Online Appendix to Intergenerational Elasticities of Housing Consumption and Income*

Lancelot Henry de Frahan† Jung Sakong‡

August 14, 2024

*The views expressed in this paper are those of the authors and do not reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

†University of Chicago (email:lancelot@uchicago.edu)

‡Federal Reserve Bank of Chicago (email: jung.sakong@chi.frb.org)

Online Appendix

A CEX Durable Service Flows

This section follows Meyer and Sullivan (2017)'s data appendix very closely. For simplicity and because there are a couple of very minor differences with what they do, we describe the computation of service flows in this section.

Housing Services: For renters, housing consumption is defined as the annual rent they pay. For Homeowners, it is the self-reported rental equivalent for their house. Homeowner housing expenses such as mortgage interest, property tax, maintenance, repairs, insurance and other expenses are substracted from total expenditures and replaced by the rental equivalent. For renters in public housing (only 2.37% of the sample), we impute a rental value using a quantile regression of the log rent on an urban indicator, region and state dummies, the SMSA status, number of bathrooms, bedrooms, half bathrooms and rooms, the presence of window or central air conditioning as well as region specific cubic trends. We impute the predicted 40th quantile of rent based on the house and geographic characteristics. This is justified by evidence from the PSID mentioned by Meyer and Sullivan (2017).

Vehicle Service Flows: The purchase price of new and used vehicles as well as finance charges are substracted from expenditures and converted into a service flow. The service flow is computed as $\delta * (1-\delta)^t * P$ where P is the original purchase price (in real terms) and t is the number of years since the vehicle was bought. The depreciation rate, δ is estimated by regressing log price of the car on its age at the time of purchase and make-model-year fixed effects. This regression is ran on a sample of all vehicles for which the purchase price is reported and which were bought within one year of the interview (this sample includes 172,160 vehicles). The coefficient from this regression is -0.17. It is then converted into a depreciation rate by taking $\delta = 1 - exp(-0.17) = 0.156$. There is a substantial number of vehicles in the sample for which the purchase price is not reported. We impute a current market price for these vehicles by running the regression described below on the same estimation sample as above (sample of cars bought within 12 months of the interview with a reported purchase price). We run two separate regressions for the years 1980-2002 and 2003-2016 because of a change in the set of variables available in those years. For years 1980-2002, we regress log purchase price on a cubic in the age of the vehicle, an indicator for new/used, automatic transmission, power brake, power steers, air conditioning, diesel, an urban indicator, a quadratic in log total family expenditures, fixed effects for family size, age and education of the head of household and make-model-year fixed effects. For years, 2003-2016, the independent variables are: a cubic in the age of the vehicle, an indicator for new/used, an urban indicator, a quadratic in log total family expenditures, fixed effects for family size, age and education of the head of household and make-year fixed effects. The coefficients from those regressions are used to predict current market price where it is missing. The prediction is scaled up by the coefficient of a regression (again in the estimation sample) ,without constant, of the reported purchase price on the predicted price. This is because by Jensen's inequality - we ran the regression on log price - the predicted price *in levels* will tend to be an underestimate.

B Discussion of Inferring Total Consumption IGE

We start with the Engel curves for housing in each generation v:

$$h_i^v = \theta_0^v + \theta_1^v c_i^v + \nu_i^v \tag{B.45}$$

where h_i^v is log housing consumption, c_i^v is log total consumption and generations $v \in \{o, y\}$ are measured in 1940 and 2015 respectively. We assume $cov\left(c_i^o, \nu_i^o\right) = 0$, $cov\left(c_i^y, \nu_i^y\right) = 0$ and $cov\left(c_i^o, \nu_i^y\right) = 0$. We form the following proxies for consumption in each generation $v \in \{o,y\}$:

$$c_i^{*v} = -\frac{\theta_0^v}{\theta_1^v} + \frac{1}{\theta_1^v} h_i^v \tag{B.46}$$

Notice that $c_i^{*v} = c_i^v + \frac{1}{\theta_1^v}\nu_i^v$ for $v \in \{o,y\}$. We also define a few correlation coefficients between variables of interest:

$$\rho_{\nu\epsilon} = \frac{cov\left(\nu_i^o, \varepsilon_i\right)}{\sqrt{var\left(\nu_i^o\right)}\sqrt{var(\varepsilon_i)}} \tag{B.47}$$

$$\rho_{\nu\nu} = \frac{cov\left(\nu_i^o, \nu_i^y\right)}{\sqrt{var\left(\nu_i^o\right)}\sqrt{var\left(\nu_i^y\right)}} \tag{B.48}$$

$$\tag{B.49}$$

Let us assume that $0 \leq \rho_{\nu\nu} \leq 1$. The fact that $\rho_{\nu\nu}$ is potentially greater than 0 reflects the fact that the portion of housing that is not explained by total consumption may be partially correlated over generations (if taste for housing is hereditary for instance). We rule out data generating processes in which $\rho < 0$ as it seems implausible. Finally, when estimating the housing Engel curve equations in the CEX data, we obtain an estimate of:

$$R_{hc,v}^2 = \frac{\left(\theta_1^v\right)^2 var\left(c_i^v\right)}{var\left(h_i^v\right)} \tag{B.50}$$

B.1 Potential OLS bias

Running OLS is impossible in our dataset because we do not have individual linkages across time. For the sake of intuition, it is however useful to consider the probability limit of the OLS coefficient (in the intergenerational elasticity of consumption regression) when using housing consumption as a proxy. The probability limit of the OLS estimator is:

$$\beta_{OLS} = \frac{cov\left(c_i^{*y}, c_i^{*o}\right)}{var\left(c_i^{*o}\right)} \tag{B.51}$$

$$= \frac{cov\left(c_i^o + \frac{1}{\theta_1^o}\nu_i^o, \beta c_i^o + \varepsilon_i + \frac{1}{\theta_1^y}\nu_i^y\right)}{var\left(c_i^{*o}\right)} \tag{B.52}$$

By definition of ε_i being the prediction error in the population regression, $cov\left(c_i^o, \varepsilon_i\right) = 0$. Also, by assumptions on the Engel curve equations, $cov\left(c_i^o, \frac{1}{\theta_1^y}\nu_i^y\right) = 0$ and $cov\left(\frac{1}{\theta_1^o}\nu_i^o, \beta c_i^o\right) = 0$. As a result, we have:

$$\beta_{OLS} = \beta \frac{var\left(c_i^o\right)}{var\left(c_i^{*o}\right)} + \frac{\frac{1}{\theta_1^o} cov\left(\nu_i^o, \varepsilon_i\right)}{var\left(c_i^{*o}\right)} + \frac{\frac{1}{\theta_1^o}\frac{1}{\theta_1^y} cov\left(\nu_i^o, \nu_i^y\right)}{var\left(c_i^{*o}\right)} \tag{B.53}$$

Let us focus on the second term which can be re-arranged as:

$$\frac{\frac{1}{\theta_1^o} cov\left(\nu_i^o, \varepsilon_i\right)}{var\left(c_i^{*o}\right)} = \rho_{\nu\varepsilon}\sqrt{1 - R_{hc,o}^2}\sqrt{\frac{var(\varepsilon_i)}{var\left(c_i^{*o}\right)}} \tag{B.54}$$

$$= \rho_{\nu\varepsilon}\sqrt{1 - R_{hc,o}^2}\sqrt{R_{hc,o}^2}\sqrt{\frac{var(\varepsilon_i)}{var\left(c_i^o\right)}} \tag{B.55}$$

Under the assumption of a stationary distribution of consumption, $var\left(c_i^o\right) = var\left(c_i^y\right)$ and the term is equal to $\rho_{\nu\varepsilon}\sqrt{1-R^2_{hc,o}}\sqrt{R^2_{hc,o}}\sqrt{1-R^2_{cc}}$ where R^2_{cc} is the R^2 of the regression of c_i^y on c_i^o. This expression underlies a specific threat to consistency of OLS estimates when using housing as a proxy: there may be a correlation between ν and ε. For instance, if housing is a vehicle for asset accumulation, families who own and consume a lot of housing relative to their total consumption (large ν_i^o) in 1940 may end up with a larger total consumption in 2015 than predicted by their 1940 total consumption. Notice however that we use both renters and owners in 1940 and there is presumably no reason to believe that high ν renters in 1940 would systematically have a larger ε in 2015. From now on, we assume $\rho_{\nu\varepsilon} = 0$ but keep in mind that housing as a mechanism of asset accumulation may generate a bias. As made clear by the expression above, the size of the bias is lower for very large $R^2_{hc,o}$ (housing is such a strong proxy for total consumption that variance in ν is small).[1] Under the assumption that $\rho_{\nu\varepsilon} = 0$:

$$\beta_{OLS} = \beta\frac{var\left(c_i^o\right)}{var\left(c_i^{*o}\right)} + \frac{\frac{1}{\theta_1^o}\frac{1}{\theta_1^y}cov\left(\nu_i^o,\nu_i^y\right)}{var\left(c_i^{*o}\right)} \tag{B.56}$$

$$= \beta\frac{\left(\theta_1^o\right)^2 var\left(c_i^o\right)}{var\left(h_i^o\right)} + \rho_{\nu\nu}\frac{\frac{1}{\theta_1^y}\sqrt{var\left(\nu_i^o\right)}\sqrt{var\left(\nu_i^y\right)}}{\frac{1}{\theta_1^o}var\left(h_i^o\right)} \tag{B.57}$$

$$= \beta R^2_{hc,o} + \rho_{\nu\nu}\frac{\theta_1^o}{\theta_1^y}\sqrt{1-R^2_{hc,o}}\sqrt{1-R^2_{hc,y}}\sqrt{\frac{var\left(h_i^y\right)}{var\left(h_i^o\right)}} \tag{B.58}$$

Let us assume that distributions are stationary (and Engel curves are time invariant) so that $R^2_{hc,o} = R^2_{hc,y} = R^2_{hc}$, $var\left(h_i^y\right) = var\left(h_i^o\right)$ and $\theta_1^o = \theta_1^y$. We obtain:

$$\beta_{OLS} = \beta R^2_{hc} + \rho_{\nu\nu}\left(1-R^2_{hc}\right) \tag{B.59}$$

This expression has two component. First, the use of a proxy for total consumption leads to attenuation bias because the proxy is a (classical) error ridden version of the true regressor. The attenuation bias stronger if the R-squared of the Engel curve regression is lower. Note that R^2_{hc} can be estimated in the CEX data and one can therefore correct for attenuation bias in OLS. Second, intergenerational correlation in ν - preference for housing for instance - leads to a bias in the estimate. Using the assumption that $0 \le \rho_{\nu\nu} \le 1$ we can bound the size of this second term by $0 \le \rho_{\nu\nu}\left(1-R^2_{hc}\right) \le 1-R^2_{hc}$. As reported in Appendix Table OA.3, the R^2 of the Engel curve regression in the CEX data ranges from 0.55 to 0.83.

C Placebo and Robustness

In this section, we discuss in more details the placebos shown on table OA.4. Comparing the first and second columns, we note that whether unadjusted surname-level correlations recover the true family-level correlations likely hinges on whether the process of intergenerational transmission of consumption is more intimately linked to race, geography, or human capital. Indeed, the surname-level correlations between housing consumption and proxies for geography and race are markedly different than the individual-level correlations. On the other hand, the surname-level correlations with proxies for human capital are of similar magnitude to the individual-level correlations.

In the last three columns, we show the estimated relationship between each proxy and the log of housing consumption in 1940 under the three different covariate adjustments presented in section 3 (i.e. race, Census region, and race by Census region). These placebo tests can be interpreted in two ways. First, it is a test of the conditions required in Propositions 1 and 3 for consistency of the estimator. The second (related) interpretation is that the placebo tests act as a "laboratory" for our adjustment: they compare the coefficients

[1] The bias is also larger when the variance of ε is large relative to consumption - that is when 1940 actual consumption is not a strong predictor of 2015 consumption.

of an individual-level regression with the corresponding last name level regression (after performing our adjustment). By the analogy principle, if surname- and individual- level estimates are close in the placebos, we also expect them to be close in the intergenerational regression.

The results from these placebos tests are somewhat mixed. They suggest that, when it comes to studying the correlation between log of housing consumption and proxies that are very closely related to the covariates in G (i.e. the covariates included in the adjustment), the adjustment closes the gap between the individual-level and surname-level correlations. For instance, adjusting for race brings the surname-level correlation between *hispanic* and log housing consumption very close to its individual-level counterpart. Similarly, adjusting for Census regions greatly reduces the gap in estimated correlation with *County Wage* (the average wage in one's county of residence). On the other hand, the covariate adjustment usually slightly increase the gap between surname-level and individual-level correlations for measures of human capital. We note however that magnitudes remain roughly comparable for most proxies of human capital (despite the increased gap). In addition, the unadjusted surname-level correlations were already very similar to their individual-level counterpart.

Our conclusion from these placebo tests is the following: the best specification (unadjusted versus adjusted) depends on whether the process of intergenerational transmission of consumption is more intimately linked to race, geography, or human capital. Since it is difficult to take a firm stand on this issue, we present the results from all four specifications. Fortunately for the sake of interpretation, the magnitude of the difference in estimated (one-generation) IGE of consumption is not too large across specifications: it varies from 0.73 to 0.79.

D Surname Group Size

In this section we refer to the intersection of races and regions as "groups" and explore the variation in group specific housing consumption across surnames. The goal is to get a sense of (i) whether the data contains enough variation in group specific log housing consumption to consistently estimate the heterogeneous parameters, and (ii) whether the answer to (i) depends on the size of surnames included in the sample. In order to do so, we perform simulations. The simulations work as follow. First, we draw (with replacement) a sample of size N from the 1940 Census - similar to what we would do for a bootstrap. From this sample, we generate the x^y data (log housing consumption in 2015) by using the following model: $x_i^y = \sum_g \left[\alpha^g D_i^g + \beta^g \left(D_i^g x_i^0\right)\right] + \eta_i$ for known values of $\{\alpha^g\}$ and $\{\beta^g\}$.[2]. Notice that because η_i is drawn independently from α^g and β^g, the condition required for consistency in Proposition 3 holds. We generate 200 different split samples and run the surname-level estimation in each of them. Tables OA.6 and OA.5 report the mean and standard deviation of the estimates across the 200 estimations.

The estimated parameters are all (almost exactly) equal to their "true" value. Estimates are unbiased and standard errors are small. The simulation exercise suggests that there is enough variation in the data. We further explore variations in the data by performing the same simulations with restricted samples: one that contains only last names with less than 50 male heads of household in 1940 (table OA.7) and another sample containing only last names with more than 50 male heads of households (table OA.8). There is not much of a difference between table OA.7 and OA.8. The estimates remain unbiased and standard errors are reasonable albeit slightly larger when restricting the sample to large surnames only.

Online Appendix References

Meyer, B. D. and J. X. Sullivan (2017): "Consumption and Income Inequality in the US Since the 1960s," Tech. rep., National Bureau of Economic Research.

[2] We draw η_i from a normal with mean zero and variance 1.4 (which is equal to the actual variance of log housing consumption in the 1940 data)

Figure OA.1
Data Schematic

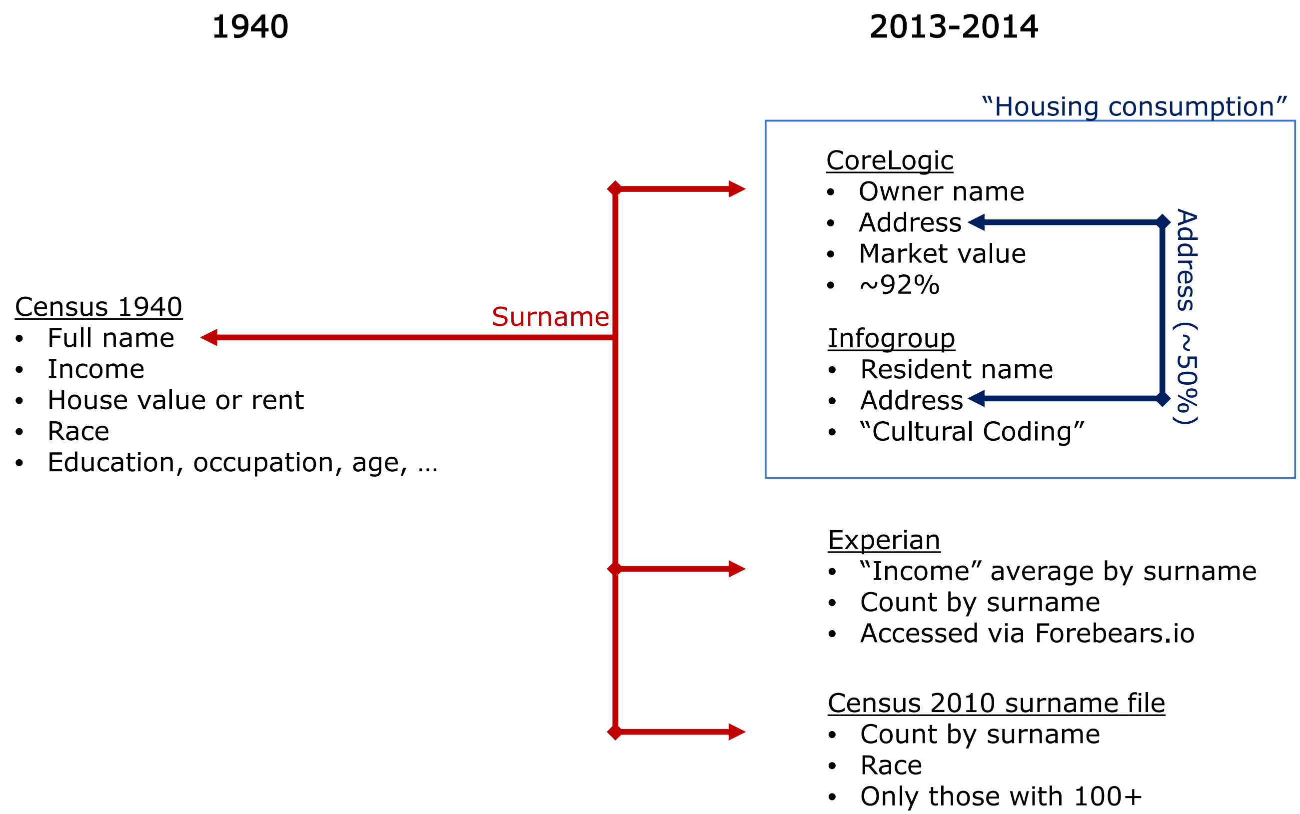
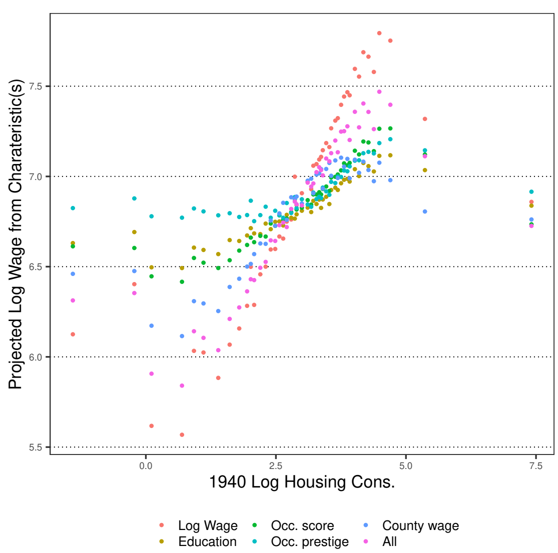
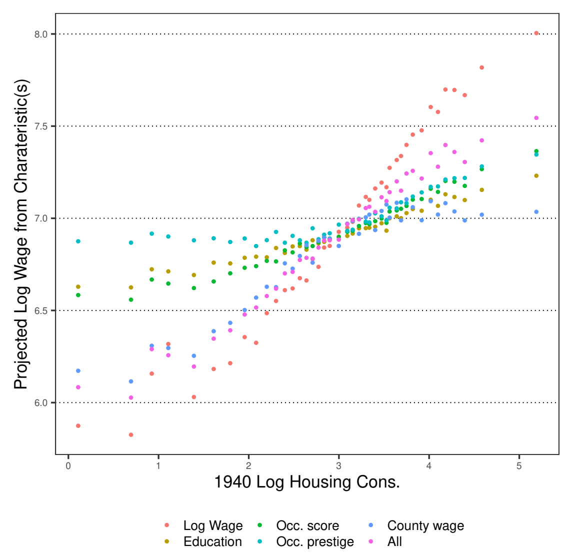


Figure OA.2
Motivating Sample Restrictions

(A) Before imposing sample restrictions

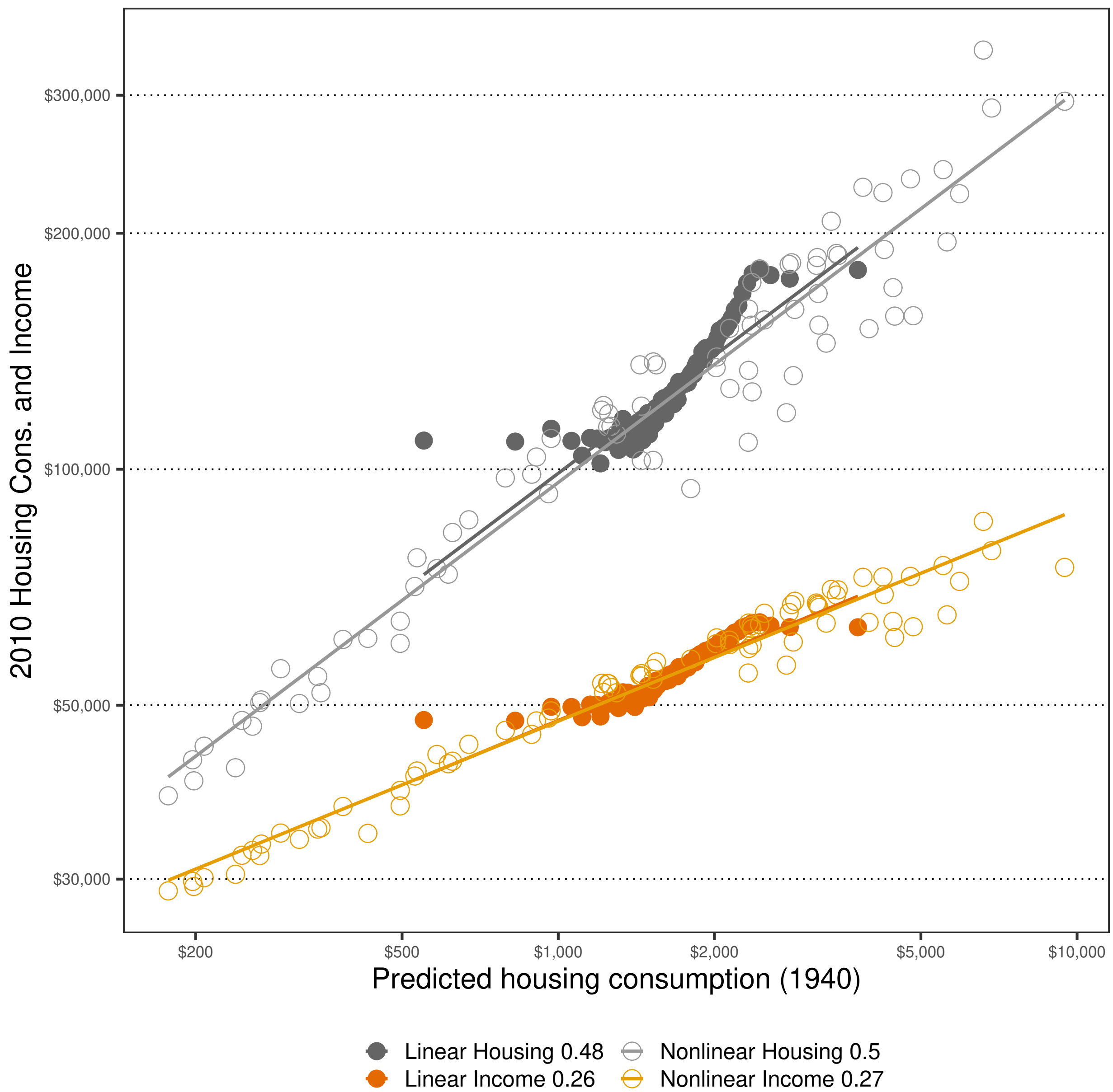


(B) After imposing sample restrictions



Notes. Panel (A) plots the mean of the individual log wage, as well as the mean of the predicted log wage based on, respectively, the occupational score, the average county wage, years of education, occupational prestige, and all these socio-characteristics at once against the percentiles of the distribution of log housing consumption in the 1940 Census *including* renters with rent above $81 and homeowners with home value below $100. Panel (B) plots the same mean outcomes against percentiles of the distribution of log housing consumption *after* dropping renters with rent above $81 and homeowners with home value below $100.

FIGURE OA.3
2SLS



Notes. The figure displays binscatters corresponding to two 2SLS specifications (linear and non-linear) for each of two outcomes: 2015 labor income and 2015 housing value. The linear specification is a simple 2SLS at the surname-level in which 1940 (surname-) average housing consumption is instrumented with 1940 (surname-) average labor income. The non-linear specification leverages the share of male heads of households, in each surname, belonging to 100 percentiles of the 1940 distribution of labor income to instrument for 1940 log housing consumption. The hollowed dots show the predicted 2015 outcome against the predicted 1940 log housing value for each percentile. The reported coefficient is the slope of a OLS regression through these 100 dots. All four reported coefficients are slopes of log-log specifications but units displayed on the x- and y-axes are converted to dollar values for ease of interpretation.

Table OA.1
Adjusted IGE: Full estimates by race

b^g Log Housing		Unadjusted		Adj. (Region)	
	$\mathbb{E}\left[x_i^0\right]$	a	b	$\hat{a}$	$\hat{b}$
Race					
White	7.504	8.354 (0.010)	0.462 (0.003)	8.790 (0.009)	0.432 (0.003)
Black	6.435	9.891 (0.021)	0.167 (0.011)	9.872 (0.026)	0.172 (0.015)
Other	6.767	8.478 (0.026)	0.477 (0.013)	8.396 (0.019)	0.529 (0.011)
All	7.407	8.362 (0.009)	0.464 (0.003)	8.979 (0.009)	0.389 (0.002)
b^g Log Income		**Unadjusted**		**Adj. (Region)**	
	$\mathbb{E}\left[x_i^0\right]$	a	b	$\hat{a}$	$\hat{b}$
Race					
White	7.005	9.837 (0.004)	0.162 (0.001)	10.239 (0.004)	0.102 (0.001)
Black	6.177	9.217 (0.014)	0.166 (0.007)	9.677 (0.015)	0.113 (0.008)
Other	6.442	9.617 (0.011)	0.177 (0.006)	9.985 (0.010)	0.154 (0.005)
All	6.935	9.784 (0.003)	0.211 (0.001)	10.189 (0.004)	0.145 (0.001)

Notes. The upper panel of this table reports the mean log housing consumption in 1940, the intercept and slope from surname-level regression of log housing consumption on log housing consumption, and the intercept and slope from the region-adjusted regression, all by race. The lower panel reports the corresponding statistics for income. Standard errors are block-boostrapped at the surname-level.

TABLE OA.2

BY SURNAME COUNTRY OF ORIGIN VS. IMMIGRATION

Country of origin	N of surname	Count in 1,000s 1940	Count in 1,000s 2015	Share post-1940	Consumption Slope	Consumption S.e.	Income Slope	Income S.e.
Ireland	9,933	7,575	15,578	0.10	0.51	(0.01)	0.49	(0.01)
Germany	48,668	9,919	18,548	0.23	0.42	(0.01)	0.24	(0.02)
Norway	2,585	558	1,070	0.29	0.28	(0.05)	0.14	(0.03)
Sweden	3,891	1,496	2,542	0.31	0.45	(0.07)	0.29	(0.06)
Finland	3,189	116	255	0.32	0.21	(0.02)	0.07	(0.01)
France	18,042	2,881	7,084	0.36	0.39	(0.01)	0.24	(0.01)
Lithuania	1,566	47	121	0.36	0.35	(0.08)	0.09	(0.04)
Denmark	511	608	1,295	0.37	0.56	(0.09)	0.34	(0.07)
Czech	3,868	496	1,019	0.46	0.23	(0.03)	0.08	(0.01)
Austria	445	533	1,019	0.46	0.48	(0.06)	0.33	(0.05)
Hungary	5,648	443	976	0.47	0.23	(0.04)	0.08	(0.02)
Netherlands	15,585	2,076	4,335	0.49	0.37	(0.01)	0.25	(0.01)
Romania	1,031	34	135	0.53	0.32	(0.08)	0.16	(0.05)
Belgium	558	93	168	0.53	0.29	(0.05)	0.27	(0.04)
Poland	25,360	1,208	2,730	0.56	0.19	(0.01)	0.04	(0.01)
Italy	48,309	3,366	9,553	0.62	0.43	(0.01)	0.12	(0.02)
Latvia	370	16	35	0.64	0.06	(0.09)	0.18	(0.05)
Estonia	170	23	41	0.75	0.41	(0.12)	0.11	(0.05)
Portugal	1,354	189	1,078	0.76	0.47	(0.06)	0.24	(0.03)
Greece	8,925	197	639	0.77	0.45	(0.08)	0.08	(0.04)

Notes. Each row of the table describes a subset of surnames whose origin can be traced to one origin country. First three columns list the number of unique surnames in each subset, and the number of individuals with that surname in 1940 and 2015. The fourth column shows the share of immigration into the US from that origin country that took place after 1940. The last four columns display the surname-level b from equation 2, for average log consumption on average log consumption (slope and standard error) and for average log income on average log income (slope and standard error).

Table OA.3
Housing Engel curves

	1960s	1970s	1980s	1990s	2000s	2010s	All Decades
Housing Consumption (incl. rent/rent equivalent, utilities, housing services, etc.)							
Slope	1.027 (0.012)	0.865 (0.012)	0.917 (0.005)	0.909 (0.004)	0.950 (0.003)	0.912 (0.004)	0.925 (0.002)
R^2	0.692	0.626	0.813	0.789	0.819	0.818	0.826
N	13,387	16,930	14,837	31,715	42,476	25,141	144,486

Notes. Results from regressing a measure of housing consumption on total consumption in CEX. All regressions include year fixed effects and control for age, race, Census region and urban/rural status. The sample includes years 1959-1961, 1972-1973, 1984-2016. Standard errors are robust to heteroskedasticity.

Table OA.4
Placebo Test - Comparison of individual-level regressions with surname-level regressions

	(1)	(2)	(3)	(4)	(5)
Race & Ethnicity					
White	0.086	0.219	.	.	.
Black	-0.077	-0.159	.	.	.
Other race	-0.009	-0.060	.	.	.
Hispanic	-0.007	-0.055	-0.008	-0.074	-0.009
Foreign	0.054	0.233	0.236	0.078	0.097
Geography					
Northeast	0.128	0.341	.	.	.
Midwest	0.033	0.134	.	.	.
South	-0.174	-0.458	.	.	.
West	0.013	-0.016	.	.	.
County Wage	0.261	0.450	0.458	0.311	0.332
Human Capital					
Education	1.480	1.353	1.167	2.163	1.896
Occ. Score	4.698	5.266	5.086	5.934	5.693
Occ. Prestige	3.615	3.502	2.761	5.559	4.748
Farmer	-0.162	-0.207	-0.231	-0.168	-0.189
Employer	0.011	0.014	0.012	0.021	0.020
Self-employed	-0.060	-0.064	-0.080	- 0.034	-0.049
Private Employee	0.066	0.073	0.088	0.036	0.052
Public Employee	-0.007	-0.018	-0.015	-0.015	-0.013

Columns:
(1) Individual-level regression
(2) Surname-level regression
(3) Surname-level regression - adjusted for race
(3) Surname-level regression - adjusted for region
(3) Surname-level regression - adjusted for race and region

Notes. Coefficients displayed are from regressing the variable named in the first column on log housing consumption in 1940 at the individual- and surname level. Columns (3) – (5) contain coefficients corresponding to formula (7) after estimating race and/or region specific parameters. All last names are weighted by their number of male heads in 1940.

Table OA.5

Simulations: Estimates versus true parameters (IGE slope)

	Region									
	Northeast		Midwest		South		West		All	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Race										
White	0.67 .	0.67 (0.001)	0.2 .	0.2 (0.001)	0.245 .	0.245 (0.002)	0.096 .	0.096 (0.002)	0.311 .	0.317 (0.001)
Black	0.071 .	0.071 (0.017)	-0.093 .	-0.092 (0.017)	0.206 .	0.206 (0.005)	0.024 .	0.023 (0.035)	0.08 .	0.082 (0.004)
Other	-0.119 .	-0.118 (0.019)	0.209 .	0.209 (0.011)	0.254 .	0.256 (0.009)	0.497 .	0.497 (0.005)	0.368 .	0.368 (0.004)
All	0.654 .	0.655 (0.001)	0.202 .	0.203 (0.001)	0.247 .	0.246 (0.002)	0.129 .	0.13 (0.002)	0.308 .	0.314 (0.001)

Columns:
(1) True parameter
(2) Between Last Names OLS - adjusted for covariates (standard error in parenthesis)

Table OA.6

Simulations: Estimates versus true parameters (α constant)

	Region									
	Northeast		Midwest		South		West		All	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Race										
White	3.248 .	3.247 (0.009)	5.889 .	5.889 (0.008)	5.575 .	5.576 (0.009)	6.583 .	6.58 (0.014)	5.181 .	5.133 (0.004)
Black	6.44 .	6.439 (0.1)	6.037 .	6.03 (0.09)	5.679 .	5.679 (0.023)	5.618 .	5.621 (0.198)	5.789 .	5.793 (0.023)
Other	7.942 .	7.936 (0.117)	5.47 .	5.47 (0.051)	5.174 .	5.167 (0.04)	4.713 .	4.714 (0.023)	5.335 .	5.351 (0.023)
All	3.39 .	3.377 (0.009)	5.891 .	5.891 (0.008)	5.59 .	5.591 (0.008)	6.456 .	6.45 (0.013)	5.237 .	5.189 (0.004)

Columns:
(1) True parameter
(2) Between Last Names OLS - adjusted for covariates (standard error in parenthesis)

Table OA.7

Simulations: Estimates versus true parameters (IGE) - rare names only

	Region									
	Northeast		Midwest		South		West		All	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Race										
White	0.67	0.67	0.2	0.2	0.245	0.245	0.096	0.096	0.311	0.343
	.	(0.002)	.	(0.001)	.	(0.002)	.	(0.002)	.	(0.001)
Black	0.071	0.071	-0.093	-0.092	0.206	0.206	0.024	0.021	0.08	0.095
	.	(0.016)	.	(0.016)	.	(0.005)	.	(0.038)	.	(0.004)
Other	-0.119	-0.119	0.209	0.209	0.254	0.254	0.497	0.497	0.368	0.364
	.	(0.016)	.	(0.01)	.	(0.01)	.	(0.006)	.	(0.004)
All	0.654	0.658	0.202	0.203	0.247	0.251	0.129	0.143	0.308	0.341
	.	(0.002)	.	(0.001)	.	(0.002)	.	(0.002)	.	(0.001)

Columns:
(1) True parameter
(2) Between Last Names OLS - adjusted for covariates (standard error in parenthesis)

Table OA.8

Simulations: Estimates versus true parameters (IGE) - Large names only

	Region									
	Northeast		Midwest		South		West		All	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Race										
White	0.67	0.67	0.2	0.2	0.245	0.245	0.096	0.095	0.311	0.31
	.	(0.004)	.	(0.005)	.	(0.004)	.	(0.009)	.	(0.002)
Black	0.071	0.077	-0.093	-0.09	0.206	0.205	0.024	0.027	0.08	0.081
	.	(0.079)	.	(0.067)	.	(0.018)	.	(0.212)	.	(0.014)
Other	-0.119	-0.116	0.209	0.213	0.254	0.253	0.497	0.497	0.368	0.368
	.	(0.092)	.	(0.067)	.	(0.032)	.	(0.013)	.	(0.01)
All	0.654	0.654	0.202	0.202	0.247	0.246	0.129	0.125	0.308	0.307
	.	(0.004)	.	(0.005)	.	(0.004)	.	(0.009)	.	(0.002)

Columns:
(1) True parameter
(2) Between Last Names OLS - adjusted for covariates (standard error in parenthesis)