Analysis of Total Household Income vs Property Value

203:02 Annual Social and Economic Supplement (ASEC)2023

Daniel S Brown Jason Scott Martin Zapata

Contents

1. Introduction

Property values influence personal wealth, social mobility, and quality of life. Analyzing the relationship between household income and property value is of interest to government agencies, policymakers, economists, and the US population.

Describing the relationship between Total Household Income and Property Value is more compelling today than ever, as the newer generations face difficulty accessing affordable housing. This analysis describes the connection between income and property value, offering insights into broader economic trends. These insights can help decision-makers to be better informed when making decisions on taxation, urban planning, and social programs.

The purpose of this analysis is to describe the bivariate relationship between Total Household Income and Property Value using data from the 2023 Annual Social and Economic Supplement (ASEC) by the U.S Census Bureau.

Research Question: *What is the relationship between total household income and property value in the U.S.?*

2. Description of the Data Source

The data used for our analysis comes from the Current Population Survey (CPS) datasets, specifically the Annual Social and Economic Supplements (ASEC). CPS is a government survey conducted by the US Census Bureau and the Bureau of Labor Statistics. It is the primary source of the labor force statistics for the US and is widely used for various economic, social, and demographic analyses. Beyond standard demographic and labor force data, the ASEC Supplement contains additional data such as work experience, income, noncash benefits, and health insurance coverage.

CPS uses a probability sample of 95,000 occupied households, selected to represent the civilian non-institutional population of the US. The sample is located in approximately 826 sample areas comprising 1,328 counties and independent cities with coverage in every US state and in the District of Columbia. The data is collected via in-person and telephone interviews. The unit of observation in our selected dataset is one household, while other datasets within CPS ASEC are at the family and individual person levels. Key features of the data that were relevant to our model include measures of total household income, property value, and whether the household received any social security income.

3. Data Wrangling & Operationalization

To adequately address the research question, we first identified our two variables of interest, Total Household Income (HTOTVAL) and Property Value (HPROP_VAL). Total Household Income represents the aggregate economic capability of a household, which is crucial for understanding wealth distribution and economic status. Property Value, as an indicator of wealth accumulation, provides insights into financial stability and investment in assets. The operational definitions are clear: HTOTVAL is the total income earned by all household members, measured in dollars, while HPROP_VAL is the estimated market value of the household's property, also measured in dollars.

Alternative ways to operationalize the concepts were considered. For instance, per capita income (total income divided by the number of household members) was an alternative but was not selected because the focus was on the overall economic capability of the household, not individual income. Similarly, using median property value in the area as a proxy for HPROP_VAL was considered but not chosen because it would not reflect individual household property values accurately.

Next, we examined our variables of interest for summary-level statistics as well as the presence of null values. We saw that Total Household Income ranged from -9,999 to 3,300,477 and Property Value ranged from 0 to 9,999,999, with 0 nulls between the two variables. We then filtered out any records where Total Household Income was not positive, which represents either no household income specified or a negative amount. Next, we also filtered out records where Property Value was zero, which represents renters who do not own their property. Our original dataset was 88,978 rows and these filtering steps brought the final processed dataset to 38,019 rows.

We also created an indicator column for whether the household received social security income, based on the variable HSS_YN = 1. As a final step, we split the overall data frame into an exploration and confirmation set, at 30% and 70% of the total rows respectively. We used this exploration set to create scatter plots and assess variable transformations, and decide which statistical tests to run. After making these decisions, we swapped to the confirmation set to calculate our model results and significance.

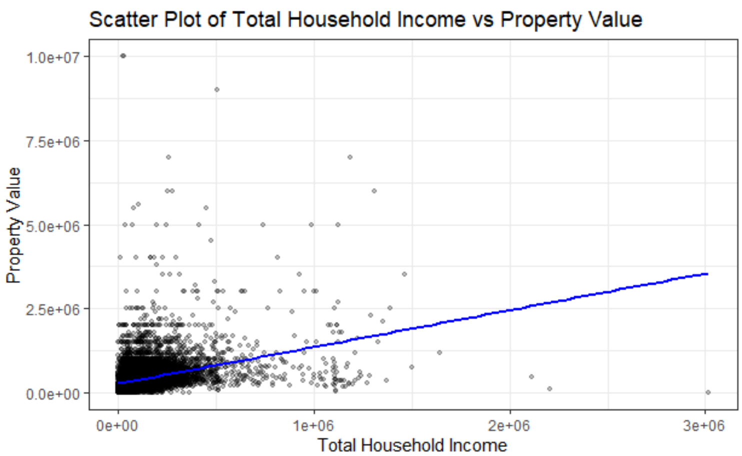


Figure 1. Total House Income (HVTOTAL) vs Property Value(HPROP_VAL) Linear Regression Model

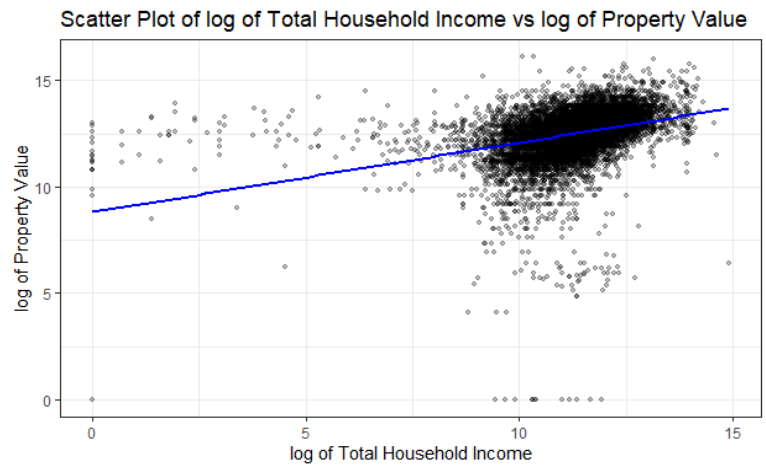


Figure 2. Logarithmic Total House Income (log(HVTOTAL)) vs Logarithmic Property Value(log(HPROP_VAL)) Linear Regression Model

4. Model Specification

In *Figure 1*, a linear regression was plotted, illustrating a positive slope between Household Property Value and Total Household Income. However, the data points are densely clustered at lower values of income and property value, suggesting non-linearity in the relationship. Additionally, the presence of extreme outliers may disproportionately affect the regression line.

To address these issues, we applied a logarithmic transformation to both variables, which helps to stabilize variance and mitigate the impact of outliers. This transformation is shown in *Figure 2*, where the logarithmic scale linearizes the relationship and reduces heteroscedasticity. The logarithmic transformation provided more interpretable insights and resulted in better model fits, as evidenced by improved residual behavior and statistically significant findings.

To address potential discontinuities in our linear regression model, we included an indicator variable for Households Receiving Social Security (HSS_YN). This variable helps us test whether households receiving social security payments experience a different relationship between income and property value compared to those that do not. It captures the impact of additional income sources on property value, which is particularly relevant for understanding how social support mechanisms influence economic outcomes.

5. Model Assumptions

Two assumptions are required for our large-sample linear model to be valid: data must come from an i.i.d. sample, and a unique BLP must exist. First, data must come from an i.i.d sample. Given the nature of the sample, there is likely a violation of i.i.d. due to clustering. For example, households within the same state or region may exhibit similar characteristics, due to geographic clustering. Households may also be affected by local economic conditions and the housing market within their region, which can produce similarities within clusters.

Second, a unique BLP must exist. This is the case when all the covariances between the different X's and between the X's and Y's are finite, and there are no heavy tails in the variable distributions. Based on the histograms of each variable, this assumption is met. For the BLP to be unique, there must be no perfect collinearity between variables.

Since HTOTVAL is the only independent variable used, there is no risk of perfect collinearity, and this assumption is met.

6. Model Results and Interpretation

Table 1: Regression Results: Total Household Income and Property Value

	Dependent variable:		
	Property Value	Log Property Value	Property Value
	(1)	(2)	(3)
Total Household Income	1.125***		1.144***
	(0.052)		(0.054)
Log Total Household Income		0.354***	
		(0.012)	
Receives Social Security			23,221.560***
			(5,740.033)
Constant	266,094.200***	8.508***	255,260.400***
	(5,909.612)	(0.134)	(7,354.865)
Observations	26,613	26,613	26,613
R^2	0.128	0.126	0.129
Adjusted R^2	0.128	0.126	0.129
Residual Std. Error	395,298.000 (df = 26611)	1.040 (df = 26611)	395,155.600 (df = 26610)
F Statistic	3,920.815*** (df = 1; 26611)	3,851.177*** (df = 1; 26611)	1,971.910*** (df = 2; 26610)
Note:			*p<0.1; **p<0.05; ***p<0.01

Our regression results can be found in Table 1 above. After running a Wald Test between Untransformed Model (1) and Untransformed Model + Indicator Variable (3) which includes the indicator variable “Receives Social Security” the results showed statistical significance, showing Untransformed Model + Indicator Variable (3) performs better. The coefficient for Total Household Income is 1.144, with a standard error of 0.054. The p-value is less than 0.01. The relationship between Total Household Income and Property Value is statistically significant at the 1% level. For every dollar increase in Total Household Income, the Property Value is expected to increase by $1.144 with a positive relationship between income and property value. The constant term is 255,260.400. Representing the estimated property value when the Total Household Income is zero. with statistical significance at the 1% level. The coefficient for the indicator variable Receives Social Security is 23,221.560 with a standard error of 5,740.003. The p-value is less than 0.01. The R-squared value is 0.129, indicating a 12.9% of the variance in Property Value is explained by the Total Household Income and Receiving Social Security. While the relationship between Total Household Income and Property Value is statistically significant the low R-squared value suggests that there are other factors not expressed in the model that significantly influence the property values. This observation is supported by the large residual standard error of 395,155.600 suggesting a considerable variability in property values not explained by the household income + Receives Social Security.

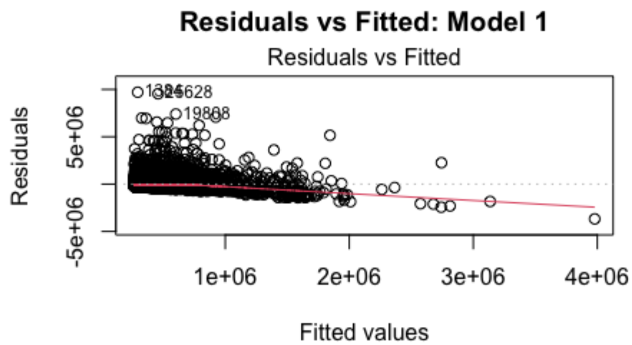
7. Overall Effect

The F-statistic 1,971.910 with a p-value of 0.01 indicates that the model is statistically significant. However, as mentioned previously the low R-squared value may indicate that its practical utility may be limited. The regression analysis shows a significant positive relationship between Total Household Income + Receives Social Security and Property value.

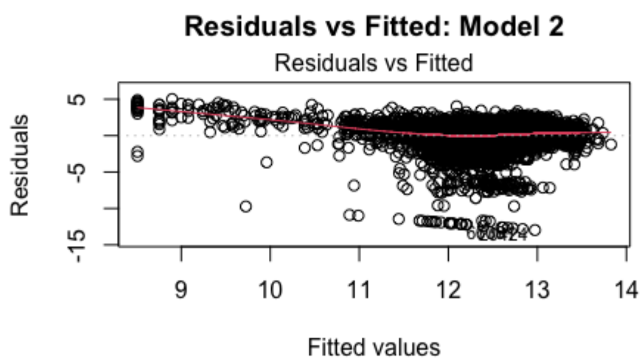
8. Appendix

Link to dataset: https://www.census.gov/data/datasets/time-series/demo/cps/cps-asec.html

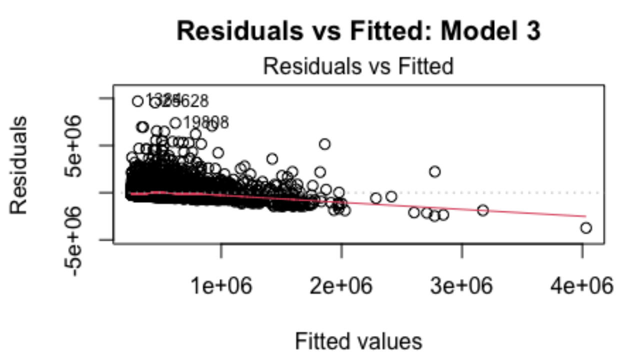
Model Specifications



Model 1: Untransformed Model, HPROP_VAL ~ HTOTVAL



Model 2: Log Transformed Model, log(HPROP_VAL) ~ log(HTOTVAL)



Model 3: Untransformed + Indicator Variable, HPROP_VAL ~ HTOTVAL + receives_social_security