Undergraduate Economic Review

Volume 18 | Issue 1 Article 8

2021

The Effect of Income on Healthy Food Options

Hannah M. Doherty
Centre College, hannah.doherty@centre.edu

The Effect of Income on Healthy Food Options

Abstract

This paper explores the effect of income per capita on the number of grocery stores and fast-food franchises in an area. Using a panel dataset to allow for the inclusion of every county in the United States across a period of three years, the results suggest that the income per capita of a county significantly impacts the number of grocery stores and fast-food restaurants in the area. Other factors such as education, age, and attributes regarding time constraints also play an important role in determining the number of grocery stores and fast-food franchises in a location.

Keywords

Income, Grocery Stores, Fast-food, Food deserts

I. Introduction

The United States Department of Agriculture reports that around 23.5 million Americans currently live in areas that are categorized as food deserts. These areas have low access to grocery stores and healthy food options, but they have a high prevalence of fast-food restaurants. Lower-income communities are disproportionately affected by the lack of access to nutritious options, which is linked with dietary problems such as high blood pressure, cardiovascular disease, obesity, and an increased prevalence of diabetes.[1]

Most previous studies have focused on a single community of interest or geographic area, while this paper will use a panel dataset to allow for the inclusion of every county in the United States across a period of three years. This is distinct from the methods of previous studies, which have tended to use cross-sectional datasets from two distinct time periods to look at the change in fast-food restaurant or grocery store numbers over time. In addition, most existing research has looked at the effect of income per capita on either the prevalence of fast-food restaurants or the lack of grocery stores, but previous research has not looked at both of these attributes together. This paper will look at both the number of fast-food restaurants and the number of grocery stores by using two separate models, one for each dependent variable, but will use the same control variables and datasets in both models. An instrumental variable regression using two-stage least squares will be conducted to control for the presence of reverse causality in the model, which many existing studies have not done.

The results, overall, suggest that the income per capita of a county significantly impacts the number of grocery stores and fast-food restaurants in the area. Other factors such as education, age, and attributes regarding time constraints also play an important role in determining the number of grocery stores and fast-food franchises.

The remainder of this paper is organized as follows. Section II reviews the relevant empirical literature regarding the effect of income on healthy food options, as well as the factors that drive franchises to choose locations. Section III discusses the economic theory of the study. Section IV describes the data along with the descriptive statistics. Section V presents the empirical model and hypotheses. Section VI analyzes the results and discusses econometric problems and sources of bias. Section VII concludes the paper, followed by a brief appendix and bibliography.

II. Literature Review

The empirical findings of existing research tend to support the hypothesis that low-income areas have a higher percentage of fast-food restaurants and have

[1] "Food Deserts: Definition, Effects, and Solutions." *Medical News Today*

decreased access to healthy food options, resulting in their classification as "food deserts." In the study, "Determinants of Food Deserts" researchers found that areas with a lower median income have a much higher prevalence of fast-food restaurants with unhealthy food options, and areas with higher minority populations were statistically more likely to be classified as a food desert.[2]

Another study, titled, "Disparities in Neighborhood Food Environments: Implications of Measurement Strategies" found that neighborhoods with higher income levels and higher proportions of white residents tend to have greater access to grocery stores, while poorer neighborhoods and those with higher proportions of Black or Hispanic residents have relatively low access to grocery stores.[3]

The impact of time constraints on households was explored in the study, "Convenience, Accessibility, and the Demand for Fast Food." Researchers looked at variables such as work hours and female labor force participation to determine the importance of household time constraints and the convenience of fast-food restaurants in the demand for fast food. This study also looked at the opportunity-cost-of-time for the preparation of healthy meals versus the near immediate service provided by fast-food franchises. The results of this study supported the researchers' hypothesis that consumer accessibility to fast-food is a significant factor in demand for fast-food, and demographic variables were found to be significant contributors as well, especially race.[4]

The paper, "Do Fast-Food Chains Price Discriminate on the Race and Income Characteristics of an Area?" which was conducted in 1997, demonstrates that fast-food franchises do take into account the demographics of an area. This study found significant differences in price based on the race characteristics of a zip-code region. Taking into account income disparities and differences in production costs for the franchise in each location, researchers found that fast-food restaurants charge 5.4% more for a 50% increase in the Black population of the area, which was statistically significant. This demonstrates that fast-food franchises do look at the demographics of an area.[5]

Most existing research has focused on the demographic statistics for a single community of interest and has looked at either the prevalence of fast-food restaurants or the lack of healthy options, but previous research has not looked at both of these attributes together. This paper will look at both the number of fast-

[2] Alviola, P. A., Nayga, R. M., Thomsen, M. R., & Wang, Z. (2013). Determinants of Food Deserts. *American Journal of Agricultural Economics*, *95*(5), 1259–1265.
[3] Bader, M. D. M., Purciel, M., Yousefzadeh, P., & Neckerman, K. M. (2010). Disparities in Neighborhood Food Environments: Implications of Measurement Strategies. *Economic Geography*, *86*(4), 409–430.
[4] Jekanowski, M. D., Binkley, J. K., & Eales, J. (2001). Convenience, Accessibility, and the Demand for Fast Food. *Journal of Agricultural and Resource Economics*, *26*(1), 58–74.
[5] Graddy, K. (1997). Do Fast-Food Chains Price Discriminate on the Race and Income Characteristics of an Area? *Journal of Business & Economic Statistics*, *15*(4), 391–401.

food restaurants and the number of grocery stores on a county level, across the entire United States.

III. Theoretical Model

It is important to consider what drives grocery stores and fast-food franchises to choose specific locations. Assuming that grocery stores and fast-food restaurants are profit maximizing entities, the central focus when choosing a location is to increase total revenue and decrease total costs, which is known as the profit maximization model.[6]

According to the *Washingtonian*, grocery stores look at the traffic levels and number of public transportation stops in order to determine if a location will be accessible. In addition, franchises consider the consumer demographics of the neighborhood.[7] Grocery stores also look at the accessibility of an area, the neighborhood demographics, and recent residential growth, but are primarily driven by household income.[8] In addition to income, grocery stores "have a set of consumer lifestyle profiles that are their customer base," and they "look for these lifestyle demographic characteristics in their prospective markets. Typically, other demographic factors that retailers would look for would be education, ethnicity and employment." [9] Whole Foods specifically is one grocery store chain that relies heavily on demographic information to determine where to place stores, specifically searching for a consumer base of "well-educated, affluent" citizens that will be interested in natural and organic foods. The education level of a community is the primary driver of where Whole Foods locates its stores, because "the chain counts on consumers who are willing to pay more because they know about the health benefits of eating organic or have a taste for less common foods." Because of this information, it is necessary to include demographic statistics such as race and educational attainment in the model.[10]

It is also beneficial to look at how grocery stores and fast-food franchises aim to maximize profits in terms of supply and demand. From a production standpoint, firms look at how easy and lucrative it is to operate a business in a certain location, taking into account the price and quality of available inputs, capital and land, and the labor market. On the demand side, firms consider what the customers in specific locations desire, for example, considering the income of an

---

[6] "Profit Maximization Model of a Firm." *Economics Discussion*

[7] Kashino, Marissa. "How Whole Foods Decides If Your Neighborhood Is Worthy." *The Washingtonian*

[8] Ungerleider, Neal. "How Fast Food Chains Pick Their Next Location." *Fast Company*, 25 Aug. 2014

[9] Loria, Keith. "How Do Grocery Stores Find the Right Location for Expansion?" *Grocery Dive* 17 Feb. 2017

[10] Kashino, Marissa. "How Whole Foods Decides If Your Neighborhood Is Worthy." *The Washingtonian*

area. Some determinants of grocery store and fast-food franchise locations, like education, could influence both supply and demand. For example, a lower-educated workforce is more attractive for employment at fast-food franchises, and firms may consider that more educated citizens tend to desire healthier food options.

In the study, "Fresh Vegetable Demand Behavior in an Urban Food Desert" researchers looked at whether the emergence of food deserts is a result of demand-side factors, such as low income, or supply-side factors that cause prohibitively high costs of operation for grocery stores and cause limited access to healthy food in these areas. This study looked at simultaneous causality regarding the lack of healthy food options in food deserts and attempted to determine whether a low demand for healthy foods caused food deserts or if limited healthy food access caused low expressed demand for healthy food. The study looked at the opening of a non-profit green-grocer in an inner-city neighborhood with a 92% African-American population and an average income that was lower than 96% of U.S. neighborhoods. They found that lower income areas have a much higher income elasticity for healthy produce; an increase in disposable income resulted in a higher increase in demand for nutritious goods than for the general population. Ultimately, the study concluded that food deserts form as a result of supply side factors: lower income areas purchase fewer fresh vegetables because they do not have access to these healthy options, not because there is not a demand for nutritious food.[11]

Based on the empirical findings of existing studies, it is expected that counties with lower average incomes will have a lower amount of grocery stores in the area and a higher prevalence of fast-food restaurants. Two separate regressions will be run, one with the number of fast-food restaurants as the dependent variable, and one with the number of grocery stores as the dependent variable. There is expected to be a positive relationship between the variable of interest, average county income, and the dependent variable, the number of grocery stores, because as the average income in a county increases, more grocery stores will be attracted to the location. There is expected to be a negative relationship between the average county income and the number of fast-food restaurants, because as the average income in a county increases, that area will be less attractive to fast-food restaurants. This is excluding the impacts of any constant network effects, where a higher prevalence of businesses attracts more firms, which will be controlled for by using time and entity fixed effects in the panel regression.

In addition, reverse causality is present in the model: higher-income areas could attract more businesses, and therefore more grocery stores, while at the same time, areas with more grocery stores could be seen as more attractive and appeal to higher-income home buyers, resulting in a higher income per capita. The same reverse causality is seen in the regression with fast-food restaurants as the

[11] Weatherspoon, D., Oehmke, J., Dembele, A., & Weatherspoon, L. (2015). Fresh vegetable demand behaviour in an urban food desert. *Urban Studies*, *52*(5), 960–979.

dependent variable. An instrumental variable regression approach will be used to eliminate bias and ensure that a causal effect can be estimated. The instrumental variable used in this study will be the percentage of households in the county that have a relative outside of the immediate family residing in the household. Multigenerational households are more likely to occur when economic resources are scarce.[12] This percentage is correlated with the income per capita of a county, but it is assumed to not be correlated with the number of grocery stores or fast-food restaurants in a county in any way except through the income per capita, therefore its use as an instrumental variable will eliminate bias in the model.

IV. Data

The data for this study was obtained from numerous government databases and merged into an unbalanced panel dataset, spanning the years 2011, 2014, and 2016, and including every county in the United States. The table below details the sources of each variable in the dataset.

Table I: Variable Sources

Variables	Source
GROC, FFR, Unemployment_rate	The Economic Research Department within the U.S. Department of Agriculture
Income_per_capita	The Bureau of Economic Affairs, Regional Economic Accounts
violentcrimeper1000	The Uniform Crime Reporting Program of the Federal Bureau of Investigation
White, Black, Hispanic, Asian American_Indian, percentMale, percentFemale, Age16to19years, Age20to24years, Age25to44years, Age45to54years, Age55to59years, Age60yearsandover, Naturalresourcesconstruction, Worknightshiftleavingfor, commuteunder30mins, commute30to59mins, commuteover60mins, Novehicleavailable, vehicleavailable, vehiclesavailable, ormorevehiclesavailable, Lessthanhighschoolgraduate, Highschoolgraduateincludese, Somecollegeorassociatesdegr, Bachelorsdegree, Graduateorprofessionaldegree, percentrelatives	The American Community Survey, from the U.S. Census Bureau

[12] "More Kids Living in Multigenerational Families." *Poverty Solutions at the University of Michigan*, 25 Sept. 2018

The dataset includes two dependent variables, which will be used in two separate regressions; *GROC*, which is the number of grocery stores per county, and *FFR*, which is the number of fast-food restaurants per county. These variables were both divided by the population of each county and multiplied by 1000 to give the number of grocery stores and fast-food restaurants per 1000 residents. In addition, the dependent variables will be included in the model in the logarithmic form. The variable of interest is the income per capita of each county, which will be included as the natural log of income per capita, *ln_incpercap*. Below, Table I presents the summary statistics for the dataset, but further descriptions of each variable can be found in the Appendix.

Table II: Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
GROC	18.811	69.922	0	2076
FFR	70.091	257.572	0	8264
Income per capita	38592	10074.703	16216	137739
Unemployment rate	6.756	2.826	1.2	29.3
White	86.125	14.792	11.121	99.155
Black	8.609	13.548	0	85.988
Hispanic	8.787	13.059	.336	96.254
Asian	1.287	2.272	0	36.503
American Indian	2.014	6.294	.015	86.275
percentMale	50.047	2.14	43.367	72.199
percentFemale	49.953	2.14	27.801	56.633
violentcrimeper1000	.5	.561	0	7.426
Age16to19years	3.951	1.36	0	19
Age20to24years	9.002	3.12	0	36.9
Age25to44years	39.569	4.6	14.9	61.8
Age45to54years	23.854	2.845	5.1	45
Age55to59years	10.652	2.109	2.2	48.7
Age60yearsandover	12.97	3.601	0	39.1
ConstructionMaintenance	12.844	4.085	2.8	52.5
Nightshift	5.267	2.529	0	31
commuteunder30mins	69.355	11.674	20.8	96.5
commute30to59mins	23.237	9.183	0	66.3
commuteover60mins	7.408	4.394	0	33.2
NoVehicle	2.285	1.765	0	34.1
OneVehicle	17.965	4.644	2.7	51.7
TwoVehicles	40.773	4.94	13.2	64
ThreeOrMoreVehicles	38.975	7.584	8.5	76.2
LessThanHS	15.153	6.751	1.9	53.7
HSgrad	34.966	6.797	8.2	78.6
SomeCollege	30.008	5.142	8.9	49
BachelorsDegree	13.078	5.335	0	43
GradDegree	6.795	3.746	0	39.1
percentrelatives	5.754	2.823	0	24

The dataset includes 7,123 observations for each variable. It is particularly interesting to note that there is a wide disparity in the number of fast-food restaurants and grocery stores in each county. Some counties have as few as 0 fast-food restaurants and grocery stores, while others have as many as 2,076 grocery stores or 8,264 fast-food restaurants. The average number of grocery stores per county is 18.811 and the average number of fast-food restaurants per county is 70.091, demonstrating that there are outliers skewing the data. The 25th percentile of the data for the variable, *GROC*, is 3 and the 75th percentile is 12, meaning that 50% of counties have between 3 and 12 grocery stores, further demonstrating that counties that have as many as 2,076 grocery stores are extreme outliers. Similarly, the 25th percentile for the variable, *FFR*, is 5 and the 75th percentile is 46, meaning that 50% of counties have between 5 and 46 fast-food restaurants, while some outliers have as many as 8,264. Because of this, these dependent variables will be included in the model as natural logs. The variable of interest, *ln_incpercap*, will also be included as a natural log because of the existence of extreme outliers. The middle 50% of counties have an income per capita between $31,986 and $42,894, but there are some counties that have an income per capita as high as $137,739, demonstrating that this variable is also significantly skewed.

V. Empirical Model

In this paper, seven separate empirical equations will be specified for each of the dependent variables. First, the impact of income on healthy food options is estimated with a simple linear equation using OLS, with the grocery stores and fast-food restaurants as the dependent variables and the variable of interest, income per capita, on the right-hand side. The number of grocery stores and the number of fast-food restaurants in each county have been divided by the total county population and multiplied by 1,000 to provide the number of fast-food restaurants and grocery stores per 1,000 residents, and both will be included as natural logs. The basic regression model estimated primarily is shown below.

Model (1)

$$lnGROCper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + u_{it}$$
$$lnFFRper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + u_{it}$$

It is expected that the coefficient on the natural log of income per capita will be negative in the model with fast-food restaurants as the dependent variable and positive in the model with grocery stores as the dependent variable. Because a panel dataset is used, the observations are not independently and uniquely distributed, so a simple OLS regression cannot be used to make a causal interpretation.

Subsequently, a fixed effects regression model is estimated to account for the panel dataset. Entity fixed effects are added to the model first, which controls for any unobserved characteristics that vary between counties but are fixed over time and corrects the standard errors for the presence of autocorrelation. This is shown below in the model as α_i.

Model (2)

$$lnGROCper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + \alpha_i + u_{it}$$
$$lnFFRper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + \alpha_i + u_{it}$$

In regression three, the time fixed effects are added as well, which is shown below as λ_t. All of the following models include time and entity fixed effects.

Model (3)

$$lnGROCper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + \alpha_i + \lambda_t + u_{it}$$
$$lnFFRper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + \alpha_i + \lambda_t + u_{it}$$

In regression four, the demographic control variables are added. This includes race, sex, age, and education variables, as shown below. In this model, the variables for less than a high school graduate, percent female, and over 60 years of age are omitted to prevent the occurrence of multicollinearity. In addition, the only included race category is *White*, therefore the other race categories will be in reference to the percentage of White residents in the county. In the following models, these demographic control variables will be referred to as θ_{it}.

Model (4)

$$lnGROCper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + \beta_\theta \theta_{it} + \alpha_i + \lambda_t + u_{it}$$
$$lnFFRper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + \beta_\theta \theta_{it} + \alpha_i + \lambda_t + u_{it}$$

Next, labor variables are added, along with the violent crime rate, as shown in the model below. In the following models, these variables will be referred to as κ_{it}.

Model (5)

$$lnGROCper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + \beta_\theta \theta_{it} + \beta_\kappa \kappa_{it} + \alpha_i + \lambda_t + u_{it}$$
$$lnFFRper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + \beta_\theta \theta_{it} + \beta_\kappa \kappa_{it} + \alpha_i + \lambda_t + u_{it}$$

In the next regression, the variables related to transportation were added. These include percent working the night shift, commute times, and the number of vehicles in each household. The variable for commute time under 30 minutes and the variable for households with three or more vehicles were omitted to prevent the

occurrence of multicollinearity. In the following model, these variables will be referred to as Ω_{it}.

Model (6)

$$lnGROCper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + \beta_\theta \theta_{it} + \beta_\kappa \kappa_{it} + \beta_\Omega \Omega_{it} + \alpha_i + \lambda_t + u_{it}$$

$$lnFFRper1000_{it} = \beta_0 + \beta_1 ln\_incpercap_{it} + \beta_\theta \theta_{it} + \beta_\kappa \kappa_{it} + \beta_\Omega \Omega_{it} + \alpha_i + \lambda_t + u_{it}$$

The final model consists of an instrumental variable regression using two-stage least squares. An instrumental variable regression technique is necessary to correct for the presence of reverse causality in the model. The presence of reverse causality in the model causes income to be correlated with omitted variables, creating bias in the estimates. The instrument, Z_{it}, is *lnpercentrelatives*, which is the natural log of the percentage of households in the county that have a relative outside of the immediate family residing in the household. The relevancy of this instrument will be demonstrated later, when it is shown that Z_{it} is correlated with income per capita. In addition, it is assumed that *lnpercentrelatives* is an exogenous instrument due to the fact that multigenerational households do not factor into grocery store and fast-food restaurant location decisions.[13] Therefore this percentage is correlated with the income per capita of a county, but it is not correlated with the error term in determining the number of grocery stores and fast-food restaurants in a county. Its use as an instrumental variable will eliminate bias in the model. The final model is shown below:

Model (7)

$$lnGROCper1000_{it} = \beta_0 + \beta_1 ln\_\widehat{incper}cap_{it} + \beta_\theta \theta_{it} + \beta_\kappa \kappa_{it} + \beta_\Omega \Omega_{it} + \alpha_i + \lambda_t + u_{it}$$

$$lnFFRper1000_{it} = \beta_0 + \beta_1 ln\_\widehat{incper}cap_{it} + \beta_\theta \theta_{it} + \beta_\kappa \kappa_{it} + \beta_\Omega \Omega_{it} + \alpha_i + \lambda_t + u_{it}$$

In the model shown above, $ln\_\widehat{incper}cap$ is the predicted income per capita after regressing on the instrumental variable, $lnpercentrelatives$.

[13] "More Kids Living in Multigenerational Families." *Poverty Solutions at the University of Michigan*, 25 Sept. 2018

Table II: Regression Results for the Effect of County Income per Capita on Grocery Stores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	lnGROCper1000	lnGROCper1000	lnGROCper1000	lnGROCper1000	lnGROCper1000	lnGROCper1000	lnGROCper1000
ln_incpercap	-0.004	-0.277***	0.028	-0.006	-0.040	-0.026	4.014***
	(0.030)	(0.042)	(0.067)	(0.069)	(0.088)	(0.093)	(1.280)
White				-0.013*	-0.011	-0.012	-0.001
				(0.008)	(0.008)	(0.009)	(0.001)
percentMale				-0.022	-0.023	-0.024	0.035*
				(0.014)	(0.016)	(0.017)	(0.020)
ln16to19				-0.014	-0.005	0.001	-0.080
				(0.020)	(0.022)	(0.023)	(0.051)
ln20to24				-0.043	-0.018	-0.018	-0.089
				(0.027)	(0.035)	(0.038)	(0.132)
ln25to44				-0.027	0.030	0.034	-1.852***
				(0.111)	(0.133)	(0.145)	(0.211)
ln45to54				0.036	0.068	0.073	-1.374***
				(0.083)	(0.098)	(0.103)	(0.303)
ln55to59				-0.051	-0.050	-0.040	-0.334**
				(0.047)	(0.052)	(0.054)	(0.144)
lnHS				-0.317***	-0.326***	-0.304***	0.607*
				(0.102)	(0.111)	(0.112)	(0.314)
lnsomecollege				-0.262***	-0.234**	-0.215**	-0.342***
				(0.087)	(0.096)	(0.093)	(0.114)
lnbachelors				-0.035	-0.035	-0.043	-0.680***
				(0.048)	(0.052)	(0.052)	(0.227)
lngraddegree				0.007	-0.007	-0.011	-0.346***
				(0.029)	(0.033)	(0.034)	(0.077)
lnUnemploymentRate					0.008	0.008	0.785***
					(0.028)	(0.029)	(0.268)
lnlaboroccupation					-0.036	-0.045	0.135
					(0.040)	(0.041)	(0.097)
lnviolentcrimeper1000					-0.004	-0.003	0.024*
					(0.007)	(0.007)	(0.012)
lnnightshift						-0.014	-0.138***
						(0.020)	(0.038)
lncommute30to59mins						-0.035	-0.068
						(0.045)	(0.042)
lncommuteover60mins						0.006	0.102**
						(0.025)	(0.044)
lnnovehicle						0.001	0.029
						(0.011)	(0.021)
ln1vehicle						0.020	0.242***
						(0.045)	(0.078)
ln2vehicles						0.013	-0.350***
						(0.072)	(0.132)
Years	11, 14, 16	11, 14, 16	11, 14, 16	11, 14, 16	11, 14, 16	11, 14, 16	11, 14, 16
State effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Time effects	No	No	Yes	Yes	Yes	Yes	Yes
Clustered standard errors	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,947	6,947	6,947	6,937	6,379	6,310	6,310
R-squared	0.884	0.884	0.885	0.885	0.878	0.876	0.876

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table III: Regression Results for the Effect of County Income per Capita on Fast-Food Restaurants

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	lnFFRper1000	lnFFRper1000	lnFFRper1000	lnFFRper1000	lnFFRper1000	lnFFRper1000	lnFFRper1000
ln_incpercap	0.353***	0.236***	0.022	0.040	0.015	0.020	-1.397
	(0.025)	(0.039)	(0.074)	(0.074)	(0.089)	(0.090)	(0.953)
White				0.006	0.002	0.002	0.003***
				(0.008)	(0.009)	(0.009)	(0.001)
percentMale				-0.006	-0.002	0.001	-0.044***
				(0.016)	(0.016)	(0.016)	(0.015)
ln16to19				0.027	0.014	0.010	0.032
				(0.020)	(0.021)	(0.021)	(0.041)
ln20to24				0.022	0.022	0.040	-0.084
				(0.044)	(0.053)	(0.055)	(0.094)
ln25to44				0.455***	0.364**	0.375**	0.687***
				(0.155)	(0.176)	(0.180)	(0.129)
ln45to54				0.234***	0.184*	0.189*	0.597***
				(0.089)	(0.102)	(0.104)	(0.223)
ln55to59				0.066	0.041	0.063	0.187**
				(0.047)	(0.051)	(0.051)	(0.090)
lnHS				0.014	0.025	0.027	-0.312
				(0.095)	(0.103)	(0.104)	(0.228)
lnsomecollege				0.016	0.055	0.051	0.206***
				(0.077)	(0.085)	(0.082)	(0.077)
lnbachelors				0.012	0.043	0.037	0.455***
				(0.048)	(0.050)	(0.050)	(0.166)
lngraddegree				0.016	0.040	0.040	0.178***
				(0.033)	(0.037)	(0.037)	(0.057)
lnUnemploymentRate					0.007	0.008	-0.306
					(0.028)	(0.028)	(0.199)
lnlaboroccupation					0.045	0.041	0.041
					(0.040)	(0.039)	(0.068)
lnviolentcrimeper1000					0.005	0.006	-0.004
					(0.007)	(0.007)	(0.008)
lnnightshift						-0.013	0.015
						(0.021)	(0.025)
lncommute30to59mins						-0.025	-0.278***
						(0.042)	(0.029)
lncommuteover60mins						0.021	-0.049
						(0.024)	(0.032)
lnnovehicle						-0.021*	0.052***
						(0.012)	(0.016)
ln1vehicle						0.011	0.366***
						(0.047)	(0.055)
ln2vehicles						-0.027	0.429***
						(0.072)	(0.100)
Years	11, 14, 16	11, 14, 16	11, 14, 16	11, 14, 16	11, 14, 16	11, 14, 16	11, 14, 16
State effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Time effects	No	No	Yes	Yes	Yes	Yes	Yes
Clustered standard error:	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,790	6,790	6,790	6,783	6,277	6,220	6,220
R-squared	0.848	0.848	0.848	0.849	0.857	0.858	0.858

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VI. Results

The tables above present the regression results for each of the dependent variables. Before discussing the results, it should be noted that several threats to internal validity were addressed in each of the models. First, omitted variable bias is minimized through the inclusion of control variables based economic theory. This issue is further accounted for through the use of a panel dataset, which allows for each entity to be observed more than once, thus minimizing omitted variable bias and heteroskedasticity. The use of an instrumental variable regression approach also controls for omitted variables that cannot be adequately controlled for, while also correcting for the issue of simultaneous causality bias. Functional form misspecification is also minimized in this paper, because tests were run on each of the control variables to determine whether the variable has a nonlinear relationship with each of the dependent variables and determine the best functional form of each variable.

Looking at the results for the model with dependent variable of grocery stores, which is found in Table II, the first column provides results for the regression that does not control for entity or time fixed effects and only includes the variable of interest, income per capita. In this specification, the coefficient is not significant and it is negative, which is contrary to the hypothesis. The coefficient remains negative but becomes significant at the 1% level when entity fixed effects are added in regression two. In the third specification, when time fixed effects are added, the coefficient becomes positive, as hypothesized, indicating that counties with a higher income per capita are associated with a higher number of grocery stores, but the coefficient is no longer significant. As the control variables are added in the fourth, fifth, and sixth regressions, the coefficient remains negative and insignificant.

In the final specification, where an instrumental variable is included in the model, the coefficient on income per capita is positive, as expected, and significant at the 1% level. In the first stage regression, which is presented in the appendix, the instrument is statistically significant at the 1% level. The coefficient demonstrates that a 1% increase in the income per capita of a county is associated with a 4.014% increase in the number of grocery stores per 1,000 residents in the county, ceteris paribus.

In this model, it is also important to note that several of the demographic control variables are significant. The variable *percentMale* is significant at the 10% level and can be interpreted as meaning that a one percentage point increase in the percentage of Males in the county is associated with a 3.5% increase in the number of grocery stores per 1000, ceteris paribus. Three of the age categories were significant as well: 25 to 44, 45 to 54, and 55 to 59. The first two variables were significant at the 1% level, while the category 55 to 59 was significant at the 5% level. This indicates that a 1% increase in the county population of citizens aged 25

to 44 is associated with a 1.852% decrease in grocery stores. A 1% increase in residents aged 45 to 54 is associated with a 1.374% decrease in grocery stores, and a 1% increase in residents aged 55 to 59 is associated with a 0.334% decrease in grocery stores.

The variables *somecollege*, *bachelors,* and *graddegree* were significant as well in this model at the 1% level. The variable *somecollege* indicates that a 1% increase in the county residents who attended some college is associated with a 0.342% decrease in the number of grocery stores per 1000 residents. A 1% increase in the number of county residents who hold a Bachelor's degree is associated with a 0.68% decrease in the number of grocery stores per 1000 residents. A 1% increase in the number of county residents who have a graduate degree was associated with a 0.346% decrease in the number of grocery stores per 1000 residents. The unemployment rate was also significant at the 1% level. A 1% increase in the unemployment rate is associated with a 0.785% increase in the number of grocery stores per 1000 residents.

Finally, looking at the transportation variables, several were significant as well. The variable *lnnightshift* is significant at the 1% level, and indicates that a 1% increase in the percentage of county residents who work the night shift is associated with a .138% decrease in the number of grocery stores. The variable for commutes over 60 minutes was also significant at the 1% level. This variable indicates that if the percentage of county residents who have a commute over 60 minutes increases by 1%, the number of grocery stores per 1000 residents will increase by 0.102%. In addition, the variable *ln1vehicle* was significant at the 1% level and indicated that a 1% increase in households with one car is associated with a 0.242% increase in the number of grocery stores per 1000 residents. A 1% increase in the households with two cars was found to be associated with a decrease of grocery stores by 0.350%, which is significant at the 1% level.

Looking at the dependent variable, fast-food restaurants, which is found in Table III, the first column provides results for the regression that does not control for entity or time fixed effects and only includes the variable of interest, income per capita. In this specification, the coefficient is significant at the 1% level, but it is positive, which is contrary to the hypothesis. The coefficient remains significant and positive when entity fixed effects are added in regression two. In the third specification, when time fixed effects are added, the coefficient remains negative but it is no longer significant. As the control variables are added in the fourth, fifth, and sixth regressions, the coefficient remains insignificant but becomes positive.

In the final specification, where an instrumental variable is included in the model, the coefficient on income per capita is negative, as expected, but it is not significant. The insignificance could indicate that there is no relationship between county income per capita and the number of fast-food restaurants in the area, but

this would be contrary to economic theory. The insignificance is likely evidence of bias in the coefficient which will be discussed further.

Similar to the model with the dependent variable, grocery stores, several of the demographic control variables are significant. The variable *White* is significant at the 1% level and indicates that a one percentage point increase in the percentage of White county residents is associated with a 0.3% increase in the number of fast-food restaurants per 1000 residents. The variable *percentMale* is significant at the 1% level, and means that a one percentage point increase in the percentage of Males in the county is associated with a 4.4% decrease in the number of fast-food restaurants. Three of the age categories were significant as well: 25 to 44, 45 to 54, and 55 to 59, which were also significant in the model using grocery stores as the dependent variable. The category, 25 to 44, is significant at the 1% level, and indicates that a 1% increase in the county population of citizens aged 25 to 44 is associated with a 0.687% increase in the number of fast-food restaurants. A 1% increase in residents aged 45 to 54 is associated with a 0.597 % increase in fast-food restaurants, which is significant at the 5% level, and a 1% increase in residents aged 55 to 59 is associated with a 0.187% increase in fast-food restaurants.

Several of the education demographic variables were significant as well. The variable *lnsomecollege* was significant at the 1% level and indicates that a 1% increase in the number of county residents who completed some college is associated with a 0.206% increase in the number of fast-food restaurants. In addition, the variable *lnbachelors* is significant at the 1% level and demonstrates that a 1% increase in the county residents who obtained a bachelor’s degree is associated with a 0.455% increase in the number of fast-food restaurants. The variable *lngraddegree* was also significant at the 1% level, indicating that a 1% increase in the county residents who have a graduate degree is associated with a 0.178% increase in the number of fast-food restaurants.

The transportation variables in this model were also very significant. The variable *lncommute30to59mins* is significant at the 1% level and indicates that when the percentage of county residents with a commute between 30 and 59 minutes increases by 1%, the number of fast-food restaurants per 1000 residents decreases by 0.278%. All three of the variables related to the number of vehicles per household are significant at the 1% level. The variable *lnnovehicle* demonstrates that a 1% increase in the county residents that do not have a vehicle is associated with 0.052% increase in fast-food restaurants. A 1% increase in the county residents that have one vehicle per household was associated with an increase of fast-food restaurants by 0.366%. Finally, an increase in the county residents that have two vehicles per household is associated with an increase of fast-food restaurants by 0.429%.

It is surprising that no relationship was found between income and fast-food restaurants, which likely indicates that there is bias in this model. One factor that

fast-food restaurants take into account when choosing a location that was not able to be controlled for with this data is the amount of public transportation available in the area. Public transportation use is an important indicator of the type of population in the area and also can predict the foot traffic an area will receive. Similarly, this model was unable to include the effects of traffic volume in each area, which also significantly impacts the desirability of an area to fast-food franchises. Another aspect that could be causing bias is the average children per household or similarly, the accessibility and affordability of childcare in the area. Both of these variables would be important considerations for fast-food franchises, as patrons of fast-food restaurants often choose to purchase fast-food because of its convenience and its child-friendly menu. These variables were not controlled for in this model and could be potential sources of bias.

The findings indicate that there is a significant relationship between income and grocery stores. A 1% increase in the income per capita of a county was found to be associated with a 4.014% increase in the number of grocery stores per 1,000 residents in the county, ceteris paribus. The number of additional grocery stores per county as a result of increasing income per capita depends on both how many grocery stores already exist in the county, as well as the population of the county. The richest 10% of counties have an average income per capita of over $50,140 and the average number of grocery stores in these counties is 56.54. Therefore, an increase of income per capita by 1% which results in a 4.014% increase in grocery stores, causes the number of grocery stores in these counties to increase by 2.27 stores. The poorest 10% of counties have an average income per capita of $28,708 or less and the average number of grocery stores in these counties is 4.99. Therefore, an increase of income per capita by 1% which results in a 4.014% increase in grocery stores, causes the number of grocery stores in these counties to increase by 0.2 and an increase of income per capita by 5% results in additional grocery store for these poorest counties. This is very significant, as it demonstrates that only a 5% increase in income can have real impacts on the food scarcity in poorer areas.

VII. Conclusion

This paper uses a unique approach to examine the relationship between the average income of a county and the prevalence of fast-food restaurants and the lack of grocery stores in the area. Unlike previous work, this study uses county-level data across the entire United States, spanning a time period of five years. The use of a panel dataset allows for the inclusion of the impact of entity and time fixed effects and the use of an instrumental variable controls for simultaneous causality bias, which is something that previous studies have not controlled for.

The results, overall, suggest that the income per capita of a county significantly impacts the number of grocery stores in the area. It should also be noted that

education, age, and attributes regarding time constraints played an important role in determining the number of grocery stores and fast-food franchises as well.

In the final specification, the effect of income per capita was highly significant in regards to the number of grocery stores in a county: a 1% increase in the income per capita of a county is associated with a 4.014% increase in the number of grocery stores per 1,000 residents in the county, ceteris paribus. This was the expected effect, and it was significant at the 1% level. The magnitude of this effect is relatively large as well; an increase of income per capita by 5% results in an additional grocery store for the poorest 10% of counties, and the number grows larger as the income of the county increases.

It is evident that income per capita does play a significant role in the formation of food deserts, as areas that are more wealthy have significantly more grocery stores and healthy food options, while less wealthy areas have more fast-food restaurants and less nutritious options. These findings have significant policy implications, because access to healthy foods is a necessity and it is evident that certain communities are disproportionately affected by a lack of healthy options.

APPENDIX
Table IV – First Stage Regression for Instrumental Variable Estimation
Dependent Variable: *ln_incpercap*

Variable	Coefficient	Standard Error
lnpercentrelatives	-0.0335735	0.0085027
White	-0.0004308	0.0002681
percentMale	-0.014989	0.001357
ln16to19	-0.020156	0.0080474
ln20to24	-0.0888397	0.0152572
ln25to44	0.1005576	0.0445329
ln45to54	.01931055	0.0431534
ln55to59	0.0760375	0.0234823
lnHS	-0.2290141	0.0310456
lnsomecollege	-0.054482	0.0232654
lnbachelors	0.1613488	0.0140484
lngraddegree	0.0427532	0.0115053
lnUnemploymentRate	-0.198877	0.0117127
lnlaboroccupation	0.0520741	0.0125624
lnviolentcrimeper1000	-0.0002841	0.0024134
lnnightshift	0.0145383	0.0068783
lncommute30to59mins	-0.0112789	0.0082353
lncommuteover60mins	-0.0255301	0.0058927
lnnovehicle	-0.0040058	0.0044041
ln1vehicle	-0.0222841	0.0140405
ln2vehicles	0.0346976	0.0255189

Table V – List of Variables and Descriptions

Variable	Description
FIPS	The Federal Information Processing Standard (FIPS) code is a five-digit code that uniquely identified counties and county equivalents in the United States
lnGROCper1000	The natural log of the number of grocery stores in the county per 1000 residents
lnFFRper1000	The natural log of the number of fast-food restaurants in the county per 1000 residents
Year	The year of the observation
ln_incpercap	The natural log of the county income per capita
lnUnemploymentRate	The natural log of the county unemployment rate
White	The percentage of White residents in the county
Black	The percentage of Black residents in the county
Hispanic	The percentage of Hispanic residents in the county
Asian	The percentage of Asian residents in the county
American_Indian	The percentage of America Indian residents in the county
percentMale	The percentage of Male residents in the county

percentFemale	The percentage of Female residents in the county
ln16to19	The natural log of the percentage of residents in the county between the ages of 16 and 19
ln20to24	The natural log of the percentage of residents in the county between the ages of 20 and 24
ln25to44	The natural log of the percentage of residents in the county between the ages of 25 and 44
ln45to54	The natural log of the percentage of residents in the county between the ages of 45 and 54
ln55to59	The natural log of the percentage of residents in the county between the ages of 55 and 59
ln60over	The natural log of the percentage of residents in the county over the age of 60
lnlaboroccupation	The natural log of the percentage of residents in the county who work a labor occupation, including agriculture, construction, and maintenance
lnnightshift	The natural log of the percentage of residents in the county who work the night shift and leave for work between the hours of 12:00 a.m. to 4:59 a.m.
lnnonenglish	The natural log of the percentage of residents in the county who speak a first language other than English

lncommuteunder30mins	The natural log of the percentage of residents in the county who have a commute under 30 minutes
lncommute30to59mins	The natural log of the percentage of residents in the county who have a commute between 30 and 59 minutes
lncommuteover60mins	The natural log of the percentage of residents in the county who have a commute over 60 minutes
lnnovehicle	The natural log of the percentage of residents in the county who do not have a vehicle
ln1vehicle	The natural log of the percentage of residents in the county who have one vehicle in the household
ln2vehicles	The natural log of the percentage of residents in the county who have two vehicles in the household
ln3ormorevehicles	The natural log of the percentage of residents in the county who have three or more vehicles in the household
lnviolentcrimeper1000	The natural log of the number of violent crimes per 1000 residents in the county
lnlessthanHS	The natural log of the percentage of county residents who have less than a high school degree
lnHS	The natural log of the percentage of county residents who graduated high school or achieved equivalency

lnsomecollege	The natural log of the percentage of county residents who attended some college or received an associate's degree
lnbachelors	The natural log of the percentage of county residents who received a Bachelor's degree
lngraddegree	The natural log of the percentage of county residents who obtained a graduate or professional degree

Bibliography

Alviola, P. A., Nayga, R. M., Thomsen, M. R., & Wang, Z. (2013). Determinants of Food Deserts. *American Journal of Agricultural Economics*, *95*(5), 1259–1265. http://www.jstor.org/stable/24476908

Bader, M. D. M., Purciel, M., Yousefzadeh, P., & Neckerman, K. M. (2010). Disparities in Neighborhood Food Environments: Implications of Measurement Strategies. *Economic Geography*, *86*(4), 409–430. http://www.jstor.org/stable/40929682

Biegert, T. (2017). Welfare Benefits and Unemployment in Affluent Democracies: The Moderating Role of the Institutional Insider/Outsider Divide. *American Sociological Review*, *82*(5), 1037–1064. https://doi.org/10.1177/0003122417727095

Card, D., Johnston, A., Leung, P., Mas, A., & Pei, Z. (2015). The Effect of Unemployment Benefits on the Duration of Unemployment Insurance Receipt: New Evidence from a Regression Kink Design in Missouri, 2003-2013. *The American Economic Review*, *105*(5), 126–130. http://www.jstor.org/stable/43821864

“Food Deserts: Definition, Effects, and Solutions.” *Medical News Today*, https://www.medicalnewstoday.com/articles/what-are-food-deserts#health-impact.

Graddy, K. (1997). Do Fast-Food Chains Price Discriminate on the Race and Income Characteristics of an Area? *Journal of Business & Economic Statistics*, *15*(4), 391–401. https://doi.org/10.2307/1392485

Jekanowski, M. D., Binkley, J. K., & Eales, J. (2001). Convenience, Accessibility, and the Demand for Fast Food. *Journal of Agricultural and Resource Economics*, *26*(1), 58–74. http://www.jstor.org/stable/40987095

Kashino, Marissa. “How Whole Foods Decides If Your Neighborhood Is Worthy.” *The Washingtonian*, 14 July 2015, https://www.washingtonian.com/2015/07/14/how-whole-foods-decides-if-your-neighborhood-is-worthy/.

Loria, Keith. “How Do Grocery Stores Find the Right Location for Expansion?” *Grocery Dive*, 17 Feb. 2017

“More Kids Living in Multigenerational Families.” *Poverty Solutions at the University of Michigan*, 25 Sept. 2018,

https://poverty.umich.edu/2018/09/25/more-kids-living-in-multigenerational-families/.

"Profit Maximization Model of a Firm." *Economics Discussion*, https://www.economicsdiscussion.net/profit/profit-maximization-model-of-a-firm-with-diagram/6129.

Ungerleider, Neal. "How Fast Food Chains Pick Their Next Location." *Fast Company*, 25 Aug. 2014, https://www.fastcompany.com/3034792/how-fast-food-chains-pick-their-next-location.

Weatherspoon, D., Oehmke, J., Dembele, A., & Weatherspoon, L. (2015). Fresh vegetable demand behaviour in an urban food desert. *Urban Studies*, *52*(5), 960–979. https://www.jstor.org/stable/26146023