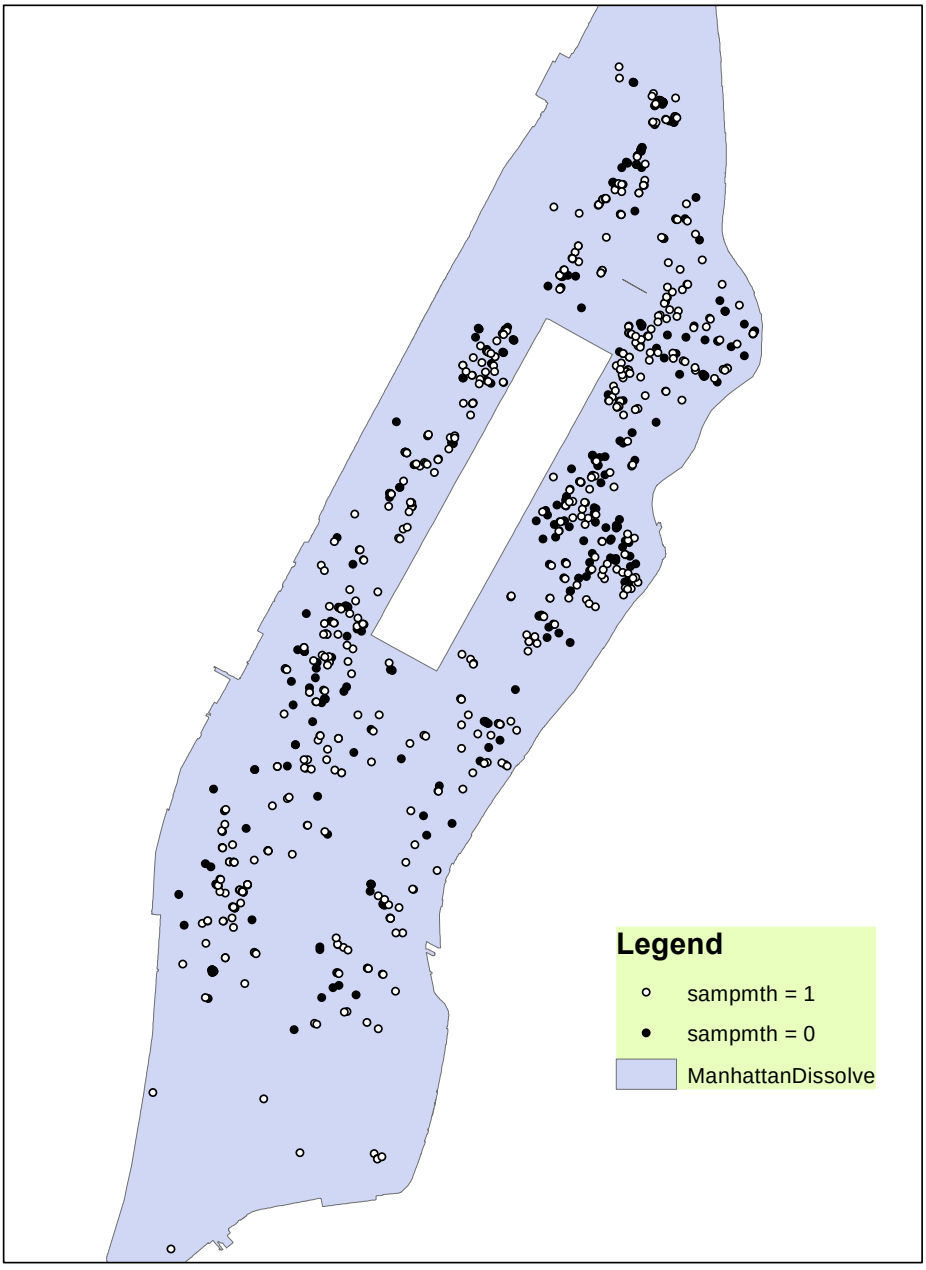
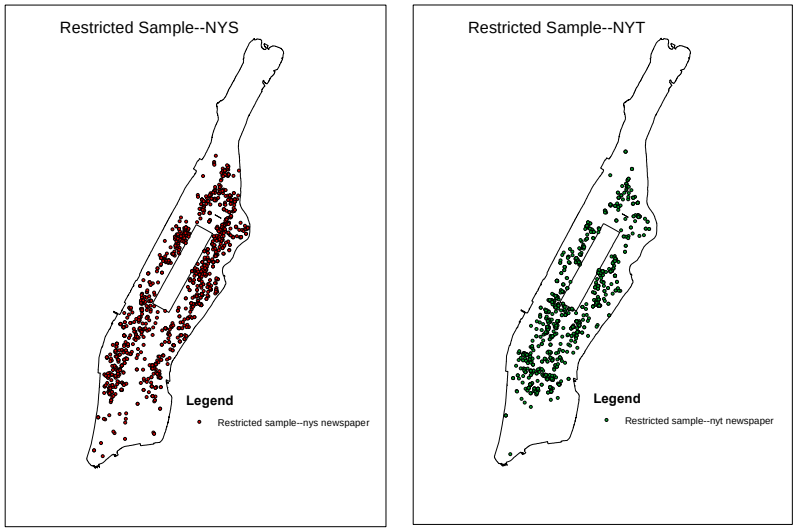


Figure 5: Pre-1900 Rents for Two Newspapers



Appendix

Examples of Rental Advertisements from Newspapers

A. New York Times, March 3, 1880

TO LET—THE ENTIRE DWELLING PART OF No. 76 6th-av., north-east corner of Waverley-place; good stand for millinery or dress-making; five rooms; rent, $500; possession May 1.

Also, the store and front basement room No. 129 Waverley-place; $30 per month; no liquor; immediate possession.

Also, on 10 years' lease, the buildings No. 46 Hudson-st. and No. 88 Thomas-st.; lot forms an L; rent, $2,500.

THORNTON M. RODMAN, Real Estate Agent, No. 696 Broadway, corner 4th-st.

STORE AND BASEMENT TO LET ON Broadway, near Bleecker-st., $2,200; bargain; also, fine corner store, near A. T. Stewart's, very low.

E. A. CRUIKSHANK & CO., No. 68 Broadway.

TO LET—THE STORE AND DWELLING NO. 8 Bowery; excellent business position; rent, $2,250; possession May 1. THORNTON M. RODMAN, Real Estate Agent, No. 696 Broadway, corner 4th-st.

B. New York World, April 9, 1889

BOARDERS WANTED.

East Side.

MADISON AVE., 155—Second floor: all light rooms; other large and single rooms; excellent board.

MADISON AVE., 73, near Madison Square—Second-story handsome large room; good table; references.

MADISON AVE., 144—With board, desirable second floor, en suite or separately; also fourth-floor square room.

MADISON AVE., 163—Desirable rooms, with large closets, running water; with board.

2D AVE., 1590—Parlor and bedroom, with or without board; suitable for a married couple or two single gentlemen or ladies.

2D AVE., 126—Three connecting rooms, suitable for 3 or 4 gentlemen, with board.

4TH AVE., 10, opposite Cooper Union—Furnished rooms, good substantial board; English cooking; $4.50 upwards.