THE IMPACT OF WAL-MART ON THE RURAL RETAIL WAGE

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

JESSICA ELAINE SMITH

Dr. Judith Stallmann, Thesis Supervisor

DECEMBER 2010

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

THE IMPACT OF WAL-MART ON THE RURAL RETAIL WAGE

presented by Jessica Elaine Smith,

a candidate for the degree of master of science,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Judith Stallmann

Professor Georgeanne Artz

Professor Anna Ball

DEDICATION

This thesis is dedicated to my siblings, Todd and Elizabeth, who have always provided me with a source of laughter, joy, support and good food. I would also like to extend thanks to my dear friend Stacey Sachs; who has always been there when I really need her. To all my friends who are also part of my family, thank you.

ACKNOWLEDGEMENTS

I would like to extend my most gracious thanks to Judith Stallmann for challenging me and providing great insight and contributions to my thesis. I feel a sense of pride having worked under her careful supervision. The same can be said of Georgeanne Artz who has a great sense of humor and the utmost patience to explain complex concepts to her students. I would like to give a special thanks to Anna Ball for being the outside member of my committee as her input was invaluable. Finally, I would like to thank Ray Bacon for all of his assistance with SAS.

TABLE OF CONTENTS

LIST OF EQUATIONS

EQUATION PAGE

LIST OF FIGURES

FIGURE PAGE

LIST OF TABLES

TABLE PAGE

ABSTRACT

The study uses panel data to test the impact of Wal-Mart on the rural retail wage. There are observations from 2,986 counties in the contiguous United States from 1990-2005. Previous studies have reported mixed findings of the impact of Wal-Mart on rural wages; the impact of Wal-Mart on rural wages is indeterminate. Wal-Mart could increase rural retail wages if it is a large labor demander relative to labor supply, if it pays higher wages than other rural retailers, and/or if it hires workers for more hours per week than other retailers (because it is open more hours). Wal-Mart could lower rural retail wages if it hires workers for fewer hours than other rural retailers, it hires a different skill mix (fewer managers per worker), and/or if it pays higher benefits that compensate for wages, and/or if it is a monopsonist. The Heckman two-step procedure is the empirical specification used to control for the endogeneity of Wal-Mart's location decision. A modified first difference model is used to test the impact of Wal-Mart on rural retail wages. The study finds that Wal-Mart selects counties where the retail wage is growing slowly. The study finds that rural retail wages grow more slowly than urban wages. While the presence of Wal-Mart slows the growth of the average weekly retail wage overall, in rural counties it increases the growth of average weekly retail wage. With the given data set the reason for the increase in the growth of the rural retail wage could not be addressed.

Chapter 1: Introduction

1.1 Background Information

In order to begin a discussion of Wal-Mart's impact on wages, employment and earnings, it is relevant to review a few statistics on the magnitude of the company's presence in the labor market. In 2008, Wal-Mart alone employed ten percent of the U.S. retail sector, and one percent of total U.S. employment (Neumark, Zhang, and Ciccarella 2008). It is the largest grocer in the U.S. and the third largest pharmacy (Neumark, Zhang, and Ciccarella 2008). In 2010, Wal-Mart employed 1.4 million U.S. employees and another 660,000 workers worldwide (Wal-Mart Stores, Inc. 2010). In addition to having a large presence in retail, Wal-Mart is highly efficient as compared to other "big box" stores and small retail stores. Wal-Mart has advantages of scale and access to capital markets over small retailers; also superior logistics, distribution and inventory control advantages over large retailers (as cited by Basker 2005a: Ghemawat 1989; Foley and Mahmood 1996; Ghemwat and Friedman 1999). Wal-Mart uses these advantages to reduce prices for consumers.

Wal-Mart's size alone generates interest in its impact on wages. According to Basker, "88 percent of Americans live within 15 miles of a Wal-Mart store, and 67 percent of retailers are located within 5 miles of a Wal-Mart store" (Basker 2007a, p.178). Wal-Mart has a significant presence in American communities. It also is seen as a low wage, low benefit employer that is a destroyer of small business (Dube, Eidlin, and Lester 2007). However, there is little empirical evidence to support the reputation.

1.2 Differences in Rural-urban Wage Impacts

Despite the popular opinion of Wal-Mart's practices as an employer, both rural and urban areas have begun to recruit mass retailers as a development strategy, with three main goals: 1) To attract sales tax revenue from local option sales taxes; 2) To bring more sales outlets into the community and entice community members to buy local; 3) To provide local employment on a large scale to utilize unemployment excess capacity (Artz and Stallmann 2006). Using a mass retailer, such as Wal-Mart, as a development strategy will not be discussed in-depth here as the following study is focused on labor effects. However, the strategy of using Wal-Mart to provide employment and wages in rural areas brings up a point stressed by Bartik (2005), if low-paying jobs are brought into an area the community may lose the benefit it expected to gain by utilizing excess labor capacity. He estimates that in metropolitan areas if jobs are providing wages at 10-15% below the industry average, which he calls the "industry premium", then few benefits, will result for the community recruiting mass retailers (Artz and Stallmann 2006, Bartik 2004). The negative effect of being below the industry average occurs because the workers are receiving low wages relative to their skills and credentials (Bartik 2004). The symptoms of providing low wage jobs are as follows: an increase or no reduction in transfer (welfare) payments, little or no increase in tax revenue, and little or no increase in property values (Bartik 2004). Stallman and Artz (2006) note that in rural areas Wal-Mart is often paying more than small retailers, and Wal-Mart distribution centers are more likely to locate in rural counties near an interstate than in urban areas (distribution center average hourly wage is $10/hour). The researchers, Stallmann and Artz, also note that the definition of a "good job" often depends on local economic conditions.

Rural areas, which start off with lower wages, may not experience the same downward pull from Wal-Mart's competition; wage conditions may actually improve. Dube, Eidlin, and Lester (2007) note that in rural areas workers are closer to the minimum wage rate. Thus, Wal-mart is not likely to have a negative impact on wages, i.e. has less of a scope for changing the wage. Alternatively, because in rural areas there is less competition from other large retailers Wal-Mart may have significantly larger effects in these counties (Neumark, Zhang and Ciccarella 2008). Dube, Eidlin, and Lester (2007) suggest that because rural areas are not as labor union dense as urban areas, there are fewer higher paying union jobs to be replaced with lower non-union jobs from Wal-Mart. Also there would be less negotiation for a higher wage in rural low union density areas. In summary, rural areas feature different labor market characteristics than urban areas, and are likely to experience different impacts of Wal-Mart's presence on labor market effects.

1.3 Research Objectives

The research question is given that rural counties have a small labor supply, lower wages, less competition, and lower union density, what is the impact of Wal-Mart on the rural retail wage? The objective of the study is to test and estimate the impact of Wal-Mart on rural average weekly retail wages. The hypothesis is that when Wal-Mart is a large labor demander relative to the labor market, it may raise average retail wages. Wal-Mart raises rural retail wages because it is entering counties that have a low wage rate and a small, but competitive labor market. A study that focuses specifically on the differences in the impact of Wal-Mart's presence on the rural retail wage should provide rural policy makers and concerned citizens with information specific to rural economies. The study uses a 'Heckit' and first differences empirical specification on a panel data set of average weekly retail wage data from 2,986 contiguous U.S. counties from 1990 to 2005.

Chapter 2: Literature Review

2.1 Analytical Framework

In competitive labor markets no one firm has the market power to reduce the average wage. Firms in competitive markets are price takers for wages (Perloff 2008). However, some recent research has found that Wal-Mart lowers the overall average retail wage, lowers the wage rate in the general merchandise and grocery subsectors, and lowers the total retail wage bill in Metropolitan Statistical Area[1] (MSA) counties (Dube, Eidlin, and Lester 2007). Neumark, Zhang, and Ciccarella (2008) find that the per capita retail payrolls (a measure of payrolls, not wages) are lowered, but increase in the general merchandise subsector. Neumark, Zhang, and Ciccarella (2008) look at retail employment in counties with a population above 21,000 and below 21,000 and find mixed results. In the 2008 article, the researchers find significant negative effects on employment for counties with population above 21,000; while insignificant results on employment are found for counties with population below 21,000. Wal-Mart's ability to lower the wage rate and total payrolls per person in urban areas is puzzling, because in a competitive market, Wal-Mart would pay the 'going' wage upon entry.

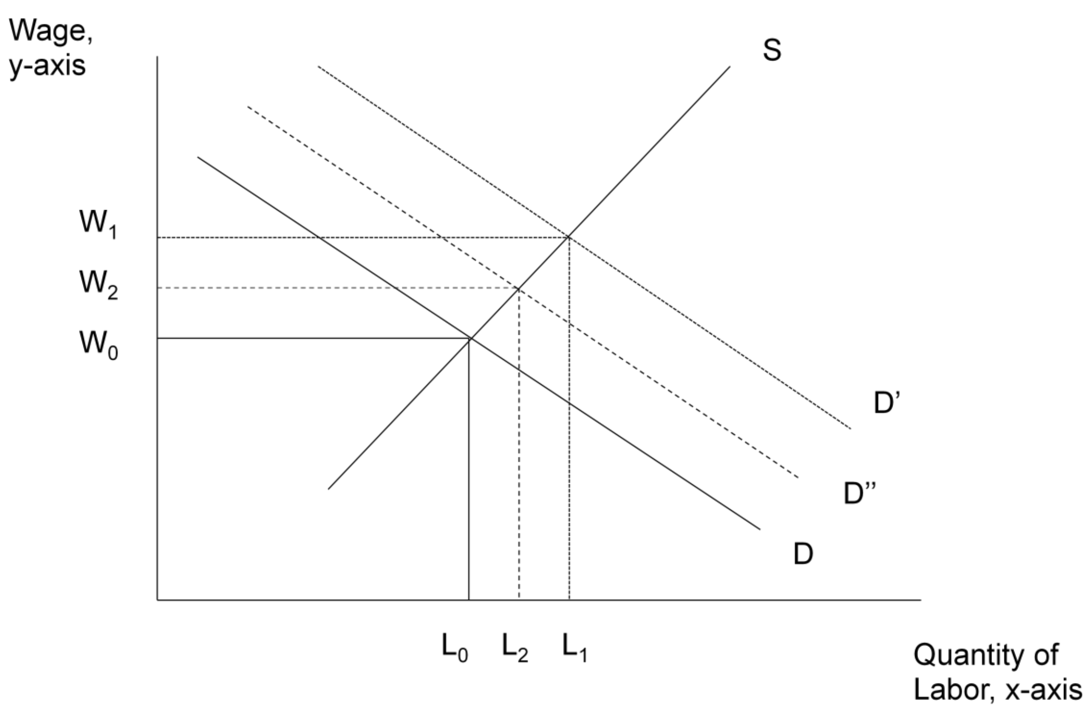
Wal-Mart employs a larger percentage of the labor supply in rural areas than in urban areas. In a rural labor market where the wage market is thin Wal-Mart might be expected to affect the average retail wage. Figure 1.1 presents a framework for the empirical analysis. The model assumptions are that there is no unemployment and the market is competitive. Initially the market is in equilibrium where labor demanded equals labor supplied, thus quantity of labor equals L_0 and wage equals W_0. When Wal-Mart first enters the labor market demand for labor shifts to D', thus quantity of labor demanded shifts to L_1 and the wage increases to W_1. The demand shifts because Wal-

[1] According to the U.S. Census Bureau a 1999 metropolitan statistical area is defined as "A metro area contains a core urban area of 50,000 or more population", else it is defined as non-MSA (U.S. Census Website, Metropolitan and Micropolitan Statistical Areas, 2010).

.

Mart is a large labor demander compared to smaller retailers in the market (Basker 2005a; Dube, Eidlin, and Lester 2007). Later, in the long run some businesses may lose jobs or go out of business (Basker 2005a). So the demand for labor decreases to D'', while the quantity of labor demanded shifts to L_2. Therefore, the new equilibrium wage ends up between W_1 and W_0 at W_2[2]. The larger the increase in the initial labor demand the larger the positive impact a Wal-Mart store will have on the retail wage. It is helpful to note that Wal-Mart stores vary in size and the amount of workers that they employ[3].

Figure 1.1 Wal-Mart Increases the Rural Retail Wage

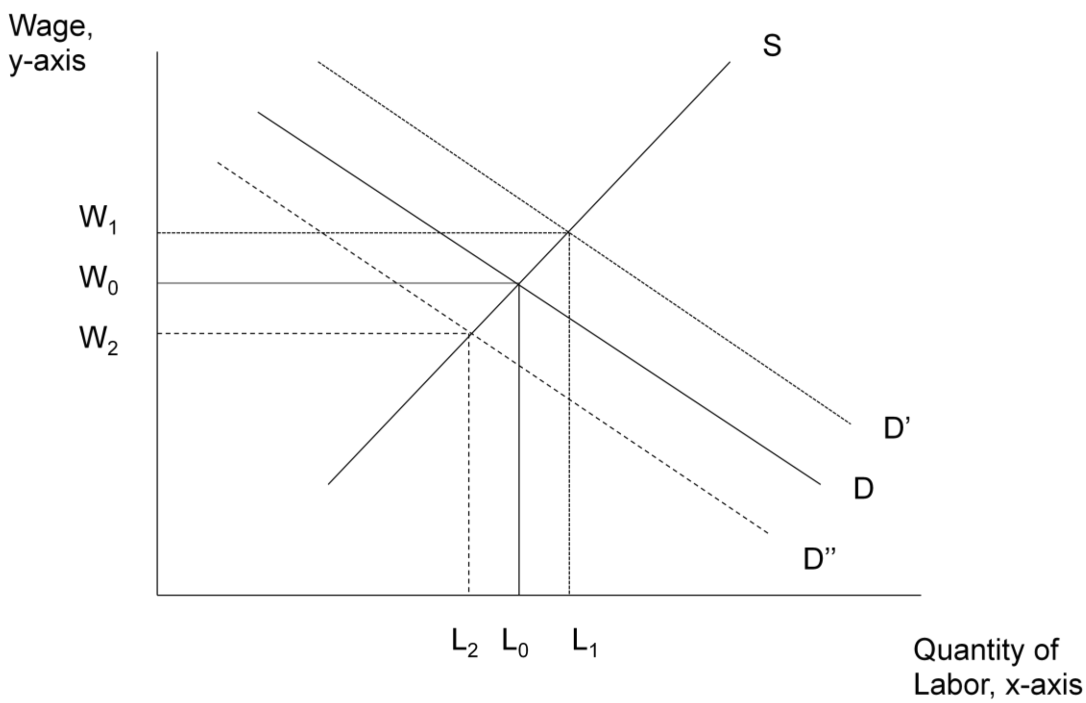


[2] Notes and email correspondence with Dr. Georgeanne Artz. May 22, 2010.

[3] Discount stores employ approximately 225 employees, supercenters approximately 350 employees, neighborhood markets approximately 95 employees (Wal-Mart Stores, Inc. 2010).

In the long-run case described by Basker (2005a) where Wal-Mart puts some establishments out of business there is an alternate scenario. As shown in figure 1.2, if the labor demanded is lower than the initial equilibrium, L_2. Wal-Mart presence could lower the rural retail wage, W_2.

Figure 1.2 Wal-Mart Decreases the Rural Retail Wage

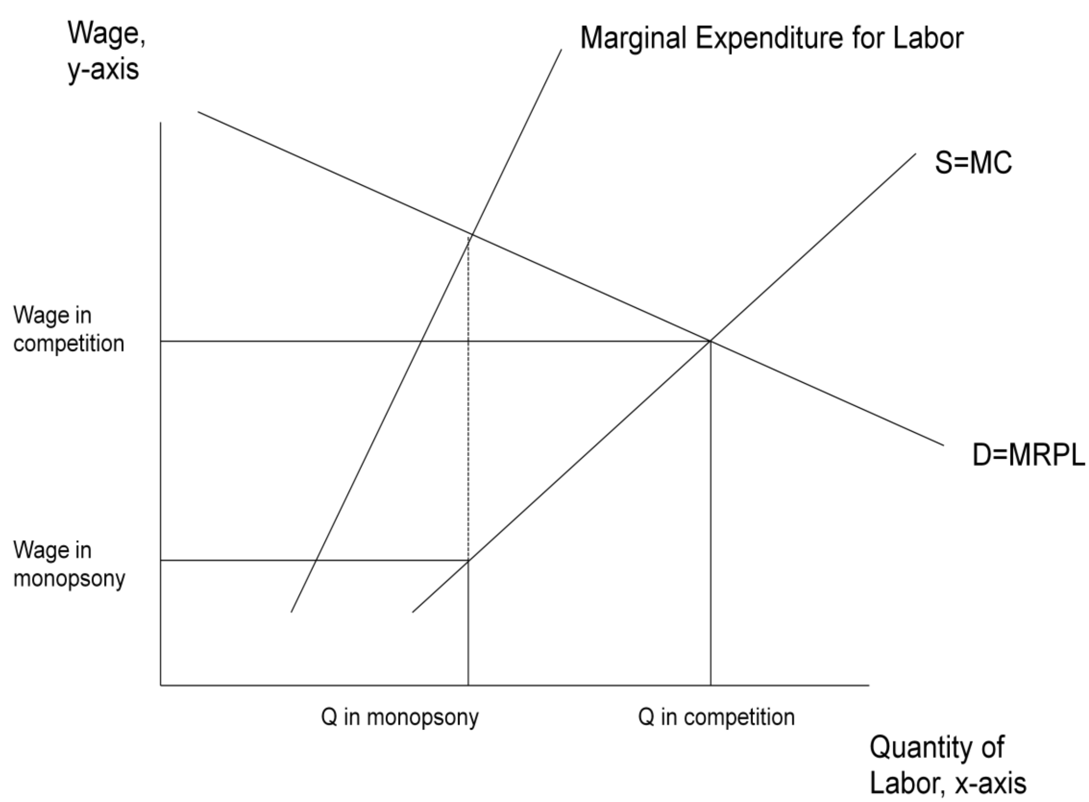


Wal-Mart presence might also lower average retail wages in rural counties if, relative to existing retailers, it does one of the following: 1) it hires more part-time labor compared to full-time labor; 2) it hires proportionately few managers per hourly workers; and or 3) it pays benefits to offset the wage. Situations one and two would lower the average wage due to the method the Bureau of Labor Statistics uses to calculate average wages in the Quarterly Census of Employment and Wages (QCEW).

The Bureau divides the total annual wages by the number of jobs not distinguishing by job type (part-time vs. full-time or managerial vs. hourly). Thus, an increase in part-time and or hourly workers would reduce the average wage per job. The Bureau of Labor Statistics does not consider the benefits that workers are receiving in the QCEW. Benefits can entice workers to lower paying jobs because they offset the lower wage. Note that figure 1.2 cannot be used to analyze average wages, as it shows marginal wages.

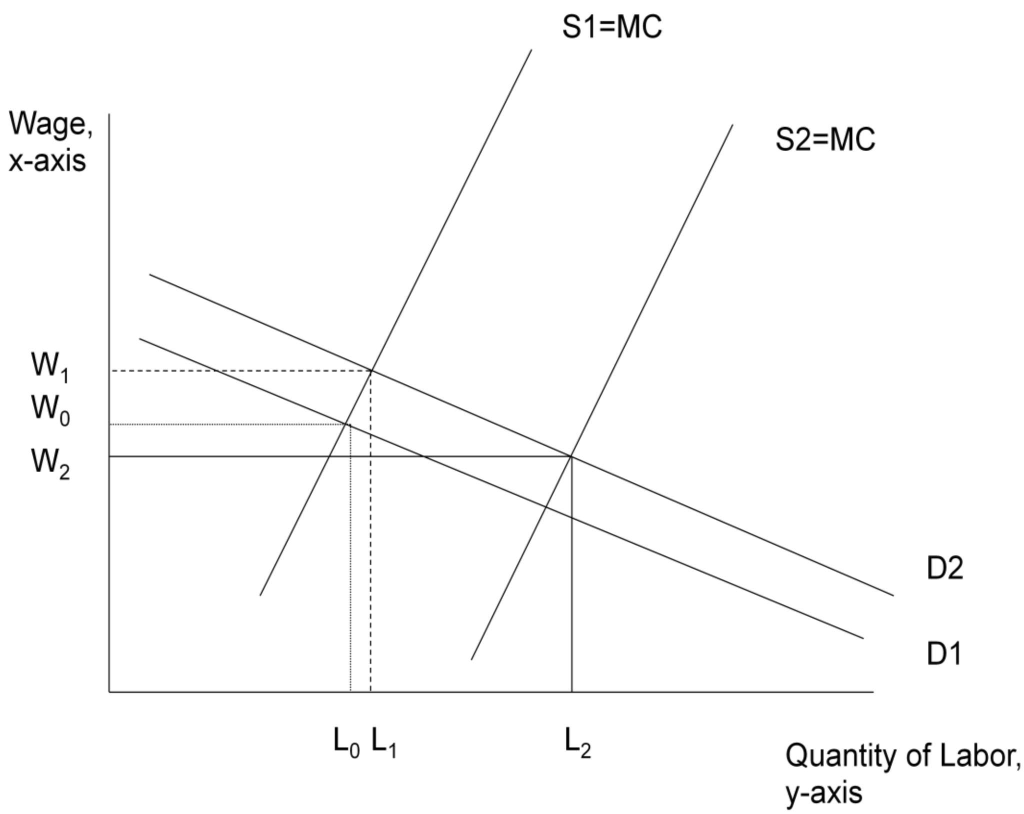
An additional situation where Wal-Mart can reduce wages is if Wal-Mart has monopsony power (Bonanno and Lopez 2008). If Wal-Mart is not a competitive player, but rather a monopsony, i.e. a single demander in the labor market then Wal-Mart would be a price-setter for wages (Perloff 2008, Bonanno and Lopez 2008). The monopsonist faces the standard supply or marginal cost of labor. Each additional unit of labor has a higher marginal cost as shown by the upward sloping curve. Because it is a monopsonist it must pay the marginal cost of labor not just to the marginal unit of labor, but to all units of labor that it hires. That means its marginal expenditure per unit of labor will be higher than the marginal cost of the unit. As shown in figure 1.3, Wal-Mart would be expected to set its marginal expenditure of labor equal to the marginal revenue product of labor (MRPL). Marginal revenue product of labor is also the firm's demand for labor. Wal-Mart would then hire the quantity of labor at the point on the supply curve underneath the intersection of the marginal expenditure of labor and marginal revenue product of labor curves. While Wal-Mart is not likely to be a true monopsonist in the retail sector or overall, if it has a large enough degree of monopsony power it could decrease the wages by some amount. Although, a positive impact on rural retail wages would not completely refute a monopsony model; it may show that Wal-Mart is a large firm relative to local labor markets so that it may increase rural retail wages.

Figure 1.3 Wal-Mart as a Monopsony



Finally, when the assumption that there is low to no unemployment is relaxed then the outcome of the model could change, as shown in figure 1.4., the new low cost employer enters the market, increasing demand. The unemployed would enter the market so more labor would be supplied, shifting the initial supply curve S1 to S2. If the supply response (S2) is larger than the increase in demand (D2) for the quantity of labor, then quantity demanded shifts to (L_2) and wage shifts downward from W_0 to W_2. The points L_1 and W_1, where the quantity of labor and wage increase when Wal-Mart increases the demand for labor (shown as D2) will not be noticeable in this situation due to the shift of the supply curve due to the unemployed entering the labor market.

Figure 1.4 High Unemployment Scenario



2.2 Wal-Mart's Impact on Retail Sales

The early Wal-Mart studies had a regional focus on retail sales and establishments. Stone (1988) used Iowa Retail Sales and Tax Reports to measure the affect of Wal-Mart on local general merchandise retails sales. He measured the change in retail pull factors for general merchandise and total retail sales as a result of Wal-Mart entry in host towns. Pull factors are defined as the ratio of per capita sales in a town to the per capita sales at the state level adjusted for income levels. Stone provided updated studies every two years through 1997 and each of these studies report generally the same findings. According to Stone (1988), the following merchandise groups show increases after the opening of a Wal-Mart store: general merchandise (includes Wal-Mart); eating and drinking; home furnishings, and total sales. Stone (1988) notes that pull factors for grocery, building materials, apparel, and specialty stores decline. The small towns that neighbor Wal-Mart host towns suffer a greater loss in total retail sales than towns that are farther away from Wal-Mart host towns.

Stone, Artz, and Myles (2002) also used a pull factor analysis with data from Mississippi sales tax reports comparing before and after effects of Wal-Mart Superstore entry on general merchandise retail sales. They find that general merchandise sales in counties hosting a Wal-Mart increased 40 percent during the first year of firm entry and peak at 59 percent of the base year in year five. They corroborate the Stone (1988) findings showing that non-host Wal-Mart counties in Mississippi lost sales after Wal-

Mart entry as did competing sectors such as grocery, building materials. Services, such as laundry, dry cleaners, beauty shops, auto-repair, electrical repair, and furniture repair shops also lost sales. Stone, Artz and Myles suggest that it may be more economical for consumers to buy new items at Wal-Mart than to have them repaired.

Artz and McConnon (2001) documented the percentage change in general merchandise sales in Maine counties and also find an overall increase in general merchandise sales for Wal-Mart host counties. The authors find a decrease for non-host counties, as well.

2.3 Wal-Mart's Impact on Retail Prices

Volpe and Lavoie (2007) examined the differences in prices of a market basket of 30 common consumer goods between Wal-Mart Supercenters and local groceries. The researchers estimated competitive effects of Wal-Mart on grocery stores in direct competition with the retailer using a unique data set of prices collected off the shelf in New England states (Connecticut, Massachusetts, and Rhode Island). Data were collected from 18 stores in a three week period (during October 2004) to avoid time trends. The stores were of three types: 1) Wal-Marts; 2) Large supermarkets within five miles of the Wal-Mart stores, considered "competing stores"; 3) Stores not in direct competition with Wal-Mart, which are used as "comparison stores". Products were organized into different categories: grocery, dairy, frozen food, health and beauty aids, meat and produce. These categories were then divided into national and private label brands. Volpe and Lavoie find that for all product categories Supercenters have the lowest price indexes and comparison stores have the highest. The researchers find that Wal-Mart Supercenters reduce national brand prices and private label prices by 6-7% and 3-8%, respectively. They conclude that Wal-Mart has a positive welfare effect on price-sensitive consumers, but note that all of Wal-Mart effects should be considered on "customers, employees, competitors, and distributors" (Volpe and Lavoie 2007, p.19).

Basker (2005b) estimated the price effect of a low-cost entrant on retail prices using the opening dates of Wal-Mart stores and average quarterly retail prices of 10 common goods from 1982-2002. The price data used were at the city level and were from the American Chamber of Commerce Research Association. Basker notes, "In most models of imperfect competition, entry of a lower-cost competitor reduces output prices. The effect is larger the smaller the initial number of firms and the higher the cross price elasticities." (Basker 2005b, p. 204) Basker also tested the idea that Wal-Mart initially reduces prices, but increases them later when the firm no longer faces competition in a market. In her instrumental variable approach she finds that Wal-Mart's long run effects on prices are negative on all products, except underwear. She finds that the decline in prices is from "1.5-3% in the short run, and four times as much in the long run-- and statistically significant" (Basker 2005, p. 226). In comparison with other retailers Wal-Mart is between 23% lower on popular drugstore items, and competitor's prices are 15% lower in Wal-Mart cities. The estimated decrease in prices is stronger for small cities where there is less competition and establishments are smaller. A decrease in prices for consumers means that their wages go farther in purchasing the consumer goods that they require, i.e. the real wage increases.

2.4 Wal-Mart's Impact on Labor Market Effects

Franz and Robb (1992) used Bureau of Labor Statistics data for 26 states where Wal-Mart operated between the years 1969-1987; the researchers make an urban-rural distinction comparing the impact on MSA vs. non-MSA counties. Some comparisons between urban and rural are made, but the focus of the study is the effect of Wal-Mart on rural counties. The study reported the effects of 689 counties with a Wal-Mart compared to 1,207 counties that do not contain a Wal-Mart during time period. Franz and Robb (1992) conducted a three part analysis. First, they compared the rates of change in economic indicators[4] before and after a Wal-Mart store enters into a county. They find that in rural areas 59% of the sample counties featured an increase in retail trade earnings and Wal-Mart counties' retail trade earnings increased faster than non-Wal-Mart counties.

In the second analysis, they created indices of the rates of change in economic indicators in a county. They estimated that rural counties hosting a Wal-Mart have more positive overall economic indicators than other rural areas. In other words, the areas that contain a Wal-Mart perform better economically than areas that do not

[4] The economic indicators in the Franz and Robb (1992) study include: population, total personal income, payrolls, non-farm proprietor's income, earnings in retail trade, earnings in SIC 52-59, total employment, total wage and salary employment, non-farm self employment, and employment in retail trade.

contain a Wal-mart. This statement does not attribute a causal relationship between Wal-Mart and positive economic indicators, as the counties with a Wal-Mart may or may not have been performing better before Wal-Mart entry. The first two steps in the study 1) quantified some of the pre-existing economic effects and 2) controlled for shocks, such as, disasters, recessions or employment patterns.

The third analysis was a series of regressions on the percentage change in retail trade earnings. The researchers do not state whether they included their indices in their regressions as an explanatory variable, but they regressed percentage change in population, percentage change in the presence of a Wal-Mart store, and percentage change in Real Personal Income (RPI) on the dependant variable of retail earnings in various sectors. The Franz and Robb (1992) regression results find that Wal-Mart has a significant impact on general merchandise and food and drink earnings in rural counties, but not on retail trade earnings in the building materials, automotive (does not specify auto part or auto service), food stores, apparel, miscellaneous, or the total retail sector. They did not present the direction of the impact on earnings of Wal-Mart's presence in the county, except for food and drinks, which is positive.

Ketchum and Hughes (1997) used a differences-in-differences-in-differences or DDD approach to examine affects of Wal-Mart entry on retail wages and employment in Maine's 16 counties. The DDD "is a comparison of relative differences between group means" (Ketchum 1995, p.28). In the case of Wal-Mart, the DDD measures the effects of Wal-Mart (the treatment) on the retail wage of counties who host Wal-Mart (those receiving the treatment) compared to wages in the manufacturing and service industries, as well as, wages in the counties that do not have Wal-Mart.

Ketchum and Hughes (1997) find that the differences in relative retail wage growth in Wal-Mart host counties compared to non-host counties are not significant, thus Wal-Mart has no significant impact on relative retail wage growth. Additionally, after correcting for time, industry, and county specific economic trends the "difference in growth between Wal-Mart and non-Wal-Mart counties for per capita retail employment and average weekly retail wages cannot be attributed to Wal-Mart's presence in Maine counties" (Ketchum 1995, p. 33). It has been noted by Neumark, Zhang, and Ciccarella (2008) that Ketchum and Hughes find noisy insignificant results due to their small sample size of only 16 Maine counties.

Hicks and Wilburn (2001) looked at the impact of Wal-Mart on personal income of counties in West Virginia. Hicks and Wilburn (2001) addressed both the impact of Wal-Mart entering the local labor market and the issue of the endogeneity (non-randomness) of Wal-Mart's location decision. To first examine whether the Wal-Mart location decision is random, the authors compared income level in host counties and non-host counties and found that they are statistically insignificant. As a further test for endogeneity they estimated a probit model on Wal-Mart's location decision using current and lagged growth rate as factors that influence Wal-Mart's decision. They find that county growth and lagged growth have no effect on Wal-Mart's decision to enter. Once they determined endogeneity is not an issue in their sample, Hicks and Wilburn

(2001) used a random effects specification. Hicks and Wilburn use U.S. Census Data for figures on employment, wages, and number of firms, and the Bureau of Labor Statistics for controls (unemployment, civilian labor force figures, personal income and consumer price index, and all urban controls). The authors find that employment, number of firms (establishments), and real wages for the retail sector (SIC 52) increase in West Virginia counties that have a Wal-Mart. The impact of Wal-Mart on real per capita wages is not statistically significant in this model. They find that their results held for alternate tests on smaller counties, although, they do not report these results and whether or not the results were statistically significant. Hicks and Wilburn (2001) suggest that the positive estimated impact of Wal-Mart on wages show that Wal-Mart employees are paid as well or better than other retail sector employees.

The following studies used national data sets and more clearly distinguish between rural and urban areas: Basker (2005a); Driewianka and Johnson (2006); Basker (2007b); Dube, Eidlin, and Lester (2007); Neumark, Zhang, and Ciccarella (2006, 2008); and Bonanno and Lopez (2008). Basker (2005a) analyzed County Business pattern data from 1,749 contiguous counties over 23 years (1964-1999) to estimate the effect of Wal-Mart expansion on retail employment at the county level. She identified three main sources of endogeneity within her study: 1) Timing of Wal-Mart entry; 2) Location of Wal-Mart stores being a non-random decision; and 3) Pre-existing employment growth at the county level. Basker used a count of planned Wal-Mart openings collected from annual reports and Rand McNally Atlases as an instrumental variable to control for endogeneity of the timing of entry of Wal-Mart stores into a county. Her ratioale behind having used planned openings is that Wal-Mart plans store entries well in advance and cannot know the economic conditions of the county on the date of entry. She also controlled for endogeneity of location or county effects by choosing those counties that are similar to those entered by Wal-Mart—counties with a population above 1,500 in base year 1964 and a positive average growth rate of total employment between 1964 and 1977. Endogeneity of employment growth was dealt with by removing counties entered by Wal-Mart before 1977 (Basker 2005a). Basker finds that with an Ordinary Least Squares regression (OLS) retail employment increases in Wal-Mart counties by approximately 40 jobs the first year of Wal-Mart entry, half of these jobs cease to exist after 5 years. Twenty jobs are estimated to have been created before Wal-Mart entry (Basker 2005a). The Instrumental Variables (IV) approach reports a gain of approximately 100 jobs, but in the years following entry there is a loss of 40-60 jobs. The net effect at the 5 year-horizon is positive and significant (Basker 2005a). Basker notes that the IV estimate is a cleaner estimation, and the IV graph she presents shows a much sharper depiction of Wal-Mart's impact.

Basker also ran IV regressions on the impact of Wal-Mart on the number of retail establishments broken down into small (fewer than 20 employees), medium (with 20-99 employees), and large establishments (100 or more employees). The number small establishments decline by four within 5 years of Wal-Mart entry. Medium establishments decline by 0.7 in the second year of Wal-Mart entry. Large establishments increase by 0.7 within the first year of Wal-Mart entry, but then there is a small subsequent decline (Basker 2005a).

Basker again used the IV approach when she looks at changes in employment in various industry sectors due to Wal-Mart entry. She finds a marginally significant decline of approximately 20 jobs in county-level retail wholesale subsector employment (Basker 2005a). She finds no significant affect on the restaurant, automobile, manufacturing sectors or total employment. Basker (2005a) tested for impact on retail sector and wholesale subsector employment in neighboring counties, but found no significant effects of Wal-Mart on neighboring counties. Basker concludes that her IV approach corrects for endogeneity and that net effects on employment and jobs are positive. However, the average impact of Wal-Mart might not be representative of an individual county's experience; she also notes that there may be larger effects at sub-county levels on employment and jobs. The major critique of Basker is that her use of store opening data that reflect the planned sequence of store openings is not a valid instrumental variable as it reflects unobservable factors that may determine Wal-Mart's location decisions (Dube, Eidlin, and Lester 2007; Neumark, Zhang, and Ciccarella 2008).

Dube, Eidlin, and Lester (2007) estimated the effects of Wal-Mart expansion on county level average wages, the total retail wage bill by county, state level retail earnings and hourly wage, state level worker benefits and state level skill composition of retail workers from 1992-2000. The Wal-Mart opening data in this study comes from a database that Wal-Mart made available on its website in late 2005. Dube, Eidlin and

Lester (2007) used a different instrumental variable than Basker (2005a). For their instrument, Dube, Eidlin, and Lester (2007) calculated the spherical distance between the geographical center of a county or state to the center of Benton County, Arkansas, the location of the first Wal-Mart store. The analysis was restricted to 1992-2000, as this was the main period of Wal-Mart expansion outside of the south and into more urban areas. Dube, Eidlin, and Lester (2007) used one expansion period to control for economic cycles that may confound the data. The average wage was constructed by dividing the total earnings, i.e. the wage bill, by total employment (job count). These data are found in the Quarterly Census of Employment and Wages (QCEW). Dube, Eidlin, and Lester (2007) find that overall retail, grocery, and general merchandise average wage is negatively impacted by Wal-Mart and that the total wage bill is reduced by Wal-Mart's presence. However, Dube, Eidlin, and Lester (2007) find mixed results in rural counties (see rural and urban distinction section), and overall there is not a clear reduction of average earnings in nonmetro counties. Dube, Eidlin and Lester (2007) explain that urban areas (MSA counties) feature average weekly earnings that are 22% higher than in non-MSA counties, thus impact of Wal-Mart on the average weekly wage is greater for higher starting values of the equilibrium wage (Dube, Eidlin, and Lester 2007).

The Current Population Survey was used to analyze the impact of Wal-Mart on hourly wages, skills and health coverage at the state level. A small significant negative impact on hourly wages is found at the state level for both Ordinary Least Squares and

Instrumental Variables estimations. They note that the impact is small because it is diluted at the state level. They also find that relatively little of the change in wage can be explained by the change in observed skill mix. In regards to employee benefits, a 0.001 reduction in Employee Supported Insurance coverage is found; the researchers state that "While this effect may be appear small, it is important to note that this represents the impact of a single Wal-Mart store. The total impact of 10 new stores is to reduce ESI coverage by 1 percentage point among total employees" (Dube, Eidlin, and Lester 2007, p. 34). The main conclusions of Dube, Eidlin, and Lester are that they find that Wal-Mart "reduces average and total retail earnings, retail wages, and health benefits for retail workers in urban areas" (Dube, Eidlin, and Lester 2007, p.36). They also note that reduced wages are not due to changes in skill sets of the labor force, thus Wal-Mart reduces labor market rents.

Neumark, Zhang, and Ciccarella (2005, 2008) utilized data provided by Wal-Mart on the location of 3,066 Wal-Mart Discount Stores and Supercenters. Like Basker (2005a), the employment and payroll data are sourced from the County Business Patterns data on count of jobs and payroll, and the analysis is at the county level from 1977-2002. The impact of Wal-Mart on employment and payroll of the retail sector and general merchandise sector was estimated using the same distance instrument as Dube, Eidlin, and Lester (2007) and the two studies were conducted concurrently. The Neumark, Zhang, and Ciccarella (2008) Instrumental Variables equation estimates that a Wal-Mart store reduces county level retail employment by about 167 workers, and general merchandise employment by 116 workers. The estimated impacts on payroll using the Instrumental Variables approach are that aggregate retail payrolls fall by approximately $1.1 to $1.7 million; while general merchandise payrolls increase.

Basker (2007b) checked the geographical instrument used in Dube, Eidlin, Lester (2005) and Neumark, Zhang, and Ciccarella (2005[5], 2008). Basker (2007b) argues that the distance instrument is not valid because it is correlated with other spatial phenomena that are related to the dependent variable of retail employment per capita, economic cycles, and location of certain industries. She ran counterfactual analysis and robustness analysis on Neumark, Zhang, and Ciccarella (2005) and the geographic instrumental variable fails both tests.

Drewianka and Johnson (2006) diverged from the instrumental variable approach and used a random growth model to determine the impact of Wal-Mart on county level employment, wages, and number of establishments. The basic tenet of the random growth method is to identify Wal-Mart's effect on the dependent variable as a change that occurs when a Wal-Mart opens in a county and then to adjust the variable for county effects. The data in Drewianka and Johnson (2006) like Dube, Eidlin, and Lester (2007) comes from the QCEW from the Bureau of Labor Statistics. To make a rural urban comparison Drewianka and Johnson broke the counties into population quartiles with data from the 2000 U.S. Census. A distinguishing feature of the study is

[5] Neumark, Zhang, and Ciccarella (2005) is working paper version of the 2008 published article, the results of these two papers appear to be identical.

that the Wal-Mart stores were analyzed by store type: discount stores, supercenters, and neighborhood markets. Store opening data was requested and obtained from Wal-Mart Stores, Inc., and contained the address, opening date, and store type for 3,194 stores in operation as of 2005.

Drewianka and Johnson (2006) ran a series of regressions (fixed effects, linear trends, and quadratic trends) on county aggregate and county retail sector employment, average weekly wage, and establishments for each type of Wal-Mart store. For the log of employment they find that significant decrease in retail employment for discount stores and supercenters when using a fixed effects regression. The linear trend equation estimates a significant increase in retail employment for discount stores and supercenters, but a negative impact for neighborhood markets. The quadratic trend is only significant positive impact for discount stores for retail employment. The only clearly significant estimate of impact on log of average weekly retail wage is for supercenters using the linear trend, and is negative. Impact of Wal-Mart on log of retail establishments features very mixed results. The researchers find negative significant impacts on log retail establishments for discount stores and supercenters when using a fixed effects model, but a positive significant impact on log of retail establishments for discount stores and negative significant impacts for supercenters using linear trend equations. Quadratic trends are not significant for log of retail establishments.

Drewianka and Johnson (2006) find a statistically significant decrease in log of average weekly wage for the second and third population quartiles for counties with

Wal-Mart discount stores; while supercenters do not have a statistically significant effect on average retail wages in these quartiles. Drewianka and Johnson had limited data in the lowest quartile, thus did not report any results except for supercenters, which were not significant. The top population quartile featured only a marginally significant decrease in average weekly wage for discounts stores, supercenters and neighborhood markets. As Drewianka and Johnson say, "the results consistently portray Wal-Mart as a fairly benign force in local labor markets. There is little here to recommend either regulations to discourage Wal-Mart's entry into a local market or subsidies to promote it" (Drewianka and Johnson 2006, p.34).

2.5 Wal-Mart's Impact on Benefits

If Wal-Mart were to pay benefits to its workers it may be paying a lower wage, but offsetting this with a health or financial package for its worker. Dube and Jacobs (2004) estimated the cost of public assistance to approximately 44,000 Wal-Mart workers in the state of California. They used wage and benefit information from Wal-Mart dated 2001, that was made public in a sex discrimination lawsuit ("Drogin" data named after the defendant) brought against Wal-Mart, compared to the rest of the retail industry data that comes from the U.S. Census March Current Population Survey. They find that "Wal-Mart employees in California use 40 percent more government funded health care than the average for families of all large retail" (Dube and Jacobs 2004, p.1).

Dube and Wertheim (2005) utilized the Drogin data from Dube and Jacobs (2004) and the March Current Population Survey to create an analysis that estimates the nationwide impact of Wal-Mart on benefits. The study also compared Wal-Mart benefits to large and small retailers. The study finds that Wal-Mart provides worse benefits than large retailers, but no worse or better benefits than small retailers. Rural areas are more likely to contain smaller retailers.

As stated previously, to examine the impact of Wal-Mart stores on employee benefits Dube, Eidlin and Lester (2007) analyzed the March Current Population Survey for Employee Supported Insurance coverage (ESI). They find that Wal-Mart stores nationwide reduce employee supported insurance coverage by 1%. There are no reported figures on the impact on benefits in rural areas versus urban areas.

2.6 Estimating Wal-Mart's Monopsony Power

As noted in the analytical framework, Wal-Mart could lower wages in rural areas if it is has monopsony power. Bonanno and Lopez (2008) used a set of simultaneous equations to measure the degree of monopsony power exerted by Wal-Mart over retail workers in U.S. contiguous counties that contain a Wal-Mart in 2004. Monopsony power was measured by calculating Buying Power Index (BPI), equation (6) in Bonanno and Lopez (2008). The researchers find the national Wal-Mart BPI level to be 2.6%, which means that on a national level Wal-Mart pays wages that are 2.6% below their marginal revenue product of labor. According to Bonanno and Lopez (2008), the

Department of Justice considers 5% BPI as an indicator of imperfect competition. Thus, Wal-Mart BPI at the national level would not support policy intervention for behaving as a monopsony. BPI is found to be higher in rural counties 3.8% than urban counties 1.8%, and highest in the West South Central Region of the U.S. (Arkansas, Louisiana, Oklahoma, and Texas). Basically, the higher the market shares of Wal-Mart in a county the greater the magnitude of BPI.

2.7 Rural and Urban Distinctions

Basker (2005a) incorporated a measure for urban and rural counties by using a dummy variable to represent the following classifications: urban, an area with an MSA distinction in 1964; suburban, an area within 25 miles of an area with an MSA distinction in 1964; rural, all other areas. While Basker (2005a) made this distinction in her data she does not report the results and instead makes generalizations based on the average in overall counties.

Neumark, Zhang and Ciccarella (2008) compared the regression model for the dependent variable of retail employment based on above county median population (over 21,000 in base year 1977) to below county median population (less than 21,000 in base year 1977), and find no statistical significance related to population density (Neumark, Zhang, and Ciccarella 2008).

Dube, Eidlin and Lester (2007) looked at MSA versus non MSA counties using 1999 definitions[6]. The authors used dummy variables to distinguish the type of county, and run the various regressions Ordinary Least Squares (OLS), Instrumental Variables (IV), and Control Function (CF) with and without population weights to determine effects in rural areas. Dube, Eidlin and Lester report rural-urban results for average earnings and the retail wage bill, summarized in table 2.1 as follows:

Table 2.1 Urban-Rural Distinctions of Dube, Eidlin, and Lester (2007)

Regression Type	**Average Earnings Urban**	**Average Earnings Rural**	**Wage Bill Urban**	**Wage Bill Rural**
Ordinary Least Squares	Not significant	Negative effects and significant	Not significant	Positive and significant
Instrumental Variables	Negative effects and significant	Not significant	Negative effects and not significant	Not significant
Control Function	Negative effects significant and marginally significant	Not significant	Negative effects and significant	Not significant

(Source: table 7 and 8 Dube, Eidlin, and Lester 2007, p.34)

Table 2.1 can be summarized as Wal-Mart entry reduces average earnings in MSA counties, but the results in non-MSA counties are not clear. The OLS, IV, and CF regressions were repeated for the retail wage bill, and while a negative effect was found in metropolitan areas no systematic effects were present in non-metro counties. The CF

[6] Estimates for earnings and the wage bill based on 1999 (2000 Census) definitions of Metropolitan Statistical Area, i.e. above and below population level 50,000. These definitions are kept constant to “avoid artificial growth in Wal-Mart stores in a MSA as counties may be added over time” (Dube, Eidlin, and Lester 2007, p.23)

and IV regressions are considered to be more accurate estimates because they control for endogeneity. Dube, Eidlin, and Lester propose that urban areas are higher wage and union dense areas where Wal-Mart is able to drive the wage down; while rural areas feature lower initial wage areas, thus the impact of Wal-Mart different.

Like Basker (2005a), Neumark, Zhang, and Ciccarella (2008), and Hicks and Wilburn (2001), the researchers Drewianka and Johnson (2006) reported the estimated impact of Wal-Mart on the number of establishments in a county. As noted previously, Drewianka and Johnson (2006) used population quartiles based on 2000 U.S. Census data to distinguish between Wal-Mart's estimated impact on labor market effects in rural versus urban areas. The readers of the Drewianka and Johnson study are reminded that the number of stores in the lowest quartile are only eight, thus one must proceed carefully with interpretation of these results. They find only marginally significant results in the top quartile for supercenters and neighborhood market type stores and these effects are small and positive. Drewianka and Johnson report positive significant effects on employment (number of jobs) at the county level in the lowest quartile for supercenter and in the third quartile for discount stores and supercenters (see table 2.2).

Table 2.2 Wal-Mart's Impact on Employment Based on Population Quartiles

Wal-Mart's Impact on Employment	**Lowest Quartile (lowest population)**	**2nd Quartile**	**3rd Quartile**	**Top Quartile (highest population)**
Discount Stores	No reported results	Not significant	Positive and Marginally significant	Not significant
Supercenters	Positive and significant	Not significant	Positive and significant	Not significant
Neighborhood Markets	No reported results	No reported results	No reported results	Not significant

(Source: table 4 Drewianka and Johnson 2006)

Drewianka and Johnson (2006) also examined Wal-Mart's estimated impact on average weekly wages based on the quartile stratifications. Significant negative impacts on average weekly wages are found for discount stores in the second quartile and for discount stores and supercenters in the third quartile. Drewianka and Johnson note that their results for average weekly wage are puzzling as the largest effects are seen in the 3rd Quartile; while the results for the 2nd quartile are smaller and only significant for discount stores. The researchers state directly, "This pattern is not consistent with either an explanation involving monopsony power or one involving the composition of the workforce (at least if we accept the evidence that a given store type has larger employment effects in smaller counties)" (Drewianka and Johnson 2006, p.28). The summary of Wal-Mart's impact on weekly wages is reported in table 2.3.

Table 2.3 Wal-Mart's Impact on Average Weekly Wages Based on Population Quartiles

Wal-Mart's Impact on Average Weekly Wages	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile
Discount Stores	No reported results	Negative and significant	Negative and significant	Not significant
Supercenters	Not significant	Not significant	Negative and significant	Not significant
Neighborhood Markets	No reported results	No reported results	No reported results	Not significant

(Source: Table 4 Drewianka and Johnson 2006)

Bonanno and Lopez (2008) also broke their estimates of Buying Power Index (BPI, their measure of monopsony) into rural versus urban distinctions and also compared results by region in the contiguous United States. While Drewianka and Johnson (2006) states that the general trend of Wal-Mart's negative impact on rural average weekly wage is not indicative of monopsony behavior by Wal-Mart; Bonanno and Lopez, however, find the BPI in rural counties is higher; 3.8% compared to 1.8% in urban counties for 2004. They find that states and regions that are more rural have a higher estimated BPI. This study consisted of one year's data (2004), the data makes no estimates how BPI changes or develops over time.

2.8 Endogeneity

A bulk of the empirical research that examines the impact of Wal-Mart on labor market effects does not control for endogeneity (thus, not reviewed in the paper). That is, researchers do not separate the affect of Wal-Mart moving into a county from unrelated economic effects. Stated differently the independent variable, i.e. average earnings, number of jobs, total wage bill etc. is correlated with the error term (unobserved effects of the area being studied). "For example, if Wal-Mart tends to expand in regions where there is stronger economic growth, it would not be surprising to find that overall employment grew faster in places where Wal-Mart entered—but the difference would reflect selection bias rather than Wal-mart's influence," state Drewianka and Johnson (2006, p.8). In addition to analyzing labor market effects, the articles in the literature review attempt to control for the endogenous nature of Wal-Mart's location decision.

To recap, Franz and Robb (1992) created indices to determine if there are significant differences between economic indicators in Wal-Mart and non Wal-Mart counties. Hicks and Wilburn (2001) identified that endogeneity is not significant in West Virginian counties by using a probit model. Ketchum and Hughes (1997) used a differences-in-differences-in-differences specification to control for endogeneity explicitly in their model. Several of the articles used Wal-Mart openings data collected from various sources combined with labor market data to calculate an Instrumental Variables (IV) estimator (Basker 2005a; Dube, Eidlin, and Lester 2007; Neumark, Zhang, and Ciccarella 2008). This IV approach corrects for endogeneity, and is compared with the results of an Ordinary Least Squares regression. Basker (2005a) controlled endogeneity by estimating planned store openings from sequential store numbers reported by Wal-Mart in a special Wal-Mart edition Rand McNally Atlas.

Concurrent studies, Neumark, Zhang and Ciccarella (2008) and Dube, Eidlin, and Lester (2007) distinguished themselves from Basker (2005a) by using a spatial correction when estimating Wal-Mart store openings from administrative data provided by Wal-Mart. The distance of each store from Bentonville, Arkansas was used to approximate the timing of planned openings, hence reducing the measurement error that Basker (2005a) incurred. Additionally, Dube, Eidlin, and Lester (2007) utilized a Control Function test (CF) to control for selection effects. If Wal-Mart were to target areas with higher than normal wages, e.g. unionized areas, the reduction in wages could put a negative bias in the IV estimator; which might over estimate the effects of Wal-Mart on wages (Dube, Eidlin, and Lester, 2007 p. 21). According to Dube, Eidlin and Lester a CF function test can correct for selectivity of Wal-Mart entry by treatment effect and confoundedness due to omitted variables (which might be related to Wal-Mart entry into a county). Drewianka and Johnson (2006) abandoned the IV approach and use a random growth model as an alternative.

The use of various techniques point to the conclusion that there is no clear solution to handling the issue of selection (endogeneity) bias in Wal-Mart's location decision, but that it must be examined and if present in the sample it an attempt to control for it should be made.

2.9 Literature Review Conclusions

A review of the literature shows that Wal-Mart has a positive impact on general merchandise sales (sales figures in general merchandise include Wal-Mart), but it decreases sales mainly in the grocery and building materials subsectors. Small non Wal-Mart neighboring towns feature the biggest decrease in retail sales upon Wal-Mart entry. Both Volpe and Lavoie (2007) and Basker (2005b) find that Wal-Mart decreases the price that consumers pay by substantial amounts. In rural areas this has a large impact on the prices competitors charge. A large reduction in the prices that consumers pay can make their wages go farther in purchasing necessary goods, i.e. increase the real wage.

The literature on retail employment is conflicting. Basker (2005a) finds a slight net increase in retail employment; while Neumark, Zhang and Ciccarella (2008) find a net decrease in retail employment. Retail payrolls overall are shown to decrease in, but general merchandise payrolls increase in the presence of a Wal-Mart (Dube, Eidlin, and Lester 2007; Neumark, Zhang, and Ciccarella 2008). The literature on Wal-Mart's impact on wages shows that Wal-Mart has a small negative impact on wages and payrolls in urban areas (Dube, Eidlin, and Lester 2007; Drewianka and Johnson 2006). Dube, Eidlin, and Lester (2007) and Drewianka and Johnson (2006) examine the impact of Wal-Mart entry in rural counties. They find that the impact in rural counties differs from urban counties, although the results are inconclusive.

Dube, Eidlin, and Lester (2007) find 1% decrease in benefits paid to employees as Wal-Mart enters a county. However, Dube and Wertheim finds that the benefits paid Wal-Mart are “better or no worse” than those paid by small retailers in rural areas. Bonanno and Lopez (2008) show that monopsony power was higher in rural areas in 2004, but this trend in firm behavior may or may not have continued. However, the study’s sample set runs from 1990-2005, thus monopsony power cannot be ruled out as an explanation of Wal-Mart’s impact on rural retail average weekly wage.

The impact of Wal-Mart on rural counties needs a clearer analysis with a nationwide data set and the classification of rural and non-rural counties needs to be more thoroughly examined. Multiple designations for rural and non-rural are used in the various articles: MSA vs. non MSA, population quartiles, population above and below 21,000. None of the studies utilize the rural continuum codes as a measure of rural and urban areas. Rural continuum codes have the potential to be broken down by metro adjacent counties; while not attempted in the study this is a possibility with this data set and future research. It is clear from the review of literature that endogeneity is a problem in some data sets, and must be tested for and controlled. No conclusions on the best methods to control for endogeneity can be discerned from a review of the previous literature. Only one article Hicks and Wilburn (2001) use a probit specification, which is the first part of the Heckit procedure used in this study. However, Hicks and Wilburn are only concerned with counties in West Virginia, and find that endogeneity is not an issue in their sample. Ketchum and Hughes (1997) are lauded for their use of a difference in difference approach to handle the complex characteristics of panel data, yet they also have a small regional sample. A main contribution of the following study is that it uses the Heckit/Modified First Difference procedure to control for endogeneity in a national data set. It also focuses on the different initial scenario in rural counties before Wal-Mart enters, and the subsequent impacts of the said firm on the rural average weekly retail wage.

Due to challenges with Bureau of Labor Statistics, County Business Patterns data wages can only be calculated consistently at the state level. Thus, there may be room for further empirical treatment of wages per worker at the county level and more work using the finer details of the rural urban continuum codes to estimate Wal-Mart's impact on rural retail wages.

Chapter 3: Methods

3.1 A General Wage Model

To formalize the model presented in the analytical frame work one must first look at the factors that typically influence the average weekly wage in a county. Wage (*Y*) can be written as a function of a vector of variables that affect wages.

$$Y = X\beta + \varepsilon \tag{1}$$

$$Y = Y(T, C_{i,} R_i \, X_{it})$$

The study's hypothesis predicts that the presence of Wal-Mart will increase the average weekly wage in rural counties. The number of stores in any county is represented by *Store* and it allows the researcher to determine the impact of Wal-Mart on the average weekly retail wage. An interaction between *Rural* and *Store* creates the variable *Rural Store*; which notes the number of stores in a rural county. The year is included to control for inflation and other changes in the economy overtime that all counties experience. The amount of education or skill level the work force has (*HS_or_less*) is also expected to affect wages. The level of unemployment in the county (*Unemp*) is a measure of excess labor supply, which is expected to affect wages. The population (*Pop_1000*) is included as a control variable.

Average weekly retail wage then becomes a function of the following variables:

$$Y = Y(store, rural, rural\ store, nyear, pop_{1000}, HS\_or\_less, unemp) \qquad (2)$$

3.2 Panel Data

The data for the dependant variable *Y* or *avg_wkwage* and the explanatory variables are drawn from panel data from various sources. Unlike cross section data panel data have two features that complicate statistical analysis. First, each of the observations for a given county is not independent of one another. Second, among groups of observations there may be variation in the error structure, i.e. heterogeneity. These two features may create variation in the data by county unit; counties may start off with very different characteristics and unobserved effects. These effects are not controlled for in the equation. In addition to the panel data issues another pertinent issue with data in Wal-Mart impact studies, is endogeneity. The fact that Wal-Mart's location decision is non-random requires researchers to test the data to ascertain if Wal-Mart's strategy leads the firm to select significantly different counties to locate their stores, i.e. creates a selection bias.

Panel data afford researchers the opportunity to look at average weekly wage observations of the same counties over time. Looking at dynamic data can remove unobserved heterogeneity bias, or the unobserved differences at the county level. There are several techniques to estimate models with panel data: pooled Ordinary Least Squares; fixed effects; first difference fixed effects and random effects. A series of tests can be used to determine the appropriate specification.

The pooled Ordinary Least Squares method requires the explanatory variables to be so complete that they capture all the relevant characteristics of the county (Dougherty 2007). In the case of Wal-Mart's impact on the county retail wage it would be difficult to find accurate measurements of the explanatory or control variables that completely capture all of the relevant characteristics of nearly every county in the United States. A Breusch Pagan Test can be used to determine if the pooled OLS is a sufficient model or if random effects are present. Random effects occur when the unexplained effects are distributed independently of the explanatory variables (Dougherty 2007). If the null hypothesis in the Breusch Pagan test of no random effects is not rejected a pooled OLS is appropriate. In this study the null is rejected (reported in Chapter 4), thus one must then choose between fixed and random effects models.

In order to use a random effects equation two conditions must be met. First, the sample must be random, and second the unobserved effects must be independent of the explanatory variables. Both the fixed effects and random effects model contain, the explanatory variables from the function equation (2). One exception is that the fixed effects model does not contain the rural variable, which is constant for each county. That means that rural is completely multicollinear with the county fixed effects.

The Durbin-Wu-Hausman test is used to determine whether a random effects or fixed effects model is a better estimation tool for a particular data set. The test has the null hypothesis that 'α' or the unobserved effects are independent from the explanatory variables (Dougherty 2007). If the null hypothesis is not rejected then the random effects model is more appropriate. In this study the null hypothesis was rejected (reported in Chapter 4), meaning "the random effects estimates are subject to unobserved heterogeneity bias and will therefore differ systematically from the fixed effects estimates" (Dougherty 2007, p. 419). Thus, the fixed effects model will be used.

A fixed effects method can then be considered. In the within-groups method of a fixed effects model the mean values of all the county level observations are calculated and subtracted from the data on that county (Dougherty 2007). The procedure of subtracting the mean values controls for variations in the error about the mean of the explanatory variables for the yearly county observations.

The fixed effects model can be written using county *i* in time period *t* as an example:

$$Y_{it} = \beta_1 + \beta_t T + \beta_i C_i + \beta_0 X_{it} + \beta_w S_{it} + \beta_r R_i + \beta_q (S_{it} * R_i) + \varepsilon_{it} \tag{3}$$

The dependent variable, Y_{it}, is the average weekly retail wage (*Avg_wkwage)*; *T* is the time trend (*Nyear)*; C_i represents the county fixed effects; *X* is a vector of covariate controls (*Pop_1000*, *HS_or_less, unemp*). S_{it} (*Store*) is a count of Wal-Mart stores for each county. The variable R_i (*Rural*) is a dummy for the rural urban continuum code for 1993. Codes 1-3 are coded zero for metropolitan areas, not rural, and codes 4-9 are coded one for rural. The interaction, $S_{it} * R_i$ *(Rural Store*), between the count of Wal-Mart stores and the rural urban continuum code creates a variable for the presence of Wal-Mart in rural counties. Finally, ε_{it} represents the error term in the equation. Covariate controls contain: population estimates divided by 1000 (*Pop_1000*), unemployment rates (*Unemp)* and the percentage of people in each county who have a high school degree or less as a proxy for education level and employee skill set (*HS_or_less*). The fixed effect model controls for the fact that any observation for a given county in the BLS data set is not independent of another observation for the same county.

Two issues occur with fixed effects models. First, variables that are constant for each individual county cannot be included in the equation, e.g. *Rural.* As noted above the rural urban continuum codes for 1993 are used to create a dummy for the measure of rural vs. urban counties, and thus the *R* (*Rural*) explanatory variable is held constant over time. The second issue with a fixed effect model is that the regression loses *n* degrees of freedom (Dougherty 2007). The data contains over 47,000 observations so losing degrees of freedom is not a concern.

3.3 Endogeneity

Endogeneity, the correlation of an independent variable to the error term is a concern in cross-sectional and time series economic impact studies due to the non-random occurrence of economic events, such as store openings. Basker (2005a) explains how endogeneity might affect outcomes in empirical research: 1) Wal-Mart may decide to move to particular counties strategically based on location and the counties pre-existing employment growth, and 2) the firm may calculate the particular time of the move. Measurement error due to difficulties in obtaining opening data from Wal-Mart may lead to more variance in the error term heightening the problem of endogeneity.

The empirical procedure employed here to estimate the impact of Wal-Mart on rural retail wages is called the Heckman or sometimes the 'Heckit' procedure, which controls for selection bias (Heckman 1979). Selection bias stems from the fact that Wal-Mart makes a non-random decision on where to locate based on characteristics of each county. Two main benefits are produced from using the 'Heckit' specification: 1) One can measure whether the Wal-Mart entry decision is random, i.e. there is a selection bias, and 2) One can create a control variable in the form of the inverse mills ratio. The first step in the Heckman procedure, and in the method of the study, is to run a probit model on the probability that Wal-Mart will enter a county.

$$Prob\ (W = 1|Z) = \varphi(Z\gamma) \tag{4}$$

The variable *W* represents the presence of a Wal-Mart in a county any time from 1990 to 2005 (*W*=1 there is a presence of a Wal-Mart in a county from 1990-2005, *W=0* no Wal-Mart present within the county in the dataset). The variable *Z* is a vector containing variables thought to influence Wal-Mart's selection of a county. The following variables are included in *Z*: the percentage change in median income from 1980-1990 (*Perchange_medincome*), percentage change in population level from 1980-1990 (*Perchange_pop*), and the population level in 1980 divided by 1000 or 10,000 to avoid scalar issues in SAS (*pop80_1000 or pop80_10000*). Phi *Φ* represents the cumulative distribution function of the standard normal distribution; while gamma *γ* represents a vector of unknown parameters (Heckman 1979).

The probabilities of each county having a Wal-Mart and an inverse mills ratio are calculated and used in the second stage OLS. The inverse mills ratio is an unchanging variable that reflects unobserved country effects.

A basic second stage Ordinary Least Squares equation is represented as:

$$Y *= X\beta + \varepsilon \tag{5}$$

The change in average weekly wage (*Wagedif*) is the dependant variable Y*, *X* is the vector of explanatory variables that impact the change in average weekly wage and ε is the error term.

The moment condition is used to calculate a measure of the "missing" explanatory variables under the assumption that the error terms are jointly normally distributed (Gibbons 2009).

$$E[Y * |X, W = X\beta + \rho\sigma\lambda(Z\gamma)] \tag{6}$$

Rho ρ is the correlation between the unobserved determinants of Wal-Mart presence and the determinants of average weekly wage. Sigma σ is the standard deviation of the error term ε and lambda λ is the inverse mills ratio (Heckman 1979). The inverse mills ratio is used in the second stage OLS and if it is statistically significant it indicates that there is selection bias in the counties that have a Wal-Mart present.

As note above, the indicated model from the specification test is a fixed effects model, but it does not allow any variables that are constant for a county. This means that the variable *Rural* cannot be included in a fixed effects model. In addition, the inverse mills ratio is fixed for each county and could not be included in a second stage model, which would not allow testing for selection bias.

A first difference model is an alternative specification of a fixed effects model (Dougherty 2007).

$$(Y_{it} - Y_{it-1}) = (\beta_t - \beta_{t-1}) + (C_i - C_i) + \beta_0(\chi_{it} - \chi_{it-1}) + \beta_w(S_{it} - S_{it}) + \beta_r(R_i - R_i) + \beta_q(R_i S_{it} - R_i S_{it}) + (\varepsilon_{it} + \varepsilon_{it-1}) \quad (7)$$

There are several things to point out about the first differences model which make it difficult to test important aspects of the data. In the first difference the time trend results in a vector of ones that becomes part of the intercept term. *Rural* is constant and becomes a vector of zeros. This standard specification of the first difference model does not contain the inverse mills ratio, but it is also a fixed value for each county and becomes a vector of 0s. These problems are similar to the problems with the fixed effects model. While a fixed effects model mathematically cannot be estimated with the constant variables a first difference model can be modified to include constant variables.

The modified first difference is an alternative approach to the first difference equation in that some of the constant variables are allowed to remain in the equation rather than being first-differenced. For example the rural variable and the inverse mills ratio are not differenced because they would become zeros. It is extremely important to have the *Rural* variable because it provides the rural-urban distinction for the hypothesis of the study. It is also important to have the inverse mills ratio to test for selection bias.

The *Store* variable, a count of Wal-Mart stores in the county is not differenced because differencing would show the entry of a new store in the county while the original variable reflects the long-term presence of Wal-Mart in the county. Retaining both *Rural* and *Store* in the equation allows their interaction to test the impact of Wal-Mart in rural counties. The education variable also is not differenced. As noted, the variable was created from the 1990 and 2000 decennial census and does not have year to year variation, so that differencing it would result in a vector of mostly zeros. With the modifications from the above paragraph the modified first differences equation becomes:

$$(Y_{it} - Y_{it-1}) = (C_i - C_i) + \beta_0\big((\chi_{popit} - \chi_{popit-1}), (\chi_{unempit} - \chi_{unempit-1}), \chi_{it}\big) + \beta_w(S_{it}) + \beta_r(R_i) + \beta_q(R_i S_{it}) + (\varepsilon_{it} + \varepsilon_{it-1}) \tag{8}$$

The dependant variable in the first and modified first differences equation becomes the change or growth in the average weekly retail wage (*Wagedif*), also noted in equation (5). The following control variables are differenced: county population divided by 1000 (*Popdif*) and the percentage of people unemployed by the county (*Unempdif*).

The second stage modified first difference equation with the inverse mills ratio included is given below:

$$(Y_{it} - Y_{it-1}) = \beta_0 (X_{it} - X_{it}) + \beta_w S_{it} + \beta_r R_i + \beta_q R_i S + \rho\sigma\lambda + (\varepsilon_{it} + \varepsilon_{it-1}) \qquad (9)$$

The county fixed effects (C_i-C_i) drop mathematically from the equation. If the coefficient on the inverse mills ratio is significant, there is evidence of selection bias that is Wal-Mart's presence in a county is non-random. The inverse mills ratio (λ) is a control for unobserved determinants of Wal-Mart's location decisions. The results of the running the modified first difference OLS (8) and the second stage modified first difference OLS (9) will be discussed in detail in Chapter 4: Results.

3.4 Data

Most interested parties are concerned about the impact of Wal-Mart at the community or county-level. A sub-state level must be used in order to differentiate between rural and urban areas. Neumark, Zhang, and Ciccarella (2008, p. 409) describe their use of the county-level thusly, "this strikes us as a reasonable geographic level of disaggregation at which to detect the effects of Wal-Mart stores—not so small that many of the effects occur outside of the geographic unit, and not so large that the effects may be undetectable."

From the review of the relevant literature it is clear that there are no comprehensive data that provides hourly wages for full and part-time workers at the county level or a count of full and part-time workers. The average hourly wage can be estimated using the data on earnings and employment from the Bureau of Labor Statistics Quarterly Census of Employment and Wages (QCEW). The average wage is calculated for the NAICS code 44-45 is given in nominal terms using the procedure described by Dube, Eidlin, and Lester (2007):

> Average annual wages per employee for any given industry are computed by dividing total annual wages by annual average employment. A further division by 52 yields average weekly wages per employee. Annual pay data only approximate annual earnings because an individual may not be employed by the same employer all year or may work for more than one employer at a time.

QCEW data are not disclosed for counties in which a firm represents more than 80% of the total industry employment (Dube, Eidlin and Lester 2007). This results in 86 counties being dropped from the data set. Of these only two have Wal-Marts.

Controls for the wage equation come from several sources. A count of Wal-Mart store openings was created from a list of Wal-Mart store opening dates posted by the retailer on their website in 2005, although the link is no longer in service (Wal-Mart

Stores, Inc., 2005). The proportion of people with a high school education or less and the population figures were found at the U.S. Census American Fact Finder (U.S. Census, American Fact Finder, 2010). Education data were constructed by summing up all people who had a high school education along with the all people who did not have a high school education and dividing by the population over 25. High school education levels were only available for the decennial census years' 1990 and 2000. Thus, the 1990 data were used for years 1990-1995 and 2000 data were used for 1996 to 2005. Population levels are from the Bureau of Economic Analysis and were available for each year (BEA, 2010). Unemployment rates are from the Bureau of Labor Statistics (BLS, 2010) and were available for every year, except for a few counties during Hurricane Katrina (BLS, 2006).

Missing data are expected issues to consider when compiling a national data set which spanned 15 years and covered U.S. contiguous counties. The original data set from the BLS data contains 3,072 counties minus Hawaiian and Alaskan counties. However, 86 counties are missing or have not reported data for at least one year from 1990-2005[7], the time span of the sample. Sixty-four of these counties are in the most rural Beale code (rural urban continuum code) designation, 9, and not likely to have a significant retail industry. Only two counties that were dropped because of missing data contain a Wal-Mart: Wyandotte, KS (rural urban continuum code 1, urban) and Menifee,

[7] These counties are listed along with their respective FIPS code in the appendix.

KY (rural urban continuum 9, rural)[8]. Despite being a large county, it is suspected that in Wyandotte Wal-Mart and Sam's Club stores make a large percentage of the market share; therefore, the BLS data is suppressed. The breakdown of Wal-Mart stores by Beale Code and Wal-Mart presence can be found along with detailed descriptions of the missing data in the appendix. Thus, 2,986 counties remain in the sample; 1,757 counties host Wal-Mart stores and 1,229 are non-host counties.

Other issues with county level data include: the creation of new counties; the merger of counties; and the creation and destruction of independent cities in VA[9]. New counties Menominee, WI and Shawano, WI were created after 1980 so population data for that year is lacking. Population levels from the old counties areas were found and combined using data from the 1980 U.S. Census, but because Menominee was lacking BLS data it was excluded on that basis.

Hurricane Katrina had an effect on QCEW data in Louisiana and Mississippi from 2005-2006; the result is imputed data produced by the BLS (BLS, 2006). BLS unemployment data was also affected by the hurricane resulting in missing data from years 2005-2006; in these cases the 2004 values were used for the missing years[10].

[8] See the Code Book in the appendix for rural urban continuum code definitions.
[9] Virginia independent cities are considered as separate counties in the BLS data; U.S. Census population data combines the cities with counties (listed in the appendix) thus the separate independent city population were constructed from the Historic Census of Virginia.

[10] The following Louisiana parishes were missing data: Jefferson, Orleans, Plaquemines, St. Charles, St. John the Baptist, and St. Tammany (BLS 2006).

For the first step of the Heckit procedure, the probit model, the dependent variable is a dummy variable equal to 1 if the count of Wal-Mart stores is greater or equal to 1 at any time between 1990 and 2005, else Wal-Mart=0. Median income levels are from the U.S. Historic Census for 1980 and 1990 and population levels are from the Bureau of Economic Analysis for 1980 and 1990 were used to create the percentage change explanatory variables for the Heckit selection procedure. Population levels for 1980 were included as an explanatory variable. Each of these variables is included in the probit because they are possible factors that Wal-Mart may have considered for locating in a county.

Chapter 4: Results

4.1 Description of the Data

The data for the results are for 2,986 cross sections (counties) from 1990-2005, thus the number of observations totals 47,796. Table 4.1 provides the mean, standard deviation, and minimum and maximum values for the dependent variables and the explanatory variables for each of the models.

Table 4.1 Descriptive Statistics of Dependent and Independent Variables

Variable Label/Description	N	Mean	Std Dev	Min	Max
Variables for pooled OLS and Fixed Effects Models					
Avg_wkwage: nominal retail sector average weekly wage (Dependant Variable)	47776	302.3529	72.37044	104.0000	1198.0000
Store: count of Wal-Mart stores in a County over time	47776	0.7924	1.2993	0.0000	36.0000
Nyear: time trend; controls for inflation	47776	8.5000	4.6098	1.0000	16.0000
Rural: a dummy variable based on a measure of rural vs. urban counties; the rural urban continuum code 1993.	47776	0.7247	0.4467	0	1
Rural store: an interaction count of Wal-Mart stores in a rural county over time.	47776	0.3424	0.5353	0.0000	5.0000
HS_or_less: percentage of people in the county who have a high school education or less	47776	0.6060	0.1167	0.1460	0.8750
Pop_1000: transformation of population values from the U.S. Census; population divided by 1000	47776	90.0588	290.0076	0.6440	9808.4900

Table 4.1 continued

Probit Variables					
Wal-Mart: a dummy, if there was a Wal-Mart store present in the county anytime between 1990-2005 then 1, else 0 (Dependant Variable)	2986	0.5884	0.4921	0	1
Perchange_medincome: the percentage change in median income from 1980 to 1990.	2986	-0.0656	11.5312	-37.7000	48.4000
Pop80_10000: transformation of population level in 1980, population 1980 divided by 10,000[11]	2986	7.5318	24.1938	0.0873	750.6544
Perchange_pop: the percentage change in population from 1980 to 1990	2986	4.4279	16.5905	-31.5650	163.0670
Modified First Difference Variables					
Wagedif: the first difference of the average weekly wage	44790	9.5709	16.1987	-705.0000	721.0000
Unempdif: the first difference of the percentage of people unemployed in each county	44790	-0.0509	1.1338	-13.9000	13.2000
Popdif: the first difference of the transformed population level in each county	44790	1.0405	5.1680	-47.0230	228.7370

[11] Population levels from 1980 were to large relative to the other variables in SAS, and thus the program would not give a readable standard error. Scalar issues are common in the Proc Probit and QLIM procedures

4.2 Ordinary Least Squares Estimates

The analysis begins with testing for the appropriate model specification for the panel data. A pooled Ordinary Least Squares (OLS) model (pooled-means there is no control for county cross sections) is estimated and used to test whether a random effects model or a pooled OLS is a more efficient specification for the panel data. As shown in table 4.2, all of the following explanatory variables in the OLS model are significant in determining the retail average weekly wage: *Store*, *Nyear*, *Rural*, *Rural Store, HS_or_less, Pop_1000, and Unemp*. While theory suggests that Wal-Mart (*Store*), should have no or a positive impact on wages, the previous empirical findings for OLS regressions on retail wages are mixed for the OLS regressions in the article, Dube, Eidlin, and Lester (2007). Neumark, Zhang, and Ciccarella (2008) find positive results on change in retail payrolls per person (not retail wages) in the OLS. The presence of a Wal-Mart (*Store*), as indicated by the count of stores in the county, has a positive impact on the average weekly retail wages; an increase of $0.9568 over estimated average weekly wage of $332.0741. The time trend (*Nyear*) has the expected effect of increasing the nominal average weekly wage. The variable is also a proxy for inflation. As expected rural areas (*Rural*) have a lower average weekly retail wage in comparison to urban areas[12].

[12] Rural counties are given by rural urban continuum codes 4-9; Urban counties are given by codes 1-3

In support of the hypothesis, the number of stores in a rural county (*Rural Store*) have a positive impact on the dependant variable and the magnitude of the parameter estimate is much larger (19.6711 vs. 0.9569) than in overall counties with a store. In Dube, Eidlin and Lester (2007) Wal-Mart presence in rural counties has a negative impact on the average retail wage in the OLS model. When other specifications are used, i.e. the Control Function and Instrumental Variables the results are insignificant. As the percentage of the population with a high school education or less (*HS_or_less*) increases the average weekly retail wage decreases, this is expected according to Dube, Eidlin, and Lester (2007) who find decreases in the average retail weekly wage for their Low Education variable in their state level estimates. As the population (*pop_1000*) increases the average wage increases. This is consistent with urban areas having higher wages, perhaps because they also have a higher percentage of retail that is unionized in comparison with rural areas (Dube, Eidlin, and Lester 2007). As unemployment increases the average weekly retail wage increases. This result is unexpected according to the analytical framework in which unemployment decreases the average weekly retail wage, reflecting low demand for labor relative to supply.

Table 4.2 Pooled OLS Regression: Impact on Average Weekly Retail Wage

Explanatory Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	332.0741	1.3260	250.43	<.0001
Store	0.9568	0.2109	4.54	<.0001
Nyear	8.7561	0.0458	191.30	<.0001
Rural	-46.8312	0.5967	-78.48	<.0001
Rural Store	19.6711	0.4523	43.49	<.0001
HS_or_less	-150.9420	2.0904	-72.21	<.0001
Pop_1000	0.0414	0.0008	48.73	<.0001
Unemp	1.7265	0.0798	21.64	<.0001
R-square 0.6383, Adjusted R-square 0.6383				

4.3 Breusch Pagan Test for Random Effects

The Breusch Pagan Test was run on the pooled OLS versus a one-way random effects model to test which specification more efficiently fits the data. As shown in table 4.3, the reported m value is large 208816 with a p-value of <.0001; the null hypothesis of no random effects is rejected. The pooled OLS model for this data set is inefficient, and fixed or random effects models must be considered.

Table 4.3 Breusch Pagan Results

Breusch Pagan Test for Random Effects (One Way)

DF	m Value	Pr > m
1	208816	<.0001

4.4 Durwin-Wu Hausman Test for Random vs. Fixed Effects

The Durwin-Wu Hausman Test for random effects was used to test the fixed effects model versus the random effects model. The null hypothesis in this case is that the unobserved effects are distributed independently of the error term. As shown in table 4.4, the reported m value 464.73 is large with a very small p-value <.0001, thus the null hypothesis is rejected. A fixed effects model is a more efficient estimate of Wal-Mart's impact on average weekly retail wage with this panel data set (Dougherty 2007).

Table 4.4 Durwin-Wu Hausman Test for Random vs. Fixed Effects

Hausman Test for Random Effects		
DF	m Value	Pr > m
6	464.73	<.0001

4.5 Fixed Effects Model Estimates

As shown in table 4.5, the R-square of the fixed effects model is much larger than that of the pooled OLS (0.9142 vs. 0.6383). The improved goodness of fit is likely due to the cross sectional variables that control for the nature of the panel data and county unobserved effects. Thus, there are 2,986 additional variables in this equation compared to the pooled OLS. Unlike the pooled OLS equation, the rural variable must be removed because it is a constant value for each county thus completely correlated with the cross section variable.

All of the explanatory variables are significant predictors of the average weekly retail wage, except the number Wal-Mart stores in a rural county (*Rural Store*). The number of Wal-Mart stores in a county (*Store*) in continues to estimate a positive impact on the average weekly retail wage, despite having controlled for county fixed effects. Using a fixed effects regression, Drewianka and Johnson (2006) found negative impacts on average retail wages on counties with a neighborhood-market-type Wal-Mart store, but found insignificant results for discount stores and supercenters. Dube, Eidlin, and Lester (2007) do not report a fixed effects model. As expected, the nominal average weekly wage increases over time (*Nyear*).

Lending no addition support to the stated hypothesis, the number of Wal-Mart stores a rural county (*Rural Store*) does not have a statistically significant impact on the average weekly retail wage. Similar to the pooled OLS, the high school education or less (*HS_or_less*) variable estimates that when the percentage of people in a county who have a high school education or less increases, the average weekly retail wage decreases. Population (*Pop_1000*) remains positive; as population increases the average weekly retail wage increases. Unemployment (*Unemp*) remains unexpectedly positive; a potential reason may be because of a lack of sufficient explanatory variables.

Table 4.5 Fixed Effects Model Results: Impact on Average Weekly Retail Wage

Explanatory Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	189.9965	6.1163	31.06	<.0001
Store	1.9332	0.2583	7.49	<.0001
Nyear	9.4976	0.0379	250.79	<.0001
Rural store	-0.6895	0.7052	-0.98	0.3282
HS_or_less	-16.2672	4.3464	-3.74	0.0002
Pop_1000	0.1910	0.0065	29.17	<.0001
Unemp	0.5457	0.0721	7.57	0.0092
R-Square 0.9142				

4.6 Modified First Difference Ordinary Least Squares

The first difference model handles the complex characteristics of panel data, and is essentially a fixed effects model (Dougherty 2007; Wooldridge 2002). The modified first difference model controls for unobserved county effects, and has the advantage of allowing for constant variables, such as *Rural,* to remain in the equation (Dougherty 2007). The *Store* and *Rural Store* variable are not differenced to retain their meaning of Wal-Mart's presence in a county; if they were differenced they would refer to the entry of Wal-Mart in a county instead. The variable *high school or less* was constructed by taking the 1990 census data forward to 1995 and the 2000 census data backward to 1996 and forward to 2005. Differencing this variable would result in a vector of 0s except for 1990 and 2000. The time trend, *Nyear,* creates a vector of ones when differenced and becomes part of the intercept. The new dependant variable is the change in the average weekly retail wage (*Wagedif*) as opposed to the level of the average weekly retail wage; likewise population is the change in population and unemployment becomes the change in unemployment.

Table 4.6 lists the results of the modified first difference model. The adjusted R-square is 0.0109, this is low. However, a majority of the explanatory variables are significant. In this model, only the *Store* variable is not a significant predictor of growth in the average weekly retail wage. The remaining explanatory variables: *Rural, Rural Store, HS_or_less, Popdif, and Unempdif* remain significant in the modified first differences first effects.

The count of Wal-Mart stores (*Store*) has an insignificant impact on the change in retail average weekly wage. This finding is different than the effect of Wal-Mart on average weekly retail wages in the previous two equations. Using the log of average retail wages, Dube, Eidlin, and Lester (2007) and Drewianka and Johnson (2006) (for supercenters only) found negative impacts when county effects are controlled. As expected rural areas (*Rural*) have a lower retail wage growth rate than urban areas, as the parameter estimate is negative and significant.

In support of the hypothesis, the number of stores in a rural county Wal-Mart (*Rural Store*) has a positive impact on the growth in the average weekly retail wage. Similar to the pooled OLS and fixed effects models, high school or less (*HS_or_less*) has a negative impact on the growth in the average weekly retail wage as expected. The modified first difference results estimate that as the growth in population (*Popdif*) increases the growth in the average retail wage increases. Finally, in this model similar to the pooled OLS and fixed effects models growth in unemployment (*Unempdif*) has a positive impact, and an unexpected result.

Table 4.6 Modified First Difference OLS Results: Impact on the Difference in Average Weekly Retail Wage

Explanatory Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	15.2385	0.4362	34.94	<.0001
Store	-0.0392	0.0767	-0.51	0.6095
Rural	-2.1625	0.2277	-9.50	<.0001
Rural store	1.5096	0.1714	8.80	<.0001
HS_or_less	-7.7618	0.7168	-10.83	<.0001
Popdif	0.0959	0.0170	5.65	<.0001
Unempdif	0.2134	0.0672	3.18	0.0015
R-Square 0.0058; Adjusted R-Square 0.0109				

4.7 Heckit Probit Results

The preceding models have not attempted to control for the non-random behavior of Wal-Mart's decision to locate in a particular county. Not accounting for the non-random element of Wal-Mart's strategy can bias coefficient estimates. This study uses a Heckit procedure to control for Wal-Mart's non-random selection of counties. The Heckit procedure estimates a probit model to account for the selection of counties. From this model the inverse mills ratio is estimated for each county. This is used as a control for county unobserved effects in a second stage equation, which is a modified first difference equation.

The first component of the Heckit procedure begins to test for the endogeneity of Wal-Mart's location strategy (Hicks and Wilburn 2001). It does this by utilizing a probit equation to determine the probability that Wal-Mart store will have a presence in a county at any time from 1990 onward. As Basker said, "This identification strategy assumes that Wal-Mart plans its store entries well in advance of entry and cannot accurately forecast exact market conditions at the time for which entry is planned" (Basker 2005a, p.178). As presented in table 4.7, the control variables: population level in 1980, population change from 1980-1990, and percentage change in median income are all found to be significant explanatory variables in whether Wal-Mart has a presence in a county. As the median income (*Perchange_medincome*) variable decreased from 1980-1990, Wal-Mart has a higher probability of having a presence in counties with a slower growth in income after 1990. As population (*Percentchange_pop*) variable increased from 1980-1990, Wal-Mart is more likely to have a presence in counties with faster population growth after 1990. Counties with higher populations in 1980 were more likely to have a Wal-Mart after 1990. As Drewianka and Johnson stated "Wal-Mart tends to expand where on-going growth is relatively strong overall, but relatively weak in the retail sector" (Drewianka and Johnson 2006, p.1)

Table 4.7 Probit Analysis: Dependant Variable Presence of a Wal-Mart Store 1990 to 2005

Parameter	DF	Estimate	Std Error	95 % Confidence Limits		Chi Square	Pr>Chi Square
Intercept	1	-0.0414	0.0071	-0.0553	-0.0275	34.01	<.0001
Perchange_median income	1	-0.0219	0.0007	-0.0231	-0.0206	1118.43	<.0001
Pop80_10000	1	0.0269	0.0006	0.0258	0.0280	2257.69	<.0001
Perchange_pop	1	0.0312	0.0005	0.0302	0.0323	3279.94	<.0001
Log Likelihood -26857.0902							

4.8 Heckit with Second Stage OLS: Modified First Difference Results

The next step of the procedure is to calculate the inverse mills ratio, λ, in the modified first difference OLS (equation 8) to determine if the Wal-Mart presence variable is non-random. Again the dependant variable is the change in the average weekly wage. The number of observations decreases to 26,355 in this equation because only the counties with Wal-Mart present (Wal-Mart=1) are included in the second stage modified first difference OLS. Thus the second stage OLS estimates the impact of the explanatory variables only on Wal-Mart host counties. In table 4.8, the results of the regression find the inverse mills ratio, λ, to be negative and statistically significant. The negative inverse mills ratio suggests that controlling for Wal-Mart selection reduces the growth in average weekly retail wages. It also suggests that the average retail wage growth is below average in Wal-Mart counties, which is consistent with the finding that there is a higher probability of Wal-Mart entering counties with slower growth in median income in the probit model. This is unlike Hicks and Wilburn (2001) who do not find endogeneity in Wal-Mart's location decision in their West Virginia sample. Their findings may be specific to unobserved characteristics of West Virginia. Sigma is the standard deviation of the inverse mill's ratio and is statistically significant.

In counties with a Wal-Mart the number of stores (*Store*) has a significant negative impact on the growth in the average weekly retail wage, table 4.8. In counties with a Wal-Mart, having more stores decreases retail wage growth. As shown by table 4.1, the maximum number of stores in one county is 36; the counties with the most stores are urban counties. If urban counties have a higher wage, as the pooled OLS suggest, then even if wages go up by the same amount as rural areas, urban areas would grow more slowly. These results are similar to Dube, Eidlin and Lester (2007) and Driewianka and Johnson (2006) (for supercenters only) who both found negative impacts on the log of average weekly retail wages. They controlled for endogeneity using Instrumental Variables and random growth models, respectively. For counties with a Wal-Mart there is no difference in average weekly retail wage growth between rural and urban counties, as shown by the insignificance of the rural urban continuum code variable (*Rural*). Perhaps the number of stores is also a control for urban, so this variable becomes insignificant, i.e. there are two controls for urban-rural without realizing it.

The results in table 4.8 support the hypothesis of the study, by finding that the number of Wal-Mart stores in a rural county (*Rural Store*) has a positive impact on the growth in average weekly retail wage. For counties with a Wal-Mart, rural counties with stores have a higher wage growth than urban counties with a store. Thus, Wal-Mart is likely to be a large labor demander in rural areas relative to the rest of retail employers, and in comparison to urban areas.

Besides the Rural variable, the remaining control variables are all statistically significant in the second stage modified first differences OLS. As expected for counties with a Wal-Mart, an increase in the county population that has a lower education level (*HS_or_less*) leads to lower growth in the average weekly retail wage. In counties with a Wal-Mart, as growth in population (*Popdif*) increases, growth in the average weekly retail wage increases; confirming the Dube, Eidlin, and Lester (2007) idea that as union density increases, typical in urban, population dense areas, average weekly wage increases. As growth in unemployment (*Unempdif*) increases in counties with a Wal-Mart, the growth in average weekly wage increases. As mentioned before this result is unexpected, and could be explained by the fact that there are few explanatory variables in the regression.

Table 4.8 Heckit/Second Stage Modified First Differences OLS Results: Impact on the Difference in Average Weekly Retail Wage

Explanatory Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	17.4456	0.5105	34.17	<.0001
Store	-0.2341	0.0756	-3.10	0.0019
Rural	-0.0511	0.3483	-0.15	0.8835
Rural store	0.7363	0.2638	2.79	0.0052
HS_or_less	-6.2762	0.8571	-7.32	<.0001
Popdif	0.0590	0.0170	3.46	0.0005
Unempdif	0.4090	0.0822	4.97	<0.0001
Sigma	14.8294	0.0922	160.76	<.0001
Inverse mills ratio	-0.4003	0.0200	-19.99	<.0001
Log Likelihood -134218				

4.9 Comparison

The results of the original modified first difference OLS are the complete set of counties in the U.S. minus the difference of one year of counties (44,790); while the second stage modified first difference OLS contains only the counties in which a Wal-Mart is present from the years 1990-2005 (26,355). The estimated impact of the number of Wal-Mart stores (*Store*) on growth in average weekly retail wage is negative and significant in the second stage OLS; while the parameter estimate for the original modified first difference OLS is insignificant. Thus, for counties with a Wal-Mart, the growth in the average weekly retail wage is slower than the growth in the complete data set. Rural and urban counties have no statistically significant difference in retail wage growth in the second stage OLS, which differs from the original first differences OLS where the rural counties have slower growth in average weekly retail wage.

Perhaps the presence of a Wal-Mart store reduces the difference in growth between rural and urban counties found in the complete data set.

The presence of Wal-Mart in rural counties (*Rural Store*) has a positive and significant impact in both models, but the parameter estimate is more positive in the original first difference model. It seems that for the Wal-Mart only data the number of stores in a rural county (*Rural Store*) has less of a positive impact on average weekly retail wage growth than the number of stores in a rural county (*Rural Store*) in the complete data set. This may be due to Wal-Mart's propensity to locate in counties with a slower growth in retail wage, but the non-Wal-Mart stores need to be examined to make more definitive conclusions. These results support the hypothesis and are similar to the results from the Drewianka and Johnson (2006) and Dube, Eidlin, and Lester (2007) where the impacts of Wal-Mart on rural log of average weekly wage are positive in direction when the non-random location strategy of Wal-Mart is controlled. The key difference in the models presented here are that the results are always positive and significant, not mixed results.

The control variables, high school or less and growth in population (*HS_or_less and Popdif*), become smaller (more negative) in the second stage OLS, but the unemployment variable (*Unemp*) gets larger (more positive). For the counties in which there is a Wal-Mart store, the impact of education and population on average weekly retail wage is slightly less than in the complete data set; while the impact of unemployment is slightly higher than in the complete data set.

4.10 What do the results say about endogeneity?

The significance of λ, the inverse mills ratio, in the second stage of the Heckit procedure is confirmation that Wal-Mart store location data is non-random, and must be corrected for when measuring the estimated impacts of the retailer. Comparing the original modified first difference model that contains the full data set with the second stage model that only contains Wal-Mart host counties it seems that the selection bias of Wal-Mart's location decision may place an upward increase on the change in the retail average weekly wage. If the model is estimated without the control for endogeneity the coefficient for number of Wal-Mart stores overall (*Store*) and the number of Wal-Mart in rural counties (*Rural Store*) are more positive. An upward bias would be indicative that Wal-Mart has a propensity to select areas that have favorable economic conditions. The direction of the inverse mills ratio or Wal-Mart's location decision shows that Wal-Mart is likely to select areas that have slower growth in the average weekly retail wage; thus the entrance of the firm increases the growth of wages. To summarize Wal-Mart is likely to select areas with favorable economic conditions, but slower than growth in the average weekly retail wage.

Chapter 5: Conclusions

5.1 Summary of Findings

A key finding in the study is that Wal-Mart's decision to locate in a county is found to be non-random by the Heckit procedure because the inverse mills ratio is statistically significant. The negative direction of the inverse mills indicates that Wal-Mart chooses areas where the growth in the average weekly retail wage is slower compared to other counties. *Store* and *Rural Store* parameter estimates become more negative in counties with a Wal-Mart, controlled for non-random selection behavior, than in the whole sample.

The number of Wal-Mart stores in a county has a statistically significant negative impact on the average weekly retail wage in the second stage modified first difference OLS. These findings are similar to the findings of Dube, Eidlin, and Lester (2007) and Drewianka and Johnson (2006) who found negative effects on the overall log of average weekly retail wages, when controlling for endogeneity. Placed in the analytical framework of the study the empirical results suggest that on average Wal-Mart is not a large labor demander relative to the overall retail sector market nationwide. Until data on the job type that Wal-Mart provides, i.e. part-time, full-time, manager or associate position, it is difficult to determine which factors are causing the positive impacts of Wal-Mart on rural retail wages.

On the other hand the rural Wal-Mart store variable (*Rural Store*) has a statistically significant positive impact on average weekly retail wage growth, which supports the hypothesis that Wal-Mart increases the rural average retail wage. Based on the empirical results, Wal-Mart may be a large enough labor demander in rural counties relative to labor supply to increase the growth of retail wages. The average weekly retail wage growth in rural counties with a Wal-Mart is an additional $0.73 more than the average weekly retail wage growth (given by the intercept in table 4.6). This means that wages in rural counties with a Wal-Mart grow at a rate that is 7.6% higher than the national average weekly retail growth rate of $9.5708 (table 4.1 mean of the *Wagedif* variable). Figure 1.1 in the analytical framework illustrates that Wal-Mart in rural counties has a large labor demand in rural counties as compared to rest of retail in the county, and increases average retail wages. It is still possible that Wal-Mart employs more part-time workers than full-time, but perhaps employees work more hours than other rural retail workers because most Wal-Mart stores are open for 24 hours, and other rural retailers have more limited hours.

Finally, a positive impact on the growth of the average weekly wage is not indicative of the behavior a monopsony would have on the wage in a rural labor market, but it does not rule out the possibility of monopsony power in its entirety

5.2 Policy Implications

The results from the study show that average weekly retail wages in rural counties with a Wal-Mart are growing 7.6% faster than the average weekly wages for the national data set. Based on the growth rate in rural retail average weekly wages Wal-Mart may have net benefits to a rural county. The results from the study can contribute to further research on the combined cost and benefits of Wal-Mart to consumers, employees, and communities. For example Basker (2005b) finds that Wal-Mart lowers prices, which benefits consumers. Lower prices also increase the real wages of Wal-Mart employees all other wage earners. At this point no research has specifically estimated the impact of Wal-Mart on real wages.

5.3 Future research directions

The research study was limited by a couple factors: 1) The inability to obtain data on hourly and full-time employees; the skill of employees and 2) A limited number of explanatory variables. One future direction for research is to analyze the full and part-time labor stratification for another test of the hypothesis that Wal-Mart raises the average weekly retail wage in rural counties. If more part-time labor is created the rural average retail wage should be decreasing. However, Wal-Mart employees could be working more part-time hours than those of other rural retail workers. Further analysis with a more detailed data set could clarify what underlying forces are increasing the average weekly retail wage in rural counties. A second direction the research on the impact of Wal-Mart on rural counties could take is to use a more detailed measure of rural to see if there are different impacts in urban, suburban, rural adjacent to urban, and more remote rural areas. These breakdowns are possible using the rural urban continuum codes of the USDA (ERS, USDA, 2010).

References

Artz, G.M., and J.C. McConnon. 2001. "The Impact of Wal-Mart on Host Towns and Surrounding Communities in Maine." Working Paper, Office of Social and Economic Trend Analysis, Iowa State University. http://www.umext.maine.edu/smallbiz/reports/Wal-Mart.pdf Accessed 10/14/2010.

Artz, G., P. Orazem, and D. Otto. 2007. "Measuring the Impact of Meat Packing and Processing Facilities in Nonmetropolitan Counties, A Difference-in-Differences Approach." *American Journal of Agricultural Economics*, 89(3), 557–570.

Artz, G. and J. Stallmann. 2006. "Recruiting Big Box Retailers and an Economic Development Strategy." National Public Policy Conference, Fayetteville, Arkansas. http://www.farmfoundation.org/news/articlefiles/392Stallman%20Artz%20NPPEC%20paper.pdf Accessed on 11/12/2010.

Bartik, Timothy J. 2005. "Solving the Problems of Economic Development Incentives." *Growth and Change* 36(20: 139-166.

Bartik, Timothy J. (2004). "Thinking about Local Living Wage Requirements." *Urban Affairs Review*. 40(2):269-299.

Basker, E. 2005a. "Job Creation or Destruction? Labor Market Effects of Wal-Mart Expansion." *The Review of Economics and Statistics* 87:174-183.

Basker, E. 2005b. "Selling a Cheaper Mousetrap: Wal-Mart's Effect on Retail Prices." *Journal of Urban Economics* 58:203-229.

Basker, E. 2007a. “The Causes and Consequences of Wal-Mart’s Growth.” *The Journal of Economic Perspectives* 21(3):177-198.

Basker, E. 2007b. “When Good Instruments Go Bad: A Reply to Neumark, Zhang, and Ciccarella (2005).” Working Paper 0706, University of Missouri-Columbia

Bonanno, A., and R. Lopez. 2008. “Wal-Mart’s Monopsony Power in Local Labor Markets.” Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Orlando, FL, July 27-29, 2008.

Drewianka, S., and D. Johnson. 2006. "Wal-Mart and Local Labor Markets, 1990-2004." https://pantherfile.uwm.edu/sdrewian/www/walmartandlocallabormarkets.pdf

Dougherty, Christopher. 2007. “Introduction to Economics.” Chapter 14: Introduction to Panel Data Models. Oxford University Press: New York

Dube, A., B. Eidlin, and B. Lester. 2005. “Impact of Wal-Mart Growth on Earnings throughout the Retail Sector in Urban and Rural Counties,” unpublished paper, University of California–Berkeley Working Paper, Institute of Industrial Relations, University of California, Berkeley. . http://laborcenter.berkeley.edu/lowwage/index.shtml (Accessed August 20, 2006.) http://laborcenter.berkeley.edu/retail/walmartwage.pdf (Direct Link)

Dube, A., B. Eidlin, and B. Lester. 2007. “Impact of Wal-Mart Growth on Earnings throughout the Retail Sector in Urban and Rural Counties.” Working Paper, Institute of Industrial Relations, University of California, Berkeley. http://laborcenter.berkeley.edu/lowwage/index.shtml Accessed 10/14/2010.

Dube, Arindrajit and Ken Jacobs (2004). “Hidden cost of Wal-Mart Jobs: Use of Safety Net Programs by Wal-Mart Workers in California.” UC Berkeley Institute for Industrial Relations and UC Berkeley Center for Labor Research and Education. August 2, 2004. http://laborcenter.berkeley.edu/lowwage/index.shtml (Accessed August 20, 2006.) http://laborcenter.berkeley.edu/retail/walmart.pdf

Dube, A., and S. Wertheim. 2005. **"**Wal-Mart and Job Quality – What Do We Know, and Should We Care?" UC Berkeley Institute of Industrial Relations, and UC Berkeley Center for Labor Research and Education. http://laborcenter.berkeley.edu/lowwage/index.shtml Accessed August 20, 2006.

Economic Research Service, United States Department of Agriculture. 2010. Briefing: Rurality. Available at the following website: http://www.ers.usda.gov/briefing/rurality/ruralurbcon/ Accessed on September 19, 2010, last updated April 28, 2004.

Franz, L., and E. Robb. 1992. "Effect of Wal-Mart Stores on Economic Environment of Rural Communities." Business and Public Administratio Research Center, University of Missouri. http://eparc.missouri.edu/Publication/Wal-Mart/Wal-Mart%20Study(1989).pdf

Gibbons, Charlie. 2009. "Non-Random Sample Selection: Doing the Two-Step from Heckman to Instrumental Variables." Political Science 239: Quantitative Methodology in the Social Sciences Seminar. University of California Berkeley http://cgibbons.berkeley.edu/Courses/PS239_S09/Selection.pdf

Heckman, J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica,* 47(1):153-161

Hicks, M.J., and K. Wilburn. 2001 "The Regional Impact of Wal-Mart Entrance: A Panel Study of the Retail Trade Sector in West Virginia." *Review of Regional Studies* 31:305–313.

Historic Census of Virginia. Demographics and Workforce. Available at the following website: http://www.coopercenter.org/demographics/census-data Accessed December 9, 2010, last update not reported.

Ketchum, Brian A. 1995. "Wal-Mart and Maine: Analyzing Wal-Mart's Effect on Employment and Wages in Maine Counties." A Senior Thesis Presented for the Completion of a Bachelor of Arts Degree in Economics.

Ketchum, Brian A. and J. Hughes. 1997. "Wal-Mart and Maine: The Effect on Employment and Wages," *Maine Business Indicators*, 42 (3):6-8.

Neumark, D., J. Zhang, S. Ciccarella. 2005. "The Effects of Wal-Mart on Local Labor markets." Working paper No. 11782. National Bureau of Economic Research, Cambridge, Massachusetts (revised 2006, 2007). http://ideas.repec.org/p/irv/wpaper/060711.html

Neumark, D., J. Zhang, and S. Ciccarella. 2008. "The Effects of Wal-Mart on Local Labor Markets," *Journal of Urban Economics*, Elsevier, 63:405-430. http://www.sciencedirect.com/science/article/B6WMG-4PDSBJ4-8/2/ceda0edb487c2db6ad5abd93aea51f85

Perloff, J.M. 2008. "Microeconomics: Theory and Applications with Calculus." Chapter 15: Factor Markets. 526-561.

Pindyck, R. and D. Rubenfield. 1998. "Econometric Models and Economic Forecasts." Chapter 9: Single-Equation Estimation Advanced Topics. Irwin/McGraw Hill

Stone, K.E., 1988. The effect of Wal-Mart Stores on businesses in host towns and surrounding towns in Iowa. Working paper. Department of Economics, Iowa State University. Available at http://www.econ.iastate.edu/faculty/stone/Effect%20of%20Walmart%20-%201988%20paper%20scanned.pdf (as of September 7, 2005).

Stone, K., G. Artz, and A. Myles. 2002. The Economic Impact of Wal-Mart Supercenters on Existing Businesses in Mississippi. Mississippi State University Extension Service, Fall. Available at http://msucares.com/pubs/misc/m1283.pdf (as of September 7, 2005).

U.S. Bureau of Economic Analysis Website. Regional Economic Accounts: Local Personal Area Income. Available at the website http://www.bea.gov/regional/reis/ Accessed September 28, 2010, last updated August 09, 2010.

U.S. Bureau of Labor Statistics Website. 2010. Bureau of Labor Statistics: Effects of Hurricane Katrina on the Quarterly Census of Employment and Wages. Available at the following website: http://www.bls.gov/katrina/qcewquestions.htm Accessed December 6, 2010, last updated April 12, 2006.

U. S. Bureau of Labor Statistics Website. 2010. Bureau of Labor Statistics: Local Area Labor Statistics. Available at the following website: http://www.bls.gov/lau/#tables Accessed December 6, 2010, last updated

U.S. Bureau of Labor Statistics Website. 2010. Quarterly Census of Employment and Wages: Employment and Wages, Annual Averages 2002. Available at the following website: http://www.bls.gov/cew/cewbultn02.htm#Wages Accessed December 12, 2009, last updated March 31, 2008.

U.S. Census Bureau Website. 2010. Metropolitan and Micropolitan Statistical Areas. Available at the following website: http://www.census.gov/population/www/metroareas/metroarea.html Accessed December 12, 2010, last updated December 9, 2009.

U.S. Census Bureau Website. 2010. County Business Patterns: Introduction. Available at the following website: http://www.census.gov/econ/cbp/intro.htm. Accessed September 28, 2010, last updated July 28, 2010.

U.S. Census Bureau, American Factfinder Website. 2010. Available at the following website: http://factfinder.census.gov/home/saff/main.html?_lang=en Accessed September 29, 2010, last update was not reported.

Volpe and Lavoie. 2007. "The Effect of Wal-Mart Supercenter on Grocery Prices in New England." *Review of Agricultural Economics* 30(1):4-26.

Wal-Mart Stores, Inc. Website. 2010. About Us: Wal-Mart. Available at the following website: http://walmartstores.com/AboutUs/7606.aspx. Accessed December 06, 2010, last updated date not reported.

Wal-Mart Stores, Inc. Website. 2010. Data Sheet - Worldwide Unit Details. Available at the following website: http://walmartstores.com/pressroom/news/10305.aspx. Accessed September 28, 2010, last updated January 31, 2010.

Wal-Mart Stores, Inc. Website. 2005. Global Wal-Mart Stores. Available at the following website: http://walmartstores.com/GlobalWMStoresWeb/navigate.do?catg=130&yearID=131&monthID=138. Accessed by Georgeanne Artz 2005. The link no longer works.

Wooldridge, J.M. 2002. "Econometric Analysis of Cross Section and Panel Data." MIT Press.

Wooldridge, J.M. 2007. "What's New in Econometrics? Lecture 10, Differences-in-Differences Estimation." NBER Summer Institute. Available at www.nber.org accessed October 30, 2010.

Appendices

I. Notes on Missing Data

A. Missing Data for the Heckit Probit Model from the Historic Census and BEA

FIPS County	Median Income 1980	Median Income 1990	Population 1980	Population 1990	BLS	Steps taken/Sources
4012 La Paz, AZ	(NA)	Available	(NA)	Available	Available	Did not become a county until 1983 Used Yuma median income for 1980 value, used 12,557 for population value. http://www.census.gov/geo/www/tiger/ctychng.html
35006 Cibola, NM	(NA)	Available	(NA)	Available	Available	Did not become a county until June 19, 1981 Used Valencia median income for 1980 value, used 30,347 for population. http://www.census.gov/geo/www/tiger/ctychng.html
Washabaugh , County, SD	(NA)	(NA)	(NA)	(NA)	(NA)	No longer a county—Merged with Jackson county in 1983 did not use the data set reported in Historic Census, but not BLS
Nansemond County, VA	(NA)	(NA)	(NA)	(NA)	(NA)	No longer a county—as of 1972 did not use the data set reported in Historic Census, but not BLS
12025-Dade County	Available	Available	(NA)	(NA)	(NA)	Dade County, Florida, officially changed its legal name to Miami-Dade County, Florida, effective November 13, 1997. To maintain the alphanumeric sequence of counties, Miami-Dade County will have a code of 086 for FIPS 6-4Combined 12025 data with 12086 data http://www.census.gov/geo/www/tiger/ctychng.html

A. Missing Data for Heckit Probit Continued

12086-Miami Dade County	(NA)	(NA)	Available	Available	Available	Dade County, Florida, officially changed its legal name to Miami-Dade County, Florida, effective November 13, 1997. To maintain the alphanumeric sequence of counties, Miami-Dade County will have a code of 086 for FIPS 6-4 Combined 12025 data with 12086 data http://www.census.gov/geo/www/tiger/ctychng.html
30113/Yellow Stone National Park	Available	Available	(NA)	(NA)	(NA)	Yellowstone National Park, Montana, shown as a county equivalent in 1990, is legally part of Gallatin County and Park County. This eliminates Yellowstone National Park (FIPS Code 30113) as a county equivalent.
Menominee County, WI 55078 Menominee Indian Reservation	Available	Available	(NA)	Available	Available	Found Data from the Census 1980
Shawano County, WI 55115	Available	Available	(NA)	Available	Available	Found Data from the Census 1980

B. Missing Data for Population, Education and Unemployment

FIPS Code	HS_or_less	Unemployment	Steps Taken	Sources
12086	1990-1996 (NA)	Available	Used Data from Dade County 12025 for 1990-1996	http://factfinder.census.gov
22051/Jefferson Parish, LA	Available	2005-2006 (NA)	Some Parishes are missing data due to Hurricane Katrina; used 2004 values for 2005-2006	http://www.bls.gov/lau/#tables
22071/Orleans Parish, LA	Available	2005-2006 (NA)	“”	“”
22075/Plaquemines Parish, LA	Available	2005-2006 (NA)	“”	“”
22087/St. Bernard Parish, LA	Available	2005-2006 (NA)	“”	“”
22089/St. Charles Parish, LA	Available	2005-2006 (NA)	“”	“”
22095/St. John the Baptist Parish, LA	Available	2005-2006 (NA)	“”	“”
22103/St. Tammany Parish, LA	Available	2005-2006 (NA)	“”	“”

C. Virginia Independent Cities

These independent cities:		Were combined with these counties:	
FIPS	**Name**	**FIPS**	**Name**
51510	Alexandria	51013	Arlington
51515	Bedford	51019	Bedford
51520	Bristol	51191	Washington
51530	Buena Vista	51163	Rockbridge
51540	Charlottesville	51003	Albemarle
51560	Clifton Forge	51005	Alleghany
51570	Colonial Heights	51041	Chesterfield
51580	Covington	51005	Alleghany
51590	Danville	51143	Pittsylvania
51595	Emporia	51081	Greensville
51600	Fairfax	51059	Fairfax
51610	Falls Church	51059	Fairfax
51620	Franklin	51175	Southampton
51630	Fredericksburg	51177	Spotsylvania
51640	Galax	51035	Carroll
51650	Hampton	51199	York
51660	Harrisonburg	51165	Rockingham
51670	Hopewell	51149	Prince George
51678	Lexington	51163	Rockbridge
51680	Lynchburg	51031	Campbell
51683	Manassas	51153	Prince William
51685	Manassas Park	51153	Prince William
51690	Martinsville	51089	Henry
51700	Newport News	51199	York
51710	Norfolk	51550	Chesapeake
51720	Norton	51195	Wise
51730	Petersburg	51053	Dinwiddie
51735	Poquoson	51199	York
51740	Portsmouth	51550	Chesapeake
51750	Radford	51121	Montgomery
51760	Richmond	51087	Henrico
51770	Roanoke	51161	Roanoke
51775	Salem	51161	Roanoke
51780	South Boston	51083	Halifax
51790	Staunton	51015	Augusta
51820	Waynesboro	51015	Augusta
51830	Williamsburg	51095	James City (County)
51840	Winchester	51069	Frederick

The BLS data contains separate info for the independent cities. The census does not so the information for the pop_1980 and pop_1990 was gathered from the Virginia Historical Census.
http://www.ers.usda.gov/Briefing/Rurality/Typology/Methods/
Source: U.S. Census of Population, 1790-2000. Prepared by Weldon Cooper Center for Public Service
http://www.coopercenter.org/demographics/census-data

D. Missing Data from the BLS

Missing Counties

FIPS	Beale Codes	County Name
5013	9	Calhoun, AR
5025	8	Cleveland, AR
8014	1	Broomfield, CO
8047	8	Gilpin, CO
8053	9	Hinsdale, CO
8109	9	Saguache, CO
8111	9	San Juan, CO
12043	8	Glades, FL
12125	8	Union, FL
13007	8	Baker, GA
13101	9	Echols, GA
13125	9	Glascock, GA
13265	9	Taliaferro, GA
16025	9	Camas, ID
16033	9	Clark, ID
17069	9	Hardin, IL
17151	9	Pope, IL
20187	9	Stanton, KS
20209	1	Wyandotte, KS
21063	8	Elliott, KY
21165	9	Menifee, KY
21189	9	Owsley, KY
21201	9	Robertson, KY
22023	8	Cameron, LA
22075	1	Plaquemines,LA
26083	9	Keweenaw, MI
26085	9	Lake, MI
28009	8	Benton, MS
28055	9	Issaquena, MS
28063	9	Jefferson, MS
30033	9	Garfield, MT
30037	8	GoldenValley,MT
30045	8	Judith Basin, MT
30069	9	Petroleum, MT
30103	8	Treasure, MT
30109	9	Wibaux, MT
31005	9	Arthur, NE
31009	9	Blaine, NE
31057	9	Dundy, NE
31073	9	Gosper, NE
31077	9	Greeley, NE
31085	9	Hayes, NE

D. Missing Data Continued

31091	9	Hooker, NE
31103	9	Keya Paha, NE
31113	9	Logan, NE
31115	9	Loup, NE
31117	9	McPherson, NE
31161	9	Sheridan, NE
31165	9	Sioux, NE
31171	9	Thomas, NE
31183	9	Wheeler, NE
32009	9	Esmeralda, NV
35021	9	Harding, NM
38007	9	Billings, ND
38013	9	Burke, ND
38063	8	Nelson, ND
38065	8	Oliver, ND
38083	9	Sheridan, ND
38087	9	Slope, ND
41021	9	Gilliam, OR
41069	9	Wheeler, OR
46017	9	Buffalo, SD
46113	7	Shannon, SD
46137	9	Ziebach, SD
47067	9	Hancock, TN
47175	9	Van Buren, TN
48033	9	Borden, TX
48045	9	Briscoe, TX
48173	8	Glasscock, TX
48261	9	Kenedy, TX
48263	9	Kent, TX
48269	9	King, TX
48301	9	Loving, TX
48311	9	McMullen, TX
48393	9	Roberts, TX
48443	9	Terrell, TX
49009	9	Daggett, UT
49031	9	Piute, UT
51036	2	Charles City, VA
51053	2	Dinwiddie County,VA
51149	2	PrinceGeorgeCounty,VA
51560	6	Clifton Forge City, VA
53023	9	Garfield, WA
53069	9	Wahkiakum, WA
55037	9	Florence, WI
55078	9	Menominee, WI

II. Code Book

Variable Name	Variable Full Name	Code/Description	Source	Additional information
bealecode (changed from rural to be more descriptive)	Ruralality	'BealeCode' Rural-Urban Continuum Codes Metro counties: 1 Counties in metro areas of 1 million population or more 2 Counties in metro areas of 250,000 to 1 million population 3 Counties in metro areas of fewer than 250,000 population Nonmetro counties: 4 Urban population of 20,000 or more, adjacent to a metro area 5 Urban population of 20,000 or more, not adjacent to a metro area 6 Urban population of 2,500 to 19,999, adjacent to a metro area 7 Urban population of 2,500 to 19,999, not adjacent to a metro area 8 Completely rural or less than 2,500 urban population, adjacent to a metro area 9 Completely rural or less than 2,500 urban population, not adjacent to a metro area	http://www.ers.usda.gov/briefing/rurality/ruralurbcon/	Rural-Urban Continuum Codes or Beale Codes form a classification scheme that distinguishes metropolitan (metro) counties by the population size of their metro area, and nonmetropolitan (nonmetro) counties by degree of urbanization and adjacency to a metro area or areas. The metro and nonmetro categories have been subdivided into three metro and six nonmetro groupings, resulting in a nine-part county codification. The codes allow researchers working with county data to break such data into finer residential groups beyond a simple metro-nonmetro dichotomy, particularly for the analysis of trends in nonmetro areas that may be related to degree of rurality and metro proximity.

Code Book Continued

Rural	Rural or Urban Counties	Recoded Rural urban continuum codes dummy 0 or 1 for if bealecode=0 or bealecode=1 or bealecode=2 or bealecode=3 then rural=0; else rural=1	http://www.ers.usda.gov/briefing/rurality/ruralurbcon/	1-Counties in metro areas of 1 million population or more; 2-Counties in metro areas of 250,000 to 1 million population; 3-Counties in metro areas of fewer than 250,000 population

Nyear	Time trend	if year=1990 then nyear=1;
		if year=1991 then nyear=2;
		if year=1992 then nyear=3;
		if year=1993 then nyear=4;
		if year=1994 then nyear=5;
		if year=1995 then nyear=6;
		if year=1996 then nyear=7;
		if year=1997 then nyear=8;
		if year=1998 then nyear=9;
		if year=1999 then nyear=10;
		if year=2000 then nyear=11;
		if year=2001 then nyear=12;
		if year=2002 then nyear=13;
		if year=2003 then nyear=14;
		if year=2004 then nyear=15;
		if year=2005 then nyear=16;
		if year=2006 then nyear=17;
		if year=2007 then nyear=18;
		if year=2008 then nyear=19;
			no source	the general course or prevailing tendency

Store	Presence of a Wal-Mart	Number of stores each year	http://walmartstores.com/GlobalWMStoresWeb/navigate.do?catg=130&yearID=131&monthID=138.	a count of Wal-Mart stores over time
rural store	Rural counties with a Wal-Mart	ruralstore= store*rural, this is an interaction, 0=urbanstore; else=ruralstore	no source	an interaction of counties which are rural and have a Wal-Mart

Code Book Continued

avg_wkwage	Average weekly wage NAICS 44-45	Dollar value rounded to the nearest dollar	http://www.bls.gov/cew/cewbultn08.htm#Wages	Average annual wages per employee for any given industry are computed by dividing total annual wages by annual average employment. A further division by 52 yields average weekly wages per employee. Annual pay data only approximate annual earnings, because an individual may not be employed by the same employer all year or may work for more than one employer at a time. Average weekly or annual pay is affected by the ratio of full-time to part-time workers, as well as by the numbers of individuals in high- and low-paying occupations. When comparing average pay levels among States and industries, data users should take these factors into consideration.
perchange_median income	Percent change median income	((Median income 1989-Median income 1979)/Median income 1979))*100	http://www.census.gov/hhes/www/income/data/historical/county/index.html	
perchange_pop	Percent change in pop 1980-90	((Pop 1990-Pop 1980)/Pop 1980))*100	http://www.bea.gov/regional/reis/	

Code Book Continued

pop80 (changed from pop_1980 by G. Artz on 10/27/2010)	population level 1980	number of people at the county level for 1980		
Wagedif	the difference of avg_wkwage over time	wagedif=dif(avg_wkwage); if year=1990 then wagedif=.;	1991-1990, 1992-1991, and so on	
Unempdif	the difference of unemp over time	unempdif=dif(unemp); if year=1990 then unempdif=.;	1991-1990, 1992-1991, and so on	
Popdif	the difference of pop_1000 over time	popdif=dif(pop_1000); if year=1990 then popdif=.;	1991-1990, 1992-1991, and so on	

Vita

Jessica Elaine Smith was born in Richmond Heights, Missouri. She completed her undergraduate education at the University of Missouri-Columbia in Animal Sciences. She was a Peace Corps Volunteer in Niger, West Africa from 2002-2004. She served as a Crisis Corps Volunteer from 2006-2007 for the UNFAO in Namibia. She continued to work in Namibia in agricultural economic development until starting her Master's program studies at the University of Missouri-Columbia.