

# When Good Instruments Go Bad: A Reply to Neumark, Zhang, and Ciccarella

Emek Basker  
University of Missouri\*

## Abstract

This note examines the instrumental variables method used by Neumark, Zhang, and Ciccarella (2005) to analyze Wal-Mart's effect on retail labor markets, and exposes major flaws in that methodology. Neumark, Zhang, and Ciccarella use an interaction between distance from Wal-Mart's headquarters and time effects to predict Wal-Mart's presence in a county, and find that each Wal-Mart store destroys, on average, approximately 200 retail jobs. These findings are in stark contrast to Basker (2005) who found a small, but positive and statistically significant, effect on jobs. I show that the IV estimates obtained by Neumark, Zhang, and Ciccarella confound Wal-Mart's causal effect with other factors. To illustrate the problem, I show that their methodology implies a large impact of Wal-Mart not only on retail employment but also on county *manufacturing* employment. Reduced-form estimates of the regressions show statistically and economically indistinguishable effects in counties with and without Wal-Mart presence, implying that other factors are most likely driving the results.

**JEL Codes:** C21, J21, L81

**Keywords:** Instrumental Variables, Wal-Mart, Retail Employment

---

\*Comments welcome to: emek@missouri.edu. I thank Daron Acemoglu, Josh Angrist, Saku Aura, David Autor, Olivier Blanchard, Patrick Buckley, David Card, Arin Dube, Jerry Hausman, David Levine, Jeff Miron, Peter Mueser and Ken Troske for helpful discussions, seminar participants at MIT and UC-Berkeley for comments, and Wal-Mart for providing the administrative data used in this paper with no strings attached.

# 1 Introduction

Several recent papers have attempted to estimate the effect of Wal-Mart's entry and expansion on retail labor markets, noting the likely endogeneity bias of OLS estimates. Since Wal-Mart is more likely to enter thriving, growing markets than foundering, declining ones, OLS estimates that show a positive relationship between Wal-Mart's entry and retail employment cannot be interpreted causally. Basker (2005) attempts to circumvent this problem using an instrument that proxies for a store's initial planning date; she argues that while the exact timing of a store's opening can be manipulated to coincide with favorable conditions, planning is done sufficiently in advance that it is not likely to be endogenous to a growth spurt — or sudden decline — exactly coinciding with Wal-Mart's entry. Her results show a small but statistically-significant long-run increase in retail jobs at the county level. A key limitation of this IV strategy is that the instrument (store planning date) is only defined for locations in which a Wal-Mart store was eventually opened, limiting inference about the impact of a Wal-Mart store on locations that have not received a store.

Two recent papers tackle this problem by noting Wal-Mart's spatial pattern of expansion (first noted by Graff and Ashton, 1994), in growing concentric circles around Bentonville, Arkansas (the location of Wal-Mart's company headquarters). In independent work, Dube, Eidlin, and Lester (2005) and Neumark, Zhang, and Ciccarella (2005) (hereafter, NZC) exploit the fact that counties enter Wal-Mart's sphere at different times, depending on their distance from Northwestern Arkansas. Dube, Eidlin, and Lester focus on Wal-Mart's effects on retail workers' earnings, and distinguish between metropolitan areas (where they find a negative effect) and rural counties (where they find mixed results). NZC estimate Wal-Mart's effects on both retail earnings and on retail employment; like Dube, Eidlin, and Lester, they find a negative effect on earnings, and also find a large and negative effect on retail employment.

The distance instrument is intuitively appealing since it is clearly exogenous to Wal-Mart's entry, and is potentially much more powerful than the planning date used by Basker

because it captures an “intent to treat” and therefore, in principle, allows estimating what Imbens and Angrist (1994) call a “local average treatment effect.” This note, however, argues that the instrument is not valid because it is correlated with other spatial patterns. As a result, coefficient estimates from these IV regressions are not interpretable, and provide *less* information about Wal-Mart’s effects than Basker’s original estimates. I show that the distance instrument fails elementary robustness checks and leads to implausible results in counterfactual exercises, including a massive increase in a county’s manufacturing employment.

Distance has been used as an instrument in several other contexts, always controversially. In the macroeconomics literature, distance to the equator has been used to identify the effect of European settlement on development (see, for example Acemoglu, Johnson, and Robinson, 2001). While there is little doubt that distance is indeed exogenous, it may be correlated with other factors that affect outcomes, and if these are not — or cannot be — properly controlled, the instrumental-variables estimate confounds the effect of the mechanism of interest (European settlement, e.g.) with the effects of other mechanisms that are also correlated with the instrument. Sachs (2003) makes the argument, for example, that distance to the equator is correlated with the incidence of malaria, which can have a direct effect on economic growth.<sup>1</sup>

The same problem arises when distance to Benton County, Arkansas, is used to identify the effect of Wal-Mart’s entry on employment and earnings outcomes. Benton County is located in the northwest corner of Arkansas, roughly equidistant from Los Angeles and Boston (1300 miles), and from Philadelphia and Phoenix (1000 miles). Population centers are gen-

---

<sup>1</sup>In labor economics, distance to schools has been used to identify the effect of schooling on earnings (see, e.g., Altonji, Elder, and Taber (2002) for Catholic schools, Maluccio (1998) for secondary schooling, Kane and Rouse (1993) for college). But in those cases the argument for exogeneity is weaker, because people can move to locate nearer to, or further from, a school; it is perhaps instructive that studies that use other instruments, such as compulsory schooling laws (e.g., Acemoglu and Angrist, 2000) find much smaller effects than studies using distance instruments. In contrast, the distance between counties, or countries, is fixed and not responsive to incentives.

erally located closer to the coasts; while the median county is 600 miles from Benton County, only four of the densest 30 counties in 1970 were located within 600 miles of Benton County. If population density affected only the *level* of retail employment per capita, this problem could be addressed by using county fixed effects. But urban and rural areas specialize in different industries (Holmes and Stevens, 2004) and have been growing at different rates (Glaeser and Kohlhase, 2003); they are also quite likely to experience different economic cycles. These differential economic cycles — hitting at different times and with different magnitudes — could spuriously show up as differences due to Wal-Mart’s expansion. I find some evidence that this, in fact, the case.

One standard approach to testing the validity of an instrument is to use overidentifying restrictions, but given the difficulty of finding plausible restrictions, I do not attempt this kind of test. There are, nevertheless, some standard tools that can be used to check instruments: counterfactual analysis being the most important. (Basker, 2005, for example, estimates the effect of Wal-Mart’s entry on employment in automobile dealerships and service stations, and in manufacturing, to help establish the validity of her instrument.) Another standard tool is robustness analysis: the validity of an instrument is called into question if results change dramatically when small changes are made to the sample or functional form. The distance instrument does not pass either of these tests.<sup>2</sup>

The rest of the note is organized as follows. Section 2 reviews Basker’s (2005) and NZC’s methodologies and summarizes their results.<sup>3</sup> Section 3 reports on my replications and

---

<sup>2</sup>The first stage regressions are invariably strong as measured by F statistics, so the problem is unlikely to be the traditional “weak instruments” problem discussed by Staiger and Stock (1997), Hahn and Hausman (2002), and elsewhere. Rather, the issue is that the instrument appears to be correlated with the error term, resulting in biased and inconsistent IV estimates. The problem of invalid instruments is discussed in detail by Hahn and Hausman (2003), who note that 2SLS estimates may be even more biased than OLS estimates when instruments are correlated with the error term. I ignore finite-sample bias in this note because sample sizes are fairly large (tens of thousands of observations).

<sup>3</sup>I focus here on NZC rather than Dube, Eidlin, and Lester (2005) because NZC use County Business Patterns data, following Basker (2005), and specifically consider the impact on jobs. Dube, Eidlin, and Lester use data from the Quarterly Census of Employment and Wages to estimate the impact on earnings only. Many, but not all, of the methodological issues discussed here apply also to Dube, Eidlin, and Lester (2005). I do not consider the effect on earnings because without controls for worker and job characteristics,

extensions of NZC’s results, and discusses their sensitivity to sample definition and functional form. I also perform some counterfactual analyses and determine, for example, that their specification implies not only a loss of 50-300 retail jobs in a county but also an increase of 500-600 manufacturing jobs in the same county. This number is highly implausible given the fact that Wal-Mart’s procurement is done on a national and international scale. Section 4 examines functional form problems in more detail and discusses possible explanations for the failure of the instruments. Section 5 concludes with new ideas for identification and lessons learned.

## 2 Review

### 2.1 Basker (2005)

Basker (2005) uses a balanced panel of 1,749 counties over 23 years from 1977-1999.<sup>4</sup> For each county, she uses retail employment as of the week of March 12 from County Business Patterns, and the number of Wal-Mart stores from various public sources, including company annual reports, the annual publication *Directory of Discount Department Stores*, and the Wal-Mart editions of the Rand McNally Road Atlas, which (since 1995) contain a list of existing Wal-Mart store locations. Because the public data sources contain errors, she uses company-assigned store numbers to proxy for stores’ planning dates. Taking the number of stores that opened each year as exogenous, she assigns “planning years” for stores in sequential order by number: the first store in 1962, the next two stores in 1964, and so on, based on the actual number of stores that opened each year. These figures are then aggregated to the county level for the instrument: the number of stores that would have

---

such as education, hours, and responsibilities, which are not available in either data set, these effects are difficult to interpret.

<sup>4</sup>The 1,749 counties are selected from over 3,000 US counties in the contiguous 48 states if they match three criteria: aggregate employment of at least 1,500 in 1964; positive aggregate employment growth between 1964 and 1977; and no Wal-Mart entry prior to 1977.

existed in a county in a given year, if stores had opened in the order in which they were numbered, holding the actual number of stores opening in each year fixed.

Basker argues that this instrument can also address endogeneity concerns, because planning is done too far in advance to foresee high-frequency employment fluctuations. Wal-Mart may take advantage of low-frequency trends in retail employment, however, which would bias any once-and-for-all estimate of Wal-Mart’s impact on retail employment. To check this, Basker uses an event-study specification with a five-year window before and after Wal-Mart’s entry: the lead coefficients are intended to capture any trend in retail employment in the years preceding entry. Specifically, Basker estimates

$$\frac{\text{retail}_{jt}}{\text{pop}_{jt}} = \alpha + \sum_j \gamma_j + \sum_t \delta_t + \theta(L) \frac{\text{Wal-Mart}_{jt}}{\text{pop}_{jt}} + \varepsilon_{jt}$$

where  $\frac{\text{retail}_{jt}}{\text{pop}_{jt}}$  is retail employment per capita in county  $j$  in year  $t$ ,  $\gamma_j$  and  $\delta_t$  are county and year fixed effects, respectively;<sup>5</sup>  $\frac{\text{Wal-Mart}_{jt}}{\text{pop}_{jt}}$  is the number of *new* Wal-Mart stores per capita in county  $j$  in year  $t$ , and

$$\begin{aligned} \theta(L) = \theta_1 F^5 + \theta_2 F^4 + \theta_3 F^3 + \theta_4 F^2 + \theta_5 F + \theta_6 \\ + \theta_7 L + \theta_8 L^2 + \theta_9 L^3 + \theta_{10} L^4 + \theta_{11} L^5 + \theta_{12} \sum_{\tau \geq 6} L^\tau \end{aligned}$$

where  $L$  is the lag operator and  $F$  is the lead operator. Because both retail employment (on the LHS) and the Wal-Mart variables (on the RHS) are given in per-capita terms, the coefficient vector  $\theta$  can be interpreted as jobs gained (or lost).<sup>6</sup>

Results from the OLS specification and the IV specification in which the counterfactual number of new Wal-Mart stores instruments for the actual number (inferred from store

---

<sup>5</sup>Basker interacts the year fixed effects with an urbanization indicator, but this does not affect the results qualitatively.

<sup>6</sup>This specification allows the year fixed effects to have a proportional effect on all counties, while the Wal-Mart variable has a constant level effect independent of county size.

lists) are presented graphically in Figures 1 and 2, respectively, reproduced from Basker (2005). In each case, the solid line shows point estimates, and the dotted lines show the 95% confidence interval. The OLS coefficients are positive both before and after Wal-Mart's entry, and show an increase in the year before Wal-Mart's entry (year 0) as well as the year after entry, followed by a modest (and statistically insignificant) decline. The fact that employment appears to be increasing *before* Wal-Mart's entry is a cause for concern, because it suggests one of two problems with the estimation: measurement error in opening dates (so that some stores appear in store lists a year after they actually opened) or endogeneity. In the latter case, Wal-Mart may observe an increase in retail employment and open a store the following year, in which case the increase in retail employment between years 0 and 1 cannot be attributed to Wal-Mart alone. The IV estimates address these concerns. With measurement error corrected, we still see an increase in employment between years (-1) and 0, but it is small relative to the post-entry increase, and statistically indistinguishable from zero. Employment increases by 100 jobs between years 0 and 1. Following Wal-Mart's entry, retail employment declines sharply as competitors close or cut back employment, then stabilizes, with a net effect five years after entry of approximately 50 retail jobs gained.

By its nature, Basker's instrument is only defined for counties with a Wal-Mart store. Basker argues that the sample selection addresses the problem of Wal-Mart's likely preference for growing counties: 75% of the counties in the sample had a Wal-Mart store open during the sample period, compared with only 13% of the excluded counties. If Wal-Mart's effect is similar across county sizes and types, then its effect on smaller and declining counties may be inferred from her estimates.

## **2.2 Neumark, Zhang, and Ciccarella (2005)**

NZC attempt to address the endogeneity in both the location of Wal-Mart's stores and the timing of entry using Wal-Mart's geographic pattern of expansion as an instrument capturing "intent to treat." To illustrate the instrument, I show Wal-Mart's store locations in 6-year

intervals — 1977, 1983, 1989 and 1995 — in Figure 3. NZC use this pattern to predict Wal-Mart’s presence in a county.<sup>7</sup> The IV strategy uses 19 instruments: distance (of each county centroid) from Benton County, Arkansas, interacted with 19 year indicators. The first stage joint F statistic is highly significant, except in some cases when the regressions are estimated separately by region.

NZC’s data differ from Basker’s in two important respects: the source of the Wal-Mart data, and the sample selection. NZC obtained a list of all Wal-Mart store locations, with their opening dates, from Wal-Mart Stores, Inc.; the resulting data set is virtually free of measurement error. (In an appendix, NZC document some of the measurement error problems with the Basker data. I note here that qualitative results are invariant to the data source; in Appendix A I document the differences between the two data sets in detail.)

Because their instrument is defined for all counties, and not only for counties that received a Wal-Mart store, their sample in principle could include all US counties. They end their sample period in 1995 because Wal-Mart’s expansion reached the coasts by that time, weakening the instrument’s predictive power. Although up to 3,032 counties are included in the sample at some point during this 19 year period, the sample is unbalanced and averages 2,699 counties per year. This restriction is due to omission of any county-year observation in which either data on retail employment or (more often) aggregate retail earnings are suppressed.<sup>8,9</sup>

---

<sup>7</sup>Holmes (2005) argues that this expansion strategy increased Wal-Mart’s profitability due to savings associated with “economies of density.”

<sup>8</sup>The paper actually lists three samples, A, B, and C. I restrict my attention to Sample A because it is the largest, and the one the authors focus on for the purpose of addressing employment.

<sup>9</sup>Data suppression is done to prevent disclosure of details regarding individual businesses. When data are suppressed, every data point is replaced with a range: e.g., 1-19 employees instead of 7 employees. Data suppression is more likely for small counties in which there are only a few retail establishments. Interestingly, Dube, Eidlin, and Lester (2005), using an empirical strategy very similar to NZC’s, but different data, find that Wal-Mart’s impact on retail earnings is *positive* in rural counties, but negative in metropolitan counties, implying that the selection based on this suppression flag may not be innocuous.

NZC estimate

$$\frac{1000 \cdot \text{retail}_{jt}}{\text{pop}_{jt}} = \alpha + \sum_j \gamma_j + \sum_t \delta_t + \phi \cdot \text{exposure}_{jt} + \varepsilon_{jt} \quad (1)$$

where all variables are as defined above, except the Wal-Mart variable, **exposure**<sub>jt</sub>, which I discuss below. The first stage regression is

$$\text{exposure}_{jt} = \alpha + \sum_j \gamma_j + \sum_t \delta_t + \sum_t \rho_t \cdot \text{distance}_j + \varepsilon_{jt} \quad (2)$$

where **distance**<sub>j</sub> is the distance in miles from county *j* to Benton County, Arkansas; the coefficient  $\rho_t$  takes on a different value every year, allowing the effect of distance to change over time.

Some functional form differences between Basker’s (2005) and NZC’s analysis are driven by the nature of the instruments. Basker’s instrument lends itself easily to a case-study specification, because it precisely predicts the opening date of stores; concerns about the imperfect nature of the instrument increase the importance of checking for increases in employment in the years leading to Wal-Mart’s entry. Because Wal-Mart’s expansion pattern consisted of “spreading out, then filling in” (Sam Walton, cited in NZC, page 3), simultaneously reaching new territory and adding stores in closer counties, NZC’s instrument is not as fine tuned. They argue that separately identifying leads and lags of the entry variable is impractical because they are high correlated (footnote 39, page 28).

The Wal-Mart variable, “exposure,” is defined as the sum of the ages of all Wal-Mart stores in the county. For example, if there are two stores in a given county, one five years old and the other twelve years old, the county’s “exposure” to Wal-Mart is  $5 + 12 = 17$ . This functional form restricts the additional effect of an existing Wal-Mart store in its 10th year of operation to be the same as the effect of an initial entry.<sup>10</sup> To interpret the results, NZC

---

<sup>10</sup>NZC justify this specification saying that the alternative measure — a count of stores — produced “IV

evaluate the predicted effect at the mean level of exposure in the sample.

Equation (1) also constrains the effect of a store (or a year of exposure) to be constant *per capita* across small and large counties. While Wal-Mart stores vary in size, the range over which they vary is several orders of magnitude smaller than the range of differences in the size of counties; in practice, all the identification is coming from smaller counties.<sup>11</sup> Interpreting the estimation results at the mean (by multiplying the coefficient by the average population in the sample), NZC’s OLS estimates of Equation (1) correspond to an increase of approximately 20 jobs per year of “exposure” to Wal-Mart; at mean county exposure, this implies a statistically-significant increase of 160 retail jobs. IV estimates, in contrast, are sharply negative: they show a decrease of approximately 35 retail jobs per year of exposure, or approximately 280 jobs for the average exposure period.

In the next section, I replicate these results with some minor changes and analyze their sensitivity to changes in functional form and sample selection criteria. I then perform a counterfactual analysis by using the same methodology to estimate Wal-Mart’s effect on county manufacturing employment.

### 3 Replication and Extension

#### 3.1 Replication

I use administrative data from Wal-Mart Stores, Inc. to estimate Equation (1). Following NZC, I limit the sample to the period 1977-1995, although I include all counties in the regressions.<sup>12</sup> I use both NZC’s measure of exposure and the more transparent number of

---

estimates of the overall employment effects were implausibly large (and negative)” (footnote 39, page 28). I return to this point in Section 4.

<sup>11</sup>In 1977, the 95th percentile county by population was 100 times larger than the 5th percentile county by population. Exact employment per Wal-Mart store is not known, but the difference is unlikely to be larger than 5-fold.

<sup>12</sup>My sample includes 57,855 observations, or 3,045 counties over 19 years. NZC’s “A sample” includes 51,274 observations. 261 counties have retail employment suppressed at some point during the sample period. Of these, 88 have a single instance of suppression, and another 56 have two instances each. In cases of data

stores as endogenous RHS variables.<sup>13</sup> Results are shown in Panel A of Table 1; both OLS and IV results conform with those reported in NZC. Remaining differences are most likely due to the slightly altered sample.

Evaluating the OLS coefficients at the sample means of population (77,700) and exposure (9.6), the average effect of a Wal-Mart store, using the exposure measure, is estimated to be a gain of 143 retail jobs. IV estimates of the same specification imply a loss of 164 jobs. The corresponding figures for the specification using the number of Wal-Mart stores in the county are 105 jobs gained (OLS) and 462 jobs lost (IV). The large difference between the exposure and stores IV results is in keeping with NZC’s contention that the stores specification results in “implausibly large” coefficients (footnote 39, page 28).

### 3.2 Normalization

Given the highly skewed distribution of both exposure and population, it is probably better to use median rather than means to evaluate the coefficients. The 1995 (the last year of the sample period) median exposure, conditional on at least one Wal-Mart store in the county, is 11; median sample population in 1995 is approximately 23,000. At the median, then, IV estimates for the exposure specification in Panel A of Table 1 imply a loss of 5 retail jobs per year of exposure, or 55 jobs for the median county with a Wal-Mart store in 1995. The stores specification implies a loss of 137 retail jobs per Wal-Mart store.

But since Wal-Mart’s per-capita effect in different markets is likely to be inversely related to the size of the market, a better solution is to use a functional form that allows stores to have a fixed level effect on employment. To do this, I normalize the Wal-Mart variable (number of stores or exposure to Wal-Mart) by population, as in Basker (2005). In Panel B

---

suppression, I replace the suppressed value with the mean employment for non-suppressed counties in the same range: 13 for 1-19 employees, 64 for 20-99 employees, and so on.

<sup>13</sup>As in NZC and Basker (2005), standard errors are clustered at the county level.

of Table 1 I repeat the above analysis using a normalized Wal-Mart variable:

$$\frac{1000 \cdot \text{retail}_{jt}}{\text{pop}_{jt}} = \alpha + \sum_j \gamma_j + \sum_t \delta_t + \phi \cdot \frac{1000 \cdot \text{Wal-Mart}_{jt}}{\text{pop}_{jt}} + \varepsilon_{jt} \quad (3)$$

where  $\text{Wal-Mart}_{jt}$  is either the number of Wal-Mart stores in county  $j$  at time  $t$  or county  $j$ 's exposure to Wal-Mart (sum of store ages) as of time  $t$ . The coefficient  $\phi$  is directly interpretable as the average effect of a Wal-Mart store on retail employment, or the average effect of one year of exposure to Wal-Mart, depending on the specification.

Focusing on the IV estimates, an average year of exposure leads to a loss of about 6 retail jobs (or about 67 jobs for the median county with a Wal-Mart store in 1995); the per-store effect is much larger — a loss of 283 jobs.

### 3.3 Counterfactual Analyses

#### 3.3.1 Timing

Since the instrument is not precise enough to allow separate estimation of pre- and post-entry effects, I employ an indirect test. I “pre-date” the opening of each Wal-Mart store by 2 years, and re-estimate Equations (1) and (3). Since two years of the pre-entry period are now confounded with the post-entry period, if the estimated decline in employment is due to Wal-Mart’s entry and not other, coincident, factors, we should see estimated coefficients drop towards zero (as we would in any case of measurement error).

Results from this exercise are shown in Table 2. Panel A shows results for the original NZC functional form which imposes a constant *relative* effect of Wal-Mart regardless of county size, and Panel B shows results that impose a constant *absolute* effect. In Panel A, OLS estimates are smaller, as expected; IV results for the exposure specification, however, show a larger coefficient. Evaluated at the median 1995 population, the perturbed regression shows a loss of 6.5 rather than 5 jobs for every year of exposure to Wal-Mart; since all stores have been aged by 2 years in the sample, the median exposure in 1995 is 13 rather than

11, implying an average job destruction of 84 jobs — 50% higher than the IV estimates of the unperturbed regression. The stores specification, in contrast, shows a smaller (absolute) point estimate compared with the baseline regression, and implies a loss of 79 rather than 137 jobs. The results for the normalized specification in Panel B are even more striking: IV estimates for both the exposure and the stores specifications are larger than those in Table 1. The exposure specification shows a loss of 6.6 rather than 6 jobs per year of exposure to Wal-Mart, or 87 jobs at the median exposure level; and the stores specification shows a loss of 430 jobs per store, 50% larger than before.

Since measurement error leads to attenuated point estimates, the *larger* absolute value of the estimates must be due to the bias introduced by changing stores’ opening dates. These results suggest that the instrument is picking up pre-existing trends that are correlated with, but not caused by, Wal-Mart’s expansion into a county.

### 3.3.2 Manufacturing

As a second counterfactual exercise, I estimate Wal-Mart’s impact on county-level manufacturing employment. Given the national and international character of Wal-Mart’s procurement (Basker and Van, 2006), we do not expect to find any effect on manufacturing employment at the county level. Basker (2005) reports no observed impact on manufacturing employment (Figure 11, page 181).<sup>14</sup>

Results are displayed in Table 3. As before, Panel A shows estimates of Equation (1) and Panel B shows estimates of Equation (3). In Panel A, OLS results for both the exposure and stores specifications are negative, indicating that Wal-Mart appears to locate in counties with relatively declining manufacturing sectors. The sign is reversed, however, for the IV estimates of the exposure specification: evaluated at the median 1995 population, each year

---

<sup>14</sup>Manufacturing employment is computed as the sum of employment across 2-digit manufacturing industries. Because each of these smaller sectors has a substantial amount of data suppression, the sum has substantial measurement error. However, there is no reason to believe this measurement error is systematically correlated with the Wal-Mart variables.

of exposure to Wal-Mart is estimated to create 46 jobs; with 11 years of exposure, this implies a gain of 511 manufacturing jobs *per county* attributed to Wal-Mart. The stores specification has smaller, but still quite large, opposite-sign results. If this specification is to be believed, each Wal-Mart store is responsible for a loss of 231 manufacturing jobs in the county. In Panel B, OLS results are positive, and both IV estimates are positive: a Wal-Mart store creates, on average, 630 manufacturing jobs, or 60 manufacturing jobs per year of exposure.

## 4 Discussion

### 4.1 Exposure

The differences in magnitudes across the exposure and stores specifications, along with the inconsistent results of the counterfactual exercises, raise questions about the validity of the baseline estimates. One reason for these problems is that the exposure variable artificially prolongs the relationship between distance and Wal-Mart’s presence. As the stock of stores in the area immediately around Bentonville ages, it continues to be over-weighted relative to areas receiving stores for the first time. Thus, older stores in proximate locations contribute to the exposure of nearby locations long after they are outnumbered by more distant stores.<sup>15</sup> NZC’s claim that the “identification strategy works much better for the exposure measure” than for a count of stores (footnote 39, page 28) has some support — first stage F statistics are roughly 50% larger for the exposure specifications — but first-stage results are strong across the board.

In any given year  $t$ , aggregate exposure (and exposure at each county) increases for two reasons: first, the existing stock of stores ages by one year; and second, new stores are added to the stock. If the existing stock of exposure is  $\mathbf{exposure}_{t-1}$ , and a total  $\mathbf{new}_t$  new stores

---

<sup>15</sup>This over-weighting of proximate counties could be made to last even longer if we used the sum of squared store ages instead of the sum of ages.

open, the contribution of new stores to the increase in aggregate exposure is given by

$$\text{newshare}_t = \frac{\text{new}_t}{\text{exposure}_t - \text{exposure}_{t-1}}$$

By construction, the first store, opened in Rogers, Arkansas, in 1962, contributed 100% of that year's increase in exposure; the share of new stores is strictly lower for all subsequent years. In 1977, for example, 158 existing stores all aged by one year, while 33 new stores opened; new stores therefore constituted 17.3% of the total increase in exposure. The annual contribution of new stores to exposure is shown, along with the number of new stores opened, in Figure 4; between 1977 and 1995, the contribution of new stores to the overall increase in exposure exceeded 25% only in 1981, the year Wal-Mart acquired 106 Kuhn's Big K stores in nine states. By 1995, the contribution of new stores to increased exposure fell below 5%.

Figure 5 plots the first-stage coefficients for both the exposure (solid line) and stores (dashed line) specifications. In each case, the coefficient  $\rho_t$  on the interaction of distance and year  $t$  indicator is given for each year 1978-1995 (the coefficient on distance interacted with the 1977 indicator is normalized to zero). The first-stage coefficients for the exposure measure correspond closely to those plotted in the bottom panel of Figure 3 (page 37) of NZC.

To interpret the first-stage coefficients for the store specification, recall that in the early years (in the 1960s and early 1970s) Wal-Mart had very few stores *anywhere*; so distance from Bentonville was a poor predictor of the locations of stores. As the chain expanded, building up its presence in Arkansas, Missouri, Oklahoma, and nearby states, distance became a fairly good predictor of store presence: the further away a location, the less likely it was to have a Wal-Mart store. As Wal-Mart's circle of stores expanded, however, this relationship between store presence and distance has once again become weak. This non-monotonic relationship between distance and the number of Wal-Mart stores shows up clearly in the figure. Until 1988, distance is increasingly negatively correlated with the number of stores in a county;

after 1988, the relationship between the number of stores and distance starts to reverse itself. By 1993, the coefficient on the interaction term is statistically indistinguishable from zero (they are normalized to zero for 1977), and by 1995 it is statistically positive: relative to 1977, being far away from Bentonville is correlated with having *more* stores rather than few stores. If we used the number of *new* stores rather than the number of existing stores, as NZC’s Figure 2 (page 36) motivates us to do, we would find a stagnant relationship between distance and new store openings for the period 1977-1986, then a strong increasing relationship between 1986-1992, followed by a stagnation and decline as new stores after 1992 are not located disproportionately in any distance ring.

The transformation of a count of stores to the exposure variable is not innocuous. Because the first-stage coefficients for exposure and stores move together for the first part of the sample, and in opposite directions for the second part, we can get IV estimates with different magnitudes and even different signs (as in the case of the manufacturing regressions in Table 3 above) depending on whether the bulk of the variation used to identify the coefficients is from the first or the second half of the sample period.

## 4.2 Identification

If the problem were merely that the RHS variable is misspecified, the manufacturing regressions shown in Table 3 would show no effect of Wal-Mart on manufacturing employment — at least for the normalized “stores” specification in Panel B. The problem, as those regressions demonstrate, is that Wal-Mart’s expansion pattern is correlated with other industry trends.

To see where the identification is coming from, I estimate the reduced-form equation

$$\frac{1000 \cdot \text{retail}_{jt}}{\text{pop}_{jt}} = \alpha + \sum_j \gamma_j + \sum_t \delta_t + \sum_t \psi_t \cdot \text{distance}_j + \varepsilon_{jt} \quad (4)$$

I show the coefficients  $\psi_{1978} - \psi_{1995}$  in Figure 6. (Since the reduced-form equation does not

include a Wal-Mart variable, these estimates do not depend on the definition of the Wal-Mart variable.) The solid line shows coefficient estimates from a regression that includes the full sample of counties (3,045 observations per year). The dashed line shows coefficients from a specification that used only the 1,537 counties with at least one Wal-Mart store by 1995; and the dotted line shows coefficients from a regression that includes only the remaining 1,508 counties with *no* Wal-Mart by 1995.

The reduced-form results show where the identification of Wal-Mart's impact comes from in these specifications. Wal-Mart opened its first stores in California, Pennsylvania, and Nevada in 1990, and its first Massachusetts, New York, Maryland, Oregon, Maine, and New Hampshire stores in 1992. At the same time, retail employment per capita, which had been high in distant counties relative to close ones for several years, began to converge. NZC are assuming that this relationship is due to Wal-Mart's entry into these distant counties. A major problem with this interpretation is the similarity between the patterns of the reduced-form coefficients for counties that never had a Wal-Mart store and counties that have had one (at least until 1991, when they diverge). Whatever caused a relative drop in retail employment per capita in "far" counties in 1985 caused it for both Wal-Mart and non-Wal-Mart counties; the divergence between "close" and "far" counties between 1986 and 1989 also did not discriminate between "Wal-Mart counties" and those that never got a Wal-Mart store.

One possibility is that different counties are growing at different rates, for reasons exogenous to Wal-Mart's presence. We could include county-specific linear trends in the regressions along with county fixed effects (which allow different levels of the intercept term), but this is impractical for three reasons. First, as NZC would argue, these county trends may be (at least partially) endogenous to Wal-Mart's presence. Second, and more problematically, they would be highly correlated with the instruments, leaving little variation in the instruments for identification. The remaining variation — after county-specific trends are "partialled out" of the instruments — is the year-by-year *deviations* of Wal-Mart's expansion path from a

linear trend: these deviations are very likely to be endogenous to economic conditions in the counties on its frontier of expansion. Finally, as the reduced-form results in Figure 6 demonstrate, trends may be insufficient to address the problem, if the differences are caused by different timing and/or extent of economic fluctuations.

If county-level trends must be excluded for the identification strategy to work, we need to look for specific culprits — variables correlated with distance from Benton County which may have a direct effect on the outcome variables, and for which we can control directly. I offer a couple of suggestions in the next section.

### 4.3 Possible Explanations

There are various possible reasons for the spurious results presented here. A major problem with the use of distance from Bentonville as an instrument is that county characteristics are spatially correlated. For example, in 1990, the correlation between a county's per-capita manufacturing employment and average per-capita manufacturing employment in counties within 100 miles was 0.58; it was 0.30 for per-capita retail employment, and 0.65 for log population. That same year, the correlation between the number of Wal-Mart stores in a county and the average number of stores in counties within 100 miles was 0.47. This, along with spatial correlation in other variables, both observable and unobservable, is likely to be a major part of the explanation.

One difference between “close” and “far” counties is their level of urbanization. As noted in the Introduction, the major population centers are concentrated in a relatively narrow distance band from Benton County — between 900 and 1,300 miles. This is shown graphically in Figure 7. (Note that the scaling of population density changes at 10,000 people per square mile.) With the exception of St. Louis City, 272 miles from Benton County with 10,000 people per square mile in 1970, virtually all dense population centers are concentrated within the band 900-1,300 miles from Benton County. Because distance from Bentonville is correlated with population density, so is the timing of Wal-Mart's entry

into a county. If business cycles reach dense and sparse counties at different times, and with different consequences for retail employment, these differential effects could be spuriously attributed to Wal-Mart's presence given its pattern of expansion.

I explore this explanation in several ways. First, I interact the year fixed effects with two urbanization dummies following Basker (2005): one if the county was inside a 1960 MSA, and another if it was outside, but within 20 miles of, a 1960 MSA. I also estimate the model interacting year fixed effects with the 1970 population density. When the Wal-Mart variable used is exposure, these alternative specifications (not shown) provide much smaller estimates of  $\phi$  (in absolute terms), and in some specifications  $\phi$  becomes statistically insignificant; the alternative specifications have little effect on the more transparent stores specification.

I also consider the possibility that “oil counties” — counties where the local economy moves with oil prices, often in the opposite direction to the rest of the country — which are mostly located close to Arkansas (primarily in Texas, Oklahoma, and Louisiana) were driving the results. I estimate the regressions omitting counties in which the 1970 employment share of the mining industry was above 10%, but found no statistically or economically significant differences with these counties omitted.<sup>16</sup>

There are other possible culprits, including differential economic trends and fluctuations due to different industry mixes, income distributions, housing stocks, climate, and many other variables. Investigating all of them is beyond the scope of this note; the main point here is that many variables are distributed unevenly in a way that is correlated with distance from Benton County.

---

<sup>16</sup>Following Buckley (2002), I use data from the 1970 Regional Economic Information System to determine the mining share of employment. See <http://www.bea.gov/beat/regional/reis/> for details.

## 5 Conclusion

The analysis presented here is not conclusive, but it demonstrates the sensitivity of the results and the correlation of Wal-Mart’s expansion pattern with factors that may directly affect the evolution of retail employment per capita. Geographic/time series patterns of retail employment are economically and statistically similar across counties with and without Wal-Mart, suggesting that the instrument is picking up other factors rather than a causal effect of Wal-Mart. A counterfactual analysis that artificially ages each store by 2 years, adding both noise and bias to the data, results in *larger* rather than smaller point estimates of Wal-Mart’s effect on retail employment. The IV estimates also show a systematic effect of Wal-Mart on county-level manufacturing employment which is substantially larger than the effect on retail employment.

NZC argue that almost any control variable in the regression could be endogenous due to “the pattern of Wal-Mart’s growth, and its extensive penetration” (footnote 23, page 12). The decision to omit control variables would be less problematic if their instruments were uncorrelated with other exogenous patterns — if, for example, the reduced-form coefficients looked significantly different for counties with and without Wal-Mart stores. Since that is not the case, it is impossible to determine econometrically the degree to which Wal-Mart has shaped the differential retail employment patterns experienced by “near” and “far” counties.

Unfortunately, this leaves us still without a definitive answer for Wal-Mart’s impact. The best evidence, though imperfect, comes from Basker (2005) and shows a small, but positive, treatment effect on the treated. More research is needed in order to uncover the true effect of Wal-Mart’s entry on local employment. In principle, a regression discontinuity specification could exploit the geographic pattern of expansion — using an instrument that “turns on” when a county enters the “frontier zone” of Wal-Mart’s reach, and “turns off” a year or two afterwards — although the first counties to be treated are likely to be different from counties treated later on, in ways that are probably correlated with Wal-Mart’s impact.

More generally, this note makes the known, but often underappreciated, point that

exogeneity does not automatically mean an instrument satisfies the exclusion restriction. This point has been made by others in similar contexts; see Altonji, Elder, and Taber (2002) for a discussion. Verifying that the instrument is not correlated with variables that can directly affect the outcome of interest is a first-order requirement.

## A Data

As noted in Section 2, Basker (2005) and NZC use different data sets to determine Wal-Mart's opening dates. In the note above, I have restricted myself to the official "administrative" data used by NZC, but since those data are not publicly available, I review the differences between the two sources. The raw data used in Basker (2005) are available at <http://economics.missouri.edu/~baskere/data/>.<sup>17</sup>

The administrative data file contains 10 stores which opened between 1977 and 1995 and which do not appear in the Basker data set. Of these, 9 have closing dates in the administrative data set, and four of them closed within 2 years of opening. Basker omitted stores that appeared in store lists for only one or two years as many of them were listed in error, which accounts for some of these omissions. The Basker data set also includes 14 stores that never appear in the administrative data; all but two of these have a closing date of 1993 in the Basker data — i.e., they appear in some editions of Chain Store Guides but never in the Rand McNally Road Atlas.

Among the 2,096 stores with (true) opening dates between 1977 and 1995 that can be matched across the two data sets, 1,198 (57%) have the same year of entry in both data sets, and an additional 627 (30%) have opening years within one year. Only 33 stores (1.6%) have an error of more than four years in the Basker opening date, although a handful of cases are egregious, with opening dates for five stores off by 10 years or more. (The raw correlation coefficient between the two sets of opening years is 0.97.) Figure A-1 shows the distribution of differences in opening dates for stores that opened between 1977 and 1995.<sup>18</sup>

Basker's use of store numbers to construct the instrument reduces this measurement

---

<sup>17</sup>Following the posting of NZC, Wal-Mart Stores, Inc. provided a spreadsheet with opening-date data on its public relations web site, <http://www.walmartfacts.com>, but later modified the spreadsheet by removing the opening dates.

<sup>18</sup>Many of these difference are due to the 1990-1993 period, for which entry dates in the Basker are imputed using state totals, as explained in the Appendix to Basker (2005). When those years are excluded from the comparison, 69% of stores have the same opening year assigned in both data sets, and 94% are within one year.

error considerably. Over the same time period (1977-1995), assigning store opening dates based on store numbers results in the correct year for 1,352 stores (65%); errors between the two are only weakly correlated (the raw correlation coefficient for the errors is 0.27).

## References

- Acemoglu, D., and J. Angrist (2000) “How Large are the Social Returns to Education? Evidence from Compulsory Schooling Laws,” *NBER Macroannual*, pp. 9–59.
- Acemoglu, D., S. Johnson, and J. A. Robinson (2001) “The Colonial Origins of Comparative Development: An Empirical Investigation,” *American Economic Review*, 91(5), 1396–1401.
- Altonji, J. G., T. E. Elder, and C. R. Taber (2002) “An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schools,” National Bureau of Economic Research Working Paper 9120.
- Basker, E. (2005) “Job Creation or Destruction? Labor-Market Effects of Wal-Mart Expansion,” *Review of Economics and Statistics*, 87(1), 174–183.
- Basker, E., and P. H. Van (2006) “Putting a Smiley Face on the Dragon: Wal-Mart as Catalyst to U.S.-China Trade,” unpublished paper, University of Missouri.
- Buckley, P. D. (2002) “The Effect of Male Earnings on Marriage Rates: Evidence from the Texas Oil Boom and Bust,” unpublished paper, MIT.
- Dube, A., B. Eidlin, and B. Lester (2005) “Impact of Wal-Mart Growth on Earnings throughout the Retail Sector in Urban and Rural Counties,” unpublished paper, University of California–Berkeley.
- Glaeser, E. L., and J. E. Kohlhase (2003) “Cities, Regions and the Decline of Transport Costs,” Harvard Institute of Economic Research Discussion Paper 2014.
- Graff, T. O., and D. Ashton (1994) “Spatial Diffusion of Wal-Mart: Contagious and Reverse Hierarchical Elements,” *Professional Geographer*, 46(1), 19–29.
- Hahn, J., and J. Hausman (2002) “A New Specification Test for the Validity of Instrumental Variables,” *Econometrica*, 70(1), 163–189.
- (2003) “IV Estimation with Valid and Invalid Instruments,” unpublished paper, MIT.
- Holmes, T. (2005) “The Diffusion of Wal-Mart and Economies of Density,” unpublished paper, University of Minnesota.
- Holmes, T., and J. J. Stevens (2004) “Geographic Concentration and Establishment Size: Analysis in an Alternative Economic Geography Model,” *Journal of Economic Geography*, 4(3), 227–250.
- Imbens, G. W., and J. D. Angrist (1994) “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 62(2), 467–475.

- Kane, T. J., and C. E. Rouse (1993) "Labor Market Returns to Two- and Four-Year Colleges: Is a Credit a Credit and Do Degrees Matter?," National Bureau of Economic Research Working Paper 4268.
- Maluccio, J. (1998) "Endogeneity of Schooling in the Wage Function: Evidence from the Rural Philippines," International Food Policy Research Institute, Food Consumption and Nutrition Division Discussion Paper 54.
- Neumark, D., J. Zhang, and S. Ciccarella (2005) "The Effects of Wal-Mart on Local Labor Markets," National Bureau of Economic Research Working Paper 11782.
- Sachs, J. D. (2003) "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income," National Bureau of Economic Research Working Paper 9490.
- Staiger, D., and J. H. Stock (1997) "Instrumental Variables Regression with Weak Instrument," *Econometrica*, 65(3), 557–586.

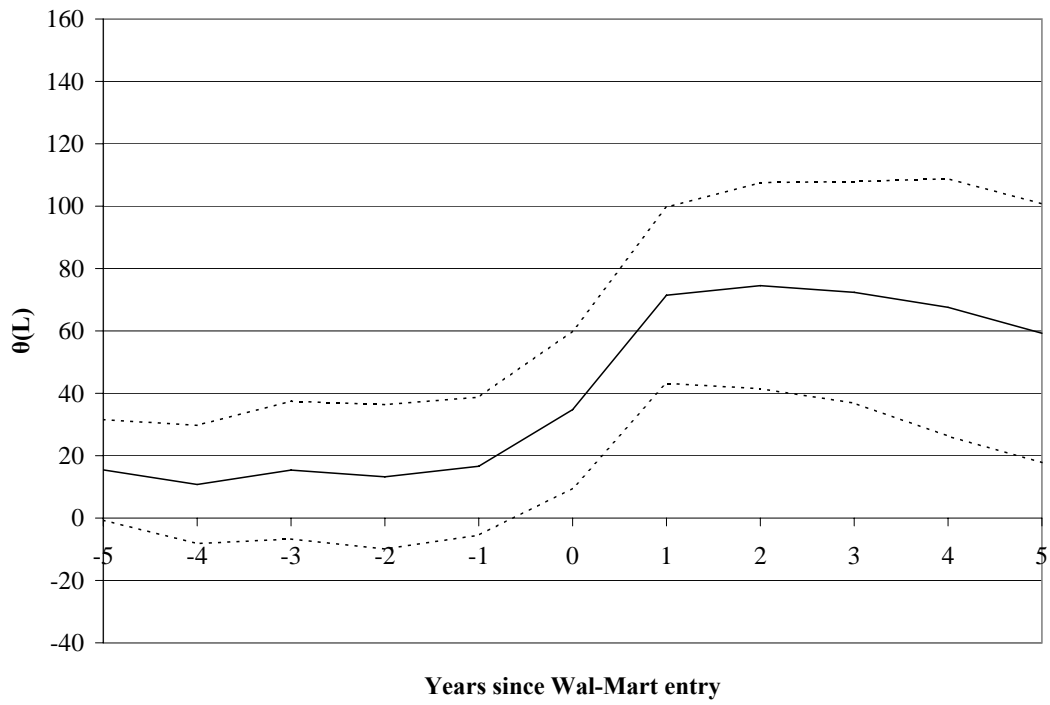


Figure 1. Basker OLS Results

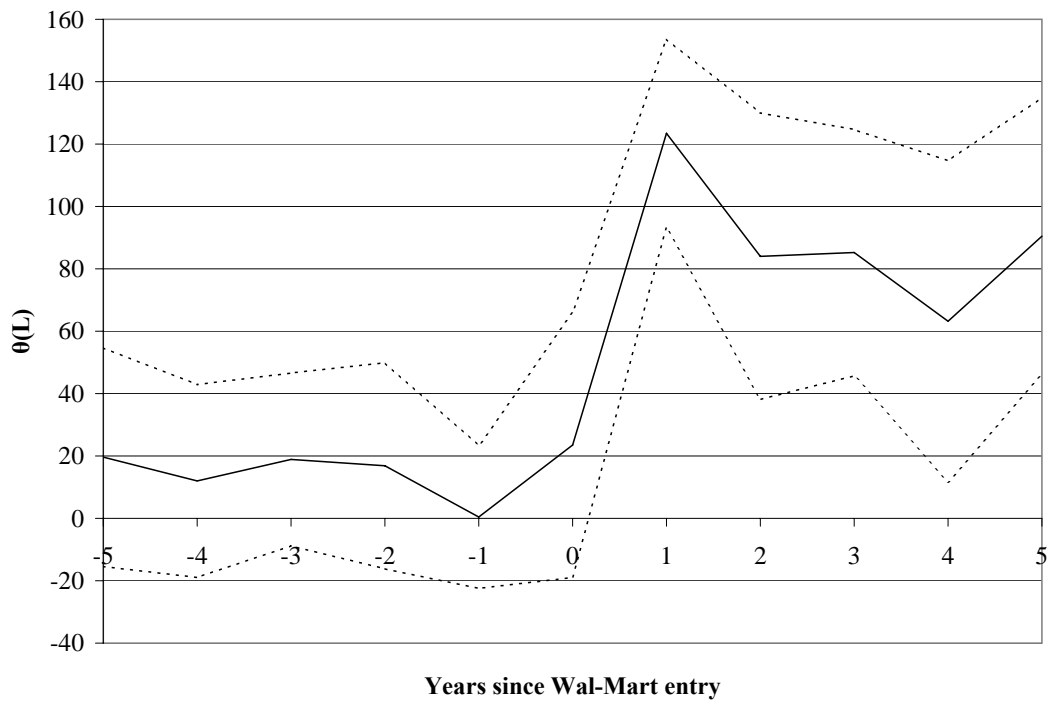


Figure 2. Basker IV Results

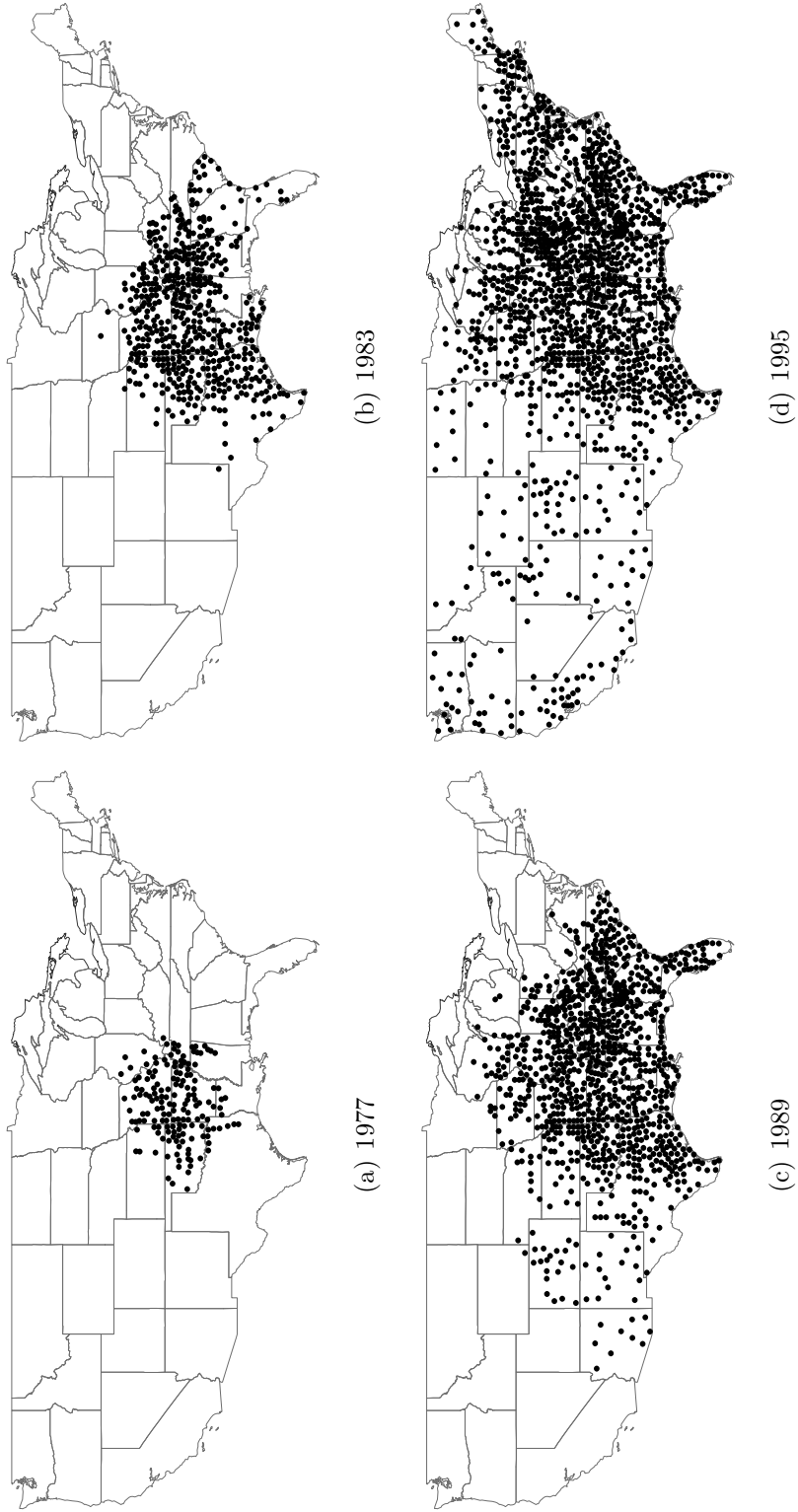


Figure 3. Wal-Mart's Store Locations, 1977-1995

Figure 4. New Stores' Contribution to Increased Exposure

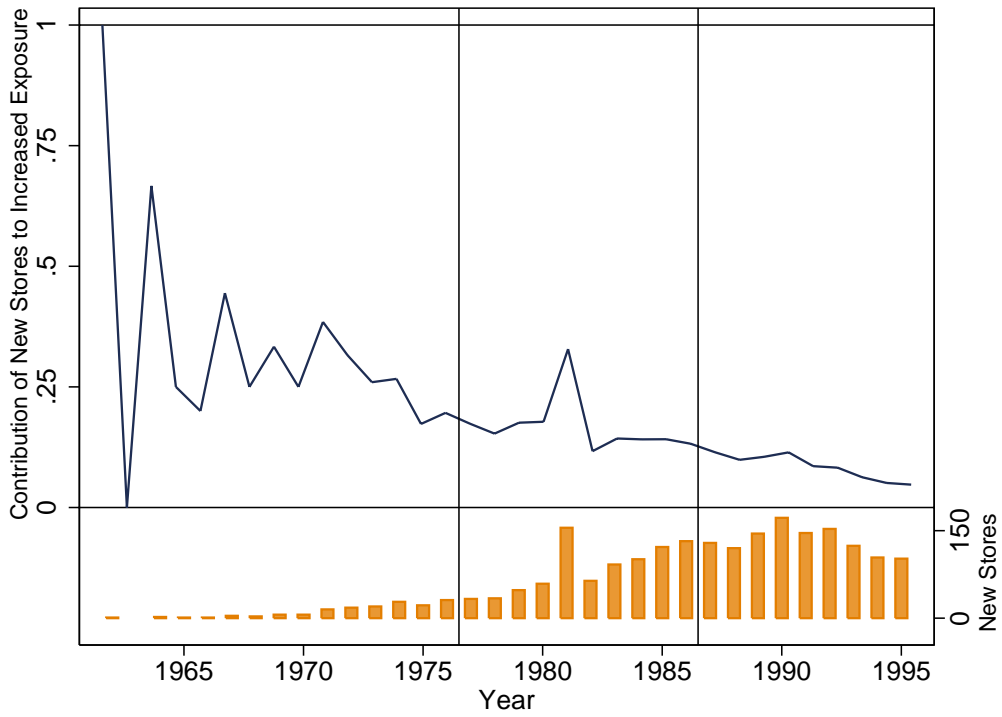


Figure 5. First Stage Coefficients on Distance by Time Interaction Terms

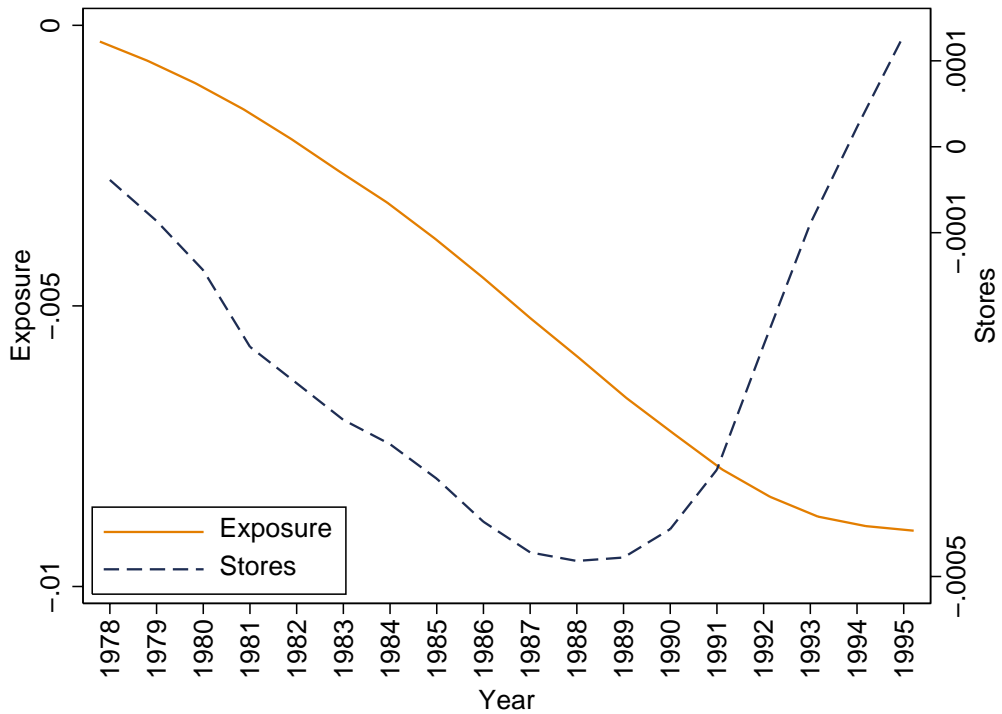


Figure 6. Retail Employment Reduced-From Coefficients on Interaction Terms

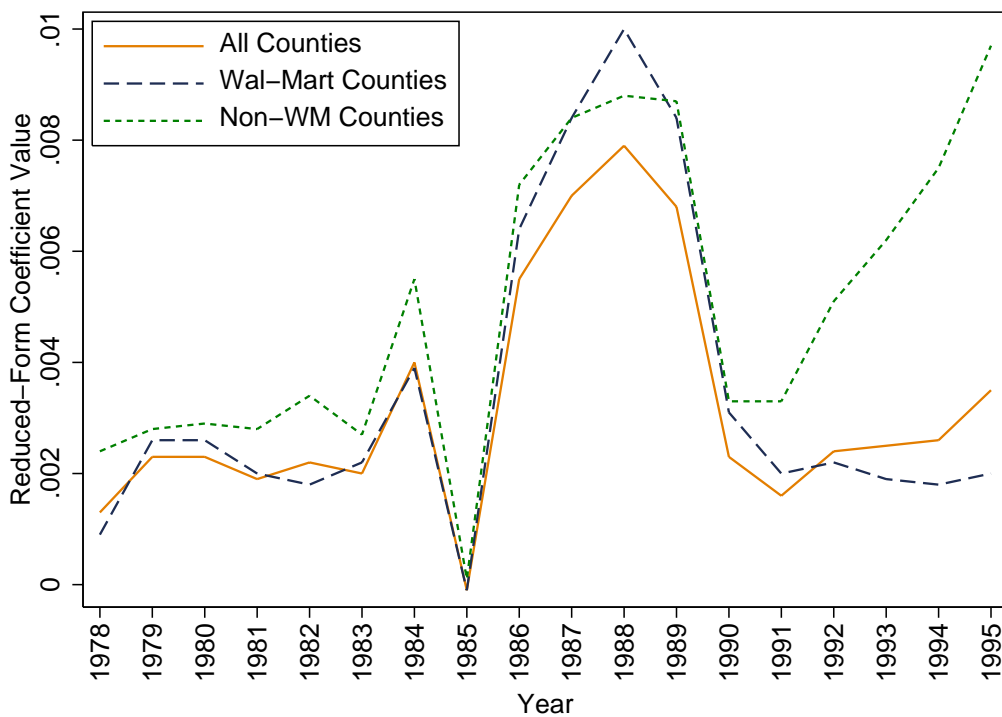


Figure 7. 1970 Population Density by Distance from Benton County

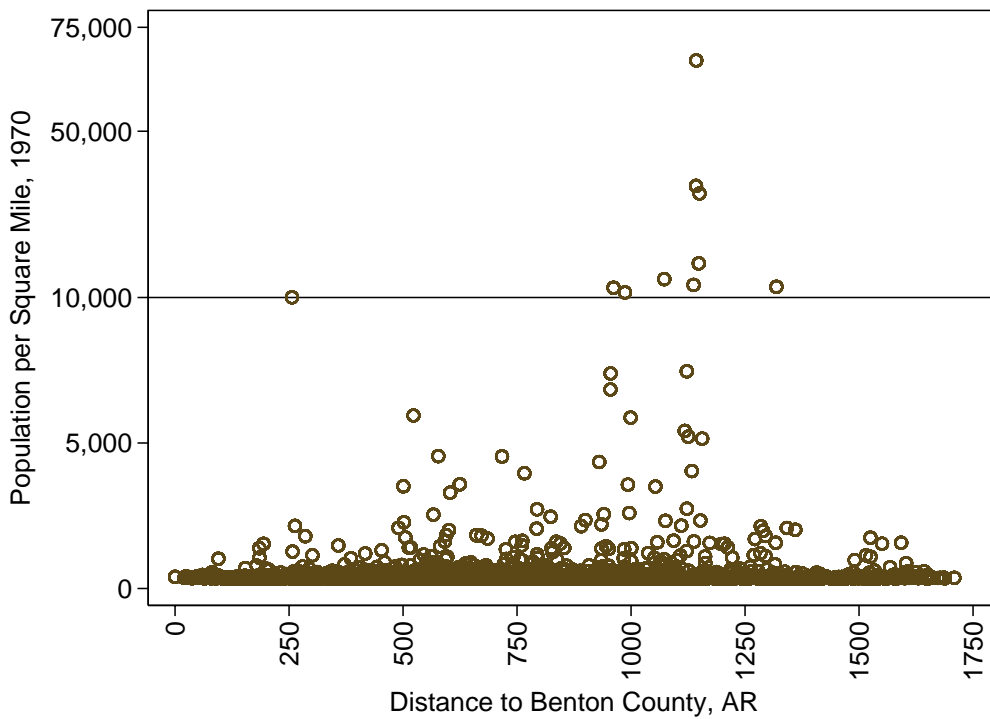


Table 1. Baseline Replication: Wal-Mart's Effect on Retail Employment

	OLS	IV	OLS	IV
<i>Panel A: Only LHS Normalized by Population</i>				
Exposure	0.1911*** (0.0312)	-0.2192** (0.1032)		
Stores			1.3537*** (0.3180)	-5.9479*** (0.9641)
First Stage F		32.6		25.0
p Value		0.0000		0.0000
<i>Panel B: LHS and RHS Normalized by Population</i>				
Exposure per Capita	6.1567*** (1.2541)	-6.1220** (2.9688)		
Stores per Capita			93.9383*** (12.8221)	-282.8821*** (52.8008)
First Stage F		38.4		22.2
p Value		0.0000		0.0000

Table 2. Counterfactual Exercise: Age Each Store by 2 Years

	OLS	IV	OLS	IV
<i>Panel A: Only LHS Normalized by Population</i>				
Exposure	0.1799*** (0.0308)	-0.2807*** (0.1067)		
Stores			1.1632*** (0.3984)	-3.4293*** (1.1588)
First Stage F		37.0		23.5
p Value		0.0000		0.0000
<i>Panel B: LHS and RHS Normalized by Population</i>				
Exposure per Capita	6.9631*** (1.2576)	-6.6503** (2.8687)		
Stores per Capita			89.6980*** (12.0032)	-429.7978*** (59.4084)
First Stage F		38.2		19.7
p Value		0.0000		0.0000

Table 3. Counterfactual Exercise: Wal-Mart's Effect on Manufacturing Employment

	OLS	IV	OLS	IV
<i>Panel A: Only LHS Normalized by Population</i>				
Exposure	-0.0932** (0.0408)	2.0183*** (0.1928)		
Stores			-2.5812*** (0.2684)	-10.0362*** (1.4780)
First Stage F		32.8		25.1
p Value		0.0000		0.0000
<i>Panel B: LHS and RHS Normalized by Population</i>				
Exposure per Capita	13.7834*** (2.1855)	59.7388*** (4.8571)		
Stores per Capita			39.7016** (16.6328)	629.6941*** (87.5268)
First Stage F		38.8		22.2
p Value		0.0000		0.0000

Figure A-1. Errors in Opening Dates

