The Determinants of Income Across Texas Counties

Pranjal Shrestha

Centre College

12 May 2024

I. Introduction

Household income plays a vital role in shaping living standards and satisfaction with the achieved standard of living (Yu et al., 2020). Income inequality has remained one of the most well-known issues in the United States (Danziger, 1976). Various studies have been conducted to identify the factors determining income in the United States at the regional, state, and metropolitan levels (Castells-Quintana et al., 2015). However, very few studies have been undertaken to identify the determinants of income in Texas at the county level.

Understanding these factors is important because they can help reveal why some areas thrive while others struggle economically. Given Texas's diverse economy and population, examining how different aspects, such as education, employment, and local resources, affect income is essential. This paper aims to fill this existing gap by analyzing and identifying the determinants of income in Texas at the county level.

II. Literature Review

Household income varies widely across state counties in the United States (US Census Bureau, 2023). The distribution of income among individuals is influenced by numerous social and economic factors. Aigner and Hines (1967) included key factors such as education level, age, race, and urban population in their analysis. Education is particularly significant because higher qualifications and knowledge generally lead to higher incomes. Considering both the highest level of education attained and household members' occupations is essential (Yu et al., 2020).

Age also plays a role in income variation. Younger individuals often earn more due to career advancement opportunities, while income typically declines with age, which can negatively affect median income levels within counties. This age-related income dynamic suggests the importance of considering demographic profiles when analyzing income data.

Additionally, race and unemployment rates are crucial factors in understanding income disparities. Minority groups often face systemic barriers that result in lower income levels and higher unemployment rates (Andolfatto et al., 2017). This highlights the need for policies aimed at addressing racial inequalities in employment and income.

The literature also identifies other significant factors affecting household income across counties, including the types of occupations held by household members. Occupations such as management, sales, construction, and production sectors play a vital role in determining income levels. For instance, managerial and sales roles are generally associated with higher incomes compared to production-oriented positions (Hovhannisyan, 2019). Therefore, understanding the distribution of these occupations within counties is vital for a comprehensive analysis of income determinants.

This multifaceted approach to examining income determinants not only sheds light on the complexities of income inequality but also provides insights that can inform policymakers seeking to improve economic opportunities across different regions. Moreover, addressing educational disparities and enhancing access to quality jobs are critical steps towards fostering a more equitable income distribution within and across counties (Wang et al., 2014).

III. Model Specification

Dependent Variables:

$LNMEDINC_i$ = The natural log of Median Household Income, in thousands of dollars, in county i in the year 2020.

MEDINC$_i$ = Median Household Income, in thousands of dollars, in county i in the year 2020.

Independent Variables:

BACHELORS$_i$ = Percentage of the population in county i with a bachelor's degree or more.

GRAD$_i$ = Percentage of the population in county i with a graduate degree or more.

UNEMPRATE$_i$ = Percentage of the labor force that is unemployed in county i.

UNEMPRATESQ$_i$ = Percentage of the labor force that is unemployed in county i squared.

FORBORN$_i$ = Percentage of the population that is foreign-born in county i.

AGE65OVER$_i$ = Percentage of the population aged 65 and over in county i.

WHITE$_i$ = Percentage of workers who are white in county i. (Not included in the final model)

MALE$_i$ = Percentage of the population that is male in county i. (Not included in the final model)

MGMTOCC$_i$ = Percentage of the employed population working in management, business, and financial operations occupations in county i.

LNSALES$_i$ = The natural log of the percentage of workers who are employed in sales and related occupations in county i.

SALES$_i$ = Percentage of workers who are employed in sales and related occupations in county i. (Not included in the final model)

CONSTRUCTION$_i$ = Percentage of workers who are employed in construction, extraction, and maintenance occupations in county i.

LNPRODUCTION$_i$ = The natural log of the percentage of workers who are employed in production, transportation, or material moving occupations in county i.

PRODUCTION$_i$ = Percentage of workers who are employed in production, transportation, or material moving occupations in county i. (Not included in the final model)

TABLE 1: Summary Statistics

Variable	Obs	Mean	Std. dev.	Min	Max
lnmedinc	253	10.87589	.2328108	10.03082	11.57078
medinc	253	54348.11	13140.66	22716	105956
bachelors	253	19.8246	8.040362	0	53.2282
grad	253	6.120206	3.238975	0	19.44651
unemprate	253	2.929379	1.38235	0	9.205209
unemprate_sq	253	10.4846	9.807125	0	84.73587
forborn	253	9.16392	6.890105	0	39.27732
age65over	253	18.15853	5.771574	8.986098	45.29914
white	253	80.82398	10.00002	42.79399	100
male	253	50.9135	3.523103	44.87398	70.94017
mgmtocc	253	13.53391	5.373258	2.80975	64.78873
lnsales	252	2.187888	.3626833	.0818301	2.804108
sales	253	9.353368	2.643126	0	16.51235
construction	253	12.74169	3.902819	0	28.18713
lnproduction	253	2.660968	.3723162	-.0099503	3.443079
production	253	15.15565	4.710879	.990099	31.28312

IV. Expected Signs of Coefficients

BACHELORS$_i$ and **GRAD$_i$**: With an increase in education levels, people tend to have higher incomes. For this reason, the coefficient on **BACHELORS$_i$** and **GRAD$_i$** should be positive.

UNEMPRATE$_i$: This should be negative. Higher unemployment rates indicate fewer jobs and lower income.

FORBORN$_i$: Immigrants often have different skill sets or motivations that influence their income positively. So, the coefficient should be positive.

AGE65OVER$_i$: The sign of the **AGE65OVER$_i$** should be negative since people retire by the age of 65 and their income falls as they get older after this point.

$WHITE_i$: Historical trends suggest that white individuals may have higher incomes compared to other racial groups due to systemic factors. So, the coefficient should be positive.

$MALE_i$: It has been found that males have higher income often, so the coefficient on **$MALE_i$** should be positive.

$MGMTOCC_i$, **$SALES_i$**, **$CONSTRUCTION_i$**, and **$PRODUCTION_i$**: Depending on the specific industries and occupations, the coefficients could be positive or negative, reflecting the relationship between employment in these sectors and household income. So, the signs of these coefficients are ambiguous.

The null and alternative hypotheses for each of these variables are:

Positive expected coefficient signs: BACHELORS, GRAD, FORBORN, WHITE, MALE

$$H_0: \beta \leq 0$$

$$H_A: \beta > 0$$

Negative expected coefficient signs: UNEMPRATE, AGE65OVER

$$H_0: \beta \geq 0$$

$$H_A: \beta < 0$$

Ambiguous expected coefficient signs: MGMTOCC, SALES, CONSTRUCTION, PRODUCTION

$$H_0: \beta = 0$$

$$H_A: \beta \neq 0$$

V. Data Collection

Data was collected from Social Explorer 2024, a platform known for providing easy access to demographic information about the United States. The dataset comprises observations from over 250 counties across Texas, providing a broad representation of geographic regions within the state. Social Explorer serves as a valuable resource for accessing and analyzing demographic, social, and economic data sourced from the U.S. Census Bureau and other reliable sources.

VI. Estimating the Equation

Model 1:

Estimated Regression Equation:

$$\begin{aligned} MEDINC_i = \ & \beta_0 + \beta_1 BACHELORS_i + \beta_2 GRAD_i + \beta_3 UNEMPRATE_i + \beta_4 FORBORN_i \\ & + \beta_5 AGE65OVER_i + \beta_6 WHITE_i + \beta_7 MALE_i + \beta_8 MGMTOCC_i + \beta_9 SALES_i \\ & + \beta_{10} CONSTRUCTION_i + \beta_{11} PRODUCTION_i + \varepsilon_i \end{aligned}$$

This model is an original model that includes all the independent variables and dependent variable MEDINCi as mentioned above. In terms of goodness of fit, Model 1 has an adjusted R^2 of .6429, which means around 64.29% of the variation of the income around its mean is explained by the model, adjusted for degrees of freedom.

This is a quite good model since most of the variables are significant, with BACHELORS and occupation variables being highly significant. This suggests that BACHELORS could potentially have a significant influence on determining median household income within the model. Specifically, a one percentage point increase in the percentage of the population in county i with a graduate degree or more is associated with an increase in median household income of 689.67 dollars, all else equal. While this indicates a reasonably good fit, however only 8 out of 11 independent variables are statistically significant. Given these factors, it suggests that the model might need adjustment for a better fit.

TABLE 2: Model 1 Regression

Source	SS	df	MS
Model	2.8653e+10	11	2.6048e+09
Residual	1.4862e+10	241	61667571.9
Total	4.3515e+10	252	172677021

Number of obs	=	253
F(11, 241)	=	42.24
Prob > F	=	0.0000
R-squared	=	0.6585
Adj R-squared	=	0.6429
Root MSE	=	7852.9

medinc	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
bachelors	689.6616	172.3123	4.00***	0.000	350.2311	1029.092
grad	661.5765	410.9249	1.61	0.109	-147.8864	1471.039
unemprate	-873.1748	378.6167	-2.31 **	0.022	-1618.995	-127.3543
forborn	-189.4522	85.69004	-2.21	0.028	-358.2493	-20.65517
age65over	-927.5093	109.1473	-8.50***	0.000	-1142.514	-712.5047
white	60.56914	58.65869	1.03	0.303	-54.98006	176.1183
male	-189.7199	163.6646	-1.16	0.248	-512.1155	132.6758
mgmtocc	1182.441	123.7237	9.56***	0.000	938.7232	1426.159
sales	891.5541	224.5132	3.97***	0.000	449.2955	1333.813
construction	969.4084	148.3778	6.53***	0.000	677.1255	1261.691
production	641.0497	125.3117	5.12***	0.000	394.2037	887.8957
_cons	16117.41	12168.92	1.32	0.187	-7853.619	40088.44

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	253	-2757.808	-2621.909	12	5267.818	5310.219

Note:
* = significant at the 10% level
** = significant at the 5% level
*** = significant at the 1% level

Model 2:

Estimated Regression Equation:

$$LNMEDINC_i = \beta_0 + \beta_1 BACHELORS_i + \beta_2 GRAD_i + \beta_3 UNEMPRATE_i + \beta_4 FORBORN_i + \beta_5 AGE65OVER_i + \beta_6 WHITE_i + \beta_7 MALE_i + \beta_8 MGMTOCC_i + \beta_9 SALES_i + \beta_{10} CONSTRUCTION_i + \beta_{11} PRODUCTION_i + \varepsilon_i$$

In Model 2, the dependent variable MEDINC is transformed into the natural log of MEDINC, as suggested by the literature review. This seemed to be a better model, considering two out of three main goodness of fit measures are leaning towards the model as the better fit. The AIC and BIC in Model 2 are much lower than in Model 1. The adjusted R^2, however, has dropped from .6429 to .6310. The theory in econometrics says the higher the R^2, the better the model. As for AIC and BIC, the lower the better the model. Additionally, several variables such as GRAD which was not statistically significant in Model 1 are now significant at the 10% level. UNEMPRATE, which was previously statistically significant at the 5% level, is now highly significant at the 1% level. This indicates that UNEMPRATE now has a more substantial impact on determining median household income within the model compared to the previous model. Specifically, if the percentage of the labor force that is unemployed increases by 1 percentage point, the median household income is expected to decrease by 2.01%, all else equal.

TABLE 3: Model 2 Regression

Source	SS	df	MS			
Model	8.87947845	11	.807225314	Number of obs	=	253
Residual	4.7791396	241	.019830455	F(11, 241)	=	40.71
Total	13.658618	252	.054200865	Prob > F	=	0.0000
				R-squared	=	0.6501
				Adj R-squared	=	0.6341
				Root MSE	=	.14082

lnmedinc	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
bachelors	.0111928	.00309	3.62***	0.000	.005106	.0172796
grad	.0129451	.0073689	1.76 *	0.080	-.0015705	.0274607
unemprate	-.0201217	.0067895	-2.96***	0.003	-.033496	-.0067473
forborn	-.0051612	.0015366	-3.36	0.001	-.0081882	-.0021343
age65over	-.017097	.0019573	-8.74***	0.000	-.0209526	-.0132415
white	.0012847	.0010519	1.22	0.223	-.0007874	.0033567
male	-.0029686	.0029349	-1.01	0.313	-.0087499	.0028128
mgmtocc	.021574	.0022187	9.72***	0.000	.0172036	.0259445
sales	.0180326	.0040261	4.48***	0.000	.0101018	.0259633
construction	.0201008	.0026608	7.55***	0.000	.0148595	.0253421
production	.0140831	.0022471	6.27***	0.000	.0096565	.0185096
_cons	10.10857	.2182177	46.32	0.000	9.678711	10.53843

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	253	10.26443	143.1034	12	-262.2067	-219.8061

In TABLE 3, since the value of adjusted R^2 is still lower, it is beneficial to revise the specification of certain independent variables to ensure a more precise depiction of their association with LNMEDINC.

<u>Model 3:</u>

$$LNMEDINC_i = \beta_0 + \beta_1 BACHELORS_i + \beta_2 GRAD_i + \beta_3 UNEMPRATE_i + \beta_4 FORBORN_i + \beta_5 AGE65OVER_i + \beta_6 WHITE_i + \beta_7 MALE_i + \beta_8 MGMTOCC_i + \beta_9 LNSALES_i + \beta_{10} CONSTRUCTION_i + \beta_{11} LNPRODUCTION_i + \varepsilon_i$$

In Model 3, the natural logarithm of the dependent variable MEDINC was taken, along with the natural logarithm of the independent variables SALES and PRODUCTION. Model 3 shows a significant improvement in fit compared to previous models, with an adjusted R-squared value of 0.6631, meaning that about 66.31% of the variation in median household income is accounted for by the model, adjusted for degrees of freedom. Additionally, both the AIC and BIC values are lower, indicating even a better model. This signifies that LNSALES now has a more substantial impact on determining median household income within the Model 3 compared to Model 2. This can be interpreted as follows: if the percentage of workers who are employed in sales and related occupations in county i increases by 1%, the median household income will increase by 0.15 %, all else equal.

TABLE 4: Model 3 Regression

Source	SS	df	MS			
				Number of obs	=	252
				F(11, 240)	=	45.92
Model	9.23671497	11	.839701361	Prob > F	=	0.0000
Residual	4.38856896	240	.018285704	R-squared	=	0.6779
				Adj R-squared	=	0.6631
Total	13.6252839	251	.054284	Root MSE	=	.13522

lnmedinc	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
bachelors	.0080603	.003168	2.54	**0.012	.0018198	.0143009
grad	.01557	.0072406	2.15	**0.033	.0013067	.0298332
unemprate	-.0185175	.0065279	-2.84	***0.005	-.0313768	-.0056582
forborn	-.0025233	.0015235	-1.66	*0.099	-.0055244	.0004779
age65over	-.013452	.0019774	-6.80	***0.000	-.0173473	-.0095567
white	.0010922	.0010163	1.07	0.284	-.0009098	.0030943
male	-.0006085	.0028761	-0.21	0.833	-.0062741	.0050571
mgmtocc	.0298205	.002857	10.44	***0.000	.0241926	.0354484
lnsales	.148694	.0282457	5.26	***0.000	.0930529	.2043351
construction	.0187984	.0025845	7.27	***0.000	.0137072	.0238897
lnproduction	.2124704	.0282861	7.51	***0.000	.1567497	.268191
_cons	9.355573	.2524594	37.06	0.000	8.858254	9.852892

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	252	10.03273	152.7812	12	-281.5623	-239.2092

Model 4:

$$LNMEDINC_i = \beta_0 + \beta_1 BACHELORS_i + \beta_2 GRAD_i + \beta_3 UNEMPRATE_i + \beta_4 UNEMPRATE\_sq_i + \beta_5 FORBORN_i + \beta_6 AGE65OVER_i + \beta_7 WHITE_i + \beta_8 MALE_i + \beta_9 MGMTOCC_i + \beta_{10} LNSALES_i + \beta_{11} CONSTRUCTION_i + \beta_{12} LNPRODUCTION_i + \varepsilon_i$$

In Model 4, a polynomial of the independent variable (UNEMPRATESQ) was introduced. A two-way scatterplot was generated to assess whether UNEMPRATE exhibits a linear or polynomial relationship, leading to its transformation accordingly. With this modification, the adjusted R-squared value has improved, and the AIC and BIC values have decreased, signifying an improved model.

TABLE 5: Model 4 Regression

Source	SS	df	MS
Model	9.33092778	12	.777577315
Residual	4.29435615	239	.017968017
Total	13.6252839	251	.054284

Number of obs	=	252
F(12, 239)	=	43.28
Prob > F	=	0.0000
R-squared	=	0.6848
Adj R-squared	=	0.6690
Root MSE	=	.13404

lnmedinc	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
bachelors	.0081559	.0031406	2.60	**0.010	.0019692	.0143427
grad	.0134367	.0072376	1.86	*0.065	-.000821	.0276944
unemprate	.0221163	.0188883	1.17	0.243	-.0150925	.0593251
unemprate_sq	-.0060812	.0026557	-2.29	**0.023	-.0113128	-.0008496
forborn	-.0019842	.0015285	-1.30	0.195	-.0049951	.0010268
age65over	-.0126564	.0019907	-6.36	***0.000	-.016578	-.0087348
white	.0009936	.0010084	0.99	0.325	-.0009929	.00298
male	.0001156	.0028685	0.04	0.968	-.0055351	.0057663
mgmtocc	.0304317	.0028446	10.70	***0.000	.024828	.0360353
lnsales	.1462079	.0280203	5.22	***0.000	.0910096	.2014062
construction	.0187185	.0025622	7.31	***0.000	.0136711	.023766
lnproduction	.206764	.0281498	7.35	***0.000	.1513106	.2622174
_cons	9.276606	.2526216	36.72	0.000	8.778957	9.774255

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	252	10.03273	155.5155	13	-285.0311	-239.1485

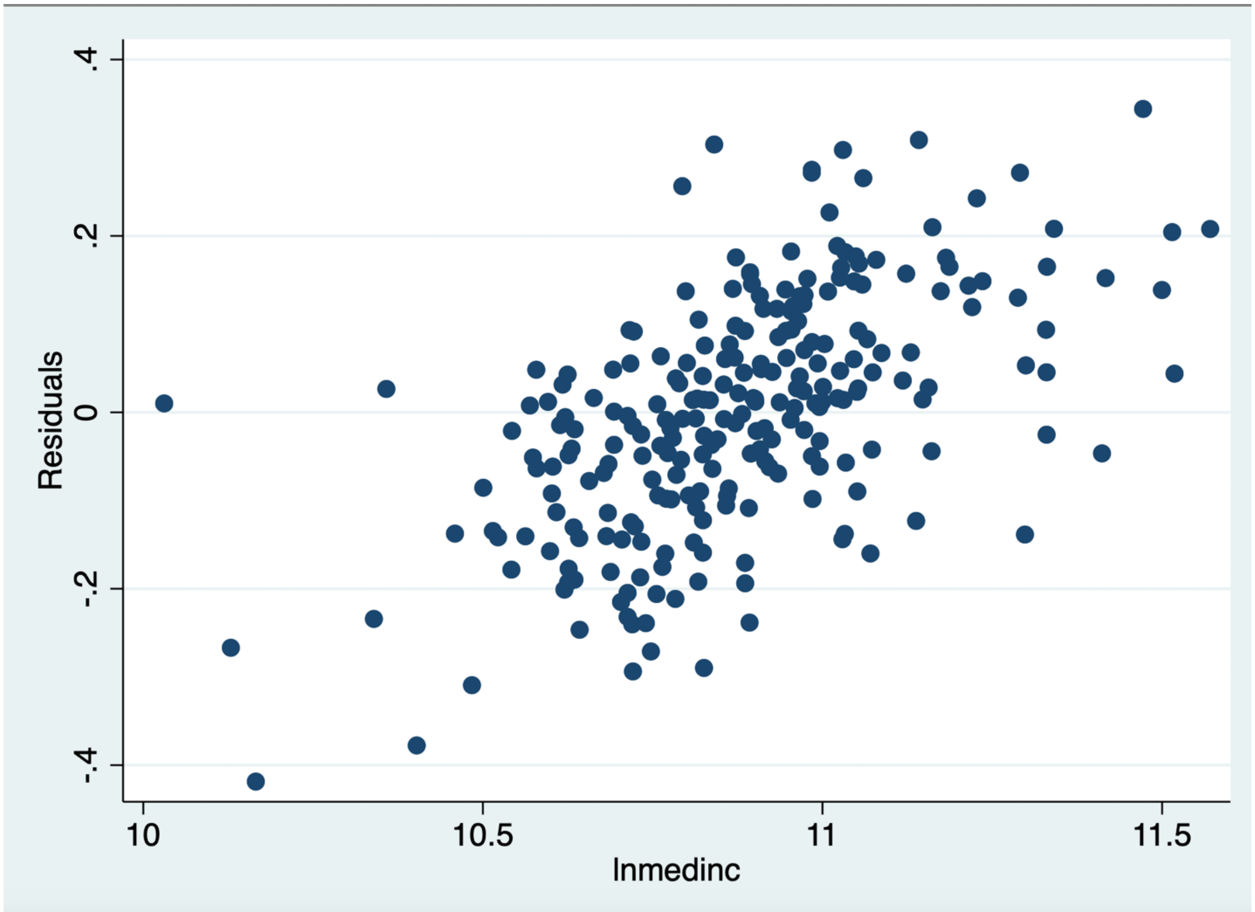
Even after the modification of the specification, certain variables continue to raise concerns. Notably, the significance of GRAD at the 10% level, alongside the persistent insignificance of WHITE and MALE, suggests the potential presence of multicollinearity within the model. To evaluate this potential concern, STATA was used to find the Variance Inflation Factors (VIFs), which provide insight into the degree of correlation among the predictors:

Variable	VIF	1/VIF
unemprate_sq	9.47	0.105592
unemprate	9.39	0.106492
bachelors	8.73	0.114597
grad	7.60	0.131617
mgmtocc	2.09	0.478754
age65over	1.69	0.592308
forborn	1.54	0.647462
lnproduction	1.52	0.656514
lnsales	1.44	0.693148
white	1.41	0.711639
construction	1.34	0.744656
male	1.25	0.801303
Mean VIF	3.96	

It appears that both BACHELORS and GRAD exhibit higher VIFs (8.73 and 7.60, respectively) compared to the threshold of 5.0, suggesting a significant degree of correlation between these two variables, as anticipated. However, when the regression was run by removing one of those two variables, the adjusted R^2 dropped with the AIC and BIC being increased. So, both BACHELORS and GRAD were kept. Furthermore, as WHITE and MALE are highly insignificant, they are excluded from another model. Doing so gave us an even higher adjusted R^2 and lower AIC and BIC.

After determining that Model 4 represented the best specification model for the data, testing multicollinearity, and removing irrelevant variables, further testing for heteroskedasticity was carried out to ensure it followed the assumptions and prevented from making errors. First, the residuals were plotted against the LNMEDINC. The presence of heteroskedasticity was indicated by a displayed pattern and the unequal deviation between the residuals which can be observed in the figure below.

FIGURE 1: Residuals Against the LNMEDINC



To further support the claim, the White Test was conducted. With the p-value of 0.0003, we rejected the null hypothesis at the 1% level. Thus, it was confirmed that there was heteroskedasticity.

```
White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

   chi2(64) = 110.81
Prob > chi2 = 0.0003

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	110.81	64	0.0003
Skewness	7.93	10	0.6362
Kurtosis	0.69	1	0.4075
Total	119.42	75	0.0008

Heteroskedasticity was corrected using the robust standard errors. The R-squared improved the overall model, indicating a better fit.

TABLE 6: Final Model Regression

```
Linear regression                               Number of obs     =        252
                                                F(10, 241)        =      53.90
                                                Prob > F          =     0.0000
                                                R-squared         =     0.6835
                                                Root MSE          =     .13376
```

lnmedinc	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
bachelors	.0083513	.0036271	2.30	**0.022	.0012065	.0154962
grad	.0122038	.0088227	1.38	0.168	-.0051756	.0295831
unemprate	.0217456	.016775	1.30	0.196	-.0112988	.0547901
unemprate_sq	-.0061811	.0021416	-2.89	***0.004	-.0103998	-.0019624
forborn	-.0023071	.0016837	-1.37	0.172	-.0056237	.0010096
age65over	-.0122377	.0022075	-5.54	***0.000	-.0165862	-.0078892
mgmtocc	.0307288	.0028924	10.62	***0.000	.0250312	.0364265
lnsales	.1497755	.0232327	6.45	***0.000	.1040105	.1955406
construction	.0185476	.0028832	6.43	***0.000	.012868	.0242271
lnproduction	.2035687	.0262503	7.75	***0.000	.1518594	.255278
_cons	9.362867	.1653429	56.63	0.000	9.037165	9.688568

With its improved goodness-of-fit, alignment with theory, and appropriate t-scores, the regression equation presented in TABLE 6, which accounts for heteroskedasticity, stands as the final model:

$$
\begin{aligned}
LNMEDINC_i = {} & \beta_0 + \beta_1 BACHELORS_i + \beta_2 GRAD_i + \beta_3 UNEMPRATE_i \\
& + \beta_4 UNEMPRATE\_sq_i + \beta_5 FORBORN_i + \beta_6 AGE65OVER_i + \beta_7 MGMTOCC_i \\
& + \beta_8 LNSALES_i + \beta_9 CONSTRUCTION_i + \beta_{10} LNPRODUCTION_i + \varepsilon_i
\end{aligned}
$$

VII. Evaluation

Omitted Variables

To prevent omitted variable bias, different measures were taken to ensure that all relevant independent variables were included in the estimated equation. This involved performing regression analysis with several variables and conducting a thorough literature review to identify and incorporate key variables into the analysis. However, there were some variables that were not found in the Social Explorer dataset. For instance, marriage could be a factor determining income.

Irrelevant Variables

When the irrelevant independent variables are included in the model, it reduces the precision of standard errors, subsequently impacting the t-scores and confidence intervals of the model. There are several ways to identify if the variables are irrelevant including but not limited to t-tests, examining multicollinearity, considering adjusted R^2, and others. There were two irrelevant variables identified in this paper: WHITE, and MALE. All of these were identified as irrelevant variables through t-tests, adjusted R^2 analysis, and consideration of their impact on other coefficients.

Incorrect Functional Form

When the functional form of the model does not precisely represent the relationship between the variables, it leads to incorrect estimates and biased results. To address this, various functional forms such as linear, semi-log, log-log, and polynomial were tested to find the best fit.

Additionally, a thorough literature review guided the selection of the appropriate functional form. For instance, based on multiple studies, taking the logarithm of MEDINC was recommended, which ultimately resulted in the best-fitting model.

Multicollinearity

Multicollinearity means when two or more independent variables are highly correlated and can affect the goodness of fit of the model. In this paper, multicollinearity was identified using the Variance Inflation Factor (VIF) in STATA. According to the VIF, the general rule is that multicollinearity exists when the value of VIF is more than the threshold of 5.0. In our model, since two variables' BACHELORS and GRAD had high VIFs and had a value greater than 5.0. However, since their coefficients were statistically significant, those variables were not removed in the final model.

Serial Correlation

Serial correlation exists in time series dataset, which is the correlation of observations of the error term. It does not apply to this study because it focuses on cross-sectional data.

Heteroskedasticity

Heteroskedasticity occurs when there is a pattern in the plot of residuals against the dependent variable or when the deviation between the residuals is not equal. As previously mentioned, heteroskedasticity existed in the data. It was identified through two different methods: the White Test and residual plot against the LNMEDINC. However, it was corrected using the robust standard errors.

VIII. Conclusion

In this study, income determinants across Texas counties were explored through regression models, with adjustments such as variable transformations enhancing model fit and explanatory power. Despite encountering challenges like multicollinearity and heteroskedasticity, the final model successfully addressed these issues, providing a comprehensive understanding of income determinants. The final model offers a reasonable explanation of income determinants, with identified variables providing valuable insights into income levels. However, further research is needed to delve deeper into the insignificant variables such as GRAD, FORBORN, and UNEMPRATE, which were included in the final model. Notably, the coefficient on FORBORN was anticipated to be positive, reflecting assumed skilled manpower and a diverse skillset contribution. However, the negative estimation suggests the prevalence of low-skilled labor in Texas, shedding light on the complexity of regional labor dynamics. Overall, this study contributes to the understanding of income determinants in Texas counties, highlighting the need for continued research to refine models and capture the nuanced factors influencing income levels accurately.

Bibliography

Castells-Quintana, David; Ramos, Raul; and Royuela, Vincenta. "Income Inequality in European Regions: Recent Trends and Determinants." Review of Regional Research 35 (2015): 123-146.

Yu, Grace B. , Dong-Jin Lee, M. Joseph Sirgy, and Michael Bosnjak. "Household Income, Satisfaction with Standard of Living, and Subjective Well-Being. The Moderating Role of Happiness Materialism." Springer Link, 2019. November 12. https://link.springer.com/article/10.1007/s10902-019-00202-x.

United States Census Bureau. "American Community Survey 1-Year Estimates: Household Income Data." September 2023. https://www.census.gov/programs-surveys/acs.

Danziger, Sheldon. "Determinants of the Level and Distribution of Family Income in Metropolitan Areas, 1969." Land Economics 52, no. 4 (1976): 467–78.

Shao, Liang Frank. "Robust Determinants of Income Distribution across and within Countries." <i>National Library of Medicine</i>, 2021. July 1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8248696/#:~:text=Second%2C%20we%20update%20the%20literature,the%20unexplained%20GDP%20are%20the.

Aigner, D. J. and Heins, A. J. "On the Determinants of Income Equality." The American Economic Review 57, no. 1 (1967): 175–84.

Wang, Chen; Wan, Guanghua; and Yang, Dan. "Income Inequality in the People's Republic of China: Trends, Determinants, and Proposed Remedies." Journal of Economic Surveys 28, no. 4 (2014): 686-708. doi: 10.1111/joes.12077.

Andolfatto, David , and Andrew Spewak. "Why Do Unemployment Rates Vary by Race and Ethnicity?" Federal Reserve Bank of St. Louis, 2017. February 6. https://www.stlouisfed.org/on-the-economy/2017/february/why-unemployment-rates-vary-races-ethnicity.

Hovhannisyan, Anna. "The Determinants of Income Inequality: The Role of Education." ResearchGate, 2019. December 1. https://www.researchgate.net/publication/347456611_THE_DETERMINANTS_OF_INCOME_INEQUALITY_THE_ROLE_OF_EDUCATION.