THE UNIVERSITY OF CHICAGO

The Grass is Greener on the Other Side
An Analysis of Consumption Inequality in Rural and Urban Areas

A BACHELOR THESIS SUBMITTED TO

THE FACULTY OF THE DEPARTMENT OF ECONOMICS

FOR HONORS WITH THE DEGREE OF

BACHELOR OF THE ARTS IN ECONOMICS

BY LONG PHAM

CHICAGO, ILLINOIS

MAY 2017

Consumption Inequality in Rural and Urban Households

Abstract

This paper explores the quantitative differences in household decision-making in rural and urban households. Both communities have their own distinct set of resources, choices, and problems. The issue of enacting policies that benefit both groups has troubled policymakers for generations. Using the U.S. Census Bureau's metric of urbanness and data from the Kilts-Nielsen Consumer Panel, I find quantitative differences in consumption of individuals in urban and rural households. Additionally, I find that the traditional definition of urbanness does not entirely account for two different types of agents and that urban households have a lower auto-correlation persistence factor than suggested by previous literature (0.68 vs. 0.98).

1 Introduction

1.1 Statement of the Research Question

Is there a difference in consumption inequality between rural and urban households? Can the existence of such a difference be explained by differences in income risk?

1.2 Motivation

The answers to the questions of inequality have direct effects on education policy and economic growth. For the most part, the literature mostly focuses on wealth and income inequality as prime motivators in problems with economic growth and education. However, the available literature ignores what these changes lead to which is consumption inequality. While a a few papers address how consumption inequalities affect the household and responses to inflation (Kaplan, et al. 2016) (Storesletten, et al. 2004), there is little to no literature on how consumption differs in metropolitan areas compared to rural areas which make up around 20 % of the U.S. population [10].

Both state and federal lawmakers seek to create catch-all policies that please all of their constituents. However, there are large differences in both social and economic conditions, such as political affiliation and special interests, between these two communities. Households face different prices, bundles, and choices based on where they reside and households often rely on this observation when determining their opinions of state and federal policies. This paper identifies of few of the economic outcomes of the aforementioned differences. Namely, it verifies the differences in income distributions of rural and urban areas and identifies the degree to which consumption inequality differs in both groups. The focus of this paper is to understand if consumption inequality differs in rural and urban households, and if they do, can they be explained by differences in income risk?

1.3 Definitions

What is rural and what is urban?

In order to stay consistent with the literature, and clarify the direction of this paper, we must define our set of terms. According to the United States Census Bureau, and subsequently every government agency, urban areas are defined as having a population density at least 1,000 persons per square mile [10]. Therefore, we shall define rural as anything less than this. Note that through this definition, we have also included suburban areas in our definition of rural populations. Later in this paper, I highlight some of the effects of grouping surburban households in the urban demographic rather than the rural one.

Formally known as urban clusters, suburban areas are defined by the U.S. Census Bureau as any any whose population density is between 500 and 1,000 persons per square mile. I include urban clusters in the rural group rather than the urban group because most suburbs of large cities (for example, Los Angeles or Chicago) tend to have an urban sprawl. That is, areas that are colloquially considered suburbs by their inhabitants actually fit the

Census Bureau's definition of an urban area. In addition, urban clusters do exist around cities but they are much further from the city's center than urban sprawl areas. Thus, it becomes increasingly difficult for them to access the city on a regular basis—whether for work, education, or leisure.

Still, urban clusters often appear between large areas of low density populations. We can think of these as major cities of sparsely populated states which act as centers of commerce and gathering [1]. These areas are not considered urban by the U.S. Census Bureau since their population density does not exceed the threshold. An investigation into the effects of grouping urban clusters with urban populations is done briefly in Section 4. Figure 1 depict these types of areas.

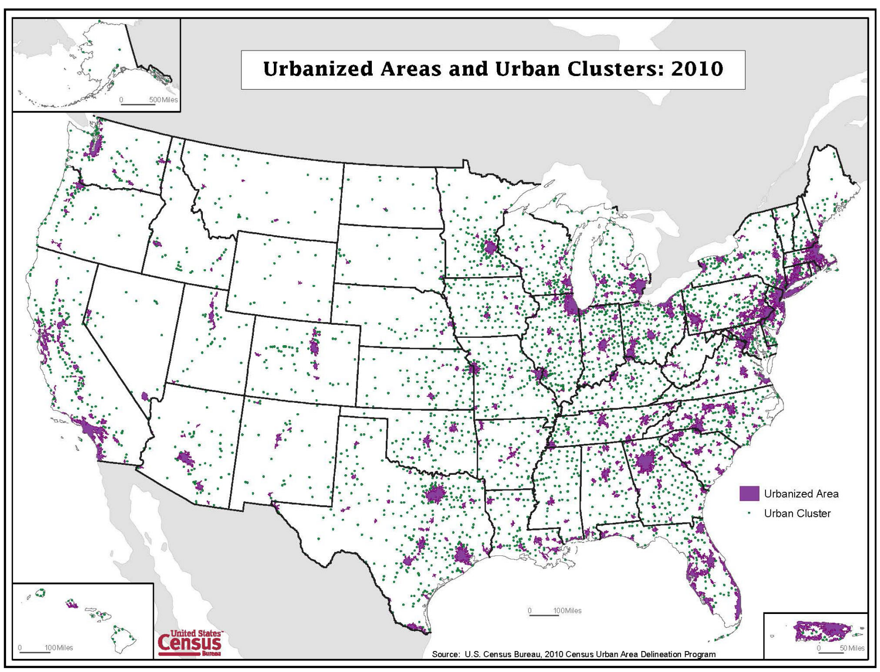


Figure 1: Urban Areas and Urban Clusters

Specific Consumption Metrics

For consumption metrics, I am, in particular, analyzing the difference in the underlying distributions and levels of non-durable goods (such as food). Due the nature of the data set,

[1]These are abundant primarily in the Midwestern United States. Examples of such places are Sioux Falls, South Dakota and Lincoln, Nebraska.

it is quite difficult to include assets like the price of an individual's home or the proportion of funds they spend transportation.

Specific Income Metrics

The model in this paper uses the assumption that there are varying income distributions between rural and urban areas. That is, the two key differences between rural and urban areas are that urban areas have a higher population density (by definition) and that income levels for urban populations has a higher variance. The data supports this assumption (see Section 2.2) and is used to calibrate the model in Section 4.

1.4 Background and Previous Work

Within the literature of inequality in the United States, there is a large focus cross-sectional wealth and income inequality in the nation as a whole. While both differences in income and wealth help prescribe decisions for policymakers, there is a small, but emerging subset of literature on the end result of the aforementioned inequality—consumption. Still, the available literature focuses on differences in consumption levels in the country at large and largely ignores the differing lifestyle choices and circumstances within the United States. Specifically, it ignores differences in population density–a large determinant of economic opportunities. This paper seeks to extend the available literature on inequality primarily relating to consumption by finding to what extent the cross-sectional levels of income and consumption differ in rural and urban areas in the United States.

In Kaplan and Schulhofer-Wohl (2016), the duo's major findings are that household inflation rates are heterogeneous and that they are independent of the aggregate inflation level. Furthermore, they found that household inflation rates vary due to differences in prices rather than differences in bundles. They find that at the aggregate geographical level, the cross-section of low incomes and older household heads experience higher inflation on average and, at the regional level, the West and Midwest experience lower inflation. However, they find these observable household characteristics to be insignificant in their ability to predict household inflation rates. Looking at a different subset of characteristics based on population, I find significant impacts on consumption—the direct result of household inflation rates.

In order to separate the effect of varying household inflation levels in areas with different population demographics, I control for the fixed effects of the household demographics. Using this method, as well as separating expenditure into its nominal and real components, Kaplan, Mitman, and Violante (2016) found that while nominal expenditures dropped significantly during the Great Recession, real consumption only dropped about 20% less. Furthermore, they reiterated the co-movement of consumption and housing prices found in Mian, Rao, and Sufi (2013). Measuring the elasticity with gross housing wealth, they found the elasticity of consumption to be 0.12 using OLS which is less than the one found in MRS (2013). Although the fall in housing prices was an aggregate shock, the cross-sectional fall in prices most likely also differs by region. Since the cost of housing differs vastly between dense and sparse population regions, both real consumption and nominal expenditure are expected to differ between the regions as well.

Another method of measuring consumption differences between different types of households compares the elasticities of a subset of goods. The example proposed in Aguiar and Bils (2015) compares low- and high- income households using non-durable entertainment (a high elasticity good) and food at home (a low elasticity good) claiming that the second-stage OLS regression on all goods tends to be inaccurate since there are a large number of small coefficients. While this is true, the use of certain elasticity ratios to determine the level of inequality may also prove to be inaccurate as it is difficult to determine which goods to select in the calculation. While the elasticity analysis may yield more accurate results between inequality between a small subset of goods in different households, the use of the LASSO and Weighted Least Squares regression methods this paper implements may yield results for a larger subset of goods and identify what principal good determine expenditure in different population communities and the income classes within them.

The model used in this paper is an extension of the model proposed in Kaplan (2012). Using data from the Panel Study of Income Dynamics (PSID) to calibrate the model, Kaplan uses a lifecycle model that closely matches the observed joint distribution of consumption, wages and hours in U.S. data. The strength of Kaplans proposed model lies with its use of a strong restriction on optimal labor supply over time. In my model, I make a modification to the labor supply decision. That is, the choice of the distribution of income over time is determined at birth by the settling in an urban or rural household. Using data from the Kilts-Nielsen Consumer Panel, I am able to estimate values for the level of income risk the household faces in each population group. This modification allows for a model that is more closely able to fit the data in different population divisions.

Based on the research done by Storesletten, Telmer and Yaron (2000) that idiosyncratic income shocks are important in the increase of both consumption and income inequality with age, I create two separate distributions of income shocks for the two population divisions. This key assumption I make on the distinctness of the two varying distributions follows from the papers finding that these income shocks are extremely persistent with an auto-correlation coefficient of greater than 0.98. Furthermore, the paper finds that inequality is not driven primarily by decisions but rather by a reduced-form statistical process. If this finding holds, both the data and my proposed variation on the life cycle model should find 1) distinct and significant distributions in steady-state income and consumption and 2) the initial endowment for each agent should not be indicative of their income and consumption in later years. The decreased risk (i.e. a higher level of risk sharing) in smaller communities results in a similar auto-correlation in the persistent shock as the auto-covariance found in Storesletten, et al. (2004). However, I find that the increased income risk present in urban households results in a much lower auto-correlation factor than in previous literature. The persistent auto-correlation factor (0.68) is about 30% lower than that found in Storesletten, et al. In addition to the decreased persistence factor of income, I also find that the variance of the shocks are higher for urban areas.

Blundell and Preston (1998) studies consumption inequality as a metric of evaluating permanent inequality. To do so, they use the growth paths of both income and consumption inequality to analyze the growth in short-term income risk. That is, they develop a method of identifying the permanent and transitory shocks to income to a household over time by using the variances and growth in variances of income and consumption. Further, they show evidence that income inequality rises at a faster rate than consumption inequality. Using their method, I solve a consumption-savings model with two separate income processes and show that higher levels of income risk in urban households cause the path of consumption to be more homogeneous that in rural households.

1.5 Organization

The paper proceeds as follows: Section 2 explores the empirical differences in consumption and income dynamics in rural and urban areas. The data shows significant differences in dynamics using population density as a dependent variable as well as population density as a separator. The analysis done in this section includes regression analysis relating consumption to population density and income. The analyses mentioned above are also shown in this section. The findings based on this analyses are discussed at the end of this section. Section 3 further describes the theoretical framework and provides a model for the behavior of households in each group. Section 4 describes alternative metrics of urbanness and their effectiveness and Section 5 concludes.

2 Data

This section describes the data collection process and how I use it to identify and verify differences in consumption and income.

2.1 Data Sources

I use data from the Kilts-Nielsen Consumer Panel, which from here on will be abbreviated KNCP. The panel follows approximately 60,000 households each year who provide information to Nielsen about their families, purchasing history as well as when and where they make these purchases. Each household scans the Universal Product Code (UPC) (bar code) using the Nielsen Homescan tool inside their homes to record all their purchases intended for home use. The products tracked include both non-durable and durable items but for the purposes of this study, we will concentrate primarily on non-durable goods, particularly food items.

Prices of the goods are recorded in two ways. If the store is covered by Nielsen, the good's price is set the the average price of the good at that store during the week the purchase occurred. Otherwise, the panelist is prompted to enter the price manually. This does not likely skew the data since we assume households have little incentive to change the price of the product dramatically.

The Nielsen data seeks to select panelists such that the proportion of households in the panel match that of the total population. However, this method of selection does not guarantee a representative sample of the population. Therefore, the panel users a "projection factor" for each household which denotes how many households in the population are estimated to share the same characteristics. The analyses below take this into account and weight accordingly using this factor. Additionally, households are dropped from the panel if they are not considered "active" by Nielsen. Nielsen defines active by a "minimum spending requirement per four week period depending on the household size". If a household meets this requirement over a 12-month period, their data is included in the panel. Nielsen also offers incentives for households such as monthly drawings and rewards upon entering monthly data, thus controlling for households that have a bias towards helping out with experimeints.

Although the panel dates back as far as 2004, the analysis only considers households that were part of the panel past 2010 (the years ranging from 2011 to 2015, inclusive on both ends). The rationale behind this decision is for the purpose of avoiding any heavy skew brought upon by the initial shock of the financial crisis. In doing so, we can make the assumption that aggregate shocks affecting all households is small.

Another reason for the use of the years ranging from 2011 to 2015 is the constancy of the real interest during this time period. The low and constant value of the real interest rate during this period (around 1.5%) [3] allows for easier calibration of the model as well as better juxtaposition between rural and urban areas without worrying about the effect on the interest rate on behavior in the two different population density categories.

While the KNCP contains information about the state and county of residence of each household in the data set, it does not contain census data at the individual county level. Therefore, using the latest census data from each state, I construct an index that maps each household to their corresponding geographical features. These features include the population of their county, their proximity to the nearest urban center, the population density as well as other factors (abundance of water area, number of total households in the county, etc...).

2.2 Methodology

Income

Due the nature of this study's differentiation of urban versus rural areas, the difference between the two income distributions must be verified. Since income comes in the form of categorical data based on range, I treat the data as a histogram with n buckets where n is the number of choices for income. Then using the "projection factors" as a weight vector, I compare the mean values of income between rural and urban areas. While the results are significant (all p-values much less than 0.01), the large sample sizes can skew the meaning of the p-value. Therefore, I calculate Cohen's $d-statistic$ to measure the effect size of income in both groups. The results are shown in Table 1. Since the effect size is greater than 0.5, then we can consider the two means to be reasonably distinct.

	Year				
	2011	2012	2013	2014	2015
Mean					
Aggregate	$ 87,216.20	$ 84,628.65	$ 85,756.39	$ 87,717.28	$ 89,242.59
Urban	$ 92,315.03	$ 90,211.78	$ 91,115.04	$ 92,211.74	$ 93,889.71
Rural	$ 83,716.23	$ 80,802.68	$ 82,054.12	$ 84,669.77	$ 86,083.59
Variance					
Aggregate	0.717	0.717	0.760	0.768	0.774
Urban	0.719	0.731	0.760	0.778	0.783
Rural	0.708	0.698	0.748	0.755	0.761

Note: Mean income is in dollars and variance is computed using log dollars.

Table 1: Income Summary Statistics

Consumption

For consumption of non-durable goods, I run a similar type of analysis. Comparing the means, I find the difference to be statistically significant. However, looking at the effect size 0.13, it looks like the difference is not as large in consumption. Furthermore, the comparison of the distributions seems inverted. For income, the urban group has a higher mean while for consumption, the rural group has higher mean. A probable cause for this is that while individuals in urban areas tend to have higher wages, it is a result of the higher cost of living and rent in these areas. While firms may compensate for the difference in housing costs, transportation costs are generally higher in urban areas. Individuals usually have to pay a premium or tax for driving their own vehicles in the form of gas or parking fees. Even if individuals do not own a vehicle, they incur the cost of using other means of transportation (i.e. public transit or ride-sharing).

	Year				
	2011	2012	2013	2014	2015
Mean					
Aggregate	$ 7,915.13	$ 7,640.10	$ 7,490.88	$ 7,615.30	$ 7,395.82
Urban	$ 7,856.29	$ 7,623.63	$ 7,451.84	$ 7,574.79	$ 7,355.78
Rural	$ 7,950.02	$ 7,649.75	$ 7,513.99	$ 7,639.44	$ 7,419.68
Variance					
Aggregate	1.049	1.036	1.042	1.003	0.984
Urban	1.107	1.077	1.107	1.085	1.063
Rural	1.001	0.999	0.987	0.934	0.918

Note: Mean consumption is in dollars and variance is computed using log dollars.

Table 2: Consumption Summary Statistics

2.3 Findings

Regressions on the Aggregate

I run a Weighted Least Squares (WLS) regression on the total annual consumption level of each household with annual income dummies, population density, a rural indicator and other household demographics (see equation below). I find that while the regression outputs a significant value for population density, the effect is very small and its result can be accredited to the vast sample size. Additionally, running population density interacted with the income variables yields very low effect sizes and can also be considered to be negligible. The specification for the regression is the following:

$$
\begin{aligned}
Consumption = \beta_0 + \beta_1 isRural + \beta_2 HouseDense + \beta \cdot \mathbf{income} + \lambda \cdot \mathbf{educ} + \gamma \cdot \mathbf{race} \\
+ \beta_{n-2} hasChildren + \beta_{n-1}(HouseholdSize - 1) \\
+ \beta n hasChildren \times HouseholdSize + u
\end{aligned}
$$

The data shows that income is very significant with a p-value much less than 0.01 as well

Table 3: Regression of Aggregate Consumption on Household Demographics, 2015

	(1)	(2)	(3)	(4)	(5)
Intercept	9852.8506***	6507.4500***	8649.3793***	5797.7182***	1014.7023***
	(575.4280)	(1387.8923)	(898.0635)	(241.9645)	(331.8866)
isRural			1214.0714***	822.2713***	150.7553
			(113.3098)	(294.5608)	(791.9220)
HouseDense					-0.7290***
					(0.1563)
PopDense	-0.8322***				
	(0.0792)				
Income					
\$20,000-\$39,999 (2)		1464.0687***		1145.5521***	-556.3962
		(179.5103)		(312.1114)	(2922.8052)
\$40,000-\$59,999 (3)		3392.5747***		2367.4628***	525.2721*
		(194.8728)		(331.8831)	(310.0699)
\$60,000-\$99,999 (4)		4209.4997***		3768.2471***	1144.8577***
		(182.3118)		(309.2109)	(293.1754)
$\geq \$100,000$ (5)		5172.9252***		4794.2006***	2051.8574***
		(175.7153)		(289.8689)	(281.5460)
(2) $\times isRural$				474.6275	699.7729**
				(381.8077)	(355.3099)
(3) $\times isRural$				1648.3628***	1452.4374***
				(408.3508)	(381.2006)
(4) $\times isRural$				764.7054**	849.4258**
				(382.0315)	(356.8430)
(5) $\times isRural$				900.2637**	811.8212**
				(365.3398)	(341.8168)
Race					
White					947.7439***
					(219.9553)
Black					-1454.4524***
					(260.3441)
Asian					-3792.9750***
					(351.4168)
Education Level					
High School Diploma					1023.3818***
					(141.2381)
Bachelor's Degree					-585.3927***
					(130.7535)
Post-Graduate Degree					-1600.0011***
					(206.6684)
Has Children					5213.0276***
					(415.4277)
Household Size					3146.9834***
					(69.6656)
hasChildren × HouseholdSize					-1468.7683***
					(114.8222)

The dependent variable is consumption of non-durable goods and standard errors are noted in parentheses. The omitted categories are household head's income $< \$20,000$ for income, "other" for race, and "household head's highest education less than a high school diploma" for education level.

as a signficant effect of the rural indicator whose average effect across the years of the panel is around 800 dollars. When controlling for more factors such as race and education level, the results of the rural indicator get much less significant. Analyzing the variance inflation factor, we see that it rises (along with the standard error) as we add more regessors to the model. This can be attributed to the notion that, all factors being equal, simply living in an area considered to be urban or rural will not change one's consumption habits. Rather, it is the unique set of environmental factors that determine outcomes of consumption in these places. For instance, income can have different effects on consumption based on where one resides due to the varying costs of living, but it is improbable that keeping things like cost of living and income constant that one can expect to have different consumption decisions solely on living somewhere the U.S. Census Bureau considers rural.

While we have determined that the choice to live in rural/suburban areas is engendered by a set of environmental variables, we'd like to know to what extent this decision affects how the other regressors are weighted on the consumption outcome. Next, instead of looking at the aggregate, we shall analyze the individual behavior between urban and rural groups. Additionally, note that when income is interacted with the indicator, the values are extremely significant. This further signals that household demographics may have varying degrees of effects given whether the household is considered rural or urban.

Regressions on Rural and Urban Groups

Given the results of Tables 4 and 5, lower incomes seem to have smaller effects on consumption in urban areas than in rural areas. An F-test between the two regressions shows that they are indeed distinct with a p-value much less than 0.01. Given higher income levels, the effects of income on consumption are similar. This supports the popular notion that the decreased cost of living in rural environments causes necessity goods be more accessible. That is, households are able to spend a larger percentage of their income on non-durable goods than on other durables and services such as housing and transportation. The effect of

Table 4: Regression of Consumption on Household Demographics (Rural), 2015

	(1)	(2)	(3)
Intercept	7891.8287***	5651.8345***	4284.8940***
	(53.1023)	(80.5718)	(179.5642)
HouseDense			-0.3133
			(0.2862)
PopDense	0.5613***		
	(0.1309)		
Income			
\$20,000-\$39,999		1307.7710***	896.0881***
		(104.1850)	(101.7565)
\$40,000-\$59,999		2611.1569***	1816.8531***
		(114.3830)	(114.5174)
\$60,000-\$99,999		3219.0615***	2191.4208***
		(107.4286)	(111.4146)
$\geq$ \$100,000		4549.7886***	3414.7032***
		(106.7273)	(116.5995)
Race			
White			889.4629***
			(147.3607)
Black			-1362.0756***
			(182.6222)
Asian			-1330.2418***
			(264.6700)
Education Level			
High School Diploma			354.5350***
			(86.5079)
Bachelor's Degree			-86.3281
			(83.1604)
Post-Graduate Degree			-605.2707***
			(139.0559)
Has Children			2208.6061***
			(268.4455)
Household Size			1550.4491***
			(46.7451)
hasChildren $\times$ HouseholdSize			-1102.3126***
			(74.9890)

The dependent variable is consumption of non-durable goods and standard errors are noted in parentheses. The omitted categories are household head's income $< \$20,000$ for income, "other" for race, and "household head's highest education less than a high school diploma" for education level.

Table 5: Regression of Consumption on Household Demographics (Urban), 2015

	(1)	(2)	(3)
Intercept	8137.0743***	5695.4385***	4326.8803***
	(56.4844)	(134.7924)	(246.8987)
HouseDense			-0.0138
			(0.0091)
PopDense	-0.0151***		
	(0.0048)		
Income			
\$20,000-\$39,999		937.4019***	494.7805***
		(174.2882)	(170.4982)
\$40,000-\$59,999		1986.9483***	1289.7292***
		(184.4724)	(182.1801)
\$60,000-\$99,999		3060.9390***	2124.6360***
		(172.2893)	(174.4153)
$\geq$ \$100,000		4009.6913***	3012.8304***
		(161.2447)	(171.4092)
Race			
White			462.8195**
			(190.4187)
Black			-448.9972**
			(216.4108)
Asian			-1458.5133***
			(276.1166)
Education Level			
High School Diploma			341.3228**
			(142.5898)
Bachelor's Degree			-120.4428
			(120.0487)
Post-Graduate Degree			-404.7905**
			(179.0803)
Has Children			2276.0615***
			(378.8023)
Household Size			1290.7221***
			(60.1015)
hasChildren $\times$ HouseholdSize			-897.7076***
			(102.4092)

The dependent variable is consumption of non-durable goods and standard errors are noted in parentheses. The omitted categories are household head's income $< \$20,000$ for income, "other" for race, and "household head's highest education less than a high school diploma" for education level.

race also seems to drive consumption differences higher in rural areas than in urban areas.

Additionally, the sign of the population density coefficient is different between rural and urban households. While the population density coefficient in the urban regression is slightly negative, the coefficient on the rural regression is about an order of magnitude higher in the positive direction. This suggests that population density increases one's consumption decision up to a threshold and once it reaches this threshold, population density's effect subsides. Further investigation of this is done in Section 4.

Estimating Household Location based on Various Characteristics

Using logit and probit regressions, I test if certain characteristics of a household can help identify where they reside. Then, factors that cause individuals to lean towards one group or another can be determined. The first regression for both the logit and probit model is of the form:

$$isRural = \beta_0 + \beta \cdot \mathbf{income} + \epsilon$$

where **income** is a vector of income dummies.

Next, I additionally control for race, education level, and household dynamics with the following regression:

$$\begin{aligned} isRural = \beta_0 &+ \beta_1 consumption + \beta \cdot \mathbf{income} + \lambda \cdot \mathbf{educ} + \gamma \cdot \mathbf{race} \\ &+ \beta_{n-2} hasChildren + \beta_{n-1}(HouseholdSize - 1) \\ &+ \beta_n hasChildren \times HouseholdSize + \epsilon \end{aligned}$$

Note: A regressor in bold means that it is categorical and therefore uses a vector of dummies.

From the output of both models, it seems the probability of being a rural household is

Table 6: Rural and Urban Logistic Regression, 2015

	Logit (1)	Logit (2)	Probit (1)	Probit (2)
Intercept	0.8951***	0.2919***	0.5532***	0.1832***
	(0.0394)	(0.0754)	(0.0237)	(0.0461)
Consumption		0.2710***		0.1653***
		(0.0139)		(0.0084)
Income				
\$20,000 - \$39,999	0.0112	-0.0892*	0.0067	-0.0526*
	(0.0465)	(0.0477)	(0.0280)	(0.0285)
\$40,000 - \$59,999	-0.1810***	-0.3574***	-0.1097***	-0.2123***
	(0.0466)	(0.0489)	(0.0282)	(0.0293)
\$60,000 - \$99,999	-0.3189***	-0.5271***	-0.1942***	-0.3161***
	(0.0447)	(0.0482)	(0.0271)	(0.0289)
$\geq$ \$100,000	-0.7033***	-0.9016***	-0.4331***	-0.5459***
	(0.0474)	(0.0526)	(0.0288)	(0.0317)
Race				
White		0.4726***		0.2901***
		(0.0612)		(0.0378)
Black		-0.5810***		-0.3628***
		(0.0703)		(0.0435)
Asian		-0.5039***		-0.3168***
		(0.0891)		(0.0551)
Education Level				
High School Diploma		0.5477***		0.3275***
		(0.0384)		(0.0229)
Bachelor's Degree		0.2647***		0.1598***
		(0.0319)		(0.0194)
Post-Graduate Degree		0.1540***		0.0916***
		(0.0452)		(0.0277)
Has Children		0.0483		0.0379
		(0.1258)		(0.0758)
Household Size		0.0500***		0.0304***
		(0.0185)		(0.0111)
HasChildren $\times$ HouseholdSize		-0.0072		-0.0059
		(0.0336)		(0.0202)

The dependent variable is consumption of non-durable goods and standard errors are noted in parentheses. The omitted categories are household head's income $< \$20,000$ for income, "other" for race, and "household head's highest education less than a high school diploma" for education level.

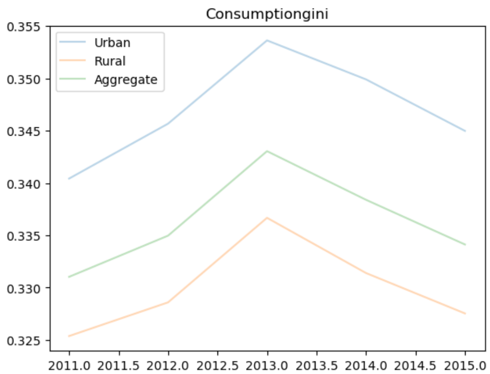
inversely related to their income level. This makes sense since a lot of the literature notes that urban households are compensated more due to the cost of living. However, this could also be due to the nature of high density population centers. Firms are often headquartered in urban areas and, thus, there are often more opportunities as well as more competition for skilled workers than in rural areas.

It also makes sense that having a larger household size results in a higher probability of living in a rural area. In rural areas, landowners often hand down the working of the property to their children and continue to reside on the property (for example, a farm). In addition, people leaving in rural and suburban areas may feel they are free to having more kids. This could be due to the popular rationale that the countryside is safer than the city and that the tradeoff between work and raising a child is much lower than in a city.

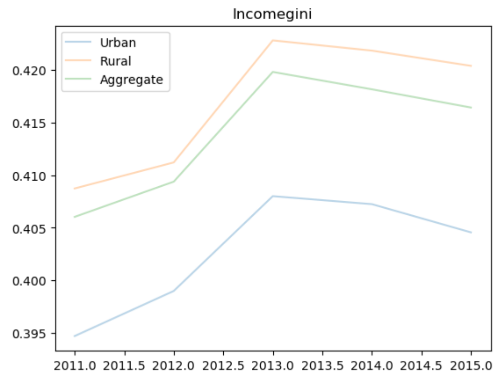
It should be noted, however, that having at least one child is not indicative of whether a household resides in a rural or urban area. One explanation could be that no matter where one resides, there are both social and cultural incentives and pressures to have children. Also, note that income levels within the \$ 20,000 - \$ 40,000 range do not indicate movement one way or the other. This supports the notion that there exists a certain set of occupations that necessary in any local economy no matter where one resides (i.e. teachers, carpenters, bakers, etc.). The above observation also motivates the earlier argument that urban populations tend to have higher compensations due to the higher percentage of service-based, white collar work and that firms are often headquartered in urban areas.

Inequality Metrics

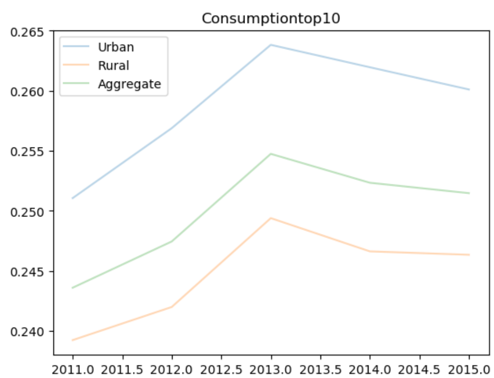
Now that we've considered how household characteristics change and affect a household's choice of consumption in rural and urban households, we can ask if these results are economically significant. Can we expect these differences in choices the household makes based on their environment to affect their well-being? The short answer is yes, we can. Figure 2 plots various metrics of inequality on rural and urban households.



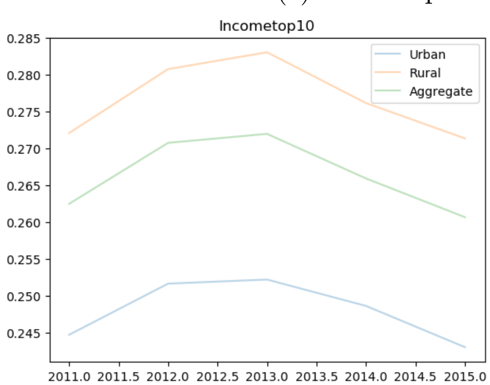
(a) GINI Coefficient (Consumption)



(b) GINI Coefficient (Income)



(c) Consumption Share of Top 10%



(d) Income Share of Top 10%

Figure 2: Various Metrics of Inequality across Time[a]

[a]Note that the values for the metrics of inequality in income is should be systematically biased down since the Nielsen data caps income reporting at any value greater than \$ 100,000. Although the value itself may be skewed, the ordering of which group experiences most inequality should remain.

Observe that, systematically, the income metrics of inequality are higher in rural areas while consumption metrics of inequality are lower in urban areas. However, as discussed in the literature, the magnitude of the consumption inequality is much less than that of income inequality. Furthermore, the difference between the level of inequality both groups are smaller for consumption than they are for income.

3 Theory

The baseline model that I will be using is an adaptation of the consumption-savings model from Kaplan (2012) with assignment to either an urban or rural household. As aforementioned, this model also includes two separate income processes which will be calibrated using the variances of consumption and income growth for each group.

3.1 The Model

Let us consider a continuum of heterogeneous households indexed i. The model period is one year. Households starting at time 0 and work until time T where they die with certainty and only in period T do they have a non-zero probability of dying. Each household has expected utility preferences over their annual consumption, $c_{i,t}$, noted by:

$$\mathbb{E} \sum_{t=0}^{T} \beta^t u(c_{i,t})$$

In any given period, a household can choose to consume his full income, $Y_{i,t}$, or invest in the economy's single asset denoted a. When choosing to invest, the agent is guaranteed a return of $(1 + r)$ in the next period. Additionally, households may not borrow. Log income follows an exogenous stochastic process that consists of four components:

$$\ln Y_{i,t} = y_{i,t} = \kappa_t + z_{i,t} + \epsilon_{i,t}$$
$$z_{i,t} = \rho z_{i,t-1} + \eta_{i,t}$$
$$\epsilon_{i,t} = \upsilon_{i,t}$$

κ_t is a non-stochastic experience effect which is assumed to be the same within all individuals of each group. $z_{i,t}$ is an $AR(1)$ process with a persistence factor ρ that experiences a shock $\eta_{i,t}$ in each period and $\epsilon_{i,t}$ is an idiosyncratic IID shock. Both shocks are independent of each other. The persistent shock η and idiosyncratic IID shock ϵ are assumed to be drawn from two separate discretized normal distributions with mean zero. Each household is born with an initial wealth endowment $a_{i,0} = 0$. Using this specification, we have identified the household budget constraint. Explicitly, it is:

$$\max_{c_t, a_{t+1}} \quad \mathbb{E} \sum_{t=0}^{T} \beta^t u(c_t) \quad \textit{subject to}$$
$$c_t + a_{t+1} \leq y_t + (1+r)a_t$$
$$a_{t+1}, c_t \geq 0$$
$$a_0 \geq 0$$
$$a_0 \quad \textit{given}$$

One can see that it is possible to identify the household's consumption and savings decision in each period given both the current period's income as well as the current period's assets. Thus, denoting the value function of each household as $V(\cdot)$, we have the following recursive

Bellman formulation:

$$V(a,y) = \max_{c,a`} u(c) + \beta\,\mathbb{E}[V(a`,y`)|y] \quad \textit{subject to}$$

$$c + a` = y + (1+r)a$$

$$a` \geq 0$$

Households are born with probability p of being in an urban household and probability $(1-p)$ of birth into a rural household. Once the decision of housing is made, the agent remains in that group for their entire lifetime. The circumstances each household faces in each group is determined by the income process noted above, albeit with difference variances and persistence factors for the shocks. I assume that urban households have a greater degree of income risk and thus, their persistence factors ρ are lower and the variance of their shocks is higher compared to rural households. Mathematically, we have

$$\rho_u \leq \rho_r$$

$$\sigma^2_{\eta,u} \geq \sigma^2_{\eta,r}$$

$$\sigma^2_{\epsilon,u} \geq \sigma^2_{\epsilon,r}$$

This assumption is supported by the data as shown below.

3.2 Calibrating the Model

First, in order to control for the effects of aggregate shocks, I regress log income on a full set of age and time effects. Formally, I denote the log earnings for household i of age a at time t as $y^l_{i,a,t}$ and estimate the following regression:

$$y^l_{i,a,t} = \phi_t + h_a + y_{i,a,t}$$

Thus, we can take the residual $y_{i,a,t}$ as our measure of log income that is controlled for time and age effects. This, along with sample moments (namely, the covariance and variance), are used to estimate ρ, σ_η^2, and σ_ϵ^2 for aggregate, rural, and urban households (see Appendix B for the proof of the identification of the model parameters). The result is shown in Table 7.

	Parameter		
	ρ	σ_η^2	σ_ϵ^2
Aggregate	0.786	0.182	0.044
Urban	0.689	0.247	0.063
Rural	0.982	0.120	0.027

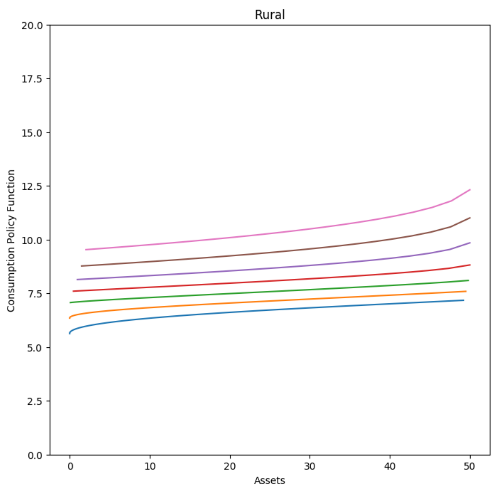
Table 7: Model Parameter Estimates

Next, I choose parameters based on existing literature and economic conditions. I set the number of working periods for each household to be 35 to represent the typical stop work age in the United States [8]. Then, as aforementioned, I set the interest rate to 0.015 to reflect the empirical real interest rate and choose the discount factor β accordingly. As advised by the literature [9], I use the isoelastic relative risk aversion utility function:

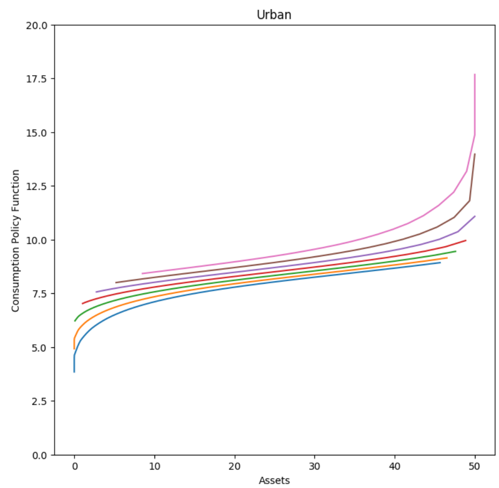
$$u(c) = \frac{c^{1-\gamma} - 1}{1 - \gamma}$$

and set γ equal to 2.

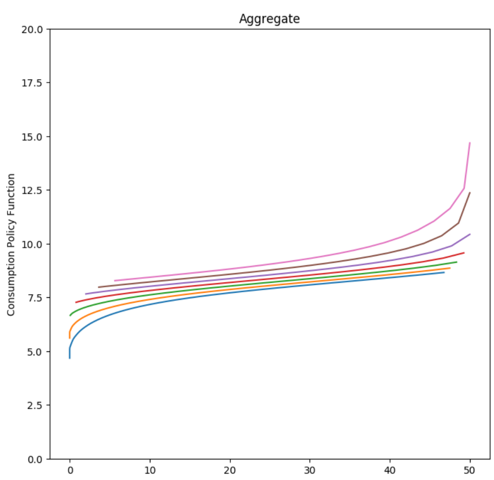
I solve each household's optimization problems based on approximation of the decision rules. In particular, I use a 60-point unequally-spaced, non-linear wealth grid along with a stochastic earnings grid discretized from the shock variances (σ_η^2 and σ_ϵ^2) and persistence factor ρ as well as 7 distinct income states. The discretization is done using the Rouwenhorst method of discretizing Markov processes [7].



(a)



(b)



(c)

Figure 3: Consumption Policy of Households based on income (lines) and Savings (x-axis)

3.3 Results of the Model

Figure 3 plots the household consumption policy given their income and level of wealth for the period. That is, this figure models consumption as a function of income and savings. We can observe that while the spread of consumption policy is greater between income levels in rural households, the overall difference in consumption policy does not vary much with wealth. On the other hand, urban households have a smaller spread in consumption policy with respect to income. However, a marginal increase in assets at lower income levels vastly increases consumption. The same is true for marginal increases in assets at the highest income levels. This can be accredited to lower income households in urban areas being less likely to be able to save in the first place. Thus, when they're able to save, they end up consuming it in the next period. This phenomenon does not occur in rural households because their high persistence factor and low variance of shocks engender a more steady stream of income so they do not necessarily need assets to smooth their consumption.

This notion is reinforced in Figures 4 and 5. The figure shows the average consumption and savings levels in each period of the working life where simulated households experience shocks based on their status as urban or rural in each period. The steady consumption policy based on savings and income engenders a higher mean level of consumption in the rural households. Additionally, the notion above that households in the rural group are not as reliant on their assets can be seen in a lower level of mean assets compared to the urban group. From the figure, we can see that rural households save much less than urban households on average. Notice that households possess no assets at the end of their working period. This is because households know they will die with certainty at the end of their working period and have no motives (i.e. bequesting) to save so they consume their full income (i.e. $c_T = y_T + (1+r)a_T$ with $a_{T+1} = 0$). Overall, the trend of both mean assets and consumption seems to be similar for both urban and rural households. For the most part, the disparity between the mean level of consumption in urban and rural households tends to occur in the middle of the working life, with the difference between the mean levels growing closer to zero as the working life ends.

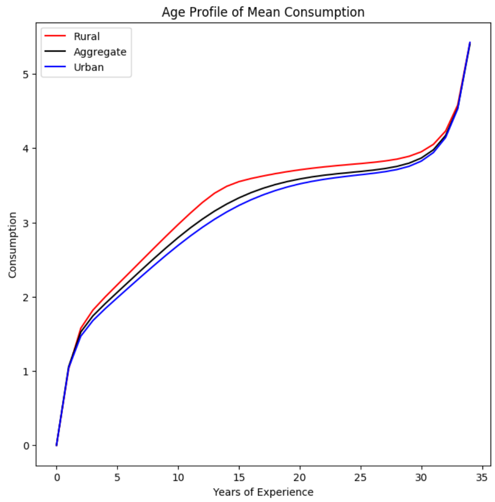


Figure 4: Consumption across Ages

Figure 5: Savings across Ages

Autocorrelation Factor

While the ρ parameter for rural areas is consistent with that of recent literature (0.98) [9], the ρ parameter for urban areas is about 30% smaller (0.689). Recall that ρ is the persistence factor of the permanent shock (η). In words, it is to what degree a household lives with a persistent shock and how quickly that shock decays. In the case of the urban household, both positive and negative shocks decay more quickly and due to the higher variance of the shock (σ_η^2), the household faces more income uncertainty, or risk. Because it is much harder to predict what income level one will be given in the next period, consumption paths tend to similar across different income levels. However, once we add the ability to save, households are guaranteed a return. Thus, saving allows the household to smooth its consumption. This can be see in Figure 3 where high savings levels leads to an increase in consumption since households will still be able to smooth any shock they receive with their savings.

On the other hand, coupled with the low variance of shocks, rural households tend to have more certainty in their income due to the high persistence factor (0.982). Once given their income draw, households have the expectation that their income is not likely to change in the next period. Thus, their consumption patterns vary less with the amount they save.

4 An Alternative Specification for Urbanness

In this section, I propose and analyze other definitions of urbanness in an attempt to obtain different images of rural and urban households.

Urban Clusters

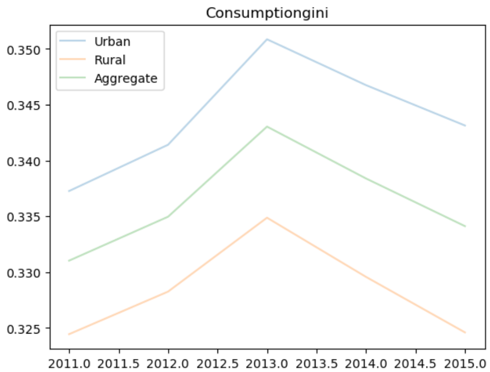
Including urban clusters into my definition of the rural household[2], I recalculate the GINI coefficient and top 10% shares of both income and consumption.

The trend of inequality seems to be the same for both including urban clusters in the urban group and in the rural group. One observation that can be made is that while the overall level of inequality for consumption and income in urban households has remained relatively the same, rural households seem to have a slightly elevated level of inequality. This difference is not large enough to draw any conclusions but it may merit future investigation.

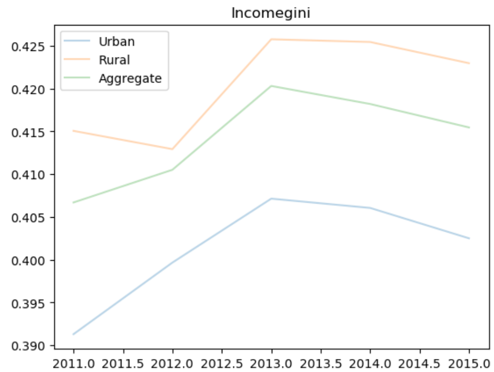
Appendix C contains the regressions from Section 2 ran on including urban clusters in the urban household group. There are no largely noticeable differences between the regressions including urban clusters in the rural group and including them in the urban group except that the R^2 values increase by 0.03 when changing the value of the rural indicator and housing density becoming statistically significant in the between group regressions. This may suggest that housing density could indicate how developed an area is even though it is more sparsely populated.

However, using housing density as a measure of "ruralness" produces similar results as using the urban cluster definition as housing density and population density are very correlated (correlation coefficient of 0.98).

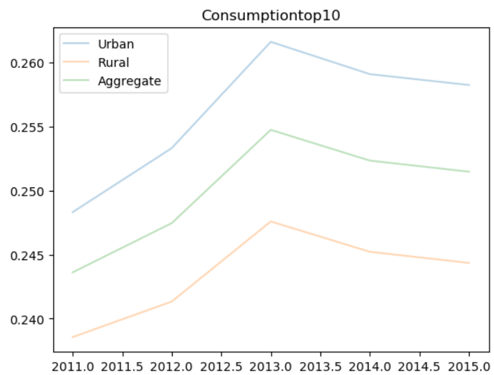
[2]Recall that urban clusters are any county who's average population density is between 500 persons per square mile and 1,000 persons per square mile



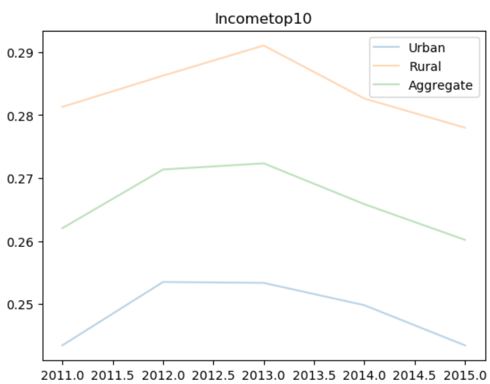
(a) GINI Coefficient (Consumption)



(b) GINI Coefficient (Income)



(c) Consumption Share of Top 10%



(d) Income Share of Top 10%

Figure 6: Various Metrics of Inequality across Time (Urban Cluster) [a]

[a]Note that the values for the metrics of inequality in income is should be systematically biased down since the Nielsen data caps income reporting at any value greater than \$ 100,000. Although the value itself may be skewed, the ordering of which group experiences most inequality should remain.

5 Conclusion

5.1 Directions for Further Work

The Kilts-Nielsen Consumer Panel has very granular, detailed and accurate data on household demographics as well as consumption of non-durable goods, primarily food and household supplies. However, the shortcomings of the data set lies in its lack of data on durable consumption such as the price of a household's home or their funds spent on transportation. Also, the use of income as a categorical variable limits the ability estimate the level of inequality since high earners are grouped into a single category (the $> \$100,000$ categorical income variable). Future work should consider merging a data set rich in income decomposition such as the Panel Study of Income Dynamics (PSID) with the Consumer Panel. Although it may be difficult to reconcile consistency among the two data sets, the result would allow for a more accurate measurement of the true difference in dynamics or urban and rural areas.

Furthermore, while it seems there are distinct groups of household's based around population density, it is difficult to clearly draw a line between rural and urban areas. Instead, further work should decompose the data into three separate groups: urban, rural and the urban cluster. Even within urban clusters, it is difficult to distinguish which are suburbs of large cities and which are population centers of rural areas.

5.2 Concluding Remarks

To conclude, let us revisit the question initially proposed: Does consumption inequality differ in rural and urban households? Can the existence of such a difference be explained by differences in income risk? In short, the answer is yes. Households do make different decisions for consumption based on where they live. At average income levels, both types of household consume similar quantities. However, at the bottom of the distribution lies the true difference. The data suggests there is a fixed cost to living in a urban household.

This limits the consumption of poor urban households such that they consume a smaller proportion of their income than a similar individual in a rural household (coefficient of 400 vs 800 of the rural household). Furthermore, the data estimates that the persistence factor of income is 30% lower in urban households than previously suggested in the literature [3]. This translates to a higher degree of income risk as urban households are less sure of their future shocks. The model then suggests that in such urban households that the consumption path is more similar across income levels than in rural households. However, the question of what factors affect the level of risk is left to further research.

5.3 Acknowledgements

I am grateful to my advisor Greg Kaplan for his constant support and wealth of knowledge in the field of economic inequality as well as the BA workshop organizers Victor Lima and Kotaro Yoshida. For non-economics related resources, I'd like to thank the University of Chicago Computer Science tech staff for access to their computer clusters as well as the University of Chicago Research Computing Center for their help in accessing the Kilts-Nielsen data. I also thank Soyoung Eom and Patrick Dougherty for providing feedback on my statistical methodology.

[3] The literature [9] finds a value of > 0.98 while I find 0.689 for urban households

References

[1] Mark Aguiar and Mark Bils, 2015. "Has Consumption Inequality Mirrored Income Inequality?" *American Economic Review,* 105 (9), 2725-2756.

[2] Richard Blundell and Ian Preston, 1998. "Consumption Inequality and Income Uncertainty," *The Quarterly Journal of Economics.* May 1998.

[3] "Effective Federal Funds Rate — FRED — St. Louis Fed." *fred.stlouisfed.org.* Federal Reserve Bank of St. Louis, 5 Mar. 2017. Web. 10 Mar. 2017.

[4] Greg Kaplan, 2012. "Inequality and the Life Cycle," Quantitative Ecoomics 3 (2012), 471-525.

[5] Greg Kaplan, Kurt Mitman, and Giovanni L. Violante, 2016. "Non-durable Consumption and Housing Net Worth in the Great Recession: Evidence from Easily Accessible Data."

[6] Greg Kaplan and Sam Schulhofer-Wohl, 2016. "Inflation at the Household Level."

[7] Karen Kopecky, Richard Suen, 2010. "Finite state Markov-chain approximations to highly persistent processes" *Review of Economy Dynamics,* 2010.

[8] "The Difference Between Retirement Age and Stop Work Age" *ssa.gov.* Social Security Adminstration. Web. Accessed 12 Apr. 2017.

[9] K. Storesletten, C. Telmer, and A. Yaron, 2004. "Consumption and risk sharing over the life cycle." *Journal of Monetary Economics,* 51 (3), 609-633.

[10] "Urban and Rural - Geography." *census.gov.* United States Census Bureau, 12 Dec. 2016. Web. Accessed 10 Mar. 2017.

Appendices

A Sample Sizes and Distributions

Table 8: Rural Proportions and Sample Sizes

	2011	2012	2013	2014	2015
Proportion Rural	0.628	0.630	0.628	0.627	0.627
	(0.657)	(0.657)	(0.655)	(0.660)	(0.658)
Sample Size	119025.956	119189.506	119549.677	120294.067	121335.234
	(62.092)	(60.538)	(61.097)	(61.557)	(61.380)

Note: Unweighted sample sizes and the unweighted percentage of rural households are denoted in parentheses. Sample sizes are in thousands of persons and denotes the number of households in the aggregate economy.

Table 9: Income Distribution per Year

income	< \$20,000	\$20,000 - \$39,999	\$40,000 - \$59,999	\$60,000 - \$99,999	≥ \$100,000
2011	0.147742	0.229682	0.184181	0.230593	0.207802
	(0.105150)	(0.247890)	(0.220528)	(0.280551)	(0.145880)
2012	0.157775	0.243678	0.189155	0.225007	0.184385
	(0.105570)	(0.240807)	(0.223579)	(0.280733)	(0.149311)
2013	0.170330	0.245028	0.176818	0.213871	0.193952
	(0.104932)	(0.242745)	(0.220846)	(0.278131)	(0.153346)
2014	0.164326	0.236212	0.170096	0.216967	0.212399
	(0.103709)	(0.238722)	(0.218367)	(0.281089)	(0.158114)
2015	0.157176	0.230188	0.167635	0.217201	0.227799
	(0.099185)	(0.235516)	(0.214011)	(0.285060)	(0.166227)

Note: Unweighted proportion distributions for each year are denoted in parentheses.

Table 10: Income Distribution per Year (Rural and Urban)

Year	Income	<$20,000	$20,000 - $39,999	$40,000 - $59,999	$60,000 - $99,999	$\geq$ $100,000
2011	Rural	0.160292	0.247079	0.191348	0.225282	0.175999
		(0.114986)	(0.268717)	(0.225557)	(0.270090)	(0.120650)
	Urban	0.126573	0.200336	0.172091	0.239552	0.261448
		(0.086332)	(0.208042)	(0.210904)	0(.300568)	(0.194154)
2012	Rural	0.170130	0.263671	0.193871	0.219409	0.152920
		(0.114513)	(0.260924)	(0.229277)	(0.272119)	(0.123167)
	Urban	0.136707	0.209585	0.181112	0.234553	0.238044
		(0.088469)	(0.202338)	(0.212681)	(0.297205)	(0.199307)
2013	Rural	0.185210	0.261482	0.181289	0.210448	0.161570
		(0.115193)	(0.262698)	(0.225266)	(0.270214)	(0.126629)
	Urban	0.145194	0.217235	0.169265	0.219654	0.248653
		(0.085412)	(0.204788)	(0.212436)	(0.293193)	(0.204171)
2014	Rural	0.176769	0.250209	0.173512	0.215971	0.183539
		(0.112852)	(0.257418)	(0.224081)	(0.274074)	(0.131575)
	Urban	0.143440	0.212720	0.164364	0.218640	0.260837
		(0.085936)	(0.202382)	(0.207259)	(0.294725)	(0.209698)
2015	Rural	0.169482	0.244602	0.171424	0.218093	0.196399
		(0.108394)	(0.254090)	(0.220377)	(0.278670)	(0.138469)
	Urban	0.136521	0.205994	0.161275	0.215704	0.280505
		(0.081455)	(0.199752)	(0.201754)	(0.297364)	(0.219675)

Note: Unweighted proportion distributions for each year are denoted in parentheses.

B Identification of the Model Parameters

Recall that the natural logarithm of income follows a first-order Markov process with four components:

$$\ln Y_{i,t} = y_{i,t} = \kappa_t + z_{i,t} + \epsilon_{i,t}$$
$$z_{i,t} = \rho z_{i,t-1} + \eta_{i,t}$$
$$\epsilon_{i,t} = \upsilon_{i,t}$$

and that the shocks in each period follow a normal distribution with mean zero.

$$\eta_{i,t} \sim N(0, \sigma_\eta^2)$$
$$\epsilon_{i,t} \sim N(0, \sigma_\epsilon^2)$$

Using the sample moments (in this case, the variances and covariances) and the assumption that $z_{i,0} \; \forall \; i$, we can identify the three unknown parameters of the model $(\rho, \sigma_\eta^2, \sigma_\epsilon^2)$.

Let us define the covariance between income at time t and time $t + d$ as $m_{t,d}^{lev}$ such that we have the following:

$$m_{t,d}^{lev} = \mathrm{Cov}(y_{i,t}, y_{i,t+d})$$

To identify the parameters of the model, we must first make a few observations and define our set of terms:

$$
\begin{aligned}
\mathrm{Var}(\kappa_t) &= 0 \qquad \forall t \\
m_{t,0}^{lev} &= \mathrm{Cov}(y_{i,t}, y_{i,t}) \\
&= \mathrm{Var}(y_{i,t}) \\
&= \mathrm{Var}(\kappa_t + z_{i,t} + \epsilon_{i,t}) \\
&= 0 + \mathrm{Var}(z_{i,t}) + \sigma_\epsilon^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(z_{i,t}) &= \mathrm{Var}(\rho z_{i,t-1} + \eta_{i,t}) \\
&= \rho^2 \mathrm{Var}(z_{i,t-1}) + \sigma_\eta^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}(z_{i,t}, z_{i,t+1}) &= \mathrm{Cov}(z_{i,t}, \rho z_{i,t} + \eta_{i,t+1}) \\
&= \rho \mathrm{Var}(z_{i,t})
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}(z_{i,t}, z_{i,t+d}) &= \mathrm{Cov}(z_{i,t}, \rho z_{i,t+d-1} + \eta_{i,t+d}) \\
&= \mathrm{Cov}(z_{i,t}, \rho(\rho z_{i,t+d-2} + \eta i, t+d-1) + \eta_{i,t+d}) \\
&= \mathrm{Cov}(z_{i,t}, \rho^2 z_{i,t+d-2} + \rho \eta i, t+d-1 + \eta_{i,t+d}) \\
&= \mathrm{Cov}(z_{i,t}, \rho^d z_{i,t} + \sum_{i=0}^{d} \rho^i \eta_{a+d-i}) \\
&= \rho^d \mathrm{Var}(z_{i,t})
\end{aligned}
$$

$$
\begin{aligned}
m_{t,1}^{lev} &= \mathrm{Cov}(y_{i,t}, y_{i,t+1}) \\
&= \mathrm{Cov}(\kappa_t + z_{i,t} + \epsilon_{i,t}, \kappa_{t+1} + z_{i,t+1} + \epsilon_{i,t+1}) \\
&= \mathrm{Cov}(z_{i,t}, z_{i,t+1}) \quad \text{since } z, u \text{ independent} \\
&= \rho \mathrm{Var}(z_{i,t}) \\
\implies m_{t,1}^{lev} &= \rho^d \mathrm{Var}(z_{i,t})
\end{aligned}
$$

Using these observations, we can first identify ρ using the slope.

$$\begin{aligned}
\frac{m_{t,3} - m_{t,2}}{m_{t,2} - m_{t,1}} &= \frac{\rho^3 \text{Var}(z_{i,t}) - \rho^2 \text{Var}(z_{i,t})}{\rho^2 \text{Var}(z_{i,t}) - \rho^1 \text{Var}(z_{i,t})} \\
&= \frac{\rho^2(\rho - 1)}{\rho(\rho - 1)} \\
&= \rho
\end{aligned}$$

Now, using ρ and the difference in levels, we can identify σ_η^2.

$$\begin{aligned}
m_{t,2} - m_{t,1} &= \rho^2 \text{Var}(z_{i,t}) - \rho \text{Var}(z_{i,t}) \\
&= \rho(\rho - 1) \text{Var}(z_{i,t})
\end{aligned}$$

$$\begin{aligned}
\text{Var}(z_{i,t}) &= \rho^2 \text{Var}(z_{i,t-1}) + \sigma_\eta^2 \\
&= \rho^2(\rho^2 \text{Var}(z_{i,t-2}) + \sigma_\eta^2) + \sigma_\eta^2 \\
&= \dots \\
&= \sigma_\eta^2 \sum_{i=0}^{t-1} \rho^{2i}
\end{aligned}$$

$$\implies \quad \sigma_\eta^2 = \frac{m_{t,2} - m_{t,1}}{\rho(\rho - 1) \sum_{i=0}^{t-1} \rho^{2i}}$$

Knowing $\text{Var}(z_{i,t})$, we identify the final parameter, σ_ϵ^2.

$$m_{t,0}^{lev} = \text{Var}(z_{i,t}) + \sigma_\epsilon^2$$

It is clear to see that the model is overidentified if we have more than three income observations for each household.

C Regressions Using Urban Cluster

Table 11: Regression of Consumption on Household Demographics (Rural), 2015

	(1)	(2)	(3)
Intercept	7941.2079***	5631.3087***	4618.8145***
	(64.4618)	(88.0067)	(201.8368)
HouseDense			-2.2295***
			(0.5924)
PopDense	0.2574		
	(0.2686)		
Income			
\$20,000-\$39,999		1385.8657***	939.6299***
		(114.1352)	(111.7815)
\$40,000-\$59,999		2618.1416***	1791.8955***
		(126.0230)	(126.8758)
\$60,000-\$99,999		3284.3993***	2203.7107***
		(119.7951)	(125.1479)
$\geq$ \$100,000		4570.9587***	3420.7425***
		(121.3272)	(132.6007)
Race			
White			1139.8877***
			(163.9746)
Black			-1611.7795***
			(209.4192)
Asian			-1843.5560***
			(319.7507)
Education Level			
High School Diploma			330.7548***
			(96.2769)
Bachelor's Degree			-39.4196
			(94.7849)
Post-Graduate Degree			-495.2580***
			(162.7667)
Has Children			2012.4534***
			(304.2162)
Household Size			1563.1816***
			(53.6758)
hasChildren × HouseholdSize			-1066.5543***
			(85.5924)

The dependent variable is consumption of non-durable goods and standard errors are noted in parentheses. The omitted categories are household head's income $< \$20,000$ for income, "other" for race, and "household head's highest education less than a high school diploma" for education level.

Table 12: Regression of Consumption on Household Demographics (Urban), 2015

	(1)	(2)	(3)
Intercept	8198.8414***	5709.7994***	4121.3343***
	(45.5664)	(111.6556)	(208.6436)
HouseDense			-0.0156*
			(0.0086)
PopDense	-0.0171***		
	(0.0045)		
Income			
\$20,000-\$39,999		943.0732***	555.8053***
		(143.8527)	(140.3598)
\$40,000-\$59,999		2138.8521***	1452.7189***
		(152.8306)	(150.7560)
\$60,000-\$99,999		3039.8192***	2131.6058***
		(141.6058)	(143.1874)
$\geq$ \$100,000		4126.6401***	3120.9590***
		(133.4909)	(141.8274)
Race			
White			316.0458*
			(163.8446)
Black			-432.4532**
			(187.0109)
Asian			-1146.8005***
			(239.1990)
Education Level			
High School Diploma			384.3096***
			(116.1001)
Bachelor's Degree			-151.4986
			(99.0167)
Post-Graduate Degree			-526.4838***
			(149.2896)
Has Children			2420.0239***
			(315.3820)
Household Size			1340.4189***
			(50.7037)
hasChildren $\times$ HouseholdSize			-974.1369***
			(85.4909)

The dependent variable is consumption of non-durable goods and standard errors are noted in parentheses. The omitted categories are household head's income $< \$20,000$ for income, "other" for race, and "household head's highest education less than a high school diploma" for education level.