NBER WORKING PAPER SERIES

THE EFFECTS OF WAL-MART ON LOCAL LABOR MARKETS

David Neumark
Junfu Zhang
Stephen Ciccarella

Working Paper 11782
http://www.nber.org/papers/w11782

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2005

Neumark is a Senior Fellow at the Public Policy Institute of California (PPIC), Professor of Economics at the University of California—Irvine, Research Associate at the NBER, and Research Fellow at IZA. Zhang is a Research Fellow at PPIC. Ciccarella is a Research Associate at PPIC. We are grateful to Emek Basker, Colin Cameron, Hilary Hoynes, Chris Jepsen, Howard Shatz, Brandon Wall, and seminar participants at PPIC and UC-Davis for helpful comments. We are also grateful to Wal-Mart for providing data on store locations and opening dates, and to Emek Basker for providing her Wal-Mart data set and other code; any requests for Wal-Mart data have to be directed to them. Despite Wal-Mart having supplied some of the data used in this study, the company has provided no support for this research, and had no role in editing or influencing the research as a condition of providing these data. The views expressed are those of the authors, and not those of PPIC, Wal-Mart, or the National Bureau of Economic Research.

The Effects of Wal-Mart on Local Labor Markets
David Neumark, Junfu Zhang, and Stephen Ciccarella
NBER Working Paper No. 11782
November 2005
JEL No. J2, J3, R1

ABSTRACT

We estimate the effects of Wal-Mart stores on county-level employment and earnings, accounting for endogeneity of the location and timing of Wal-Mart openings that most likely biases the evidence against finding adverse effects of Wal-Mart stores. We address the endogeneity problem using a natural instrumental variable that arises from the geographic and time pattern of the opening of Wal-Mart stores, which slowly spread out from the first stores in Arkansas. In the retail sector, on average, Wal-Mart stores reduce employment by two to four percent. There is some evidence that payrolls per worker also decline, by about 3.5 percent, but this conclusion is less robust. Either way, though, retail earnings fall. Overall, there is some evidence that Wal-Mart stores increase total employment on the order of two percent, although not all of the evidence supports this conclusion. There is stronger evidence that total payrolls per person decline, by about five percent in the aggregate, implying that residents of local labor markets earn less following the opening of Wal-Mart stores. And in the South, where Wal-Mart stores are most prevalent and have been open the longest, the evidence indicates that Wal-Mart reduces retail employment, total employment, and total payrolls per person.

David Neumark
Public Policy Institute of California
500 Washington Street, Suite 800
San Francisco, CA 94111
and NBER
neumark@ppic.org

Stephen Ciccarella
Public Policy Institute of California
500 Washington Street, Suite 800
San Francisco, CA 94111
ciccarella@ppic.org

Junfu Zhang
Public Policy Institute of California
500 Washington Street, Suite 800
San Francisco, CA 94111
zhang@ppic.org

I. Introduction

Wal-Mart is more than just another large company. It is the largest corporation in the world, with total revenues of $285 billion in 2005. It employs over 1.2 million workers in the United States, at about 3,600 stores.[1] To put this in perspective, the Wal-Mart workforce represents just under one percent of total employment in the United States, and just under ten percent of retail employment. It exceeds the number of high school teachers or middle school teachers, and is just under the size of the elementary school teacher workforce. Wal-Mart is reported to be the nation's largest grocer, with a 19 percent market share, and its third-largest pharmacy, with a 16 percent market share (Bianco and Zellner, 2003).

During the past two decades, as Wal-Mart started to compete with a wider range of retailers and pushed into more areas, it increasingly encountered resistance from local communities. Opponents of Wal-Mart have tried to block its entry on many grounds, including the prevention of urban sprawl, preservation of historical culture, protection of the environment and "main-street" merchants, and avoidance of road congestion.[2] Yet two of the most commonly-heard criticisms are that Wal-Mart eliminates more jobs than it creates for a community and that Wal-Mart's wage levels pull down standards for all workers—not simply creating low-wage jobs, but driving down wages.[3] Wal-Mart executives dispute these claims. For example, its Vice President Bob McAdam has argued that there are many locations where Wal-Mart helps revitalize rundown communities and creates jobs in other businesses in addition to what Wal-Mart itself offers (PBS, 2004). And Lee Scott, Wal-Mart President and CEO, has asserted that "there are some who say that Wal-Mart's wages and benefits have some kind of negative impact on wages across the board. That's just plain wrong" (Scott, 2005). Of course Wal-Mart offers other potential benefits in the form of lower prices for consumers (Basker, forthcoming; Hausman and Leibtag, 2004).

The argument that Wal-Mart not only creates low-wage jobs but drives down wages (or more importantly, earnings) of other workers is politically potent, fueling charges that Wal-Mart increases the

---

[1] See http://www.walmartfacts.com/newsdesk/wal-mart-fact-sheets.aspx#a125 (as of September 8, 2005).
[2] See, for example, Bowermaster (1989), Rimer (1993), Nieves (1995), Kaufman (1999), Ingold (2004), and Jacobs (2004).
[3] See, for example, Quinn (2000), Norman (2004), and Wal-Mart Watch (2005).

burden on taxpayers. A report by the Democratic Staff of the Committee on Education and the Workforce of the U.S. Congress (Miller, 2004) claims that because of Wal-Mart's low wages, an average Wal-Mart employee costs federal taxpayers an extra $2,103 in the form of tax credits or deductions, or public assistance such as healthcare, housing, and energy assistance. There are many heroic assumptions needed to construct such estimates, and this is not the place to dissect them. However, a key implicit assumption is that in the absence of Wal-Mart, employees of the company would have higher-paying jobs, rather than, for example, no jobs. Thus, whether Wal-Mart's entry into a labor market increases or decreases employment is of interest. On the other hand, if Wal-Mart openings put downward pressure on wages of other employers, then estimates based only on the levels of wages paid by Wal-Mart could be understated.[4] As a consequence, the effects of Wal-Mart on earnings are perhaps the key question.

In this paper, we seek to provide a definitive answer to two central questions about the effects of Wal-Mart on local labor markets: Does Wal-Mart create or eliminate jobs? And does Wal-Mart indeed push down earnings? We believe that our evidence improves substantially on existing studies of these and related questions, most importantly by implementing an identification strategy that accounts for the endogeneity of store location and timing and how these may be correlated with future changes in earnings or employment. Indeed, it has been suggested that Wal-Mart's explicit strategy was to locate in small towns where the population growth was increasing (Slater, 2003, pp. 92). If Wal-Mart tends to enter fast-growing areas in booming periods, then we might expect to observe employment and wages or earnings rising in apparent response to Wal-Mart's entry, even when the stores actually have negative effects on both outcomes.

Our identification strategy is driven by a systematic pattern in the openings of Wal-Mart stores. Sam Walton, the founder of Wal-Mart, opened the first Wal-Mart store in 1962 in Rogers, Arkansas, in Benton County. Five years later, Wal-Mart had 18 stores with $9 million of annual sales. Wal-Mart first grew into a local chain store in the northwest part of Arkansas. It then spread to adjacent states such as Oklahoma, Missouri, and Louisiana. From there, it kept expanding to the rest of the country after closer

[4] Dube and Jacobs (2004) report similar types of estimates of taxpayer burden in California. They also consider the implications of wages falling at other employers.

markets were largely saturated (Slater, 2003, pp. 28-29). The relationship between Wal-Mart stores' opening dates and their distance to the headquarters is primarily a result of Wal-Mart's growth strategy. In his autobiography, Sam Walton describes the expansion of Wal-Mart as follows:

> "[Our growth strategy] was to saturate a market area by spreading out, then filling in. In the early growth years of discounting, a lot of national companies with distribution systems already in place—Kmart, for example—were growing by sticking stores all over the country. Obviously, we couldn't support anything like that. … We figured we had to build our stores so that our distribution centers, or warehouses, could take care of them, but also so those stores could be controlled. We wanted them within reach of our district managers, and of ourselves here in Bentonville, so we could get out there and look after them. Each store had to be within a day's drive of a distribution center. So we would go as far as we could from a warehouse and put in a store. Then we would fill in the map of that territory, state by state, county seat by county seat, until we had saturated that market area. … So for the most part, we just started repeating what worked, stamping out stores cookie-cutter style" (Walton, 1992, pp. 110-111).

Wal-Mart's practice of growing by "spreading out" geographically means that distance from Benton County, Arkansas, and time—and more specifically their interaction—is a good predictor of when and where stores opened.[5] Although a standard model of employment or earnings might well have time effects, as well as county effects (which are perfectly collinear with distance from Benton County), there is good reason to believe that the distance-time interaction can be excluded, and hence serve as a valid instrument for store openings. Thus, the key innovation in this paper is to instrument for the opening of Wal-Mart stores with interactions between time and the distance between Wal-Mart host counties and Benton County, Arkansas, where Wal-Mart headquarters are located.

II. Literature Review

There are a number of studies that address claims about Wal-Mart's impacts on local labor markets. However, we regard much of this literature as uninformative. First, some of the existing work is by advocates for one side or the other in local political disputes regarding Wal-Mart's entry into a particular market. These studies are often hastily prepared, plagued by flawed methods and arbitrary assumptions, and sponsored by interested parties such as Wal-Mart itself, its competitors, or union groups (e.g., Bianchi and Swinney, 2004; Freeman, 2004; and Rodino Associates, 2003), and can hardly be expected to provide impartial evidence on Wal-Mart's effects. Hence, they are not summarized here.

[5] For example, Wal-Mart's expansion did not reach California until 1990. It first entered New England in 1991. In 1995 Wal-Mart opened its first store in Vermont and finally had a presence in all 48 contiguous states.

There is also an academic literature on the impact of Wal-Mart stores, focusing on the effects of Wal-Mart openings on local employment, retail prices and sales, poverty rates, and the concentration of the retailing industry, as well as the impact on existing businesses. This research is limited by three main factors: the restriction of much of it to small regions (often even sub-areas of small states); its lack of focus on employment and earnings effects; and its failure to account for the endogeneity of Wal-Mart locations, either at all or (in our view) adequately.

Many of these studies, especially the early ones, focus on the effects of Wal-Mart at the regional level, spurred by the expansion of Wal-Mart into a particular region. The largest number of studies focus on the effects of Wal-Mart on retail businesses and sales, rather than on employment and earnings. The earliest study, which is typical of much of the research that has followed, is by Stone (1988). He defines the "pull factor" for a specific merchandise category as the ratio of per capita sales in a town to the per capita sales at the state level, and examines the changes in the pull factor for different merchandise categories in host and surrounding towns in Iowa after the opening of Wal-Mart stores. Stone finds that in host towns, pull factors for total sales and general merchandise (to which all Wal-Mart sales belong) rise after the arrival of Wal-Mart. Pull factors for eating and drinking and home furnishing also go up because Wal-Mart brings in more customers. However, pull factors for grocery, building materials, apparel, and specialty stores decline, presumably due to direct competition from Wal-Mart. He also finds that small towns surrounding Wal-Mart towns suffer a larger loss in total sales compared to towns that are further away.[6] Related results for other regions—which generally, although not always, point to similar conclusions—are reported in Keon, et al. (1989), Barnes, et al. (1996), Davidson and Rummel (2000), and Artz and McConnon (2001). All of these studies use administrative data, and employ research designs based on before-and-after comparisons in locations in which Wal-Mart stores did and did not open.[7]

---

[6] Stone's study was updated regularly (see, for example, Stone, 1995, 1997), but its central message remained the same: Wal-Mart pulls more customers to the host town, hurts its local competitors, but benefits some other local businesses that do not directly compete with it. Using the same methods, Stone, et al. (2002) show similar results regarding the effects of Wal-Mart Supercenters on existing businesses in Mississippi.

[7] A couple of studies rely on surveys of local businesses rather than administrative data. McGee (1996) reports results from a small-scale survey of small retailers in five Nebraska communities conducted soon after Wal-Mart stores entered. He finds that 53 percent of the responding retailers reported negative effects of Wal-Mart's arrival

The studies reviewed thus far do not address the potential endogeneity of the location and timing of Wal-Mart's entry into a particular market. In addition, these studies do not focus on the key questions with which this paper is concerned—the effects of Wal-Mart on earnings and employment. A few studies attempt to rectify these shortcomings. Ketchum and Hughes (1997), studying counties in Maine, recognize the problem of the endogenous location of Wal-Mart stores in faster-growing regions. They attempt to estimate the effects of Wal-Mart on employment and earnings using a difference-in-difference-in-differences (DDD) estimator that compares changes in retail employment and earnings over time in counties in which Wal-Mart stores did and did not locate, compared to changes for manufacturing and services. However, virtually none of their estimated changes are statistically significant, so one cannot learn much from these data (and the data appear very noisy). More important, their approach does not address the key endogeneity questions of whether Wal-Mart location decisions were based on prior trends in retail that were already different, or anticipated changes after stores opened (despite the authors posing these questions).[8] Hicks and Wilburn (2001), studying the impact of Wal-Mart openings in West Virginia, estimate positive impacts of Wal-Mart stores on retail employment and the number of retail firms. They do not explicitly account for endogeneity, although they do address the issue. In particular, they report evidence suggesting that Wal-Mart location decisions are independent of long-term economic growth rates of individual counties in their sample, and that current and lagged growth have no significant effect on Wal-Mart's decision to enter. However, these results come from a very small sample of 14 counties in the state (with seven Wal-Mart stores), and they do not explicitly address endogeneity with respect to future growth. The latter, in particular, could generate apparent positive impacts of Wal-Mart stores.

---

on their revenues while 19 percent indicated positive effects. In a survey of Nebraska and Kansas retailers, Peterson and McGee (2000) find that less than a third of the businesses with at least $1 million in annual sales reported a negative effect after Wal-Mart's arrival, while close to one half of the businesses with less than $1 million in annual sales indicated a negative effect, with negative effects most commonly reported by small retailers in central business districts. The research design in these surveys fails to include a control group capturing changes that might have occurred independently of Wal-Mart openings. In addition, reported assessments by retailers may not reflect actual effects of these openings.

[8] The second question can only be addressed via an instrumental variables approach, and the first requires looking at changes in growth rates, not changes in levels—which is all their study does.

In more recent work, Basker (2005) studies the effects of Wal-Mart on employment (but not earnings) using nationwide data. Basker attempts to account explicitly for endogeneity by instrumenting for the actual number of stores opening in a county in a given year with the planned number. The latter is based on numbers that Wal-Mart assigns to stores when they are planned; according to Basker, these store numbers indicate the order in which the openings were planned to occur. She then combines these numbers with information from Wal-Mart *Annual Reports* to measure planned openings in each county and year. Her results indicate that county-level retail employment grows by about 100 in the year of Wal-Mart entry, but declines to a gain of about 50 jobs in five years as other retail establishments contract or close. In the meantime, possibly because Wal-Mart streamlines its supply chain, wholesale employment declines by 20 jobs. Thus, on net, Wal-Mart stores appear to create 30 jobs for the host county.[9]

The principal problem with this identification strategy, however, is that the instrument is unconvincing. For the instrument to be valid, two conditions must hold. The first is that planned store openings should be correlated with (predictive of) actual openings; this condition is not problematic. The second condition is that the variation in planned openings generates exogenous variation in actual openings that is uncorrelated with the unobserved determinants of employment that endogenously affect location decisions. This second condition holds if we assume, to quote Basker, that "the number of *planned* Wal-Mart stores … for county *j* and year *t* is independent of the error term … and *planned* Wal-Mart stores affect retail employment per capital only insofar as they are correlated with the *actual* construction of Wal-Mart stores" (2005, p. 178). The second part of the assumption is potentially problematic; actual stores should, of course, be the driving influence, although planned stores—even if they do not materialize or do so only with delay—may still affect decisions of other businesses. The first part of the assumption is a more serious concern, though, as it seems most likely that planned openings

[9] Using the same instrumental variables strategy, Basker (forthcoming) estimates the effects of Wal-Mart entry on prices of consumer goods at the city level, finding long-run declines of 8-13 percent in prices of several products including aspirin, detergent, Kleenex, and toothpaste, although it is less clear to us why location decisions would be endogenous with respect to price. Ordinary least squares (OLS) estimation also finds long-run negative effects of Wal-Mart on prices for nine out of ten products, although only three are significant and all are smaller than the IV results. Hausman and Leibtag (2004) study the effects of Wal-Mart on food prices.

will reflect the same unobserved determinants that drive endogenous location as are reflected in actual openings, and we cannot think of an argument to the contrary (nor does Basker offer one).[10]

Another potential problem is that there is substantial measurement error regarding store openings in Basker's data (see Appendix A). This measurement error arises because Basker (and Goetz and Swaminathan, 2004, discussed below), had to collect information about Wal-Mart locations and opening dates from a variety of sources including Wal-Mart editions of the *Rand McNally Road Atlas*, annual editions of the *Directory of Discount Department Stores*, and Wal-Mart *Annual Reports*, which together do not always pin down the timing of each store opening. Basker also motivates the instrumental variable (IV) as correcting for bias from measurement error in the actual opening dates of Wal-Mart stores. But the same argument against the validity of planned openings as an instrument applies.

Using various data sources, Goetz and Swaminathan (2004) study the relationship between Wal-Mart openings between 1987 and 1998 and county poverty rates in 1999, conditional on 1989 poverty rates (as measured in the 1990 and 2000 Censuses of Population). They also use an IV procedure to address the endogeneity of Wal-Mart entry, instrumenting for Wal-Mart openings during 1987-1998 in an equation for county poverty rates in 1999. Their IVs include an unspecified pull factor, access to interstate highway, earnings per worker, per capita property tax, population density, percentage of households with more than three vehicles, and number of female-headed households. The results suggest that county poverty rates increase when Wal-Mart stores open, perhaps because Wal-Mart lowers

---

[10] In this case if the OLS estimate of the effect of Wal-Mart stores on retail employment is biased upward, the IV estimate will also be biased upward, possibly by more than the OLS estimate. To take a simple example, suppose that the simultaneous model for employment (E), planned openings (P), and actual openings (A) is

$$E = \alpha A + \varepsilon$$
$$P = \beta E + \upsilon$$
$$A = \gamma E + \upsilon + \eta,$$

with ε, υ, and η uncorrelated, and υ and η uncorrelated with E. Employment depends on actual openings. Both planned and actual openings are simultaneously determined with employment, but actual openings reflect additional information captured in η, orthogonal to υ either because it reflect additional information not in the information set when P was determined, or simply random variation due to construction delays, zoning disputes, etc., which make actual and planned openings deviate. In this case the asymptotic bias in the OLS estimate of α is

$$(\gamma/(1\text{-}\alpha\gamma))\cdot\mathrm{Var}(\varepsilon)/\mathrm{Var}(A)$$

which is positive assuming $\alpha < 0$ (Wal-Mart reduces employment) and $\gamma > 0$ (employment encourages Wal-Mart openings). The IV estimate of α using P as an instrument for A is asymptotically biased upward by

$$(\beta/(1\text{-}\alpha\gamma))\cdot\mathrm{Var}(\varepsilon)/\mathrm{Cov}(A,P).$$

We know that $\mathrm{Cov}(A,P) < \mathrm{Var}(A)$. Thus, if $\beta > \gamma$, then the upward bias in the IV estimate is larger, and this holds for some values of $\beta < \gamma$. Given that A is based on more information than P (contained in η), we would expect $\beta > \gamma$.

earnings (although the authors offer other explanations as well). However, why the IVs should affect Wal-Mart openings only, and not changes in poverty directly (conditional on Wal-Mart openings), is not the slightest bit clear. And it is easy to construct stories in which invalid exclusion restrictions would create biases towards the finding that Wal-Mart openings increase poverty.[11]

Our research addresses the four principal shortcomings of the existing research on the effects of Wal-Mart on local labor markets. First, we estimate the effects of Wal-Mart openings on earnings, which is a central question. Second, we have—we believe—a far more convincing strategy to account for the potential endogeneity of Wal-Mart openings, which seems most likely to bias upward any estimated effects of Wal-Mart stores on earnings and employment. Third, we are able to use administrative data on Wal-Mart openings that eliminate the measurement error in recent work. And finally, we use a data set that is national in scope.

III. Data

Our empirical analysis relies on data from various sources, and computations of our own, which we describe in this section.

Employment and Payroll Data

Employment and payroll data are drawn from the U.S. Census Bureau's County Business Patterns (CBP). CBP is an annual series that provides economic data by industry and county. The series includes most economic activity, but excludes data on self-employed individuals, employees of private households, railroad employees, agricultural production workers, and most government employees. CBP data are extracted from the Business Register, which is a U.S. Census Bureau file of all known single and multi-establishment companies in the United States. The Business Register includes payroll and employment

[11] For example, consider the use of the number of female-headed households as an IV, and suppose that this variable is positively correlated with changes in poverty rates (because of rising inequality over this period), and also positively correlated with Wal-Mart openings (because they locate in lower-income areas). In this case the IV estimate of the effect of Wal-Mart openings on changes in poverty rates is biased upward because of the positive correlation between the instrument (female-headed households) and the error term in the equation for the change in the poverty rate. In addition, although the authors do not specify how they construct their pull factor, we assume it is similar to the measure described above—a ratio of county to statewide retail sales. Given that this is a dependent variable in other studies of the effects of Wal-Mart, it is hard to justify using it as an IV for Wal-Mart openings.

data from multiple sources including Census Bureau surveys, the Internal Revenue Service, the Social Security Administration, and the Bureau of Labor Statistics.

Payroll in the CBP includes salaries, wages, reported tips, commissions, bonuses, vacation allowances, sick-leave pay, employee contributions to qualified pension plans, and taxable fringe benefits, and is reported before deductions for Social Security, income tax, insurance, etc.[12] It does not include profit or other compensation earned by proprietors or business partners. Payroll is reported on an annual basis. Employment covers all full- and part-time employees, including officers and executives, as of the pay period including March 12 of each year. Workers on leave are included, while proprietors and partners are not. Because the breakdown of employment into full-time and part-time workers is not available, a wage cannot be computed, even assuming a given number of hours for full-time and part-time workers. Nonetheless, by putting together information on changes in payroll and changes in employment, and by studying earnings per person as well as per worker, we are able to draw some meaningful conclusions regarding the local labor market effects of Wal-Mart on total earnings or earnings of subgroups of workers.[13]

We downloaded CBP data by two-digit SIC major group (three-digit NAICS subsector since 1998) from 1977 through 2002, from the Geospatial and Statistical Data Center at the University of Virginia (through 2001) and the U.S. Census (for 2002).[14] We began with 1977 because CBP data are not continuously machine-readable for the years 1964-1976, and ended with 2002 because that was the last year available. As explained below, most our analysis goes through 1995—a period for which our identification strategy is most compelling—although the CBP data extend further. We were interested in a few different industry aggregations of the retail sector (as well as totals). For the retail sector, we wanted to construct data for the sector as a whole, for General Merchandising in particular, and for retail as a whole excluding Eating and Drinking Places and Automotive Dealers and Gasoline Service Centers

---

[12] Given that it excludes non-taxable fringe benefits, of which the most important is health insurance, we refer to the CBP measure as earnings, rather than compensation.

[13] A good description of the CBP data and its differences relative to other data sources is available at http://www.calmis.ca.gov/FILE/ES202/CEW-About.htm (as of September 6, 2005).

[14] Available at http://fisher.lib.virginia.edu/collections/stats/cbp/ (as of April 5, 2005) and http://www.census.gov/epcd/cbp/download/cbpdownload.html (as of April 5, 2005).

(following Basker, 2005)—subsectors that are least likely to compete directly with Wal-Mart. However, some complications arise in working with the CBP, because by federal law no data can be published that would disclose the operations of an individual employer. As we look at more disaggregated subsets of industries, it is more likely that data are not disclosed and so our sample becomes smaller. Consequently, we constructed three samples with which we could consistently compare at least some industry sectors for the same set of observations, as follows:

- <u>A sample</u>: all county-year observations with complete (non-suppressed) employment and payroll data for aggregate retail, and in total.
- <u>B sample</u>: all observations in the A sample that also have complete data for the General Merchandising retail subsector (SIC 53 or NAICS 452) to which Wal-Mart belongs.
- <u>C sample</u>: all observations in both the A and B samples that also have complete data for all retail major groups (SIC 52-59 or NAICS 441-454) except for Eating and Drinking Places (SIC 58) and Automotive Dealers and Gasoline Service Centers (SIC 55 or NAICS 441, 447).

Wal-Mart Store Data

Wal-Mart provided us with administrative data on 3,066 Wal-Mart Discount Stores and Supercenters. The data set contains every Discount Store and Supercenter still in operation in the United States at the end of fiscal year 2005 (January 31, 2005).[15] Variables in the data set include store number, street address, city, state, ZIP code, square footage, store type, opening date (month/day/year), store hours (e.g., open 24 hours), latitude, longitude, county FIPS code, and Metropolitan Statistical Area (MSA) code for each store. After dropping stores in Alaska and Hawaii, we used 2,211 stores in our main analysis through 1995, and 2,795 stores when we use the full sample period through 2002. By 2005, Wal-Mart also had 551 Sam's Club stores in the United States (the first opened in 1983), on which we also obtained data, although the data were less complete (for example, lacking information on square

[15] We also received data identifying a small number of stores (54, as of 2005) that closed. We return to this issue in some of our robustness analyses. This data set also indicated some store relocations within counties, which are treated as continuing stores.

footage). We do most of our analysis considering the Wal-Mart stores other than Sam's Clubs, but also some analysis incorporating information on the latter.[16]

County-Year File

We constructed a county-year file by first collecting county names and FIPS codes for the 3,141 U.S. counties from the U.S. Census Bureau.[17] We then created time-consistent geographical areas which accounted for merges or splits in counties during the sample period. For counties that split during the sample period we maintained the definition of the original county, and for counties that merged during the sample period we created a single corresponding county throughout.[18] This leads to a file of 3,094 counties over 19 years (26 years when we use the full sample), to which we merge the CBP and Wal-Mart data.

County Population Data

County population data for each year were collected from the U.S. Census Population Estimates Archives.[19] These were assigned to the counties.

Distance Construction

We compiled latitude and longitude data for each county centroid from the U.S. Census Bureau's Census 2000 Gazetteer Files.[20] Using the Haversine distance formula, we constructed distance measures from each county to Wal-Mart headquarters in Benton County, Arkansas, for reasons explained below.[21]

IV. Empirical Approach and Identification

Basic Framework

We estimate county-level models for various measures of employment rates and payroll (denoted *Y*), as functions of exposure to Wal-Mart stores in the county (denoted *W*). Because counties may have

---

[16] Sam's Clubs are different because one has to become a member, like with Costco. The small number of Wal-Mart Neighborhood Markets are not included in any data we have, but the first one did not open until 1998, beyond the sample period used for most of our analysis.
[17] Downloaded from http://www.census.gov/datamap/fipslist/AllSt.txt (as of April 5, 2005).
[18] The code for creating consistent counties over time through 2000 was provided by Emek Basker, and supplemented by us.
[19] Downloaded from http://www.census.gov/popest/archives/ (as of April 5, 2005).
[20] Downloaded from http://www.census.gov/tiger/tms/gazetteer/county2k.txt (as of April 5, 2005).
[21] The Haversine distance formula is used for computing distances on a sphere. See Sinnott (1984).

systematic differences, which may be correlated with entry of Wal-Mart stores, all models include county fixed effects (denoted *C)*. In addition, the models include fixed year effects (denoted *A*) to account for aggregate changes in the dependent variables that might be correlated with growing exposure to Wal-Mart stores, which increases over time. So indexing by county j ($j = 1,\ldots,J$) and year t ($t = 1,\ldots,T$), letting each C_j denote a county dummy variable, and similarly each A_t denote a year dummy variable, and defining α, β, γ_j, and δ_t as scalar parameters, our generic model for each observation jt is:

$$(1) \qquad Y_{jt} = \alpha + \beta W_{jt} + \sum_j C_j \gamma_j + \sum_t A_t \delta_t + \varepsilon_{jt}.$$

We begin by reporting OLS estimates of this equation. Throughout, we report standard errors that allow for arbitrary heteroscedasticity of the error across counties and autocorrelations within counties.

Dependent and Independent Variables

We look at a number of dependent variables. We first estimate models for retail employment per 1,000 residents,[22] and retail payrolls per worker. To assess the overall effects of Wal-Mart, we also estimate models for total employment per 1,000 residents, and total payrolls per worker and per person.

We use two different variables to capture exposure to Wal-Mart stores, both of which account for how long stores have been open. One is simply the number of years the first store in a county has been open, and the second is the number of stores weighted by the number of years each has been open. A priori, it is not clear whether the key effect comes from the first store in a county or from all stores. But both of these specifications of the exposure variable allow the effects of Wal-Mart stores to evolve over time as the stores remain open. Note that we do not include other variables capturing economic conditions, because of their potential endogeneity.[23]

[22] We define the variable per 1,000 rather than per 100—which would be a conventional employment rate—so the regression estimates are simpler to report. Also, the CBP data provide a count of jobs, not the number of people employed (in one or more jobs), so this employment measures corresponds more to payroll-based measures of employment.

[23] Given the pattern of Wal-Mart's growth, and its extensive penetration, even state economic conditions should not be regarded as exogenous.

Endogeneity of Wal-Mart Location Decisions and Identification

Consistent estimation of equation (1) requires that $E(\varepsilon_{jt} \mid W_{jt}, C_1, ..., C_J, A_1, ..., A_T) = 0$, or that the residual ε_{jt} is uncorrelated with the Wal-Mart exposure variable, conditional on the other observables. This condition (or its asymptotic equivalent) is a strong one. In particular, if Wal-Mart location decisions are based in part on changes in employment or payroll outcomes, then this condition is violated. This endogeneity is natural, since it would be surprising if a company as successful as Wal-Mart did not make location decisions (which include the timing of store openings) in a systematic fashion related to current conditions and future prospects that might be related to both employment and payroll.

Our identification strategy in light of this potential endogeneity is based on the geographic pattern of Wal-Mart store openings over time. Figure 1 illustrates quite clearly how—through 1995—Wal-Mart stores spread out geographically over the United States, beginning in Arkansas as of 1965, expanding to Oklahoma, Missouri, and Louisiana by 1970, Tennessee, Kansas, Texas, and Mississippi by 1975, much of the South and the lower Midwest by 1985, more of the Southeastern seaboard, the plains, and the upper Midwest by 1990, and then, in turn, the Northeast, West Coast, and Pacific Northwest by 1995. After 1995, when the far corners of the country had been entered, there was only filling in of stores in areas that already had them. This pattern is what we would expect based on Wal-Mart's growth strategy of expanding outward by leapfrogging distribution centers, as cited earlier in the quote from Sam Walton.

This pattern of growth is significant because it generates an exogenous source of variation in the location and timing of Wal-Mart store openings. In particular, Figure 1 clearly indicates that time and distance from Arkansas (in particular, Benton County) predict where and when Wal-Mart stores will open. However, this does not necessarily imply that time and distance can serve as instrumental variables for exposure to Wal-Mart stores. To see this, suppose we appeal to Figure 1 to posit an equation for the generic exposure measure W of the form

(2) $$W_{jt} = \kappa + \lambda DIST_j + \sum_t A_t \mu_t + \eta_{jt},$$

where *DIST* is a measure of distance from Benton County. Then distance and time give us no identifying information, because time is already captured in the year fixed effects included in equation (1), as is *DIST*, which is perfectly collinear with the county fixed effects in equation (1).

However, the model in equation (2) has a specific implication that is belied by the data. In particular, the additivity of the distance and year effects in the model implies that conditional on distance from Benton County, exposure to Wal-Mart stores grows at the same rate everywhere. Figure 2 shows the locations of Wal-Mart openings rather than Wal-Mart stores. That is, a point appears in the maps only in the period during which a store opens (whereas Figure 1 shows all stores in existence). The maps through 1995 are of greatest interest, capturing the period during which Wal-Mart stores spread to the borders of the continental United States. It is clear that openings are first concentrated around Arkansas, and then by the 1981-1985 period are more concentrated further away from Arkansas and less concentrated there. This pattern becomes more obvious in the 1986-1990 and 1991-1995 maps, where openings thin considerably in the area of Wal-Mart's original growth, and move first to Florida, the Southeast, and the lower Midwest, and then to California, the upper Midwest, and the Northeast. As suggested by Figure 1, after 1995 Wal-Mart's growth consists more of filling in within areas to which it had already expanded.

The fact that the rate of openings slows considerably in the Southeast, for example, in the later years, and increases in areas further away—in a rough sense spreading out from Benton County like a wave (albeit irregular)—contradicts the implication of the additivity of distance and time effects in equation (2) that, conditional on distance from Benton County, exposure to Wal-Mart stores grows at the same rate everywhere. Instead, it implies that the model for exposure to Wal-Mart should have a distance-time interaction, with exposure growing more quickly with time in locations near Benton County in the early part of the sample period, but more quickly further away from Benton County later in the sample period. Because this relationship holds through 1995, when Wal-Mart had begun to saturate border areas, we restrict most of our analysis to this period, although we also report results using the full sample through 2002. The most flexible form of this interaction, expanding on equation (2), is

$$(3) \qquad W_{jt} = \kappa + \lambda DIST_j + \sum_t A_t \mu_t + \sum_t (DIST_j \times A_t)\theta_t + \eta_{jt}.$$

Given this specification, the endogenous effect of exposure to Wal-Mart is identified in equation (1) by using the distance-time interactions as instruments for exposure to Wal-Mart stores.[24,25] One might argue that equation (3) simply exploits a non-linearity to identify the effect of Wal-Mart stores. But this non-linearity arises naturally from Wal-Mart's growth strategy, rather than arbitrarily as an artifact of the data.

In addition to these interactions predicting exposure—which the maps in Figures 1 and 2 suggest, and which are borne out by the estimation—the other key condition for our identification strategy to be valid is that we can exclude distance-time interactions from the employment and payroll models. This is, of course, intimately related to the identification strategy, since—as equation (3) makes clear—identification of the effects of Wal-Mart stores comes from differential changes over time in the dependent variables for counties at different distances from Benton County, Arkansas. Given that points at a given distance from Benton County are located on a circle with Benton County at its center, it is not immediately clear why this condition should not be satisfied. A particular area—say, 500 miles straight east from Benton County—may have economic conditions or structure that differ from those in Benton County and hence also have a systematically different trend. But there is no obvious reason why all points on the circle with radius of 500 miles should exhibit systematically different trends relative to Benton County, aside from the effects of Wal-Mart that—roughly speaking—influence points at a common distance from Benton County at the same time.

---

[24] Of course, when we implement the IV estimator, the county dummy variables are included in the first-stage regression, subsuming the linear distance variable.

[25] Just after this paper was completed, we discovered very recent research done concurrently by Dube, et al. (2005), which also exploits the geographic pattern of Wal-Mart openings to identify their effect on retail earnings growth. This latter paper differs in a number of ways, including: using Basker's data on Wal-Mart stores rather than administrative data; focusing on the retail sector and various subcategories of it; restricting the analysis to the 1992-2000 period; and conditioning on total earnings, which we regard as a dependent variable potentially influenced by Wal-Mart stores, and hence endogenous. The discussion of the maps in Figures 1 and 2, which suggest that the distance-time quasi-experiment is unlikely to be powerful after 1995 or so, raises concerns about Dube, et al. using data only from 1992-2000. They rationalize this as speaking more to the contemporaneous effects of Wal-Mart stores, which is the current policy concern. However, we are concerned that the more recent data do not provide reliable identification of the effects of Wal-Mart stores.

On the other hand, county fixed effects are highly collinear with distance—in our case perfectly so because we measure distance based on county centroids. So if we thought that the model should include a complete set of interactions between county fixed effects and the year dummy variables, the distance-time interaction would provide no identifying information. Of course even in the absence of attempts to account for endogeneity, a model with unrestricted time effects for each county would be fully saturated, and not permit the identification of the effects of Wal-Mart stores. Thus, whether or not we are attempting to correct for endogeneity, we have to assume that if there are important region-specific time effects, they occur at a more aggregated level than counties. That is, we can allow for more restricted versions of geographic variation in the time pattern of change in the outcomes, or—more flexibly—simply estimate the models for the corresponding geographic regions of the country. The maps in Figures 1 and 2, however, indicate that the geographic pattern of changes over time in Wal-Mart openings occurs over relatively broad areas, so we can only allow for region-specific time effects at a relatively high level of aggregation—such as Census regions. We consider this issue in some detail below.

V. Results

Descriptive Statistics

Descriptive statistics for population, employment, and payroll are reported in Table 1. The first and third rows, for population and retail employment, indicate that counties in the B and C samples are larger, as we would expect since data are less likely to be suppressed for larger counties and counties with larger retail sectors. On the other hand, the sample sizes, plus the information on the number of counties represented in each sample reported in the second row, suggest that the C sample is quite selective and hence estimates from this subsample may be less reliable.[26]

Average aggregate retail employment per 1,000 residents is about 60 across the three samples. General merchandising employment, which is defined for the B and C samples, is just over one-tenth of this, while retail subsector employment, defined for the C sample, is just over half. Average payrolls per worker are about $14,000 in aggregate retail, $13,000 in general merchandising, and $15,000 in the retail

[26] The share of observations in the C sample is considerably lower than the share of counties represented, because counties are frequently in the C sample for some years but not others.

subsector. In contrast, average manufacturing payrolls per worker are about $28,000, and average payrolls per worker overall are about $23,000. On average across counties, the employment rate is approximately 0.26 (or 260 per 1,000).[27]

Table 2 provides some descriptive statistics on Wal-Mart stores. We first report these figures for all counties, followed by county-year observations with at least one Wal-Mart store (approximately 29 percent of the A sample, 30 percent of the B sample, and 33 percent of the C sample, reflecting the location of Wal-Mart stores in larger counties, on average). These descriptive statistics are useful in interpreting the regression results discussed below using the different exposure measures. For counties with at least one store, the average number of stores opened over the sample period ranges from 1.33 to 1.59 across the three samples; the minimum possible value is one because we condition on having at least one store open, and the means are greater than one because there are sometimes multiple stores in counties. In these county-year cells, at least one store had been open an average of 6.6 years in the A sample, 5.8 years in the B sample, and 5.6 years in the C sample, with the differences indicating that some of the earliest stores opened in smaller counties where more data are suppressed.

The final rows of the table report the distribution of number of stores per county (for counties with a store). In the A sample, around 81 percent of counties with Wal-Mart stores have only one store, about 12 percent have two stores, and around four percent have three stores. There is then a smattering of counties with more stores (with a maximum of 17 stores in Harris County, Texas, which includes Houston, not shown in the table).

First-Stage Estimates

We begin by describing some of the results for the first-stage (equation (3)). Specifically, Figure 3 shows estimates of the relationship between our weighted exposure measure and distance, year, and the distance-time interactions. When we actually implement two-stage least squares, we use county dummy variables in place of distance in equation (3). The second and third panels show the estimated year effects

[27] Overall, excluding categories not covered by CBP, the national employment rate is around 0.3. But if employment rates vary by county and counties differ in population, we would not expect the average employment rate across counties to match this exactly.

and distance-time interactions from this specification. However, to provide a simple look at the distance-exposure relationship, the first panel shows results from a specification corresponding to equation (3) exactly—i.e., using distance instead of county dummy variables.

In the first panel, for the exposure-distance relationship, the effect is linear, so we simply plot the relationship, which is strongly negative—as the maps in Figures 1 and 2 indicate. In the second panel, we plot the estimated coefficients of the year dummy variables. In this case the relationship is positive, as both years of exposure and the number of stores rise with time. Finally, in the third panel we plot the estimated coefficients of the interactions of the year dummy variables with distance, evaluated at distances of 100 and 1,000 miles from Benton County, Arkansas. The interaction is apparent, as the gap in exposure between near and far locations widens sharply with time, reflecting stores opening at the farther locations later in the sample period, but at the nearer locations early in the sample period. It also turns out that the F-statistic for the distance-time interactions in the first stage is always very large, as reported in the tables that follow.[28]

Baseline Estimates of Effects on Retail Sector Employment

Table 3 turns to results on the effects of Wal-Mart stores on employment in the retail sector, reporting estimates from our baseline specification (equations (1) and (3)). The units of the regression estimates require some interpretation. For example, the OLS estimate of 0.26 in the top row of column (1) indicates that an additional year since the first Wal-Mart store opened results in an additional 0.26 retail employees per 1,000 persons. However, it is perhaps more meaningful to interpret this estimate in light of the average number of years a store is open in the sample, which Table 2 reports as 6.56 years for the A sample. Multiplied by the coefficient estimate, this implies a 1.71 increase, or an increase of 3.1 percent (1.71/55). For the weighted exposure measure that accounts for all stores, the corresponding estimate of 0.12 should be multiplied by 8.21, implying a somewhat smaller increase of 0.99. Table 2 suggests that, as a crude rule of thumb, multiplying the estimates for the unweighted exposure measure by about six and

[28] Although not shown in Figure 3, when the estimation is extended through 2002, we find that the distance-time interactions flatten out after 1995. This reflects the issue discussed earlier—that the distance-time interaction is much less predictive of store location and timing after 1995.

the weighted exposure measure by about eight will provide approximately comparable magnitudes that capture the average effect of Wal-Mart stores at the county level. Turning to the evidence, the OLS estimates generally point to increases (or sometimes no effect) in the aggregate retail sector and in general merchandising. In the retail subsector (which excludes food and auto) there is no evidence of employment effects, except for evidence of a negative effect for the weighted exposure measure.

However, the IV estimates are more interpretable as causal effects of Wal-Mart openings on retail employment. The evidence for these estimates is relatively clear. Nearly all of the estimates point to employment declines in the aggregate retail sector, and this is always true for the A and B samples that are more representative. The weighted exposure measure—which we prefer because it accounts for number of stores open as well as the number of years since stores opened—indicates that, evaluated at the means, Wal-Mart openings reduce employment by 2.51 for the A sample, and 1.15 for the B sample (multiplying the estimates by 8.21 and 7.77, respectively, from Table 2), or between 1.9 and 4.6 percent. In contrast, the evidence generally points to increases in employment in general merchandising (for the B sample, of about the same magnitude as the decline in aggregate retail employment)—increases that we would expect since this is the sector in which Wal-Mart is classified. For example, focusing on the B sample, the weighted exposure estimate of 0.12 implies an increase of 0.93 workers per 1,000 persons. These estimates are consistent with Wal-Mart reducing overall retail employment, although shifting its composition toward general merchandising. This is what we might expect, as Wal-Mart stores compete with some retail businesses that are not in general merchandising, and if the efficiency gains—in terms of staffing—from Wal-Mart stores outweigh any growth of employment from opening a store.

Finally, the finding that the OLS estimates of employment effects for the aggregate retail sector and the retail subsector are generally positive, and the IV estimates negative, is consistent with Wal-Mart endogenously locating stores in places where retail growth is increasing, as we might expect. (We do not necessarily expect this type of endogeneity bias for the general merchandising sector, since Wal-Mart stores seem more likely to lead to the expansion of this sector, rather than to follow its expansion.) As shown in the table, for the A and B samples there is always statistically significant evidence of endogeneity bias in the aggregate retail sector.

Robustness Analyses

We next describe a series of robustness analyses. These are described here in detail with respect to estimation of the effects of Wal-Mart stores on retail employment, and are then also carried out for the IV estimation of the effects of Wal-Mart stores on the other outcome variables. In all of these robustness analyses, we focus on the specification using the weighted exposure measure. The first row of Table 4 repeats the baseline results from the previous table, for purposes of comparison, and the robustness analyses follow.

First, the baseline estimates do not allow for spatial autocorrelation across counties, but by clustering on counties allow for arbitrary autocorrelation across time. One flexible way to introduce spatial autocorrelation is to cluster the observations on state-year cells, which allows for arbitrary correlations in the error across all counties in the state-year cell, while ruling out autocorrelations over time across observations in the same county. As the table shows, this leads to similar standard errors and hence similar statistical conclusions.[29]

Second, we sharpen the identification by dropping counties that never had a Wal-Mart store during the sample period. In this case, identification of the effects of Wal-Mart stores comes only from the time-series variation in store openings for the set of counties that got a store. This provides a potentially "cleaner" control group that consists only of early observations on counties where stores later opened, rather than also including counties in which stores never did open. The estimates are similar, although the evidence of negative employment effects in the aggregate retail sector weakens, while remaining significant for the most representative (A) sample.

[29] Standard errors based on the i.i.d. assumption were about one-third as large for the employment measures we study, and about one-half as large for the payroll measures. We could also cluster just by state, which allows arbitrary correlations in the error across observations on any county (in a state) in any year. This generally led to standard errors two to three times as large as those clustering on county or on state and year, with most of the IV estimates insignificant as a result. This is a very conservative approach that imposes virtually no structure on the errors.

Third, we consider slightly different sample definitions from the baseline. We first report estimates dropping the small number of counties with stores that closed; these estimates are very close to the baseline estimates.[30] Following that, we report estimates extending through the entire period covered by the data—ending in 2002 rather than 1995—even though the identifying relationship between time, location, and store openings is strongest through 1995, as discussed earlier. As it turns out, for the longer period the evidence of employment declines in aggregate retail disappears, and actually reverses for the C sample.[31]

Fourth, we have thus far ignored information on Sam's Clubs. We do not necessarily want to treat these as equivalent to other Wal-Mart stores, so we study them in two ways. First, we omit all counties that had a Sam's Club at some point during the sample period; 93 percent of these counties also had a Wal-Mart. Second, we use the full sample, but recalculate the exposure measure treating each Sam's Club store like other Wal-Mart stores. As the table shows, the results are insensitive to either of these changes.

Fifth, to this point we have simply counted the number of stores in a county, and multiplied by the number of years they have been open, to get an exposure measure. One consideration is that Wal-Mart stores may have relatively more impact in a smaller market. We therefore constructed a per capita exposure measure, by dividing the weighted exposure measure by county population/100,000. We re-estimated the models using this exposure measure, and then rescaled the estimated coefficients and standard errors by the ratio of the mean of this new exposure variable for county-year observations with stores open to the mean of the variable used in the baseline specification for this same subsample. This rescaling ensures that the comparison with the baseline estimates is for the same "average" effect of Wal-Mart stores at the county level. A second consideration is that store size may vary, and the exposure measure should perhaps take this into account. For example, counties with smaller populations may get

[30] An alternative is to incorporate information on these counties resetting exposure to zero when a store closes. However, we are skeptical that store closings and openings have symmetric (opposite-sign) effects, and therefore chose instead to report this sensitivity analysis.
[31] Consistent with the weaker validity of the IV for the full sample period, the F-statistic for the distance-time interactions in the first-stage regression is always lower for the full sample. Results for the other outcomes reported below are more robust to extending the sample through 2002.

smaller stores,[32] in which case simply normalizing the exposure measure for population size may go too far. Thus, we also computed an exposure measure that weights by store size as well as normalizing by county population, by first weighting stores by their square footage, computing a measure of exposure to square footage of Wal-Mart stores, rather than just number of stores, and then, again, normalizing by county population. Again, we rescaled the estimates to provide the same comparison with the average effect. These estimates are reported in the last two rows of Table 4, and reveal very similar effects to those estimated using the baseline specification. Overall, then, we view the evidence regarding retail employment as robust. In almost every case, we find declines in aggregate retail employment on the order of two to four percent, accompanied by some shift within the retail sector to general merchandising.

Using Within-Region Variation for Identification of Effects on Retail Employment

Next, we consider relaxing the key identifying assumption that aggregate time effects—rather than region-specific ones—are sufficient to capture other influences on the dependent variables. Again, we first consider this issue with regard to the estimated effects of Wal-Mart stores on retail employment, and then below consider the same issue with respect to the other outcomes. As noted above, our identification strategy rules out a fully flexible specification allowing for different time patterns in the dependent variables across counties. But the maps in Figures 1 and 2 suggest that we may be able to use our identification strategy within broad geographic regions. When we estimate the models for sub-regions of the country, we effectively relax the identifying assumption that the year effects cannot differ by region. On the other hand, we potentially throw out a good deal of identifying information in the form of strong differences across different regions of the country in the timing of the opening of Wal-Mart stores.

To motivate what we do more clearly, Figures 4-7 show the locations of Wal-Mart openings in the four Census regions—South, Midwest, West, and Northeast—from the first opening in each region through 1995. Figure 4 shows that, within the South, Wal-Mart openings expanded outward from the beginning of the sample period to about 1990. Note, for example, that in the 1986-1990 map there are

[32] This may in part reflect whether or not the Wal-Mart store is a Supercenter.

many openings in Florida, on the Gulf Coast, and in Virginia and North Carolina, and fewer in interior states of the South, compared with the 1981-1985 or 1976-1980 maps. Figures 5 and 6 indicate that the same is true for the Midwest and the West from about 1981 to 1995. For the West, however, things are perhaps a bit more complicated because of the vast stretches of sparsely populated land between Arkansas and California, so we might think that the quasi-experiment based on geographic variation in the time patterns of store openings is problematic for this region.

For the Northeast, shown in Figure 7, the problem appears clearly worse, as there is less geographic variation to exploit, and nearly all of the Wal-Mart stores opened in a very short window (1991-1995). To see why this is relevant, suppose that all stores in the Northeast opened in a single year, and all were the same distance from Benton County, Arkansas. In that case, time dummy variables and distance would completely determine Wal-Mart exposure (see equation (2)), and we would get no identifying information from the distance-time interactions in the Northeast. Of course this is not the precise situation, but Figure 7 suggests that this is not too far from the truth, and as it turns out the IV estimation using observations from the Northeast only is completely uninformative, with F-statistics for the first-stage regression near one, and standard errors 20, 30, or more times as large as for the other regions.

As a consequence, we estimated models using variation within rather than across Census regions only for the South, Midwest, and West. These results are reported in Table 5.[33] For the South and the Midwest the estimates are quite similar to the full sample, with the only qualitative difference for general merchandising. Both regions indicate quite strongly that Wal-Mart stores reduce aggregate retail employment. Although of less interest, the results for general merchandising are less consistent, which could stem from differences in the types of retailers with which Wal-Mart competes in different regions, in part because of differences in the timing of Wal-Mart's growth in these regions and the growth of other general merchandising companies. For the West, most of the estimates are not statistically significant, although there is no evidence of adverse effects. In addition, note that the IV estimates are less precise

[33] Corresponding to the first store openings depicted in Figure 6, for the West we begin the sample period in 1981.

for the West, with standard errors increasing by a factor of two or more. These latter results suggest that there may be some concerns about the validity of our quasi-experiment for the West as well as the Northeast, but this is less clear from both the maps and the estimates, and we want to avoid discarding information from this region just because it is less consistent with the sharper results for the South and the Midwest.

Nonetheless, the main conclusion we draw from these within-region estimations in Table 5 is that for the two regions for which we have the best chance of successfully running the same quasi-experiment as we do for the national data, the results are quite consistent in indicating that Wal-Mart stores reduce retail employment.

Effects on Retail Sector Earnings

Table 6 turns to effects on retail payrolls per worker. Note, as pointed out earlier, that these are not effects on wages, as the CBP data do not distinguish part-time from full-time workers. Indeed, if there is a full-time wage premium, then employment could shift from full-time to part-time, part-time wages could increase, and payrolls per worker could fall. Nonetheless, a decline in payrolls per worker, for example, is significant, because it indicates that, on average, retail workers are taking home less pay. Also, if coupled with an employment decline (or no increase), it indicates overall declines in earnings in the retail sector, although earnings in other sectors could be rising. Below, we address the latter issue by looking at total payrolls.

In Table 6, the OLS estimates suggest no declines or modest declines in payrolls per worker in the aggregate retail sector, accompanied by modest increases in general merchandising, and no evidence of effects for the retail subsector in the last column.[34] The IV estimates for general merchandising are similar. However, the estimates for the aggregate retail sector and the retail subsector point more strongly to negative effects of Wal-Mart on payrolls per worker, with all coefficient estimates negative and statistically significant. To interpret the units, the estimate of –0.06 at the bottom of column (1) implies that, at the sample means, Wal-Mart stores lower aggregate retail payrolls per worker by 0.49, or 3.5

[34] Because we do not have data on wages, these differences need not imply that there are barriers to mobility of workers across sectors of the retail industry.

percent. The estimates for the other samples are similar in magnitude. In contrast, in general merchandising the IV estimates point to modest wage gains in the B sample, and no change in the C sample. As shown in the table, for the aggregate retail sector and the retail subsector the evidence of endogeneity bias is always statistically significant. As for employment, the IV estimates indicate sharper declines in aggregate retail payrolls per worker than do the OLS estimates, again exactly what we would expect from endogenous decisions regarding the location and timing of store openings in areas with strong retail growth (and hence rising retail labor costs).[35]

Paralleling the earlier robustness analyses of effects on retail employment, Table 7 reports similar analyses of the effects of Wal-Mart on retail payrolls. Having described the various analyses and their motivation earlier, these results can be described more succinctly. In this case, all of the analyses yield results very similar to the baseline estimates. The conclusions are similar when we cluster on state-year cells rather than counties. The evidence dropping either the counties that never had a store, or where stores closed, and using the data through 2002, is robust, always yielding significant negative effects on payrolls per worker in aggregate retail and in the retail subsector, and generally some weak evidence of positive effects in general merchandising.[36] The results are also robust to dropping counties with Sam's Clubs, or treating these like the other stores. Finally, the results are very robust to normalizing the store exposure measure by county population, and to weighting by store size as well.

Finally, in Table 8, as before we report estimates for the South, Midwest, and West using within-region variation in the location and timing of Wal-Mart store openings, rather than across-region variation. In this case, the results are less consistent, as we find weaker evidence of payroll declines in the West and especially the South, and the estimates for the Midwest point to payroll increases. Thus, results from the national analysis using across-region variation in Wal-Mart store openings indicate that Wal-Mart stores lead to reduced retail payrolls per worker. This result is robust to many variations in the specification and the sample. But it is less robust to looking within-regions—an analysis that rests on a weaker identifying

[35] We would not necessarily expect this endogeneity bias with respect to general merchandising, if Wal-Mart primarily competes with other sectors of the retail industry.
[36] Note that the estimates through 2002 indicate stronger negative effects on retail payrolls, while the adverse retail employment effects were not robust to this extension of the sample.

assumption by effectively allowing the year effects to differ across regions, but which also potentially throws out important identifying information. Consequently, it is not possible to draw as firm a conclusion regarding the effects of Wal-Mart stores on retail payrolls.

Our findings for retail payrolls per worker do not necessarily carry over to wages. First, there may be shifts in the skill composition of the workforce in the retail sector. Second, we cannot distinguish full-time from part-time workers.[37] The popular perception is probably that Wal-Mart results in more employment of less-skilled, part-time workers. However, while this comparison is likely true relative to the overall economy, it may not be true relative to the retail sector. If in fact Wal-Mart results in higher skills or higher hours in the retail sector, then stable payrolls per worker could mask wage declines. At this point, though, all we can conclude is that Wal-Mart exerts some downward pressure on retail labor costs overall (since employment falls, and payrolls per worker may fall). Whether this is generated by declining employment and steady hours and wages, or a combination of employment declines and hours increases, perhaps accompanied by declining wages, the results are consistent with Wal-Mart entry resulting in total earnings in the retail sector declining, which is presumably a source of Wal-Mart's efficiency gains.

Total Employment and Payrolls

From a policy perspective, the effects of Wal-Mart stores on total employment and payrolls are probably of greatest interest. To the extent that decisions to encourage or deter Wal-Mart openings are based on economic considerations, policymakers are presumably not interested in retail jobs or earnings per se, but in improving overall employment prospects and earnings. Even though Wal-Mart may reduce retail employment and earnings, the effects on overall employment and earnings are ambiguous. If retail earnings declines depress demand, or if Wal-Mart stores also lead to declines in other sectors (such as

[37] In principle, one could study these issues using data from the Current Population Survey (CPS) or the Decennial Census of Population. However, in the CPS most county identifiers are suppressed for reasons of confidentiality. Census data are less attractive because they are not available for each year but only once a decade. Furthermore, in downloadable Census estimates by county from all long-form respondents, neither hours, education, nor income are available by industry. This leaves the option of using public use samples from the Census. But these offer relatively few observations per county, which coupled with having data only for one year per decade implies that the resulting estimates would likely be uninformative.

wholesale), then overall employment and earnings may fall. On the other hand, if wages are reduced in other sectors because of labor supply shifting out of retail, then employment may expand in some of those other sectors, although we would not expect the shift out of retail to be fully absorbed as long as labor supply curves slope upward. But if Wal-Mart brings about significant price reductions on consumption purchases, there can be stimulative economic effects that boost demand for other goods and services, many of which may be locally produced, leading to employment and earnings increases.

The estimates in Table 9 report results for the effects of Wal-Mart on the total employment rate. In this case, some of the OLS and IV estimates point to positive effects. Furthermore, the OLS and IV estimates do not differ systematically, suggesting that Wal-Mart openings are not very endogenously related to aggregate employment growth. The estimates indicate non-trivial effects. For example, using the weighted exposure measure, the estimate for the A sample is 0.42, implying that the representative exposure to Wal-Mart stores results in an increase in employment per 1,000 persons of 3.46, or 1.4 percent. On the other hand, neither IV estimate for the A sample, which is most representative, is statistically significant. Finally, the earlier estimates suggest that the overall employment increase, if there is one, occurs outside of the retail sector.

In contrast, the payroll estimates in Table 10 strongly suggest that Wal-Mart stores lead to wage declines, shifts to lower-paying jobs (or less-skilled workers), or increased use of part-time workers. All of the IV estimates in Table 10 indicate significant negative effects of Wal-Mart stores on payrolls per worker *as well as* payrolls per person—and the latter better capture total earnings effects. For example, again focusing on the A sample results using the weighted exposure measures, the estimates imply that the average effect of Wal-Mart stores entails payrolls per worker declining by 1.9 percent, and payrolls per person declining by 5.4 percent.[38] The implied effects are similar for the other estimates and samples.[39]

[38] Recall that changes in payrolls per person can differ from changes in payrolls per worker because of full-time/part-time shifts, skill shifts, etc., as well as employment changes.

[39] Throughout, we have used exposure measures that grow continually with the number of years Wal-Mart stores are open. In principle, specifications using exposure variables of this form imply that any non-zero effect will grow without bound over time. Of course, it is well-known that this sort of out-of-sample prediction based on regression

Robustness analyses are reported in Table 11 for total employment and total payrolls per person, the latter of which we regard as the more significant total payroll measure. For both employment and total payrolls per person, the results are very robust to the variants of the sample and specification that we consider. Overall, then, the evidence from the baseline specification points to positive overall employment effects, but declines in earnings per person.

In Table 12 we present the analyses estimating the model separately by region. For the South, the evidence of declines in payrolls per person is even stronger. For the Midwest and West the evidence is somewhat more ambiguous, with a significant negative estimate for the largest (A) sample in the Midwest, and two out of three estimates for the West negative but not significant. Thus, both the national estimates, and the estimates for the South and perhaps the Midwest, indicate that Wal-Mart reduces total payrolls per person, while the evidence for the West is consistent with this conclusion, but not significant.[40] When we look at total employment, the evidence from the specific regions is less consistent with the national evidence, as the estimates for the South and the Midwest (for the A sample) indicate total employment

---

estimates should not be done, and that the regression estimates instead provide evidence over the range of variation in the data. One could imagine functions that more flexibly allow the effects of Wal-Mart stores to change depending on how long they have been open—such as large sets of dummy variables for number of years open. That would be a simple matter for OLS estimation of our models. However, with IV estimation this becomes impractical, as the fitted values of the dummy variables capturing different lengths of exposure to Wal-Mart stores are highly correlated. As an alternative, we estimated specifications that capped the exposure to Wal-Mart stores at five or ten years. The simplest version of this corresponds to the specification using the number of years since the first store opened in a county, where instead of letting this grow to its maximum in the sample (34) we cap it at ten for the tenth year after the store is opened and all subsequent years. This effectively imposes a spline function in which the effect of a Wal-Mart store can grow over ten years, and then stays fixed for ten or more years. When we re-estimated the models using this alternative form of the variable, the qualitative results were almost always identical in terms of sign and were always identical in terms of statistical significance, but were larger in absolute value (by approximately 50 to 100 percent). This suggests that where we find evidence of effects of Wal-Mart stores, the shorter-run effects may be somewhat sharper than the longer-run effects we estimate. But in no case do we find statistical evidence of these shorter-run effects but not longer-run effects.

We also experimented with using a Wal-Mart measure that simply recorded whether a store had opened or the number of stores opened (without regard to how long they had been open). The payroll results (retail and total) were quite robust to using these alternative measures (but less precise), as were the changes in retail employment (although the estimates were larger, relative to the specifications using the exposure measures, than could simply be explained by the difference in units of the variable capturing Wal-Mart stores). But the IV estimates of the overall employment effects were implausibly large (and negative) for these latter two instruments. The problems with these "count" rather than "exposure" specifications likely arise because the identification strategy works much better for the exposure measures, given that the variation in how long stores have been open is in part explained by the distance-time interactions.

[40] The results for total payrolls per worker were very similar, with the only difference that for the Midwest some samples indicated a positive and significant effect.

declines rather than increases. The strong evidence of employment declines for the South helps to explain the sharper declines in payrolls per person for this region.

Overall, then, from this somewhat mixed evidence we conclude that the evidence that Wal-Mart reduces total payrolls per person is stronger than the evidence that Wal-Mart increases total employment. The quasi-experiment using across-region (as well as within-region) variation supports both of these conclusions. But only for total payrolls do we see similar evidence at the regional level, most notably for the South where Wal-Mart stores are concentrated (with 1.3 stores per 100,000 people, compared with 1.0 in the Midwest, 0.4 in the West, and 0.3 in the Northeast, as of 1995), and where they have been open the longest. In contrast, some of the regional evidence—especially for the South—indicates that Wal-Mart stores also lead to overall employment declines. Of course the total employment effects may be considered moot, since if total payrolls per person decline, then Wal-Mart reduces earnings regardless of its effects on employment.

Finally, returning to the question of endogeneity bias, in every case in Table 10 the comparison between the IV and OLS estimates indicates that the OLS estimates are biased upward, and the endogeneity bias is statistically significant. Thus, overall, we find upward endogeneity bias for retail employment, retail earnings, and total earnings, but not total employment. The results for the retail sector can be easily explained as stemming from Wal-Mart endogenously locating where retail growth is strong. However, why this endogeneity bias shows up with respect to aggregate payrolls but not aggregate employment is less clear. As just explained, however, our evidence on aggregate (total) employment effects is somewhat ambiguous. The most robust findings are that Wal-Mart reduces retail employment and total payrolls per person, and for both of these dependent variables the endogeneity bias is consistently in the direction we would expect if Wal-Mart chooses to locate in counties with strong retail growth and increasing total payrolls.

VI. Conclusions

Motivated in large part by local policy debates over Wal-Mart store openings, and the large size of Wal-Mart relative to the U.S. labor market, we estimate the effects of Wal-Mart stores on employment and earnings. Critics have charged that Wal-Mart's entry into local labor markets reduces wages and employment, and the company (and others) have countered that these claims are false.

Our analysis focuses on estimating the effects of Wal-Mart on employment and earnings, emphasizing the importance of accounting for the endogeneity of the location and timing of Wal-Mart openings that—in our view, and as borne out by the data—is most likely to bias the evidence against finding adverse effects of Wal-Mart stores. Our strategy for addressing the endogeneity problem is based on a natural instrumental variable that arises because of the geographic and time pattern of the opening of Wal-Mart stores, which slowly spread out in a wave-like fashion from the first stores in Arkansas.

On balance, the evidence is more consistent with the claims of Wal-Mart's critics, although questions remain. In the retail sector, the representative Wal-Mart presence (about eight years) reduces employment by two to four percent. There is some evidence that payrolls per worker also decline, by about 3.5 percent, but this conclusion is less robust. Either way, though, retail earnings fall. Looking at total employment, some of the evidence points to employment increases, although the regional evidence for the South and perhaps the Midwest points to employment declines. At the same time, there is stronger evidence that total payrolls per worker and per person decline, by about two and five percent, respectively, implying that residents of a local labor market do indeed earn less following the opening of Wal-Mart stores. Finally, we find clear evidence of adverse effects of Wal-Mart stores on retail employment, total employment, and total payrolls per person in the South, where Wal-Mart stores are most numerous on a total and per capita basis, and where they have been open the longest.

The earnings declines associated with Wal-Mart do not necessarily imply that Wal-Mart stores worsen the economic fortunes of residents of the markets that these stores enter. Wal-Mart entry may also result in lower prices that increase purchasing power, and if prices are lowered not just at Wal-Mart, but elsewhere as well, the gains to consumers may be widespread. Although this paper provides what we believe to be the best evidence to date on the effects of Wal-Mart on local labor markets, we do not address questions of changes in real earnings or purchasing power, and how these changes might vary for lower- and higher-income families. In addition, while not endorsing the findings of the studies cited in the Introduction that estimate the taxpayer burden imposed by Wal-Mart (through, for example, increased use of Medicaid and Food Stamps), even if Wal-Mart lowers earnings as well as prices, the lower earnings will likely, in fact, increase this burden. Finally, as we noted earlier, we cannot with the data used in this paper pinpoint the mechanism leading to earnings declines, including factors such as changes in part-time work and shifts to less-skilled workers. There are, in short, numerous remaining questions of considerable interest regarding the effects of Wal-Mart on labor markets, consumption, and social program participation and expense. The identification strategy developed in this paper may prove helpful in estimating the effects of Wal-Mart stores on these other outcomes as well.

References

Artz, Georgeanne M., and James C. McConnon, Jr. 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. Available at http://www.seta.iastate.edu/retail/publications/artz_narea_paper.pdf (as of September 7, 2005).

Barnes, Nora Ganim, Allison Connell, Lisa Hermenegildo, and Lucinda Mattson. 1996. "Regional Differences in the Economic Impact of Wal-Mart." *Business Horizon* 39(4), 21-5.

Basker, Emek. 2005. "Job Creation or Destruction? Labor Market Effects of Wal-Mart Expansion." *Review of Economics and Statistics* 87(1), 174-83.

Basker, Emek. "Selling a Cheaper Mousetrap: Wal-Mart's Effect on Retail Prices." Forthcoming in *Journal of Urban Economics*.

Bianchi, Dan, and Dan Swinney. 2004. "Wal-Mart: A Destructive Force for Chicago Communities and Companies." Center for Labor and Community Research. Available at http://www.clcr.org/publications/pdf/walmart_impact_analysis.pdf (as of September 7, 2005).

Bianco, Anthony, and Wendy Zellner. 2003. "Is Wal-Mart Too Powerful." *Business Week*, October 6. Available at http://209.157.64.200/focus/f-news/990922/posts (as of October 31, 2005.)

Bowermaster, Jon. 1989. "When Wal-Mart Comes to Town." *New York Times*, April 2, Section 6, Part 2, p. 28.

Davidson, Sharon M., and Amy Rummel. 2000. "Retail Changes Associated with Wal-Mart's Entry into Maine." *International Journal of Retail & Distribution Management* 28(4/5), 162-9.

Dube, Arindrajit, and Ken Jacobs. 2004. "Hidden Cost of Wal-Mart Jobs: Use of Safety Net Programs by Wal-Mart Workers in California." Briefing Paper Series, UC Berkeley Labor Center, August 2.

Dube, Arindrajit, Barry Eidlin, and Bill Lester. 2005. "Impact of Wal-Mart Growth on Earnings throughout the Retail Sector in Urban and Rural Counties." Unpublished paper, University of California, Berkeley.

Freeman, George. 2004. "Wal-Mart Supercenters: What's in Store for Southern California?" Los Angeles County Economic Development Corporation, January.

Goetz, Stephan J., and Hema Swaminathan. 2004. "Wal-Mart and County-Wide Poverty." AERS Staff Paper No. 371, Department of Agricultural Economics and Rural Sociology, Pennsylvania State University. Available at http://cecd.aers.psu.edu/pubs/PovertyResearchWM.pdf (as of September 7, 2005).

Hausman, Jerry. 1978. "Specification Tests in Econometrics." *Econometrica* 46(6), 1251-71.

Hausman, Jerry, and Ephraim Leibtag. 2004. "CPI Bias from Supercenters: Does the BLS Know that Wal-Mart Exists?" Working Paper No. 10712, National Bureau of Economic Research, Cambridge, Massachusetts.

Hicks, Michael J., and Kristy 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(3), 305-13.

Ingold, John. 2004. "Disdain for Wal-Mart Pulls Neighbors Together; Opposition Becomes a Social Phenomenon." *Denver Post*, April 18, A Section, p. A-01.

Jacobs, Andrew. 2004. "In a Historic Black Hamlet, Wal-Mart Finds Resistance." *New York Times*, March 6, Section A, p. 8.

Kaufman, Leslie. 1999. "Wal-Mart Casts Eye Northward." *New York Times*, February 16, Section C, p. 1.

Keon, Thomas, Edward Robb, and Lori Franz. 1989. "Effect of Wal-Mart Stores on the Economic Environment of Rural Communities." Working Paper, Business and Public Administration Research Center and College of Business and Public Administration, University of Missouri.

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

McGee, Jeffery E. 1996. "When Wal-Mart Comes to Town: A Look at How Local Retailers Respond to the Arrival of the Retailing Giant." *Journal of Business and Entrepreneurship* 8(1), 43-52.

Miller, George. 2004. "Everyday Low Wages: The Hidden Price We All Pay for Wal-Mart." Democratic Staff of the Committee on Education and the Workforce, U.S. House of Representatives. Available at http://edworkforce.house.gov/democrats/WALMARTREPORT.pdf (as of September 7, 2005).

Nieves, Evelyn. 1995. "Our Towns; Taking on Wal-Mart, Ahead of Time." *New York Times*, August 1, Section B, p. 5.

Norman, Al. 2004. The Case Against Wal-Mart (Atlantic City, New Jersey: Raphel Marketing).

PBS. 2004. "One, Two, Three, Four … We Don't Want Your Superstore." Frontline program: "Is Wal-Mart Good for America?" November 16. Available at http://www.pbs.org/wgbh/pages/frontline/shows/walmart/transform/protest.html (as of September 7, 2005).

Peterson, Mark, and Jeffery E. McGee. 2000. "Survivors of 'W-day': An Assessment of the Impact of Wal-Mart's Invasion of Small Town Retailing Communities." *International Journal of Retail & Distribution Management* 28 (4/5), 170-80.

Quinn, Bill. 2000. How Wal-Mart is Destroying America and the World and What You Can Do about It (Berkeley, California: Ten Speed Press).

Rimer, Sara. 1993. "Around New England, Main Street Faces Peril." *New York Times*, February 28, Section 1, p. 22.

Rodino Associates. 2003. "Report on Research for Big Box Retail/Superstore Ordinance." Prepared for Industrial and Commercial Development Division, Community Development Department, City of Los Angeles. Available at http://www.lacity.org/council/cd13/houscommecdev/cd13houscommecdev239629107_04262005.pdf (as of September 7, 2005).

Scott, Lee. 2005. "Wal-Mart and California: A Key Moment in Time for American Capitalism." Address at Town Hall Los Angeles, February 23. Available at http://www.walmartfacts.com/docs/981_leescottspeechattownhall2-23-05_1492524199.pdf (as of September 7, 2005).

Sinnott, R.W. 1984. “Virtues of the Haversine,” *Sky and Telescope* 68(2), 159.

Slater, Robert. 2003. The Wal-Mart Decade: How a Generation of Leaders Turned Sam Walton’s Legacy into the World’s #1 Company (New York, New York: Portfolio).

Stone, Kenneth 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, Kenneth E. 1995. “Impact of Wal-Mart Stores on Iowa Communities: 1983-93.” *Economic Development Review* 13(2), 60-9.

Stone, Kenneth E. 1997. “Impact of the Wal-Mart Phenomenon on Rural Communities.” In David P. Ernstes and Dawne M. Hicks, eds., Increasing Understanding of Public Problems and Policies (Chicago, Illinois: Farm Foundation), 189-200.

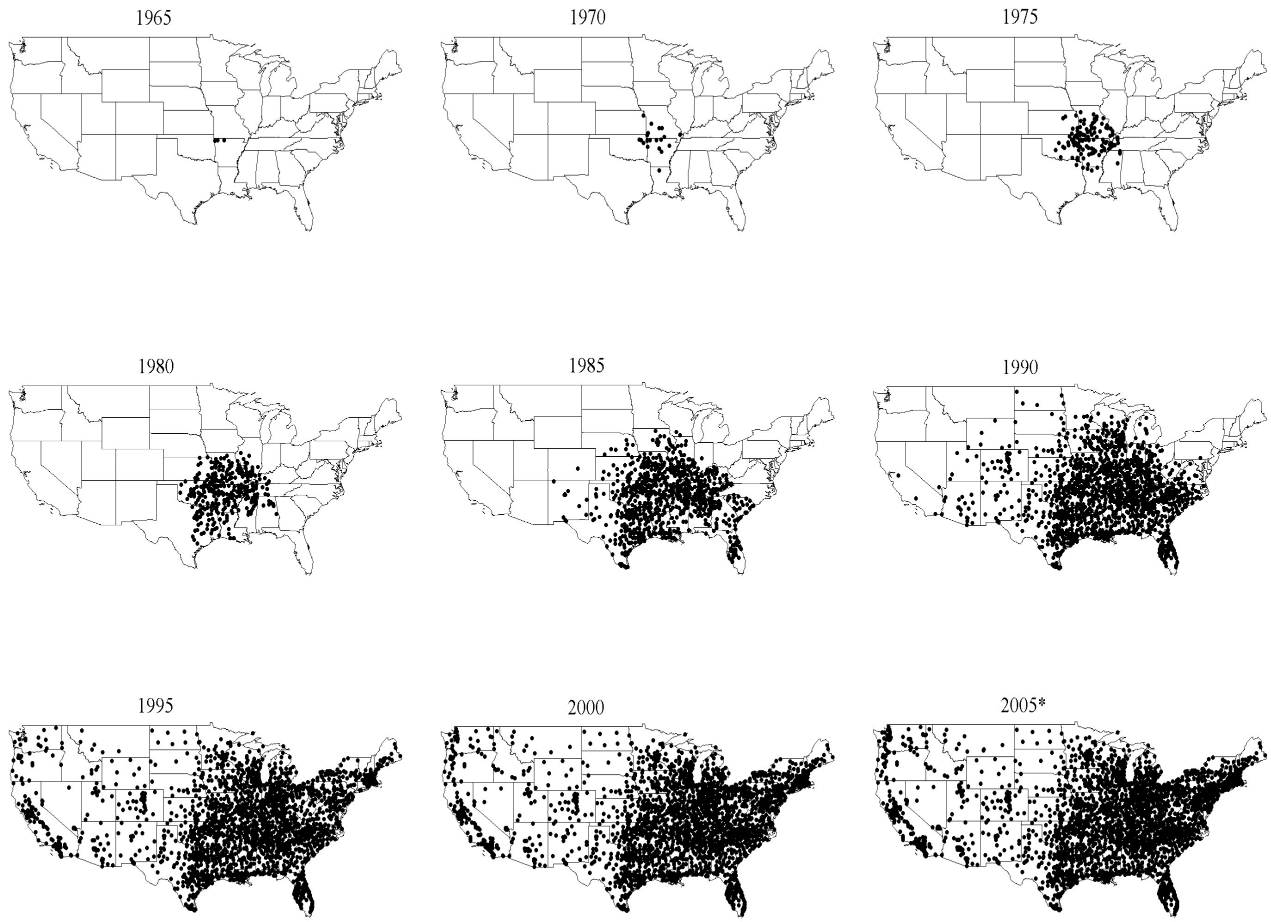
Stone, Kenneth E., Georgeanne Artz, and Albert Myles. 2002. “The Economic Impact of Wal-Mart Supercenters on Existing Businesses in Mississippi.” Mississippi State University Extension Service, Fall. Available at http://msucares.com/pubs/misc/m1283.pdf (as of September 7, 2005).

Vance, Sandra S., and Roy V. Scott. 1994. Wal-Mart: A History of Sam Walton's Retail Phenomenon (New York, New York: Twayne Publishers).

Wal-Mart Watch. 2005. “Low Prices at What Cost?” Annual Report. Available at http://fivestones.sitestream.com/pdf/2005-annual-report.pdf (as of September 7, 2005).

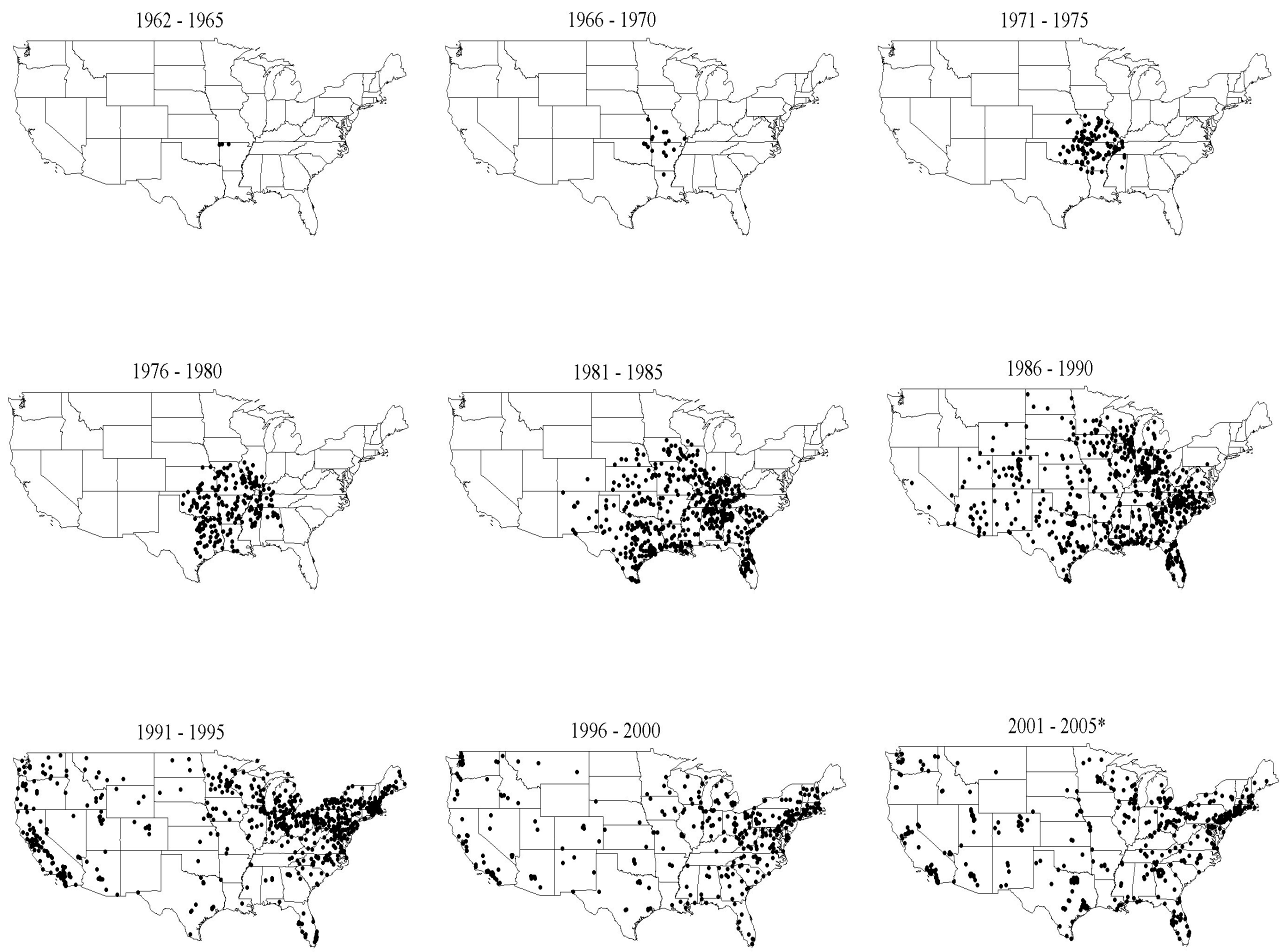
Walton, Sam. 1992. Made In America: My Story (New York, New York: Doubleday).

Figure 1: Location of Wal-Mart Stores, 1965-2005



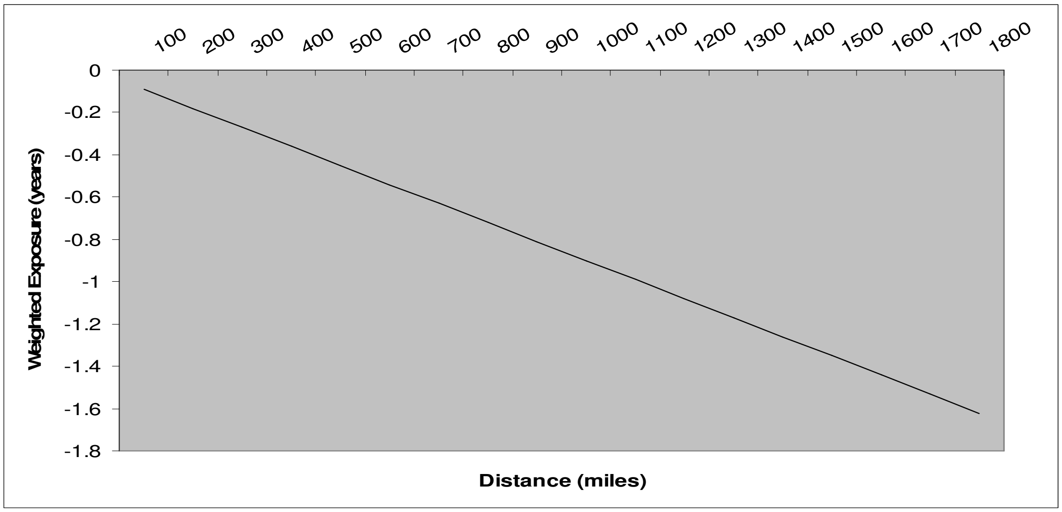
[*] Includes all locations as of January 31, 2005.

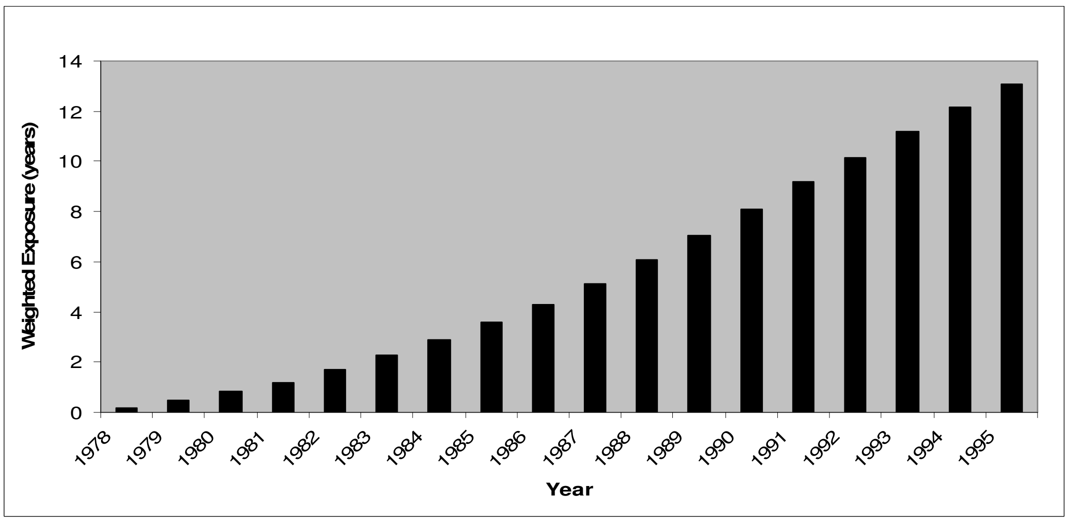
Figure 2: Location of Wal-Mart Openings, 1962-2005



* Includes all openings as of January 31, 2005.

Figure 3: First-Stage Results for Weighted Wal-Mart Exposure





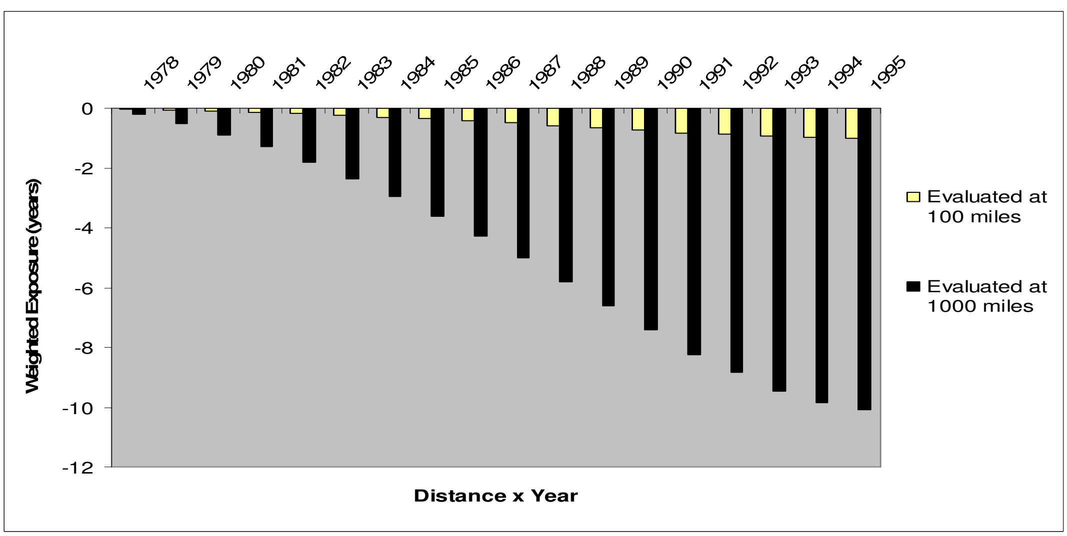


Figure 4: Location of Wal-Mart Openings – South, 1962-1995

Figure 5: Location of Wal-Mart Openings – Midwest, 1966-1995

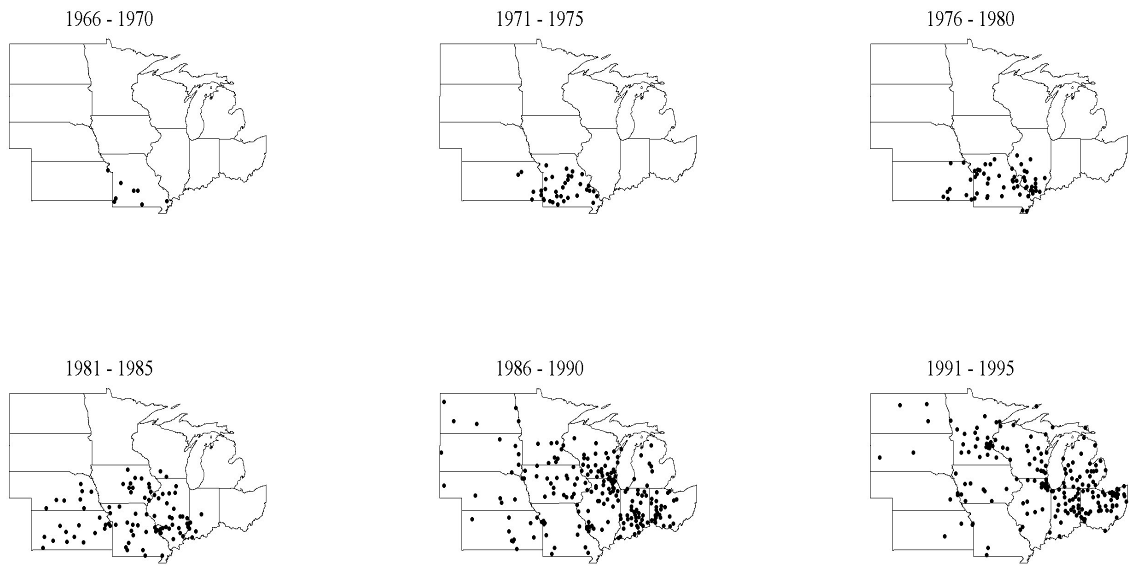


Figure 6: Location of Wal-Mart Openings – West, 1981-1995

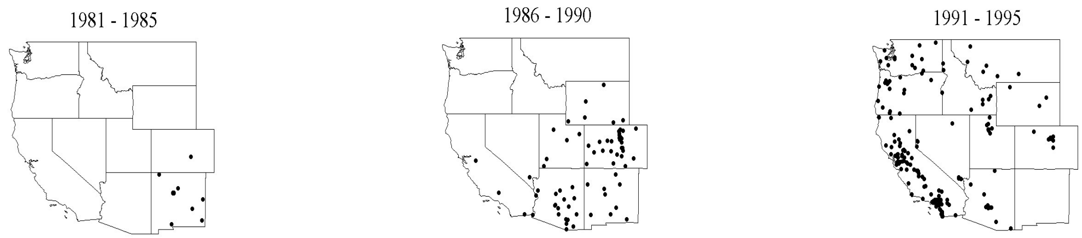


Figure 7: Location of Wal-Mart Openings – Northeast, 1986-1995

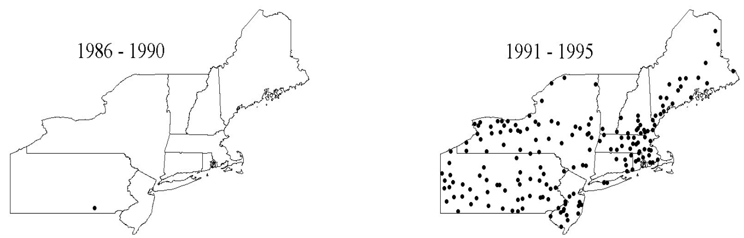


Table 1: County-Level Summary Statistics, Population, Employment, and Payroll, 1977-1995

	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*
Population	87,282 (271,431.10) [N = 51,274]	118,631 (322,398.70) [N = 35,222]	153,377 (385,570.20) [N = 17,247]
Number of Counties/Share of Total	3,032 / 0.98	2,828 / 0.91	2,091 / 0.68
Aggregate Retail Employment per 1,000 Residents	55 (27.01) [N = 51,274]	60 (26.43) [N = 35,222]	63 (26.54) [N = 17,247]
Aggregate Retail Payrolls ($1,000) per Aggregate Retail Worker	13.86 (2.46) [N = 51,250]	14.26 (2.37) [N = 35,198]	14.40 (2.22) [N = 17,227]
General Merchandising Employment per 1,000 Residents	---	7 (4.68) [N = 35,222]	7 (4.42) [N = 17,247]
General Merchandising Payrolls ($1,000) per General Merchandising Worker	---	13.09 (2.87) [N = 33,621]	13.17 (2.78) [N = 17,127]
Retail Subsector Employment per 1,000 Residents	---	---	34 (13.50) [N = 17,247]
Retail Subsector Payrolls ($1,000) per Retail Subsector Worker	---	---	15.26 (2.26) [N = 17,216]
Aggregate Manufacturing Payrolls ($1,000) per Manufacturing Worker	27.48 (8.56) [N = 50,718]	28.98 (8.54) [N = 34,871]	29.93 (8.43) [N = 17,210]
Total Employment per 1,000 Residents	243 (117.41) [N = 51,274]	265 (120.01) [N = 35,222]	282 (122.57) [N = 17,247]
Total Payrolls ($1,000) per Worker	22.34 (5.16) [N = 51,274]	23.26 (5.21) [N = 35,222]	23.78 (4.95) [N = 17,247]
Total Payrolls ($1,000) per Person	5.71 (3.72) [N = 51,274]	6.45 (3.96) [N = 35,222]	7.01 (4.27) [N = 17,247]

Means are reported first, with standard deviations in parentheses. A, B, and C samples are defined in the text. “Retail subsector” includes all retail two-digit SIC major groups (three-digit NAICS subsectors) except for Eating and Drinking Places (SIC 58) and Automotive Dealers and Gasoline Service Centers (SIC 55; NAICS 441, 447). “General Merchandising” comprises SIC 53 and NAICS 452. Payrolls are in 1999 constant dollars, and are expressed in thousands. Numbers of observations can differ slightly for the payroll per worker cells even for the same sample because sometimes zero employment is reported.

Table 2: County-Level Summary Statistics, Wal-Mart Stores, 1977-1995

	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*
Wal-Mart Stores (all counties):			
Number of Stores Opened	0.38 (0.79)	0.43 (0.88)	0.54 (1.07)
Number of Years Since First Store Opened	1.88 (4.15)	1.76 (3.86)	1.89 (3.91)
Number of Stores Weighted by Years Opened	2.35 (6.08)	2.36 (6.34)	2.82 (7.45)
Number of Observations	51,274	35,222	17,247
Wal-Mart Stores (county-year observations with at least one store open):			
Number of Stores Opened	1.33 (0.94)	1.42 (1.07)	1.59 (1.32)
Number of Years Since First Store Opened	6.56 (5.42)	5.79 (5.07)	5.61 (4.93)
Number of Stores Weighted by Years Opened	8.21 (9.01)	7.77 (9.50)	8.36 (10.86)
Number of Observations	14,678	10,713	5,820
Number of Stores Opened (shares of county-year observations with at least one store open):			
1 Store	0.811	0.772	0.716
2 Stores	0.119	0.136	0.153
3 Stores	0.037	0.046	0.058
4 Stores	0.015	0.021	0.031
5 Stores	0.010	0.013	0.020
6 Stores	0.003	0.004	0.008
7 Stores	0.002	0.003	0.005
8 Stores	0.001	0.001	0.003
9 Stores	0.001	0.002	0.003
10 Stores	0.000	0.000	0.001
11 Stores	0.000	0.000	0.001

See notes to Table 1 for details. In the last panel, a handful of counties with more than 11 stores are not reported.

Table 3: The Effect of Wal-Mart Stores on Retail Employment/1,000 Residents, Baseline Estimates

	Agg. Retail	Agg. Retail	Agg. Retail	Gen. Merch.	Gen. Merch.	Retail Subsector
Store Measure	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*	(4) *B Sample*	(5) *C Sample*	(6) *C Sample*
Years Since First Store Opened						
OLS	0.2568*** (0.0485)	0.1321** (0.0575)	0.0457 (0.0703)	0.1580*** (0.0129)	0.1595*** (0.0157)	-0.0297 (0.0379)
R^2	0.9288	0.9354	0.9568	0.8560	0.8870	0.9416
IV	-0.3981*** (0.1050)	-0.2279** (0.1112)	-0.0049 (0.1396)	0.1645*** (0.0265)	0.2259*** (0.0317)	-0.1512** (0.0767)
F-Statistic, 1st stage	62.30	53.17	31.62	53.17	31.62	31.62
Endogeneity test, p-value	<0.001	<0.001	0.675	0.779	0.016	0.068
N	51,274	35,222	17,247	35,222	17,247	17,247
Number of Stores Weighted by Years Opened						
OLS	0.1205*** (0.0311)	0.0431 (0.0314)	-0.0094 (0.0319)	0.0334*** (0.0095)	0.0151 (0.0104)	-0.0336** (0.0171)
R^2	0.9287	0.9353	0.9568	0.8525	0.8834	0.9417
IV	-0.3061*** (0.0830)	-0.1482* (0.0800)	0.0521 (0.0908)	0.1187*** (0.0197)	0.1506*** (0.0224)	-0.0749 (0.0499)
F-Statistic, 1st stage	31.01	23.59	16.05	23.59	16.05	16.05
Endogeneity test, p-value	<0.001	0.009	0.469	<0.001	<0.001	0.378
N	51,274	35,222	17,247	35,222	17,247	17,247

Standard errors are shown in parentheses below the estimates, and are robust to heteroscedasticity across counties and arbitrary autocorrelations within counties. The sample period covers 1977-1995. See notes to Table 1 for additional details. ‘*’, ‘**’, and ‘***’ indicate the estimate is statistically significant at the ten-, five-, or one-percent level, respectively. All specifications include county and year fixed effects. The Hausman (1978) test for endogeneity, in the case of a single coefficient, is based on the difference between the squared IV and OLS estimates, divided by the difference in variance of the estimates, the square root of which is distributed N(0,1). The table reports the p-value for the test of the null hypothesis of no endogeneity bias. Test results were very similar using a bootstrapped distribution for this test statistic, given that the OLS estimates are not efficient under the null of no endogeneity bias if the errors are not i.i.d.

Table 4: Additional Estimates of Effects of Wal-Mart Stores on Retail Employment/1,000 Residents, IV Estimates Using Number of Stores Weighted by Years Opened, Robustness Analyses

	Agg. Retail	Agg. Retail	Agg. Retail	Gen. Merch.	Gen. Merch.	Retail Subsector
Specification	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*	(4) *B Sample*	(5) *C Sample*	(6) *C Sample*
Baseline: Table 3						
IV	-0.3061***	-0.1482*	0.0521	0.1187***	0.1506***	-0.0749
	(0.0830)	(0.0800)	(0.0908)	(0.0197)	(0.0224)	(0.0499)
Cluster by State-Year						
IV	-0.3061***	-0.1482	0.0521	0.1187***	0.1506***	-0.0749*
	(0.0953)	(0.0909)	(0.0863)	(0.0136)	(0.0148)	(0.0439)
Drop Counties that Never Have a Store						
IV	-0.1461**	-0.0313	0.0912	0.1146***	0.1349***	-0.0429
	(0.0721)	(0.0741)	(0.0859)	(0.0183)	(0.0209)	(0.0466)
N	32,075	25,977	13,797	25,977	13,797	13,797
Drop Counties with Closed Stores						
IV	-0.3178***	-0.1461*	0.0570	0.1339***	0.1673***	-0.0845
	(0.0864)	(0.0838)	(0.1016)	(0.0195)	(0.0247)	(0.0559)
N	50,443	34,595	16,937	34,595	16,937	16,937
Through 2002						
IV	0.0496	0.0852	0.1612**	0.1710***	0.1826***	0.0253
	(0.0719)	(0.0736)	(0.0709)	(0.0227)	(0.0221)	(0.0439)
N	70,178	48,138	24,279	48,138	24,279	24,279
Drop Counties with Sam's Clubs						
IV	-0.3528***	-0.2525**	-0.1042	0.1584***	0.2357***	-0.1636**
	(0.0955)	(0.1035)	(0.1430)	(0.0229)	(0.0321)	(0.0784)
N	45,453	29,535	13,752	29,535	13,752	13,752
Combining Wal-Mart Stores with Sam's Clubs						
IV	-0.2890***	-0.1278*	0.0681	0.1110***	0.1335***	-0.0558
	(0.0799)	(0.0744)	(0.0797)	(0.0186)	(0.0203)	(0.0437)
Normalized by County Population						
IV	-0.2455***	-0.1203*	-0.0012	0.1005***	0.1143***	-0.0746*
	(0.0688)	(0.0668)	(0.0706)	(0.0159)	(0.0158)	(0.0387)
Weighted by Store Size and Normalized by County Population						
IV	-0.2879***	-0.1415*	-0.0054	0.1093***	0.1282***	-0.0860**
	(0.0775)	(0.0735)	(0.0794)	(0.0173)	(0.0177)	(0.0436)

See notes to Tables 1 and 3. Sample sizes are the same as in Table 3 unless otherwise noted. For the estimates labeled "normalized by county population," we divide the weighted exposure measure by the ratio county population/100,000 (the approximate average county size). This converts the exposure measure to a per capita basis. For the estimates labeled "weighted by store size and normalized by county population," we also weight stores by their square footage, computing a measure of exposure to square footage of Wal-Mart store, rather than just number of stores. In both cases we then rescale the estimated coefficients and standard errors by the ratio of the mean of this new exposure variable for county-year observations with stores open to the mean of the variable used in the baseline specification for this same subsample; this ensures that the comparison with the baseline estimates is for the same representative exposure to Wal-Mart stores.

Table 5: Additional Estimates of Effects of Wal-Mart Stores on Retail Employment/1,000 Residents, IV Estimates Using Number of Stores Weighted by Years Opened, Within-Region Identification

	Agg. Retail	Agg. Retail	Agg. Retail	Gen. Merch.	Gen. Merch.	Retail Subsector
Specification	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*	(4) *B Sample*	(5) *C Sample*	(6) *C Sample*
Baseline: Table 3						
IV	-0.3061*** (0.0830)	-0.1482* (0.0800)	0.0521 (0.0908)	0.1187*** (0.0197)	0.1506*** (0.0224)	-0.0749 (0.0499)
South						
IV	-1.1604*** (0.1746)	-1.1122*** (0.1959)	-0.7443*** (0.1530)	0.0371 (0.0457)	0.1150*** (0.0377)	-0.3477*** (0.0788)
F-Statistic, 1st stage	22.33	14.75	7.90	14.75	7.90	7.90
N	23,783	16,631	7,815	16,631	7,815	7,815
Midwest						
IV	-0.5971*** (0.1418)	-0.4271*** (0.1472)	-0.4481* (0.2424)	-0.0499 (0.0326)	-0.0915* (0.0472)	-0.3379*** (0.1139)
F-Statistic, 1st stage	12.42	6.70	4.55	6.70	4.55	4.55
N	17,092	10,786	5,298	10,786	5,298	5,298
West (1981-1995)						
IV	0.7840 (0.5332)	0.4616 (0.4281)	0.4956 (0.3237)	0.0976 (0.0903)	0.1969** (0.0874)	0.0193 (0.1761)
F-Statistic, 1st stage	5.48	4.79	3.77	4.79	3.77	3.77
N	5,052	3,226	1,713	3,226	1,713	1,713

See notes to Tables 1 and 3.

Table 6: The Effect of Wal-Mart Stores on Retail Payrolls ($1,000)/Worker, Baseline Estimates

	Agg. Retail	Agg. Retail	Agg. Retail	Gen. Merch.	Gen. Merch.	Retail Subsector
Store Measure	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*	(4) *B Sample*	(5) *C Sample*	(6) *C Sample*
Years Since First Store Opened						
OLS	-0.0160***	-0.0099	-0.0100	0.0614***	0.0554***	0.0055
	(0.0053)	(0.0061)	(0.0079)	(0.0086)	(0.0131)	(0.0083)
R^2	0.8064	0.8273	0.8880	0.5358	0.5751	0.8656
IV	-0.0751***	-0.0785***	-0.0842***	0.0418**	0.0073	-0.0447**
	(0.0123)	(0.0127)	(0.0165)	(0.0196)	(0.0275)	(0.0176)
F-Statistic, 1st stage	62.30	53.17	31.62	53.17	31.62	31.62
Endogeneity test, p-value	<0.001	<0.001	<0.001	0.266	0.047	0.001
N	51,250	35,198	17,227	33,621	17,127	17,216
Number of Stores Weighted by Years Opened						
OLS	-0.0087***	-0.0060**	-0.0059*	0.0182***	0.0100**	-0.0050
	(0.0032)	(0.0030)	(0.0033)	(0.0042)	(0.0047)	(0.0031)
R^2	0.8063	0.8274	0.8881	0.5346	0.5741	0.8657
IV	-0.0599***	-0.0582***	-0.0554***	0.0279**	0.0006	-0.0309***
	(0.0097)	(0.0092)	(0.0106)	(0.0142)	(0.0182)	(0.0112)
F-Statistic, 1st stage	31.01	23.59	16.05	23.59	16.05	16.05
Endogeneity test, p-value	<0.001	<0.001	<0.001	0.475	0.593	<0.001
N	51,250	35,198	17,227	33,621	17,127	17,216

See notes to Tables 1 and 3.

Table 7: Additional Estimates of Effects of Wal-Mart Stores on Retail Payrolls ($1,000)/Worker, IV Estimates Using Number of Stores Weighted by Years Opened, Robustness Analyses

	Agg. Retail	Agg. Retail	Agg. Retail	Gen. Merch.	Gen. Merch.	Retail Subsector
Specification	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*	(4) *B Sample*	(5) *C Sample*	(6) *C Sample*
Baseline: Table 6						
IV	-0.0599*** (0.0097)	-0.0582*** (0.0092)	-0.0554*** (0.0106)	0.0279** (0.0142)	0.0006 (0.0182)	-0.0309*** (0.0112)
Cluster by State-Year						
IV	-0.0599*** (0.0100)	-0.0582*** (0.0098)	-0.0554*** (0.0102)	0.0279** (0.0131)	0.0006 (0.0181)	-0.0309*** (0.0107)
Drop Counties that Never Have a Store						
IV	-0.0429*** (0.0079)	-0.0502*** (0.0082)	-0.0516*** (0.0099)	0.0319*** (0.0115)	-0.0030 (0.0162)	-0.0313*** (0.0105)
N	32,075	25,977	13,797	25,958	13,792	13,797
Drop Counties with Closed Stores						
IV	-0.0639*** (0.0104)	-0.0628*** (0.0099)	-0.0607*** (0.0119)	0.0281* (0.0153)	-0.0008 (0.0204)	-0.0339*** (0.0125)
N	50,419	34,571	16,917	32,998	16,818	16,906
Through 2002						
IV	-0.0804*** (0.0097)	-0.0748*** (0.0090)	-0.0552*** (0.0086)	0.0326*** (0.0124)	0.0191 (0.0117)	-0.0269*** (0.0083)
N	70,146	48,106	24,251	45,861	24,126	24,240
Drop Counties with Sam's Clubs						
IV	-0.0560*** (0.0114)	-0.0553*** (0.0118)	-0.0573*** (0.0166)	0.0598*** (0.0196)	0.0279 (0.0306)	-0.0124 (0.0185)
N	45,429	29,511	13,732	27,949	13,636	13,721
Combining Wal-Mart Stores with Sam's Clubs						
IV	-0.0580*** (0.0094)	-0.0545*** (0.0085)	-0.0482*** (0.0092)	0.0253* (0.0132)	-0.0007 (0.0161)	-0.0272*** (0.0097)
Normalized by County Population						
IV	-0.0499*** (0.0081)	-0.0445*** (0.0076)	-0.0403*** (0.0084)	0.0269** (0.0119)	0.0052 (0.0141)	-0.0201** (0.0089)
Weighted by Store Size and Normalized by County Population						
IV	-0.0553*** (0.0091)	-0.0502*** (0.0084)	-0.0459*** (0.0095)	0.0289** (0.0130)	0.0058 (0.0158)	-0.0231** (0.0100)

See notes to Tables 1, 3, and 4. Sample sizes are the same as in Table 6 unless otherwise noted.

Table 8: Additional Estimates of Effects of Wal-Mart Stores on Retail Payrolls ($1,000)/Worker, IV Estimates Using Number of Stores Weighted by Years Opened, Within-Region Identification

	Agg. Retail	Agg. Retail	Agg. Retail	Gen. Merch.	Gen. Merch.	Retail Subsector
Specification	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*	(4) *B Sample*	(5) *C Sample*	(6) *C Sample*
Baseline: Table 6						
IV	-0.0599*** (0.0097)	-0.0582*** (0.0092)	-0.0554*** (0.0106)	0.0279** (0.0142)	0.0006 (0.0182)	-0.0309*** (0.0112)
South						
IV	-0.0147 (0.0141)	-0.0181 (0.0131)	-0.0222 (0.0138)	0.0556*** (0.0177)	0.0471** (0.0204)	0.0133 (0.0127)
F-Statistic, 1st stage	22.33	14.75	7.90	14.75	7.90	7.90
N	23,777	16,625	7,813	16,021	7,789	7,813
Midwest						
IV	0.0680*** (0.0163)	0.0699*** (0.0168)	0.0673*** (0.0175)	0.0943*** (0.0211)	0.0948*** (0.0343)	0.0878*** (0.0169)
F-Statistic, 1st stage	12.42	6.70	4.55	6.70	4.55	4.55
N	17,074	10,768	5,280	10,204	5,241	5,272
West (1981-1995)						
IV	-0.0575 (0.0483)	-0.0695** (0.0352)	-0.0308 (0.0338)	-0.0868 (0.0721)	-0.0798 (0.0790)	0.0200 (0.0393)
F-Statistic, 1st stage	5.48	4.79	3.77	4.79	3.77	3.77
N	5,052	3,226	1,713	2,884	1,680	1,710

See notes to Tables 1 and 3.

Table 9: The Effect of Wal-Mart Stores on Total Employment/1,000 Residents, Baseline Estimates

	Total	Total	Total
Store Measure	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*
Years Since First Store Opened			
OLS	0.4497** (0.1885)	0.2475 (0.2233)	0.0217 (0.2957)
R^2	0.9432	0.9512	0.9641
IV	0.4907 (0.3867)	0.7969* (0.4092)	1.5203*** (0.5587)
F-Statistic, 1st stage	62.30	53.17	31.62
Endogeneity test, p-value	0.903	0.109	0.002
N	51,274	35,222	17,247
Number of Stores Weighted by Years Opened			
OLS	0.3416*** (0.1234)	0.2283* (0.1355)	0.1068 (0.1462)
R^2	0.9432	0.9513	0.9641
IV	0.4217 (0.3039)	0.6588** (0.2939)	1.2622*** (0.3679)
F-Statistic, 1st stage	31.01	23.59	16.05
Endogeneity test, p-value	0.773	0.099	<0.001
N	51,274	35,222	17,247

See notes to Tables 1 and 3 for details.

Table 10: The Effect of Wal-Mart Stores on Total Payrolls ($1,000)/Worker and Total Payrolls ($1,000)/Person, Baseline Estimates

	Total Payrolls/ Worker	Total Payrolls/ Worker	Total Payrolls/ Worker	Total Payrolls/ Person	Total Payrolls/ Person	Total Payrolls/ Person
Store Measure	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*	(4) *A Sample*	(5) *B Sample*	(6) *C Sample*
Years Since First Store Opened						
OLS	-0.0136 (0.0106)	-0.0079 (0.0130)	0.0048 (0.0178)	-0.0003 (0.0062)	-0.0081 (0.0078)	-0.0148 (0.0105)
R^2	0.8260	0.8095	0.9170	0.9359	0.9422	0.9616
IV	-0.0649** (0.0280)	-0.1005*** (0.0291)	-0.1444*** (0.0411)	-0.0485*** (0.0153)	-0.0570*** (0.0171)	-0.0586** (0.0234)
F-Statistic, 1st stage	62.30	53.17	31.62	62.30	53.17	31.62
Endogeneity test, p-value	0.048	<0.001	<0.001	<0.001	0.001	0.036
N	51,274	35,222	17,247	51,274	35,222	17,247
Number of Stores Weighted by Years Opened						
OLS	-0.0008 (0.0052)	0.0026 (0.0055)	0.0063 (0.0065)	0.0060 (0.0040)	0.0032 (0.0045)	0.0013 (0.0053)
R^2	0.8260	0.8095	0.9170	0.9359	0.9422	0.9616
IV	-0.0525** (0.0221)	-0.0755*** (0.0209)	-0.0991*** (0.0270)	-0.0377*** (0.0121)	-0.0397*** (0.0123)	-0.0328** (0.0151)
F-Statistic, 1st stage	31.01	23.59	16.05	31.01	23.59	16.05
Endogeneity test, p-value	0.016	<0.001	<0.001	<0.001	<0.001	0.016
N	51,274	35,222	17,247	51,274	35,222	17,247

See notes to Tables 1 and 3 for details.

Table 11: Additional Estimates of Effects of Wal-Mart Stores on Total Employment/1,000 Residents and Total Payrolls ($1,000)/Person, IV Estimates Using Number of Stores Weighted by Years Opened, Robustness Analyses

	Total Employment	Total Employment	Total Employment	Total Payrolls/ Person	Total Payrolls/ Person	Total Payrolls/ Person
Specification	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*	(4) *A Sample*	(5) *B Sample*	(6) *C Sample*
Baseline						
IV	0.4217 (0.3039)	0.6588** (0.2939)	1.2622*** (0.3679)	-0.0377*** (0.0121)	-0.0397*** (0.0123)	-0.0328** (0.0151)
Cluster by State-Year						
IV	0.4217 (0.3058)	0.6588** (0.2952)	1.2622*** (0.2966)	-0.0377*** (0.0110)	-0.0397*** (0.0104)	-0.0328*** (0.0104)
Drop Counties that Never Have a Store						
IV	0.2603 (0.2496)	0.6711** (0.2611)	1.0035*** (0.3303)	-0.0327*** (0.0095)	-0.0250** (0.0103)	-0.0208 (0.0128)
N	51,274	35,222	17,247	32,075	25,977	13,797
Drop Counties with Closed Stores						
IV	0.4584 (0.3226)	0.7466** (0.3136)	1.3698*** (0.4096)	-0.0407*** (0.0129)	-0.0425*** (0.0132)	-0.0372** (0.0168)
N	50,443	34,595	16,937	50,443	34,595	16,937
Through 2002						
IV	0.2545 (0.3415)	0.4621 (0.3240)	1.3505*** (0.3137)	-0.0664*** (0.0169)	-0.0697*** (0.0177)	-0.0414** (0.0189)
N	70,178	48,138	24,279	70,178	48,138	24,279
Drop Counties with Sam's Clubs						
IV	0.5692 (0.3543)	0.6914* (0.3764)	1.3313** (0.5603)	-0.0305** (0.0140)	-0.0419*** (0.0159)	-0.0483** (0.0236)
N	45,453	29,535	13,752	45,453	29,535	13,752
Combining Wal-Mart Stores with Sam's Clubs						
IV	0.4303 (0.2917)	0.6701** (0.2734)	1.2204*** (0.3247)	-0.0358*** (0.0117)	-0.0358*** (0.0115)	-0.0262** (0.0131)
Normalized by County Population						
IV	0.4123 (0.2546)	0.5891** (0.2475)	0.7932*** (0.2844)	-0.0302*** (0.0100)	-0.0310*** (0.0103)	-0.0283** (0.0118)
Weighted by Store Size and Normalized by County Population						
IV	0.3896 (0.2846)	0.5875** (0.2703)	0.8651*** (0.3190)	-0.0351*** (0.0113)	-0.0358*** (0.0113)	-0.0325** (0.0133)

See notes to Tables 1, 3, and 4. Sample sizes are the same as in Tables 9 and 10 unless otherwise noted.

Table 12: Additional Estimates of Effects of Wal-Mart Stores on Total Employment/1,000 Residents and Total Payrolls ($1,000)/Person, IV Estimates Using Number of Stores Weighted by Years Opened, Within-Region Identification

	Total Employment	Total Employment	Total Employment	Total Payrolls/ Person	Total Payrolls/ Person	Total Payrolls/ Person
Specification	(1) *A Sample*	(2) *B Sample*	(3) *C Sample*	(4) *A Sample*	(5) *B Sample*	(6) *C Sample*
Baseline						
IV	0.4217 (0.3039)	0.6588** (0.2939)	1.2622*** (0.3679)	-0.0377*** (0.0121)	-0.0397*** (0.0123)	-0.0328** (0.0151)
South						
IV	-2.0914*** (0.5728)	-2.4567*** (0.5831)	-1.3439** (0.5546)	-0.1015*** (0.0200)	-0.1060*** (0.0212)	-0.0791*** (0.0205)
F-Statistic, 1st stage	22.33	14.75	7.90	22.33	14.75	7.90
N	23,783	16,631	7,815	23,783	16,631	7,815
Midwest						
IV	-1.2400*** (0.4700)	0.1145 (0.4746)	0.3887 (0.7797)	-0.0267* (0.0140)	0.0036 (0.0148)	0.0139 (0.0257)
F-Statistic, 1st stage	12.42	6.70	4.55	12.42	6.70	4.55
N	17,092	10,786	5,298	17,092	10,786	5,298
West (1981-1995)						
IV	2.3270 (1.7062)	0.7325 (1.3854)	0.7034 (1.1953)	-0.0158 (0.0528)	-0.0520 (0.0457)	-0.0447 (0.0434)
F-Statistic, 1st stage	5.48	4.79	3.77	5.48	4.79	3.77
N	5,052	3,226	1,713	5,052	3,226	1,713

See notes to Tables 1 and 3.

Appendix A: Assessment of Non-Administrative Wal-Mart Data

Our original intention was to use the Wal-Mart data from Basker (2005) for our empirical analysis, as that was, at the time, the only available source of data on Wal-Mart openings. Basker generously provided the Wal-Mart data that she compiled from the several sources listed in Table A1 of her article.[41] Our initial efforts focused on making some improvements to these data in preparation for our own analysis. However, as described in this appendix, there are substantial errors in the timing of Wal-Mart store openings in these data. This led us to request administrative data directly from Wal-Mart, which we use in our analysis. This appendix does two things. First, it briefly documents the measurement problems in Basker's data.[42] Our intention is to document the measurement error, rather than to criticize Basker's work. Basker was aware of measurement problems with these data, and, as described in Section II, used planned openings as an IV for actual openings in part to address this issue.[43] Second, anecdotally, until very recently Wal-Mart was unwilling to supply data to researchers, so Basker may well have done the best job possible at the time. Thus, this appendix suggests how the data might be further modified to substantially reduce these measurement problems, which may prove useful to researchers who cannot access the administrative Wal-Mart data.

We began by extracting the set of stores for which we were confident of location information. There were 2,705 observations (stores) in her data file, with the following information (sometimes missing) provided for each store: state, city, county, ZIP Code, Supercenter classification, Wal-Mart assigned store number, and indicators of the source in which each store appeared. We compared Basker's data set to store listings in the 2004 Wal-Mart edition of the *Rand McNally Road Atlas* (*Atlas*) and on the

---

[41] These sources, and the periods covered, are: 1962-1969—Vance and Scott (1994); 1972 and 1974-1978—Wal-Mart *Annual Reports*; 1972, 1973, 1976, and 1979-1982—*Directory of Discount Department Stores*; 1983-1986—*Directory of Discount Stores (Chain Store Guide)*; 1987-1995 and 2000—*Directory of Discount Department Stores*; 1995-2001—Wal-Mart edition of *Rand McNally Road Atlas*. The dates listed reflect the sources used in the entire data set, rather than just those used in Basker's analysis, and therefore do not correspond exactly to the dates listed in Table A1 of Basker (2005).

[42] We report results for the slightly modified version of her data set, but they are virtually the same with her original data.

[43] We have raised doubts in the main text, however, about her IV. In addition, the measurement error we document below appears non-classical, raising questions about whether IV eliminates the measurement error bias.

Wal-Mart store locator website.[44] We found 2,348 exact matches between stores in Basker's data set and the 2004 *Atlas*, and 210 stores for which there were inconsistencies on at least one item but for which we could use the store locator website to resolve the contradiction. An additional 147 stores from the data set were either not found in the 2004 *Atlas* or the store locator website, or were found in the 2004 *Atlas* but were identified as Wal-Mart Neighborhood Markets on the store locator website. We dropped these latter stores only, leaving us with 2,558 valid observations from Basker's raw data set.

We next turned to measurement of store opening dates. To assess this, we compared opening dates from Basker's data to those available on a nationwide basis from Wal-Mart *Annual Reports* covering Basker's sample period. (Note that for much of the period the *Annual Reports* did not provide location information.) Using her data, we assigned opening dates to each store with a one year lag, such that stores which first appeared in one of her sources in year t were assigned the year $t - 1$. Basker had multiple sources for store information for some time periods, but chose to use only one for each period. To improve on this slightly, when stores appeared in at least one source (e.g., the 1995 *Chain Store Guide* or the 1995 *Atlas*, even if Basker used the latter), we assigned the relevant opening date. We required store openings to appear in at least two consecutive years to be considered real.

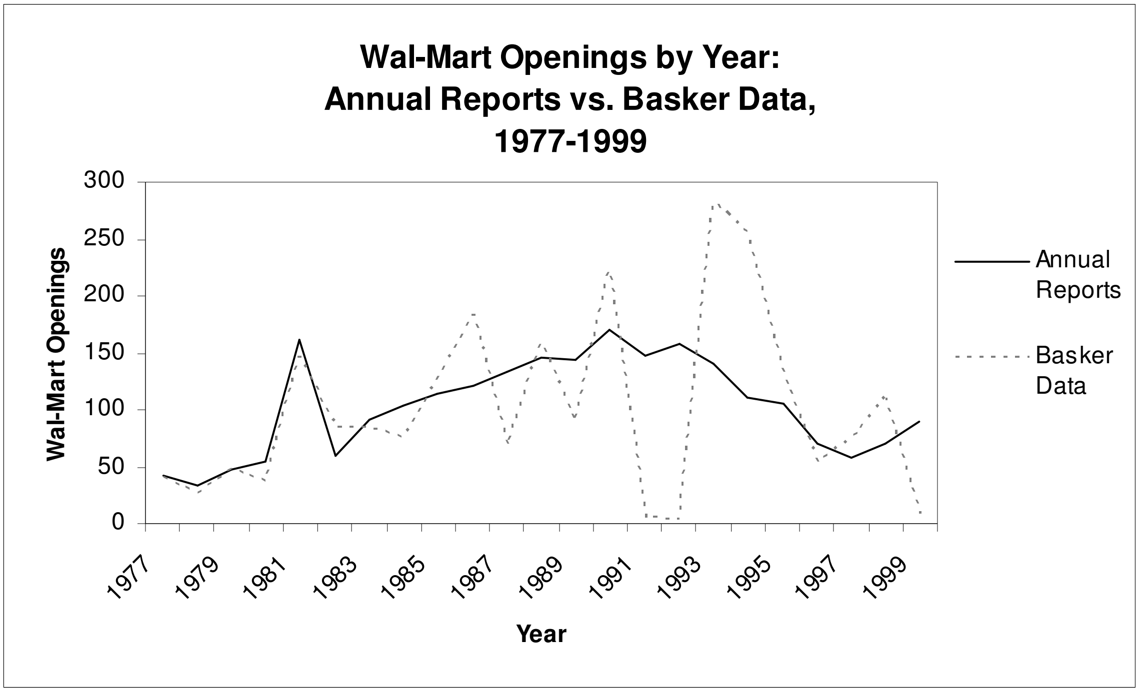
Appendix Figure A1 compares the time-series of store openings in Basker's data and in the Wal-Mart *Annual Reports*. While Basker acknowledges in her paper that the most egregious examples of measurement error in her data set occur during the period 1990-1993, the figure provides a more complete view of these errors over the entire sample period.[45] The number of store openings in Basker's data set tracks the openings indicated in Wal-Mart annual reports quite well for the period 1977-1981, but diverges significantly during the periods 1984-1995 and 1998-1999. Since Basker's data set relies solely on the *Directory of Discount Department Stores* and the *Chain Store Guide* to indicate new store openings for the period 1978-1993, differences between Basker's data and the *Annual Reports* are driven mainly by the unreliability of these data sources. Another way to quantify the fairly large extent of

[44] See http://www.walmart.com/cservice/ca_storefinder.gsp?NavMode=7 (viewed November 24, 2004).
[45] The series labeled "Basker data" in the figure is for the raw data. For 1990-1993 Basker does use an imputation procedure to try to improve upon the assignment of store openings to years, based on information from the *Annual Reports* on openings by state and year. But accurate dating of store openings, by location, remains a problem.

measurement error in Basker's data is to note that the cumulative absolute difference between the number of store openings each year in each data source is 1,123, and the correlation across years is only 0.43.

One way to correct these data and to use them (in the absence of the Wal-Mart administrative data, which may not be available on an unrestricted basis) is suggested by Appendix Figure A1, which indicates that the data errors over the sample period are consistently offsetting over two or more years, so that overcounts in one or more years are typically offset in the next year or two, and vice versa. In particular, if we group the years 1980-1982, 1984-1987, 1989-1990, 1991-1994, 1996-1997, and 1998-1999—that is, if we sometimes specify broader periods for store openings—then the errors in timing of store openings are largely eliminated. This is shown in Appendix Figure A2, and is also reflected in the cumulative absolute difference, which drops from 1,123 to 131, and the correlation, which rises from 0.43 to 0.996. Of course, the disadvantage of having to use the data grouped in this fashion is that a fair amount of variation in the timing of store openings is lost, and one also has to decide how to use or combine observations covering different numbers of years. Thus, use of the administrative Wal-Mart data is clearly preferable, if at all possible.

Appendix Figure A1: Comparisons of Time-Series of Store Openings



Appendix Figure A2: Comparisons of Time-Series of Store Openings with Grouped Years