

HOW FAR FOR A BUCK? TAX DIFFERENCES AND THE LOCATION OF RETAIL GASOLINE ACTIVITY IN SOUTHEAST CHICAGOLAND

Mark D. Manuszak and Charles C. Moul*

Abstract—We exploit variation in gasoline and cigarettes taxes in adjacent political jurisdictions for northern Illinois and Indiana to examine consumers' trade-off between prices and travel. We develop a model that relates activity in the retail gasoline industry around the tax borders to consumer locations. Our results indicate that the willingness of a typical Chicagoland consumer to travel an additional mile to buy gasoline corresponds to about \$0.065 to \$0.084 per gallon. According to our estimates, the observed area of Chicago, the jurisdiction with the highest taxes, is missing approximately 40% of the capacity that would exist were taxes equalized.

I. Introduction

MEASURING the effects of government policy on market outcomes is a general concern in empirical economics. Often, measurement of such effects is non-trivial. Unobserved heterogeneity across jurisdictions facing different policies can hinder cross-sectional analysis of the impact of policy differences, while co-movements of policy with other variables can make problematic the use of time series data. To deal with these problems, a broad literature has looked to estimate the effects of various government policies by exploiting different types of exogenous variation in those policies (Card & Krueger, 1994; Holmes, 1998; Milyo & Waldfogel, 1999; Jin & Leslie, 2003; Doyle & Samphantharak, 2008).

A more fundamental issue concerns the economic primitives that drive the effects of government policies. Rather than using exogenous policy variation simply to estimate the effects of a policy, one may be able to use that variation to link the effects of policy to underlying primitives. Clearly, this requires a model that describes the determinants of market outcomes and how government policy influences the market. Devising such a model may be more or less difficult depending on how closely one can tie the market outcomes to government policy and the underlying behavior of consumers and firms.

In this paper, we undertake an exercise that uses variation in government policies to uncover primitives that address the following questions: How much is the typical consumer

willing to pay to avoid traveling an additional mile? and How important is consumer heterogeneity along this dimension? To examine these issues, we employ data reflecting the relationship between activity in the retail gasoline industry and differences in local taxes on gasoline and cigarettes across a number of adjacent political jurisdictions in a small geographical area. The different taxes in the area that we examine provide exogenous variation in conditions facing firms in different locations: two gasoline stations could fall in close proximity to one another and yet face substantially different taxes on their primary and ancillary products. These different taxes can imply systematic variation in prices across tax regions. Because these taxes tend to be stable over long periods of time, consumers are presumably aware of the resulting price differences, as well as the travel required to obtain a lower price. In equilibrium, the entry decisions of firms and resulting concentration of economic activity will reflect the tax regime applicable to particular locations, as well as the potential consumer base, the willingness of those consumers to travel, and the proximity of the locations to different tax regions. We develop a simple model that reflects these concerns in order to exploit the observed variation in activity induced by local tax differences.

Our data come from four tax regions around Chicago and northern Indiana: the city of Chicago, Illinois; Cook County, Illinois, excluding Chicago¹; Will County, Illinois; and Lake County, Indiana. Each political jurisdiction levies unit taxes on both gasoline and cigarettes, and percentage sales taxes also apply. As table 1 documents, these tax differences are substantial. At 2001 prices, the tax per gallon on Chicago gasoline was almost 20 cents higher than that on Indiana gasoline, and the comparable tax difference per pack of cigarettes was about 89 cents. Our unit of observation within these regions is a census tract. For each tract, we gathered data on the number and size of gasoline stations. In addition, we have information from the U.S. Census about the locations of consumers, as well as other demographic and geographic information about census tracts, including their distances from various political boundaries.

Figure 1 offers two possible two-dimensional simplifications (high-tax Chicago to the left, low-tax Indiana to the right) of how we expect the concentration of retail gasoline capacity to vary with respect to the tax border. We generally anticipate a sharp discontinuity in activity in both regions that coincides with the political boundary accompanied by a decline in activity on the high-tax side as we approach the border. As we document in the next section, our raw data

Received for publication August 11, 2006. Revision accepted for publication March 27, 2008.

* Federal Reserve Board of Governors (Manuszak); Miami University (Moul).

The research of Charles C. Moul was supported in part by a grant from the Weidenbaum Center on the Economy, Government and Public Policy at Washington University. We thank the editor, George Borjas, and an anonymous referee for helpful comments. In addition to many conference and seminar participants, we thank Victor Aguirregabiria, George Deltas, George Gayle, Ron Goettler, James Prieger, Bruce Petersen, Russell Kohn, and Dan Akerberg for helpful comments and discussions. We also thank Xuehua Chen, Dan Dube, and Seung Bum Shin for their excellent research assistance and Stephen Manuszak and Forrest Moul for their data collection. The views expressed here are those of the authors and do not necessarily reflect the views of the Board of Governors of the Federal Reserve System or its staff.

¹ Although Chicago falls within Cook County, throughout we will use "Cook County" to refer to Cook County exclusive of Chicago.

TABLE 1.—TAXES IN NORTHERN ILLINOIS AND INDIANA FOR 2001

		Chicago	Cook County	Will County	Indiana
Gasoline per gallon taxes (cents per gallon)					
Federal	Federal Highway Trust Fund	18.3	18.3	18.3	18.3
	Federal Leaking Underground Storage Tank (LUST)	0.1	0.1	0.1	0.1
State	Illinois Motor Fuel	19.0	19.0	19.0	—
	Illinois Underground Storage Tank	0.3	0.3	0.3	—
	Illinois Environmental Impact	0.8	0.8	0.8	—
	Indiana Motor Fuel Excise	—	—	—	15.0
	Indiana Motor Fuel Inspection	—	—	—	0.8
Local	Cook County Motor Fuel	6.0	6.0	—	—
	City of Chicago Motor Fuel	5.0	—	—	—
Total per gallon tax		49.5	44.5	38.5	34.2
Cigarette per pack taxes (cents per pack)					
Federal		39	39	39	39
State	Illinois	58	58	58	—
	Indiana	—	—	—	15.5
Local	Cook County	18	18	—	—
	City of Chicago	16	—	—	—
Total per pack tax		131	115	97	54.5
Sales taxes					
State	Indiana sales tax	—	—	—	5.00%
	Illinois sales tax	6.25%	6.25%	6.25%	—
Local	Cook–Regional Transportation Authority	0.75%	0.75%	—	—
	Cook County sales tax	0.75%	0.75%	—	—
	City of Chicago sales tax	1.00%	—	—	—
	Will–Regional Transportation Authority	—	—	0.25%	—
		8.75%	7.75%	6.50%	5.00%
Total sales tax		8.75%	7.75%	6.50%	5.00%

Note: After January 1, 2003, the Indiana Motor Fuel Excise increased to 18 cents per gallon (cpg). After December 1, 2002, the Indiana sales tax increased to 6%. In Illinois, only 70% of the price on gasoline is subject to sales tax, implying an effective rate of 4.375%. In 2002, Indiana raised its state tax on cigarettes to 55.5 cents per pack (cpp). In June 2002, Illinois raised its state tax on cigarettes to 98 cpp. On April 1, 2004, Cook County raised its cigarette tax to 100 cpp, and raised that tax again on March 1, 2006, to 200 cpp. In 2005, the City of Chicago raised its cigarette tax to 48 cpp, and raised that tax again on January 1, 2006, to 68 cpp.

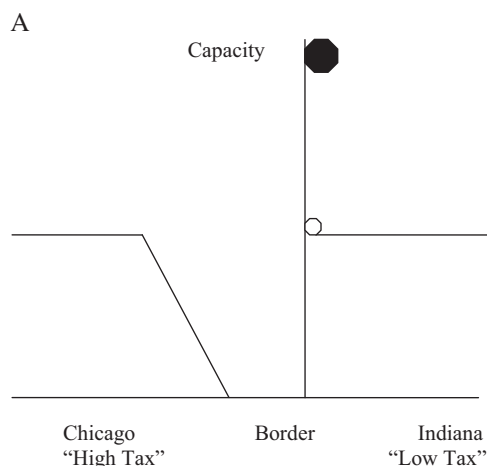
clearly exhibit such patterns. If cigarette taxes do not enter into smokers' decisions of where to buy gasoline, figure 1A, where a single slope transitions from normal capacity to no capacity, should be sufficiently descriptive. If, however, smokers consider not only lower gasoline taxes but also lower cigarette taxes when choosing how far to travel, we expect something resembling the second kink in figure 1B, where smokers are willing to travel farther than nonsmokers are.

To the extent that the tax border induces such entry patterns, our goal is to estimate some of the economic factors underlying figure 1 by observing how deep into a high-tax region one must travel to find the normal concentration of capacity and how much capacity exists in the mass point on the low-tax side of the border relative to the interior of the low-tax region. These patterns would then intuitively identify the answers to both of our questions. The slope and nature of the decline in capacity from the high-tax interior to the border reveal the extent of consumer heterogeneity, both in trade-offs between money and travel and in whether cigarette taxes matter. In an extreme case, a completely vertical drop-off would show that all consumers face the same trade-off between travel and money and that smokers do not incorporate their cigarette purchases when deciding where to buy gasoline. The typical consumer's trade-off is then approximated on the graph by the distance from the border to the midpoint of the downward-sloping region of capacity.

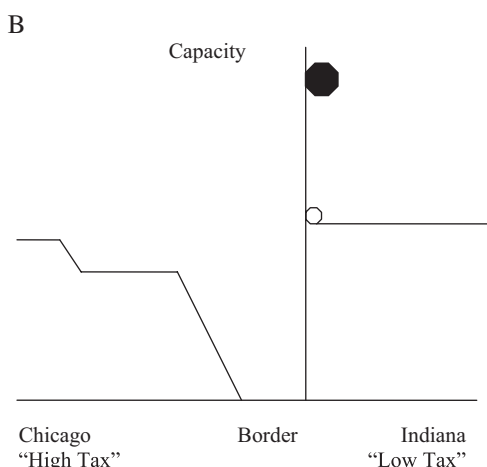
Reality, however, involves a number of complications that the stylized depiction in figure 1 ignores. First, our data are observed at discrete points that correspond to census tracts. Hence, without distributional assumptions, we could at best bound the distribution of willingness to travel at intermediate distances that do not coincide with observed locations in our discretized data, even in the two-dimensional world of figure 1. Second, locations are heterogeneous in dimensions other than their proximity to different borders. The distribution of households is not uniform throughout our data region, and other household demographic characteristics also vary across tracts. Third, our application involves a multidimensional choice problem as some consumers must choose between multiple low-price alternatives. As a result, allocation of absent high-tax activity to the low-tax region is no longer a straightforward calculation.

To deal with these problems, we propose and estimate a simple demand model that allows us to characterize the behavior of a consumer in Chicago, for example, who faces the trade-off of traveling some distance to Indiana or Cook County in order to save the amount implied by the tax differences. Our analysis is facilitated by the fact that our own gasoline price data, prior research, and institutional detail suggest that it is reasonable to assume that, in the long run, differences in average consumer prices across the region solely reflect tax differences. Consumers are heterogeneous in their willingness to travel and whether they

FIGURE 1.—POSSIBLE ENTRY PATTERNS ALONG A TAX BORDER
A. ONLY HETEROGENEITY IN PREFERENCES



B. HETEROGENEITY IN PREFERENCES AND SMOKING



smoke, and the total number of consumers in each location can depend on observable demographic characteristics of that location. Our outcome variable is the amount of capacity in each census tract measured as the total number of fueling positions at the stations in the tract. In our estimation, we relate our model of consumer demand to these observed capacity levels across locations.

The particular primitives of interest in our study—those reflecting the willingness of consumers to travel—are a first-order concern for a number of policy issues as has been recognized at least since the seminal work of Hotelling (1929). In industrial organization, this differentiation will be critical to market definition in the context of merger analysis of retailers. It will also have direct bearing on many types of tax competition among states and municipalities in public finance. Generally, the optimal structure of any system involving a geographically dispersed population will depend on the extent and nature of geographical differentiation.

Reflecting the relevance of these primitives of consumer behavior, a number of recent papers in industrial organiza-

tion have examined the importance of geographical differentiation. By extending the approach of Berry, Levinsohn, and Pakes (1995), these studies link aggregate information about consumer locations to prices and sales volumes across geographically dispersed theaters (Davis, 2006), gasoline stations (Manuszak, forthcoming), and fast food outlets (Thomadsen, 2005) in order to estimate the distribution of consumer disutility for travel and prices. In spirit, our analysis is similar, although our data are more aggregate as we observe the longer-run outcome variables of presence and size of gasoline stations rather than actual sales.² These earlier studies consider the implications of strategic pricing by firms in more detail, but they do not exploit any clear exogenous variation in factors that affect prices or location patterns. These papers instead identify their estimates of willingness to travel and the associated price elasticities by assuming exogenous product locations.³ Our estimates then offer a comparison that does not rely on this common assumption by examining the way in which product locations respond to exogenous and asymmetric shocks to the economic environment. The trade-off is that we must make strong assumptions, albeit ones that are supported in our data, about pricing behavior, as well as assumptions about the relationship between observed capacities and underlying quantities.

II. The Retail Gasoline Industry in Northern Illinois and Indiana

Table 1 documents the various taxes that are applicable in the region.⁴ When average self-service regular unleaded prices from three weekends in July and August 2001 are used, the combination of excise and specific taxes implies total taxes per gallon purchased of \$0.3971 in Indiana, \$0.4575 in Will County, \$0.5295 in Cook County, and \$0.5918 in Chicago. Assuming a seller's price of \$3.25 per pack, the comparable total tax on a pack of cigarettes is \$0.7075 in Indiana, \$1.1813 in Will County, \$1.4019 in Cook County, and \$1.5944 in Chicago.⁵ These tax differ-

² Our focus on the presence of firms is related to the literature on endogenous market structure initiated by Bresnahan and Reiss (1991) and Berry (1992). Unlike those studies, we consider variation in conditions within a single market rather than across multiple markets and do not consider strategic behavior on the part of firms. However, we attempt to uncover features of consumer demand at a more primitive level than do those studies.

³ The validity of this assumption relies on the desire of firms to locate near highly favorable locations canceling out the desire to avoid more competitive environments.

⁴ Other than Chicago, no local municipalities in this area levy their own gasoline or cigarette taxes. At the time of our data collection in mid-2001, almost all of the state and local taxes had been adopted in 1990 or earlier. As the notes to table 1 indicate, a number of tax changes occurred after 2001. However, the legislative history surrounding those tax changes and the associated coverage in the popular press suggest that those changes would not have been foreseen in 2001. For example, the earliest reference to a permanent change in the Indiana gas tax was proposed in January 2002 in HB 1317-2002, a bill that was later dropped.

⁵ Given 3% nominal price increases, this seller's price would yield prices roughly equal to those observed in 2006 in this area given current taxes.

ences are the exogenous variation that we look to exploit in order to examine the determinants of capacity location in this region.⁶

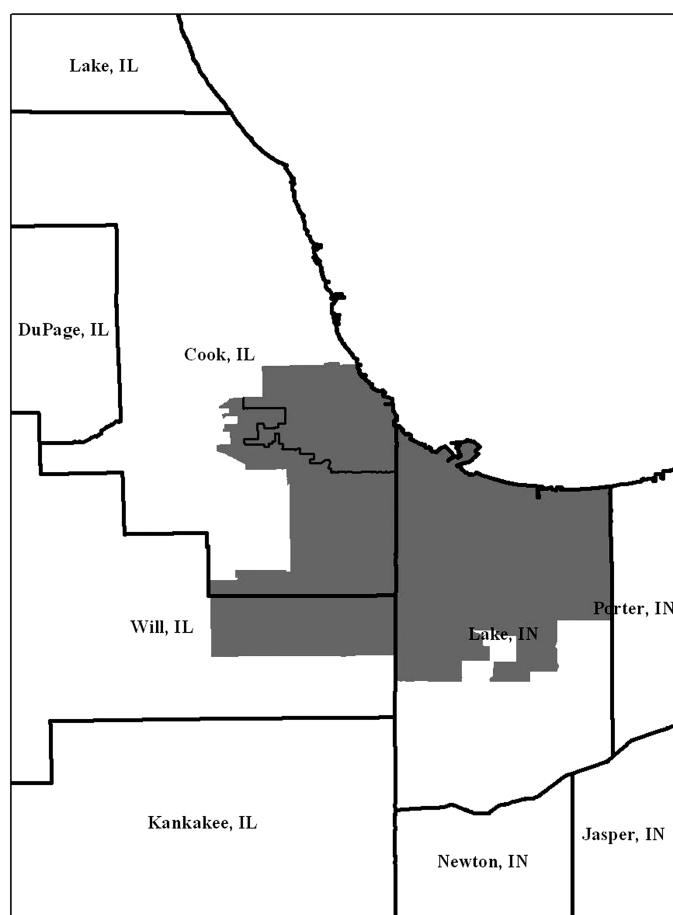
As noted in section I, we take the unit of observation to be a census tract in our empirical analysis. This definition of a location is convenient because census tracts do not cross political boundaries. We can also obtain demographic information about the consumers who reside in the tracts from the 2000 Census. In general terms, our raw data contain information on the presence and size of gasoline retailers across census tracts in different political jurisdictions, as well as characteristics of the consumers who reside in those tracts.

To obtain information on the presence of gasoline stations, we performed a comprehensive census of gasoline stations in this region during June 2001. We compiled a preliminary list of stations and their locations using local business directories. A physical canvass of the region, however, revealed that these listings were incomplete. As a result, we drove on nearly every road in the area to ensure that we had a complete list of all active stations. The region is sufficiently compact (approximately 580 square miles) and navigable that except for the occasional traffic jam, this method of data collection was not too onerous.

Figure 2 depicts the border region of northern Illinois and Indiana. The shaded parts of the figure indicate the data collection area.⁷ In the data collection area, the outer regions of southeast Chicago border Indiana to the east and Cook County to the south and west, while Cook County abuts Chicago to the north, Indiana to the east, and Will County to the south. While we looked to travel deep enough into the various tax regimes to get information on the impact of the borders at locations that are more or less close to them, the extent of our census was guided by time constraints. As a result, the boundaries at which we stopped collecting information on more stations are largely arbitrary. In addition to Chicago, some of the municipalities in our data are Gary and Hammond, Indiana, and Calumet City and Chicago Heights, Illinois.⁸

Throughout this area, we visited gasoline stations. Overall, we located 485 stations. For each station, we recorded the exact location using a handheld global positioning system; the address, if available; and the cross streets at which the station lies. We also recorded various character-

FIGURE 2.—A MAP OF NORTHERN ILLINOIS AND INDIANA



Note: Bold lines illustrate boundaries between counties. The shaded region indicates the data collection area. The black line on the interior of Cook County, Illinois, illustrates the Chicago city border in that area. The interior gap in the southern portion of Lake County, Indiana, is primarily a lake.

istics of the stations, such as brand, service levels provided, and the presence of ancillary services such as automotive care, a car wash, or a convenience store. An important characteristic for our later analysis is the number of fueling positions at each station, which corresponds to the total number of separate cars that can be served simultaneously at the station.

Table 2 provides descriptive statistics of the individual gas stations in our data and reveals some features of the retail gasoline industry in this area. Importantly, stations are generally similar across jurisdictions, as most stations are self-service only and almost always have convenience stores and the relatively new pay-at-the-pump technology. The prevalence of other services, such as automotive care and car washes, does not differ much across regions. The average number of fueling positions at stations in Indiana, however, is larger than that for stations in other regions, reflecting substantial skewness in the size distribution of Indiana stations. Although many Indiana stations are similar in size to their counterparts in other regions, Indiana contains a disproportionate number of very large stations, with the largest being almost two and a half times as big as the

Since this seller's price only matters through the implications of the sales tax, our results are not very sensitive to our choice of this seller's price.

⁶ All of these jurisdictions face the same requirements for reformulated gasoline, so production costs should not systematically vary across the region. Other policies may differ across the jurisdictions, but we do not expect these policies to asymmetrically affect retail gasoline activity in a way that depends on the borders.

⁷ The outer portions of our data collection area in Chicago and Cook County are far enough from the two other main political jurisdictions in the region, DuPage County, Illinois, and Lake County, Illinois, that no consumer in our later model would travel to them to buy cheaper gasoline given the closer and even cheaper available alternatives.

⁸ The exact boundaries of the data collection area are available on request from the authors.

TABLE 2.—STATION CHARACTERISTICS BY MARKET

Variable		All Markets	Market			
			Indiana	Will County	Cook County	Chicago
Self-service only	Mean	0.95	0.95	0.92	0.92	0.97
Full service only	Mean	0.01	0.01	0.00	0.01	0.01
Both service levels	Mean	0.04	0.04	0.08	0.07	0.02
Pay at pump	Mean	0.776	0.74	0.77	0.84	0.76
Auto service	Mean	0.098	0.1	0.08	0.11	0.08
Car wash	Mean	0.124	0.15	0.00	0.15	0.07
Restaurant	Mean	0.073	0.1	0.23	0.03	0.05
Number of self-service fueling positions	Mean	8.58	9.36	8.46	8.14	7.73
	s.d.	3.91	5.26	4.15	2.32	1.71
	25th	6	6	6	7	6
	Median	8	8	8	8	8
	75th	8	12	10	8	8
	Max	48	48	20	16	12
Number of full-service fueling positions	Mean	2.19	1.91	4.00	1.82	4.00
	s.d.	1.47	0.83	—	0.87	3.46
Number		485	208	13	139	125

largest station elsewhere. As we document below, much of this heterogeneity arises due to the proximity to borders with higher tax regions.

While our data set contains a great deal of information about individual gas stations, we aggregate the data up to the level of census tracts in order to obtain demographic characteristics of the areas around stations. In total, our survey covered 420 tracts. The on-site latitude-longitude readings allow us to place each station in a particular tract for which we can determine the total number of stations and fueling positions. This latter variable provides the basic measure of service capacity that we will emphasize in the analysis that follows. We choose fueling positions rather than station counts as our outcome variable because, as our analysis of the raw data will document, the former provides a better measure of the actual capacity that a location possesses.

Table 3 provides demographic information about the census tracts in the different regions from Summary Tape Files 3 and 4 of the 2000 Census. One tract characteristic is the total number of households as a measure of market size. We also gathered data on other demographic characteristics, such as median income, median house values, and poverty measures, to account for differences in consumer characteristics across the tracts. Using tract size, we computed population density using households for population. Finally, we recorded the number of interstate exits in each of the tracts from maps of the region, distinguishing between freeways and tollways.

As table 3 indicates, the tracts are not drastically different in many of the demographic characteristics. The average number of households is roughly the same across all regions, reflecting the Census Bureau's attempts to make tracts similar along that dimension, but substantial variation does exist. The sample average of median income is similar for Indiana and Chicago, though the fraction of households in poverty is somewhat higher in Chicago. Both Will County and Cook County have higher incomes, higher

property values, and lower poverty rates, particularly in the case of Will County. Given the relatively affluent nature of the Chicago suburbs, these differences are not surprising.⁹ One notable difference across these regions involves variation in population density. Given its more urban nature, Chicago naturally has a higher population density than the other areas. To the extent that differences in these variables affect outcomes in the market, we need to take them into account.¹⁰ This table, however, suggests that the three key regions—Chicago, Cook County, and Indiana—are not radically different from one another.

We also obtained the geographical centroid point of each tract from the census. This location information, which we later use as an approximation for the location of consumers and gasoline capacity in each tract, allows us to determine the distance from the various tracts to other tracts, including those that fall in different political jurisdictions. We measured these distances using the straight line or “great circle” distance. As can be seen by the scale of figure 2, our data are sufficiently dispersed to imply a fair amount of variation in these distances.

Table 3 also provides summary statistics on the number of stations and fueling positions for the tracts by region. This table provides additional insight into the variation in industry activity across markets. Tracts in Indiana are more likely to have stations, and its tracts with stations tend to have more stations and fueling positions than locations in other regions. Conversely, Chicago tracts are the least likely to have stations and tend to have fewer and smaller stations in the event that they have any stations. Cook County is an

⁹ Despite the common view that the South Side of Chicago is more impoverished than surrounding areas, two of the more destitute towns that we visited were Dixmoor, Illinois, in Cook County and Hammond, Indiana. Certain areas in Chicago where we gathered data appear to be fairly affluent.

¹⁰ In our empirical analysis, we also include region-specific effects to allow for differences arising due to factors such as access to public transportation.

TABLE 3.—CENSUS TRACT CHARACTERISTICS BY MARKET AND NUMBER OF STATIONS

Variable		Indiana			Will County		
		All	No Stations	Positive Stations	All	No Stations	Positive Stations
Stations	Mean	1.98	0	2.45	1.44	0	1.86
	s.d.	1.69	—	1.55	1.13	—	0.90
Fueling positions	Mean	18.54	0	22.91	12.22	0	15.71
	s.d.	22.06	—	22.40	9.72	—	7.87
Households (000s)	Mean	1.62	1.21	1.71	1.54	1.16	1.65
	s.d.	0.70	0.59	0.69	0.67	0.05	0.73
Median income (\$0000s)	Mean	3.88	2.89	4.12	5.62	6.59	5.35
	s.d.	1.61	1.42	1.56	1.25	2.07	0.98
Poverty rate (%)	Mean	13.72	21.93	11.79	4.07	2.15	4.61
	s.d.	12.91	13.73	11.99	3.38	2.05	3.60
Median house value (\$00,000s)	Mean	0.85	0.65	0.90	1.27	1.62	1.17
	s.d.	0.43	0.30	0.44	0.46	0.97	0.28
Population density (000s/sq mile)	Mean	1.38	1.88	1.27	0.45	0.24	0.52
	s.d.	1.06	1.45	0.92	0.55	0.16	0.61
Any freeway exits	Mean	0.28	0.05	0.33	0.11	0	0.14
Any tollway exits	Mean	0.05	0.05	0.05	0	0	0
Number		105	20	85	9	2	7

Variable		Cook County			Chicago		
		All	No Stations	Positive Stations	All	No Stations	Positive Stations
Stations	Mean	1.43	0	2.11	0.60	0	1.42
	s.d.	1.41	—	1.22	0.83	—	0.69
Fueling positions	Mean	11.67	0	17.15	4.62	0	10.98
	s.d.	12.26	—	11.25	6.47	—	5.42
Households (000s)	Mean	1.63	1.56	1.66	1.23	1.08	1.43
	s.d.	0.50	0.55	0.47	0.86	0.79	0.91
Median income (\$0000s)	Mean	4.49	4.51	4.48	3.52	3.57	3.45
	s.d.	1.12	1.19	1.10	1.79	2.06	1.32
Poverty rate (%)	Mean	8.82	9.63	8.44	20.59	21.07	19.92
	s.d.	8.13	9.22	7.62	14.08	14.86	13.00
Median house value (\$00,000s)	Mean	1.04	1.00	1.06	1.02	1.05	0.97
	s.d.	0.30	0.33	0.29	0.47	0.57	0.29
Population density (000s/sq mile)	Mean	1.78	1.76	1.80	4.31	4.25	4.41
	s.d.	1.04	0.93	1.09	2.60	2.88	2.18
Any freeway exits	Mean	0.23	0.10	0.29	0.13	0.12	0.14
Any tollway exits	Mean	0.03	0	0.05	0.03	0.02	0.03
Number		97	31	66	209	121	88

intermediate case between these two extremes. This table does not provide information about the extent to which within-market variation is associated with proximity to different borders, but it is suggestive of some differences in outcomes, reflecting underlying economic conditions such as the tax differences.

The natural problem with a simple comparison of activity across regions is pointed out by Holmes (1998). The presence of other region-specific characteristics confounds the ability to draw conclusions about the relationship between a specific policy and economic activity. This problem leads Holmes to exploit the existence of a border to account for other unobservable factors. Extending Holmes's analysis, we wish to explicitly exploit proximity to the border in order to separate region-specific effects from those associated with distance from different tax regimes. In other words, rather than using the border as a convenient and arbitrary division of regions that are otherwise similar except for the discrete differences such as the policies that

Holmes examines, we wish to learn about and exploit the importance of the border itself.¹¹

Tables 4 and 5 present regression estimates that investigate the importance of the tax differences and borders in more detail. The use of observations on either side of the border allows us to estimate the differential impact of the border, while observations within the regions allow us to separate these border effects from overall region-specific effects. In table 4, we consider the number of stations as the outcome variable, whereas the number of fueling positions is the outcome in table 5. Since the borders may induce discontinuities in activity, we capture border effects with crude dummy variables indicating whether the tract in

¹¹ Our use of the border and surrounding areas is similar to that of Campbell and Lapham (2004) in which they investigate the impact of exchange rate fluctuations on economic activity around the U.S.–Canada border. Card and Krueger (1994) and Milyo and Waldfogel (1999) also employ adjacent states in their analysis to partly account for unobserved heterogeneity, although they do not focus on border effects.

TABLE 4.—REGRESSIONS FOR NUMBER OF STATIONS

	(1)	(2)	(3)	(4)	(5)
Constant	1.939 (0.155)	1.072 (0.069)	1.802 (0.154)	1.029 (0.189)	1.040 (0.197)
Chicago	-1.341 (0.165)		-1.195 (0.165)		-0.398 (0.172)
Cook County	-0.506 (0.210)		-0.447 (0.233)		-0.221 (0.238)
Indiana × Chicago border		5.829 (0.198)	5.099 (0.241)		5.168 (0.402)
Indiana × Cook County border		2.006 (0.800)	1.276 (0.811)		1.366 (0.680)
Indiana × Will County border		0.970 (0.900)	0.239 (0.911)		-0.619 (0.996)
Chicago × Indiana border		-0.966 (0.112)	-0.613 (0.063)		-0.945 (0.140)
Chicago × Cook County border		-0.423 (0.177)	0.023 (0.173)		-0.541 (0.179)
Cook County × Indiana border		-0.700 (0.256)	-0.935 (0.289)		-1.306 (0.290)
Cook County × Chicago border		1.166 (0.300)	0.893 (0.336)		0.810 (0.330)
Cook County × Will County border		-0.258 (0.302)	-0.507 (0.347)		-0.719 (0.369)
Will County × Indiana border		-1.667 (0.451)	-1.667 (0.451)		-1.228 (0.413)
Will County × Cook County border		0.595 (0.457)	-0.135 (0.477)		-0.204 (0.409)
Households				0.680 (0.104)	0.607 (0.104)
Households below poverty level				-1.379 (0.412)	-1.161 (0.392)
Population density				-0.115 (0.026)	-0.074 (0.024)
Median house value				-0.020 (0.130)	-0.008 (0.113)
Median income				-0.078 (0.039)	-0.037 (0.031)
Any freeway exits × Indiana				1.279 (0.321)	1.145 (0.336)
Any freeway exits × Cook County				0.224 (0.323)	0.388 (0.271)
Any freeway exits × Chicago				-0.119 (0.272)	0.098 (0.258)
R^2	0.181	0.162	0.284	0.241	0.392

Note: $N = 420$. Standard errors robust for heteroskedasticity in parentheses.

question is on a particular border.¹² Because we expect asymmetric effects of borders, our regressors include interactions of border dummies with region dummies.

These regressions provide several consistent insights. All of the regressions indicate differences across the regions regardless of other tract characteristics.¹³ This suggests differences in the interior behavior of the regions that may be associated with the overall tax level, access to public transportation, or other regionwide factors. Other demographic characteristics, specifically the number of house-

holds and number of households living below the poverty level, also account for some of the differences in activity across the tracts.

One notable feature of these tables is the difference in the impact of freeway exits across regions.¹⁴ Throughout the tables, exits in Indiana have a large positive value, while the estimates for Cook County are smaller and the Chicago estimates are negligible. One may be concerned that exits in the interior of Indiana are an avenue that Chicago consumers use to access cheaper Indiana gasoline. However, almost all of the interstate exits in Indiana that drive this coefficient are located along Interstate 65, which runs parallel to the state border and is approximately 10 miles from the border. As a result, the activity at these exits likely reflects consumers who are traveling

¹² In the few cases that a tract borders on more than one regime, we specify the border variable as the percentage of the tract that borders each regime. For example, one tract in Indiana is adjacent to both Chicago and Cook County. Because this Chicago–Cook County border roughly bisects the Indiana tract, we set the Chicago border dummy to 0.55 and the Cook County border dummy to 0.45.

¹³ We do not include a separate Will County effect or an interaction between Will County and the number of freeway exits since estimates for those variables were invariably small and imprecise.

¹⁴ In unreported results, the impact of tollway exits was negligible across all regions.

TABLE 5.—REGRESSIONS FOR FUELING POSITIONS

	(1)	(2)	(3)	(4)	(5)
Constant	18.044 (1.995)	8.605 (0.585)	14.802 (1.399)	8.740 (2.445)	7.060 (1.664)
Chicago	-13.422 (2.044)		-10.031 (1.481)		-3.311 (1.507)
Cook County	-6.374 (2.348)		-4.133 (1.999)		-2.268 (2.015)
Indiana × Chicago border		130.515 (24.751)	124.318 (24.783)		124.880 (26.577)
Indiana × Cook County border		29.361 (11.022)	23.164 (11.095)		23.706 (9.856)
Indiana × Will County border		17.493 (3.955)	11.295 (4.155)		2.578 (5.611)
Chicago × Indiana border		-7.523 (1.054)	-4.611 (0.476)		-7.482 (1.061)
Chicago × Cook County border		-4.329 (1.212)	-0.641 (1.160)		-6.075 (1.493)
Cook County × Indiana border		-6.298 (1.957)	-8.016 (2.238)		-10.803 (2.288)
Cook County × Chicago border		10.908 (2.878)	8.915 (3.119)		8.025 (3.086)
Cook County × Will County border		-2.276 (2.352)	-4.095 (2.697)		-5.539 (2.718)
Will County × Indiana border		-13.667 (3.784)	-13.667 (3.784)		-11.596 (3.068)
Will County × Cook County border		5.062 (3.828)	-1.135 (4.034)		-2.327 (2.923)
Households				6.588 (1.159)	5.268 (1.046)
Households below poverty level				-14.893 (4.269)	-11.412 (3.707)
Population density				-1.185 (0.287)	-0.608 (0.212)
Median house value				0.949 (1.480)	1.047 (1.237)
Median income				-0.858 (0.480)	-0.211 (0.302)
Any freeway exits × Indiana				9.541 (3.130)	9.714 (3.135)
Any freeway exits × Cook County				-0.729 (2.702)	1.638 (2.244)
Any freeway exits × Chicago				-1.479 (2.243)	0.864 (2.069)
R^2	0.155	0.389	0.465	0.186	0.541

Note: $N = 420$. Standard errors robust for heteroskedasticity in parentheses.

in a much broader sense than simply making a trip from Chicago to buy gas. While we will need to account for the effect of exits, the behavior of these consumers is not a first-order concern in our analysis.

The most striking estimates in these tables correspond to the border-region interactions. First, these border effects explain a great deal of variation in the observed outcomes, more than the demographic variables. Second, the estimates for Indiana with respect to both the Chicago and Cook County borders are positive and large, particularly in the former case. Conversely, the border effects for Chicago and Cook County with respect to Indiana are negative and substantial. Chicago also experiences a decline in activity along the Cook County border, while Cook County experiences a corresponding spike. The evidence with respect to the Will County border is less clear, likely due to the small number of observations for that region. Nevertheless, the estimates suggest that ac-

tivity in Cook County declines near the Will County border, while the latter experiences a drop-off close to Indiana. Overall, the entry patterns along the various borders indicate that the border effects are increasing in the tax difference between the regions.

Our regressors in tables 4 and 5 can explain much more variation in fueling positions than variation in the number of stations. This finding is intuitive because the station regressions discard information by treating a 48-position station the same as a 4-position station. While the number of stations in Indiana is higher near the tax borders when compared to the interior, the difference of capacity is pronounced when one accounts for the unusually large size of the border stations. Given that proper interpretation of our later results hinges on a correct measure of capacity, we interpret these regressions as indicating that number of fueling positions is a more appropriate variable in our subsequent model.

TABLE 6.—STATION-LEVEL REGRESSIONS FOR SELF-SERVICE REGULAR PRODUCER PRICE

	(1)	(2)	(3)	(4)	(5)	(6)
July 27 and 28	0.890 (0.003)	0.891 (0.003)	0.894 (0.003)	0.892 (0.004)	0.891 (0.003)	0.890 (0.004)
August 19 and 20	1.146 (0.002)	1.147 (0.002)	1.149 (0.002)	1.148 (0.003)	1.147 (0.002)	1.145 (0.004)
August 26 and 27	1.271 (0.002)	1.271 (0.002)	1.274 (0.003)	1.272 (0.004)	1.271 (0.002)	1.270 (0.004)
Chicago		0.010 (0.004)		0.011 (0.006)	0.010 (0.004)	0.014 (0.006)
Cook County		−0.012 (0.003)		−0.015 (0.004)	−0.012 (0.003)	−0.014 (0.004)
Will County		0.023 (0.008)		0.024 (0.008)	0.023 (0.008)	0.022 (0.008)
Chicago × Distance-to-Indiana					−0.005 (0.004)	−0.003 (0.004)
Cook County × Distance-to-Indiana					−0.001 (0.001)	−0.001 (0.001)
Will County × Distance-to-Indiana					0.000 (0.011)	0.001 (0.011)
Chicago × Distance-to-Cook County					0.001 (0.003)	0.002 (0.003)
Cook County × Distance-to-Will County					0.003 (0.003)	0.004 (0.003)
Indiana × Distance-to-Chicago					0.002 (0.002)	0.001 (0.002)
Indiana × Distance-to-Cook County					−0.003 (0.003)	−0.002 (0.003)
Indiana × Distance-to-Will County					0.003 (0.002)	0.002 (0.002)
Will County × Distance-to-Cook County					0.003 (0.009)	0.004 (0.010)
Cook County × Distance-to-Chicago					0.003 (0.003)	0.003 (0.003)
Demographics	N	N	Y	Y	N	Y
Interstate exits by regime	N	N	Y	Y	N	Y
R ²	0.9486	0.9509	0.9495	0.9515	0.9514	0.9520

Note: $N = 911$. GLS allowing for differing variances across weeks and random effect for stations. Noninterstate exit demographics are demeaned. Distances are demeaned at regime level.

In short, these simple regressions indicate that the borders matter.¹⁵ In addition, the data display the type of discontinuous jump in activity on the low-tax side of the border that figure 1 depicts, as well as the drop in activity on the high-tax side. The problem with this analysis is that while it indicates the importance of the border effects, it provides little insight into the factors leading to those effects. In what follows, we look to develop a model that exploits this variation to estimate information about the trade-offs that consumers face and other factors that affect the concentration of capacity.

We also have some limited information about 2001 gasoline prices in the area. We collected self-service regular gasoline prices over three weekends in the summer of 2001 for 307 of the 485 stations in our complete data.¹⁶ Table 6 presents regression results relating seller prices net of taxes at the station level to tax regime effects, time effects, and the demographic characteristics that we included in our capac-

ity regressions.¹⁷ We also include interactions between regimes and distances from various borders to examine whether prices exhibit systematic patterns around the borders. We use distance measures in these regressions rather than border dummies as in our earlier analysis for two reasons. First, while station presence may exhibit sharp discontinuities along the borders, prices may display smoother variation at different distances from the border. Second, from a practical point of view, the absence of stations and corresponding price observations along the border in some regions makes inclusion of border dummies problematic.

These regressions reveal a number of patterns in prices across the region. Price movements over time are by far the most important source of variation in our data, as time dummies alone explain almost 95% of the variation in seller prices. In contrast, demographics have little explanatory power. The estimates do suggest that seller prices vary slightly across regions as the point estimates indicate that compared to Indiana, Chicago and Will County retailers respectively receive an additional 1 to 2 cents per gallon

¹⁵ The presence and magnitude of these effects are apparent through casual observation. Borders are only haphazardly marked in this region by official signs, but the Indiana border is always recognizable by the sudden appearance of large gasoline stations.

¹⁶ The excluded stations in the price data primarily fall in the outer portions of our data collection area.

¹⁷ The regressions involve GLS allowing for station-specific random effects and different variances across weeks.

while Cook County retailers receive 1 cent less.¹⁸ However, the economic significance of these estimates is arguably marginal. Finally, prices appear to exhibit little systematic variation associated with the borders, as the estimates of the distance effects are always small and statistically insignificant.

The basic conclusion that we take from these results is that most of the significant variation in gasoline prices that consumers across this area face at any time arises due to differences in taxes. In light of these findings, we will later assume that all prices within a tax region are the same and that any price differences across regions reflect differences in taxes. While our raw data contradict the former assumption as variation in gasoline prices does exist within regions, table 6 suggests that this variation is not systematic.¹⁹ Our findings and subsequent assumption are consistent with the results of Chouinard and Perloff (2004) and Alm, Sennoga, and Skidmore (2009) who, in more comprehensive analyses of gasoline prices across the country, cannot reject that, on average, state and local taxes are fully passed on to consumers.²⁰ When combined with the common claim by gas station operators that absolute margins on gasoline are roughly constant across regions and over time,²¹ full tax pass-through would generate consumer prices that solely reflect tax differences. From a theoretical viewpoint, the assumption of constant margins and total pass-through does not necessarily run against the findings of some market power among gasoline retailers in Borenstein and Shepard (1996) and elsewhere.²²

III. A Model of Local Capacity Determination

Rather than considering a complex model involving the strategic station location decisions of petroleum companies and individual station operators, we rely on our assumptions about prices to develop a substantially simplified, but not wholly unrealistic, approach to examine the determinants of activity in the retail gasoline industry in this region. We begin by positing a demand structure for consumers who may exhibit unobserved heterogeneity along discrete (smoker versus nonsmoker) and continuous (relative distaste for travel) dimensions. We then propose that the

observed capacity levels measure, with error, some transformation of quantities, which suggests a natural estimation approach in which we exploit this proposed relationship between our observed outcome data and the predictions of our model.

A. Local Demand for Gasoline

We consider data aggregated to the level of census tracts. Hence, the demand model concerns consumer choices between these different tracts as well as the base amount of gasoline demanded at each tract. Following Manuszak (forthcoming), we assume that each consumer makes a discrete choice among different possible sources of gasoline. Consumers dislike higher prices, but they also dislike traveling to purchase gasoline. Furthermore, consumers are heterogeneous in their evaluation of this trade-off. Smokers consider the tax savings from their cigarette purchases as well as their gasoline purchases but are otherwise identical to nonsmokers.

Formally, gasoline capacity exists in the K census tract locations. Consumers are also distributed across the K locations. Consumer i 's utility from purchasing gasoline at location k is given by

$$u_{ik} = A_i - \beta_i(q_i^G p_k^G + q_i^C p_k^C) - \delta_i d_{ik}, \quad (1)$$

where p_k^G and p_k^C are the respective prices of gasoline and cigarettes at location k , d_{ik} is the round-trip distance from consumer i to location k , q_i^G and q_i^C are the respective sizes of consumer i 's gasoline and cigarette purchases, β_i and δ_i are consumer i 's disutility of price and distance respectively, and A_i is consumer i 's base utility of a purchase.²³

Due to data constraints and identification concerns, we make a number of assumptions about consumer heterogeneity in the market. First, we do not know the total amount that could be sold at each location. As a result, our model does not include the outside option of no purchase, implying that the base utility, A_i , is not identified. While this value would be of interest for aggregate gasoline demand, it is less crucial in our application where we wish to learn about substitution across the different locations.

Second, we have little information about the actual size of consumer gasoline purchases. Without such information, heterogeneity in gasoline purchase size is indistinguishable from heterogeneity in price sensitivity. As a result, we assume that $q_i^G = q^G$ for all i .²⁴ We further assume that

¹⁸ In our later estimation, if we incorporate these estimated price differentials in addition to taxes, our results are largely unchanged.

¹⁹ When we discussed potential price differences with consumers in this area, almost all individuals were aware of both the tax differences and how prices might reflect those taxes even if they were unaware of actual prices that exist across or within regions at a particular time.

²⁰ In their analysis of price responses to temporary moratoria of gasoline sales taxes in Illinois and Indiana in the summer of 2000, Doyle and Samphantharak (2008) find that the short-run pass-through is somewhat less than 100%, along with some weak evidence of variation in pass-through associated with borders of states not subject to the tax repeals.

²¹ A per-gallon margin of \$0.08 to \$0.10 over the wholesale price is an industry consensus.

²² Using a model with geographical differentiation and strategic pricing, Manuszak (forthcoming) obtains estimates for the retail gasoline industry in Hawaii that imply both some market power with approximately constant absolute margins and approximately full pass-through of any taxes.

²³ While gasoline stations typically sell cigarettes, this model does not require that smokers buy their cigarettes from the same establishment that they buy their gasoline, only that they do both within the same census tract.

²⁴ As long as fill-up size does not vary systematically with location, this assumption should not be too problematic. We do not have disaggregate information about automobile holdings or actual purchase sizes by location, so we cannot verify whether such correlation exists. If fill-ups in low-tax border regions are larger than the average fill-up, then our analysis will tend to understate aversion to travel. This bias would arise because, in order to justify the observed high capacities along the border, our model

$q^G = 13$, the national average fill-up size (NPD Group, 2000). For similar reasons, we assume that $q_i^C = 0$ for nonsmokers and $q_i^C = q^C = 10$ for smokers. That is, we assume that smokers will buy a carton (ten packs) of cigarettes when they travel for their gasoline. If we interpret our model as reflecting a weekly fill-up decision for a household, then our assumption about cigarette purchases would imply slightly higher cigarette consumption by smokers than the U.S. average of one pack per day (Centers for Disease Control, 2005; Capeheart, 2001).²⁵

Third, in our demand model, consumer purchase decisions will always involve the ratio $(\frac{\delta_i}{\beta_i})$. Any assumption about the individual terms of this ratio implicitly involves an assumption about the overall ratio. Hence, we view the total ratio $(\frac{\delta}{\beta})_i$ as the item that will involve consumer heterogeneity. Of course, the ratio $(\frac{\delta}{\beta})_i$ represents the dollar amount that a consumer would have to be compensated in order to travel one mile and, as such, represents exactly the price-distance trade-off that we wish to measure. We further assume that $(\frac{\delta}{\beta})_i = (\frac{\delta}{\beta})v_i$ where $\ln(v_i) \sim F(v)$ with parameters μ and σ_v .²⁶ The distributions that we consider are log-uniform and log-normal. Finally, we normalize $E[v] = 1$, which implies that $\frac{\delta}{\beta}$ measures the mean willingness to travel across consumers. This normalization implies that $\mu = -\frac{\sigma_v^2}{2}$ in the log-normal case and $\mu = \ln(2\sigma_v) - \ln(e^{\sigma_v} - e^{-\sigma_v})$ in the log-uniform case.

To implement the model, we make some additional assumptions. First, we assume that all consumers in a particular tract are located in the center of the tract. The distance from a consumer to his or her home location is zero, so consumers do not need to travel in order to purchase locally. The distance to an alternative location is then the distance from this centroid point to the centroid point of the alternative.²⁷ As noted, we use the straight-line distance to compute the relevant travel distances.²⁸

would require more consumers with a fixed smaller fill-up size to travel than actually do.

²⁵ Heterogeneity in purchase size for cigarettes could cause difficulties similar to those for gasoline. A further problem could occur if smokers stockpile cigarettes, since their purchase behavior at any time would depend on past purchases and outstanding stock of cigarettes. While our data contain no information about this possibility, we expect a primary effect would be a distortion of our estimated measure of the share of smokers in the population.

²⁶ We consider log distributions to avoid the perverse result that some consumers might travel some distance to obtain a higher price.

²⁷ Our results are not sensitive to the use of geographical or population-weighted centroid points. The use of distances to actual stations would be problematic as we do not know exactly which station a consumer would patronize.

²⁸ Davis (2006), Manuszak (forthcoming), and Thomadsen (2005) use similar information about consumer locations and travel distances, although they compute distances from consumers to individual retail outlets. These previous studies used one-way rather than round-trip distances. The basic assumptions about travel patterns, however, are the same, so this distinction affects only the interpretation of the dollar-distance trade-off implied by estimated coefficients.

Because much of our analysis rests on distances, this assumption warrants some discussion. If consumers routinely commuted to Indiana from Chicago, our inferences would be hopelessly distorted. The most common travel patterns, however, are travel from a low-tax region (Indiana or Cook County) to a high-tax one (Chicago). Thus, we do not anticipate that this assumption will substantially bias our results. Moreover, aside from a few casinos and the ubiquitous fireworks stores in Indiana along the Illinois border, we did not observe any clear attractions in Indiana, other than the gas stations and cheap cigarettes, that would draw consumers from Illinois. To the extent that Chicago consumers, for example, incidentally find themselves in Indiana and, furthermore, just happen to buy gasoline immediately before returning to Chicago, our analysis will tend to understate consumers' aversion to travel.²⁹

Next, we assume that the price of gasoline and cigarettes at each location in a tax region is the tax for that region applied to a base producer price that is the same across all regions. Gasoline price differences in our data indicate that this assumption is roughly satisfied on average, although some price heterogeneity obviously exists within a tax region.³⁰ To the extent that tax differences are the primary source of price differences that consumers consider when deciding whether to travel from a high-tax region to a low-tax one, we do not view this assumption as a bad first approximation even absent the price patterns that we do observe.

Faced with the alternatives, each consumer in our model purchases from the location that yields the highest utility. The previous assumptions immediately imply some restrictions on consumer purchase behavior. First, a consumer will never travel to a higher tax region to buy gasoline. In other words, a consumer in Indiana will never travel elsewhere to purchase her gasoline, while a consumer in Cook County will never travel to Chicago. Second, if a consumer purchases in her tax region, she always does so at her own location. Purchasing from a more distant location in the same tax region would imply a distance cost with no commensurate price savings. Third, a consumer in a higher tax region would only consider traveling to the nearest available locations in each of the lower tax regions. Traveling to the interior of a lower tax region would be pointless, as only an additional distance cost would be incurred. Finally, a consumer who faces two lower tax regions would consider purchasing in the higher of the two only if her

²⁹ A similar bias would arise if a Chicago consumer must travel a positive distance to purchase locally or could purchase at the immediate edge of a border tract in Indiana rather than traveling to the interior of that tract. In either case, we will be overstating the distance difference required to purchase in Indiana rather than Chicago, which will tend to understate aversion to travel.

³⁰ To be more consistent with our findings about prices in our data, we could alternatively assume that consumers evaluate expected prices when choosing among different alternatives with the same expected price within each region and variation in expected prices across regions arising solely due to tax differences.

distance to that region was less than her distance to the lower tax region. This implies that a consumer in Chicago who is closer to Indiana than Cook County would always purchase in Indiana if she were to travel.

More formally, partition the set of K locations according to their tax regimes. Let R_t denote the set of locations in regime t , and order the regimes such that $t = 1$ corresponds to the lowest tax region (Indiana) and $t = 4$ corresponds to the highest tax region (Chicago).³¹ Our assumption about prices implies that $p_k^G = p_t^{G*} \equiv p^G (1 + \tau_{2t}^G) + \tau_{1t}^G$ for all $k \in R_t$ where τ_{1t}^G and τ_{2t}^G are the specific and ad valorem taxes, respectively, on gasoline in region t . Our assumption about cigarette prices yields a similar expression for the price of that product in region t .

A consumer chooses the alternative that maximizes her utility. To characterize the consumer's decision, define

$$k_{it} = \{k : d_{ik} = \min_{m \in R_t} d_{im}\}. \quad (2)$$

In other words, k_{it} provides the index of the closest location in tax region t to an individual in location i . The assumption about prices within tax regions implies that each consumer effectively chooses among four alternatives that correspond to the closest locations in each region. Letting ϕ_{ik}^{NS} denote the probability that nonsmoker i will purchase from location k , utility maximization implies that for any $k \in R_t$,

$$\phi_{ik}^{NS} = \begin{cases} 0 & \text{if } k \neq k_{it} \\ P(A_i - \beta q^G p_t^{G*} - \delta v_i d_{ik} \\ \quad \geq \max_s \{A_i - \beta q^G p_s^{G*} - \delta v_i d_{ik_{is}}\}) & \end{cases} \quad (3)$$

An analogous expression can be derived for smokers with an adjustment for the monetary implications of buying cigarettes in different tax regions. Denoting this probability for smokers as ϕ_{ik}^S and letting γ be the fraction of smokers, the probability that a consumer at location i will purchase from location k is

$$\phi_{ik} = (1 - \gamma)\phi_{ik}^{NS} + \gamma\phi_{ik}^S. \quad (4)$$

In many cases, these purchase probabilities simplify depending on the distances to various alternatives. Table 7 lists the various distance relationships and the related choice probabilities for nonsmokers where, to economize on notation, we use indexes t to denote alternative k_{it} in the distances and $P(\text{Chicago})$, for example, to denote $\phi_{ik_{i4}}^{NS}$. This table indicates that an alternative is irrelevant if a closer alternative from a lower tax region exists. For example, the top row states that a consumer will always purchase from Indiana if the closest alternative in that region is nearer than the other three regions. In the event that she is closer to Will County than Indiana, as in the second row, her decision involves a choice between the closest locations in those two regions, with her decision depending on her disutility from the higher prices in Will County

balanced against the distance she would have to travel to reach Indiana. The last row describes the behavior of an individual who is closest to a Chicago alternative, followed by Cook County and Indiana alternatives.³² This individual will travel to Indiana if the smallest distance-weighted price cost is large relative to her disutility of distance or may stay in Chicago if any price savings is insufficient to offset the travel costs. The probability that she purchases in Cook County is then the residual probability of her behavior relative to Chicago and Indiana. If Cook County were more distant than Indiana for this consumer, she would simply choose between Chicago and Indiana, as indicated in the fifth row of table 7.

The log-distribution assumption on v facilitates computation of these probabilities and illustrates where the distributional assumptions and normalizations assist in identification of the various parameters. For example, the log-normality assumption implies that the probability of a local purchase for a nonsmoking consumer in Chicago who is closest to a Chicago alternative followed by an Indiana alternative is

$$\phi_{ik_{i4}}^{NS} = \Phi \left[\frac{-\ln \left(\frac{\delta}{\beta} \right) + \ln(q^G) + \ln \left(\frac{p_4^{G*} - p_1^{G*}}{d_{ik_{i1}} - d_{ik_{i4}}} \right) - \mu}{\sigma_v} \right]. \quad (5)$$

Similar expressions result for the other choice probabilities in table 7 and the smoker choice probabilities. In these expressions, seven parameters determine the choice probabilities: δ , β , q^G , q^C , γ , μ , and σ_v . As noted earlier, we assume that the fill-up size is the average fill-up size for American consumers, so that $q^G = 13$, and likewise assume that smokers who travel buy a carton of cigarettes, so that $q^C = 10$. In addition, the normalization that $E[v] = 1$ implies that μ must be restricted relative to σ_v . These normalizations imply that four parameters govern the choice probabilities. Because the issue of interest is really $\frac{\delta}{\beta}$ and because the importance of the scale parameter σ_v and the smoker population share γ will vary depending on the relevant distances and price differences, the parameter $\frac{\delta}{\beta}$ and the measures of heterogeneity are separately identified. Intuitively, as these parameters change, the choice probabilities for consumers at different locations will change. These changes will be asymmetric depending on the proximity of the consumers to different borders and their respective price differences. In the final model, this will alter the predicted capacity outcomes for the different tracts, which will consequently alter the ability of our econometric model to characterize the observed outcomes. The parameters capturing unobservable consumer heterogeneity are econometrically identified by their discrete versus continuous natures.

³¹ As table 1 documents, this ordering is the same across all taxes.

³² Table 7 reflects the fact that an individual in Chicago will always be closer to Indiana than Will County.

TABLE 7.—CHOICE PROBABILITIES AND DISTANCES FOR NONSMOKERS

Distance Relationships	Choice Probabilities			
	$P(\text{Indiana})$	$P(\text{Will County})$	$P(\text{Cook County})$	$P(\text{Chicago})$
$d_1 < \{d_2, d_3, d_4\}$	1	0	0	0
$d_2 < \{d_1, d_3, d_4\}$	$P \left[\ln(v_i) > \ln \left(\frac{\delta}{\beta} \right) - \ln(q) + \ln \left(\frac{d_1 - d_2}{p_2^* - p_1^*} \right) \right]$	$1 - P(\text{Indiana})$	0	0
$d_3 < d_1 < \{d_2, d_4\}$	$P \left[\ln(v_i) > \ln \left(\frac{\delta}{\beta} \right) - \ln(q) + \ln \left(\frac{d_1 - d_3}{p_3^* - p_1^*} \right) \right]$	0	$1 - P(\text{Indiana})$	0
$d_3 < \{d_1, d_4\} < d_2$				
$d_3 < d_2 < \{d_1, d_4\}$	$P \left[\ln(v_i) > \ln \left(\frac{\delta}{\beta} \right) - \ln(q) + \ln \left(\max \left\{ \frac{d_1 - d_2}{p_2^* - p_1^*}, \frac{d_1 - d_3}{p_3^* - p_1^*} \right\} \right) \right]$	$1 - P(\text{Indiana})$ $- P(\text{Cook County})$	$P \left[\ln(v_i) < \ln \left(\frac{\delta}{\beta} \right) - \ln(q) + \ln \left(\min \left\{ \frac{d_1 - d_3}{p_3^* - p_1^*}, \frac{d_2 - d_3}{p_3^* - p_2^*} \right\} \right) \right]$	0
$d_3 < \{d_4, d_2\} < d_1$				
$d_4 < d_1 < \{d_2, d_3\}$	$P \left[\ln(v_i) > \ln \left(\frac{\delta}{\beta} \right) - \ln(q) + \ln \left(\frac{d_1 - d_4}{p_4^* - p_1^*} \right) \right]$	0	0	$1 - P(\text{Indiana})$
$d_4 < d_3 < d_1 < d_2$	$P \left[\ln(v_i) > \ln \left(\frac{\delta}{\beta} \right) - \ln(q) + \ln \left(\max \left\{ \frac{d_1 - d_3}{p_3^* - p_1^*}, \frac{d_1 - d_4}{p_4^* - p_1^*} \right\} \right) \right]$	0	$1 - P(\text{Indiana})$ $- P(\text{Chicago})$	$P \left[\ln(v_i) < \ln \left(\frac{\delta}{\beta} \right) - \ln(q) + \ln \left(\min \left\{ \frac{d_1 - d_4}{p_4^* - p_1^*}, \frac{d_3 - d_4}{p_4^* - p_3^*} \right\} \right) \right]$

Note: Each row presents an ordering of distances to the various tax regimes where d_1 corresponds to distance to Indiana, d_2 is distance to Will County, d_3 is distance to Cook County, and d_4 is distance to Chicago. Choice probabilities for smokers are similar but require adjustments for the monetary implications of buying cigarettes in different tax regions.

Hence, our model places substantial, but not unreasonable, restrictions on consumer behavior. These restrictions effectively imply that given the current tax structure, we know where consumers purchase gasoline if they choose to travel. The assumptions about price differences reflecting taxes along with the common purchase sizes imply that we know the monetary gain from travel. We must still, however, estimate the size of the consumer base and the fraction of those consumers who travel.

Thus far, our framework and assumptions provide a simple characterization of the gasoline purchase behavior of consumers when faced with alternative sources of gasoline and cigarettes. We can aggregate the decisions of all consumers to derive the demand at some location. However, we do not know the actual number of consumers at each location, and differences in the consumer base may serve as an important determinant of overall gasoline demand and, by extension, observed capacity levels. Specifically, the actual size of a location's demand may reflect observable characteristics of the consumer base.³³ Formally, we assume that each location k has a population of consumers given by

$$M_k = X_k \alpha_1, \quad (6)$$

where α_1 are parameters to be estimated and X_k are demographic and other characteristics of location k .

Given our model of consumer behavior along with our specification for the number of consumers at each location, we can determine the amount of gasoline demanded at location j as the weighted sum of the consumer base across locations where the weights are the probabilities that the respective consumers would purchase at location j :

$$Q_j^D = q^G \sum_{k=1}^K \phi_{kj} X_k \alpha_1. \quad (7)$$

In addition to the characteristics of location k that reflect the number of consumers at that location who may travel to j , additional characteristics of location j may affect its demand in a way that does not reflect consumers who travel from other tax regions. For example, the presence of freeway exits may alter the number of consumers at a location in a way that is difficult to directly attribute to consumers from other locations. As a result, we include the presence of variables Z_j to reflect such factors. Our final demand specification is thus

$$Q_j^D = q^G \sum_{k=1}^K \phi_{kj} X_k \alpha_1 + q^G Z_j \alpha_2, \quad (8)$$

³³ In effect, our approach involves estimating the market size at each location. As such, we could view this market size as implicitly capturing the importance of the outside alternative, namely no purchase, across locations. Notably, this market size does not depend on prices.

where α_2 reflects the impact of Z_j on demand in location j .

This framework allows us to characterize the total number of consumers who would purchase at a location. Demand depends on observed features of consumers at the various locations and parameters to be estimated. Given particular location demographics and parameters governing choice behavior, the parameters α will adjust the number of consumers in the different locations in ways that depend on the proximity of the consumers to various alternatives and their price differences. The parameters that govern the choice probabilities will then alter the importance of the price and distance differentials. However, the latter parameters will be relevant only in situations where consumers do not have a clearly dominant alternative or, in other words, when they might choose to travel.

B. Gasoline Demand and Observed Capacity

Our previous model describes the amount of gasoline demanded at each location as a function of the characteristics of various locations and the probabilities that consumers at those locations would purchase from different alternatives. We do not, however, observe the actual amount of gasoline sold at different locations. Instead, we observe capacity levels at each of those locations. We propose that these capacity levels should be closely related to the demand at each location in that the observed capacity at each location reflects a transformation of the amount demanded with error. Formally, we assume that

$$CAP_j = \rho(Q_j^D + \varepsilon_j), \quad (9)$$

where $\rho(\bullet)$ is an unknown increasing transformation describing how quantities are translated into capacities, and ε_j is independent and identically distributed (i.i.d.) with mean zero and variance σ_j^2 independent of all market characteristics.³⁴ This transformation should be interpreted as the way in which the predictions of our model, which concerns purchase behavior for a single fill-up, aggregate over time, for example, into an observed level of the more long-run capacity variable.

In our estimation, we assume that $\rho(\bullet)$ is linear. Because ρ will consequently always appear as a product with the parameters α_1 and α_2 , these parameters are not separately identified, and as a result, we subsume this parameter into the other linear parameters. We have also assumed that q^G is a known constant, so we absorb this term into α_1 and α_2 as well. Using our demand specification from the previous section, we obtain

$$CAP_j = \sum_{k=1}^K \phi_{kj} X_k \alpha_1 + Z_j \alpha_2 + \varepsilon_j. \quad (10)$$

³⁴ An attractive feature of our approach is that we require weak distributional assumptions about the error term. We do not exploit uneven variability in the errors for our estimates, but our standard errors correct for arbitrary heteroskedasticity and spatial autocorrelation.

TABLE 8.—PARAMETER ESTIMATES FOR ADDITIVE MARKET SIZE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Linear parameters							
Constant	7.07 (1.67)	12.39 (2.9)	6.78 (1.74)	7.61 (2.26)	6.96 (2.03)	6.74 (1.95)	6.95 (1.91)
Chicago	-3.32 (1.51)	-0.45 (1.94)	-2.24 (1.47)	-0.66 (1.66)	-1.16 (1.62)	-1.48 (1.59)	-1.58 (1.62)
Cook County	-2.28 (2.02)	-2.36 (2.00)	-0.72 (1.99)	-1.58 (2.06)	-1.16 (2.05)	-0.82 (2.06)	-1.10 (2.05)
Households	5.27 (1.05)	5.82 (1.00)	5.91 (0.99)	7.17 (1.16)	7.02 (1.15)	6.74 (1.05)	6.57 (1.07)
Households below poverty level	-11.41 (3.71)	-15.22 (4.09)	-13.70 (3.92)	-14.08 (4.75)	-14.28 (4.58)	-14.80 (4.16)	-14.60 (4.17)
Population density	-0.61 (0.21)	-1.10 (0.27)	-0.63 (0.20)	-0.89 (0.22)	-0.83 (0.22)	-0.75 (0.22)	-0.74 (0.21)
Median house value	1.05 (1.24)	-0.10 (1.52)	1.50 (1.49)	0.52 (1.41)	0.99 (1.49)	1.08 (1.46)	1.12 (1.46)
Median income	-0.21 (0.30)	-1.25 (0.62)	-0.51 (0.35)	-0.84 (0.54)	-0.74 (0.48)	-0.65 (0.41)	-0.67 (0.42)
Any freeway exits \times Indiana	9.70 (3.14)	9.44 (3.22)	10.05 (3.12)	9.58 (3.08)	9.68 (3.07)	9.92 (3.07)	10.03 (3.10)
Any freeway exits \times Cook County	1.64 (2.24)	3.51 (2.16)	2.80 (2.14)	2.73 (2.05)	2.43 (2.16)	2.38 (2.17)	2.87 (2.07)
Any freeway exits \times Chicago	0.86 (2.07)	0.35 (2.20)	1.00 (2.08)	0.38 (2.13)	0.62 (2.15)	0.81 (2.14)	0.97 (2.14)
Nonlinear parameters							
δ/β	—	0.320 (0.004)	0.435 (0.007)	0.484 (0.130)	0.597 (0.220)	0.510 (0.132)	0.423 (0.058)
σ	—	—	—	1.01 (0.29)	0.73 (0.21)	0.70 (0.34)	0.27 (0.14)
γ	—	—	0.09 (0.02)	—	—	0.10 (0.04)	0.08 (0.03)
Distribution for heterogeneity	—	—	—	Log-uniform	Log-normal	Log-uniform	Log-normal
median (δ/β)	—	0.320	0.435	0.410	0.546	0.470	0.417
R^2	0.541	0.533	0.576	0.564	0.569	0.576	0.576

Note: $N = 420$. Standard errors robust for heteroskedasticity and spatial autocorrelation in parentheses. Column 1 reproduces the regression with border dummies from column 5 of table 5.

This relationship serves as the basis of our econometric estimation.³⁵

The model in the previous section provides a relatively simple approach to link observed gasoline retail activity in this region to characteristics and potential travel patterns of consumers. We can apply this model to our raw capacity data by selecting the parameters that minimize the sum of squared differences between the observed and predicted outcomes for each tract. This estimation is facilitated by the fact that given values of the nonlinear parameters $\frac{\delta}{\beta}$, γ , and σ_v , a closed-form solution exists for minimizing the values of the other parameters. After specifying the nonlinear parameters, we compute choice probabilities and interact those choice probabilities with the observable covariates X_k . The minimizing value of α_1 and α_2 then involves the standard linear least squares expressions. Hence, we can concentrate the linear parameters α_1 and α_2

out of the objective function and confine our search to nonlinear parameters.

IV. Estimation Results

Tables 8 and 9 present estimates from our model relating consumer behavior to observed gasoline capacity across census tracts. These tables differ in the functional form that we use for the market size, M_k . In table 8, we use an additive specification for underlying population with $M_k = X_k\alpha_1$ where X_k includes a constant, regime dummies, and the demographic characteristics that we used in our earlier descriptive analysis. Table 9 uses what we will call the multiplicative specification in which we interact households with the X_k variables excluding households so that, in an abuse of notation, $M_k = \text{Households}_k \times (X_k\alpha_1)$. Beyond using these two specifications to check the robustness of our results, we have practical reasons for using them. The additive specification in table 8 is similar to the border regressions in table 5, so much so that we present results from the last of those regressions in column 1 of table 8 to compare the performance of an atheoretical descriptive model to one that is more parsimonious but involves substantially more structure. The additive model in table 8, however, has the unfortunate implication that a census tract

³⁵ In an alternative specification, we know the consumer base in each location up to some unknown error term so that $M_k = X_k\alpha_1 + u_k$. This formulation would imply spatial autocorrelation in the total error terms ϵ_j in equation (10) because the unobserved consumer base in one location would affect the total error relevant for another location. Notably, this spatial autocorrelation would diminish with distance due to the presence of the choice probabilities, which implies that we could use the estimation approach of Conley (1999). While we do not explicitly incorporate such autocorrelation in our estimation, we do correct our standard errors for it.

TABLE 9.—PARAMETER ESTIMATES FOR MULTIPLICATIVE MARKET SIZE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Linear parameters							
Households	7.40 (0.84)	7.39 (0.75)	7.19 (0.73)	7.41 (0.77)	7.28 (0.76)	7.11 (0.75)	7.09 (0.75)
Households × Chicago	-2.64 (1.08)	-2.56 (1.03)	-2.14 (0.99)	-1.67 (1.20)	-1.68 (1.14)	-1.62 (1.12)	-1.67 (1.13)
Households × Cook County	-2.20 (1.33)	-3.46 (1.30)	-1.31 (1.21)	-1.86 (1.27)	-1.69 (1.27)	-1.38 (1.28)	-1.53 (1.25)
Households × Poverty rate	-2.93 (5.49)	-8.77 (6.38)	-5.77 (5.31)	-8.02 (6.01)	-7.36 (5.82)	-7.46 (5.44)	-7.10 (5.56)
Households/area	-0.46 (0.13)	-0.63 (0.16)	-0.49 (0.14)	-0.63 (0.16)	-0.60 (0.16)	-0.57 (0.15)	-0.56 (0.15)
Households × Median house value	2.53 (1.41)	3.71 (1.81)	3.43 (1.57)	3.47 (1.87)	3.34 (1.78)	3.13 (1.63)	3.17 (1.62)
Households × Median income	-0.10 (0.61)	-1.44 (0.64)	-0.59 (0.56)	-0.99 (0.67)	-0.77 (0.63)	-0.64 (0.57)	-0.63 (0.58)
Any freeway exits × Indiana	9.68 (2.87)	9.22 (2.91)	9.91 (2.85)	9.05 (2.87)	9.32 (2.85)	9.59 (2.83)	9.70 (2.85)
Any freeway exits × Cook County	2.15 (2.36)	6.55 (2.00)	3.88 (2.27)	3.39 (2.24)	3.15 (2.30)	3.07 (2.30)	3.48 (2.21)
Any freeway exits × Chicago	1.10 (1.86)	2.33 (1.72)	0.70 (1.96)	0.23 (1.91)	0.31 (1.94)	0.39 (1.92)	0.45 (1.93)
Nonlinear parameters							
δ/β	—	0.245 (0.006)	0.435 (0.029)	0.603 (0.270)	0.652 (0.266)	0.547 (0.157)	0.459 (0.090)
σ	—	—	—	1.32 (0.43)	0.82 (0.22)	0.84 (0.38)	0.39 (0.19)
γ	—	—	0.12 (0.02)	—	—	0.12 (0.06)	0.09 (0.04)
Distribution for heterogeneity	—	—	—	Log-uniform	Log-normal	Log-uniform	Log-normal
median(δ/β)	—	0.245	0.435	0.459	0.464	0.487	0.425
R^2	0.550	0.545	0.586	0.578	0.583	0.588	0.587

Note: $N = 420$. Standard errors robust for heteroskedasticity and spatial autocorrelation in parentheses. Column 1 presents the results of a regression with border dummies analogous to that in column 5 of table 5.

could have underlying population reflected in positive M_k even if it has no households. The multiplicative specification in table 9 provides an alternative that addresses this shortcoming.³⁶

The columns in each of these tables reflect different assumptions about consumer heterogeneity to examine the implications of allowing richer heterogeneity and different distributional assumptions. Column 2 involves no consumer heterogeneity, so that all consumers have the same price-distance trade-off. Column 3 maintains the restriction that this trade-off is constant but allows a fraction of the population to take into account cigarette price differences. Columns 4 and 5 allow heterogeneity in the price-distance trade-off under different distributional assumptions but do not account for smokers. Finally, columns 6 and 7 allow both forms of heterogeneity and consider different distributions for the price-distance trade-off.

Although the coefficients associated with the impact of tract characteristics on market size are not the central issue in our study, we begin with a discussion of them. It should be noted, in interpreting the estimates of α_1 , that the reported estimates are really $q^G \rho \alpha_1$, where $q^G = 13$ reflects the assumed fill-up size and ρ reflects the rate at which quantities are transformed into capacity. Hence, a rough

interpretation of the reported coefficients in table 8 would be the impact on the number of fueling positions that a tract could support given a one unit change in the variables. In table 9, the interpretation would be the way in which the various characteristics alter the ability of additional households to support new fueling positions.

The results for the explanatory variables are similar across the various specifications and are broadly consistent with the findings in our earlier descriptive analysis, although the precision of various results does differ across models. As before, the number of households is clearly an important determinant of the amount of capacity in a tract as 1,000 additional households can support roughly five to seven additional pumps. Conversely, more poor households have a substantial negative effect on capacity levels, as do higher levels of population density. The presence of freeway exits in Indiana continues to account for around nine to ten pumps, while Cook County experiences a smaller effect and Chicago experiences none. Finally, as our earlier analysis suggested, the point estimates indicate that both Chicago and Cook County tend to support less capacity. Notably, however, the economic and statistical significance of those effects decline once we account for the fact that some people in those jurisdictions may travel to lower-cost areas to buy their gas, thereby providing an explanation for the relative dearth of capacity in Chicago and Cook County.

³⁶ Column 1 of table 9 presents a border regression analogous to that in the last column of table 5 in which we use the multiplicative specification for the explanatory variables.

The inclusion of smokers improves the fit of every model. In some cases, such as the difference between columns 2 and 3 in table 9, this improvement can be fairly substantial, reflecting the relatively precise estimates of the share of smokers. Overall, the point estimates suggest that the share of gasoline purchases that involve buyers who also consider cigarette prices is between 8% and 12% of the population.³⁷

What do the location and size of gas stations in our data tell us about consumer trade-offs between distance and money? In terms of the average trade-off, the results from the last two columns of tables 8 and 9 indicate a value between \$0.423 and \$0.547, with the model that yields the latter estimate providing the best overall fit to the data. The literal interpretation of these estimates is that the average consumer is indifferent between traveling an extra mile (and returning) and an amount in the range of \$0.85 and \$1.09.³⁸ Recall that the assumed fill-up size is 13 gallons; these results then imply that the typical nonsmoking consumer must save between roughly \$0.065 and \$0.084 per gallon in order to travel to and return from a location one mile away. If we assume a fuel economy of 20 miles per gallon and a price per gallon of \$1.50 (the Indiana average from our price data), \$0.15 of the necessary savings are attributable to recovering the expense of the gasoline used traveling, and the remainder can be interpreted as the cost of time. Assuming an average travel speed of 30 miles per hour yields a value of approximately \$10.50 to \$14.10 per hour in terms of time spent to obtain the lower price, values that are comparable to the 2004 median pretax wage of \$15.50 per hour for Cook County (Illinois Department of Employment Security, 2005).

We can also directly compare our estimates of this trade-off to previous estimates in the literature. Davis (2006) is especially pertinent as, in principle, his national data allow him to estimate a trade-off that is specific to the Midwest for consumer choices among movie theaters. In an earlier version of his study, he obtained a value of \$0.56 to travel an additional mile and return for the Midwest, which is markedly lower than our own estimates. How much of this difference is due to the different empirical approaches (rather than perhaps different average traveling speed) is an issue for future research. Both estimates, moreover, are substantially different from the very small estimate of \$0.26 found in Manuszak (forthcoming) for gasoline in Hawaii and the very high estimate of around \$3.00 found by Thomadsen (2005) for fast food in Palo Alto, California. Comparison with these last two is obviously complicated by regional and income differences between the Midwest and Palo Alto or Hawaii.

³⁷ As a point of reference, the share of smokers in the U.S. population is approximately 20% (Centers for Disease Control, 2005). When we impose $\gamma = 0.2$ in our model, our estimates of $\frac{\delta}{\beta}$ range from 0.85 to 0.934. However, our estimates of $\frac{\delta}{\beta}$ and σ become imprecise.

³⁸ Recall that our data involve distances measured in round-trip terms, so that travel to a station one mile away involves a total trip of two miles.

A comparison of the columns of table 8 or table 9 reveals some interesting interplay between the various forms of heterogeneity in our model. When we allow for smokers but no other heterogeneity, as in column 3 of both tables, our estimates of the price-distance trade-off, $\frac{\delta}{\beta}$, rise in both cases, implying that the average consumer must save more to travel a given distance.³⁹ This increase is intuitive. Without smokers, the model requires a larger fraction of consumers to travel in order to justify the observed capacity levels on the borders of low-tax regions. With smokers, however, the larger savings experienced by smokers offsets the need to have a low price-distance trade-off. Allowing heterogeneity in the price-distance trade-off has similar implications for the estimated mean value. When consumers can differ in $\frac{\delta}{\beta}$ as in columns 4 and 5, the estimates of the mean price-distance trade-off rise since, with heterogeneity, atypical consumers can account for some of the spikes in capacity in border areas of the low-tax regions that otherwise were attributed to average behavior.

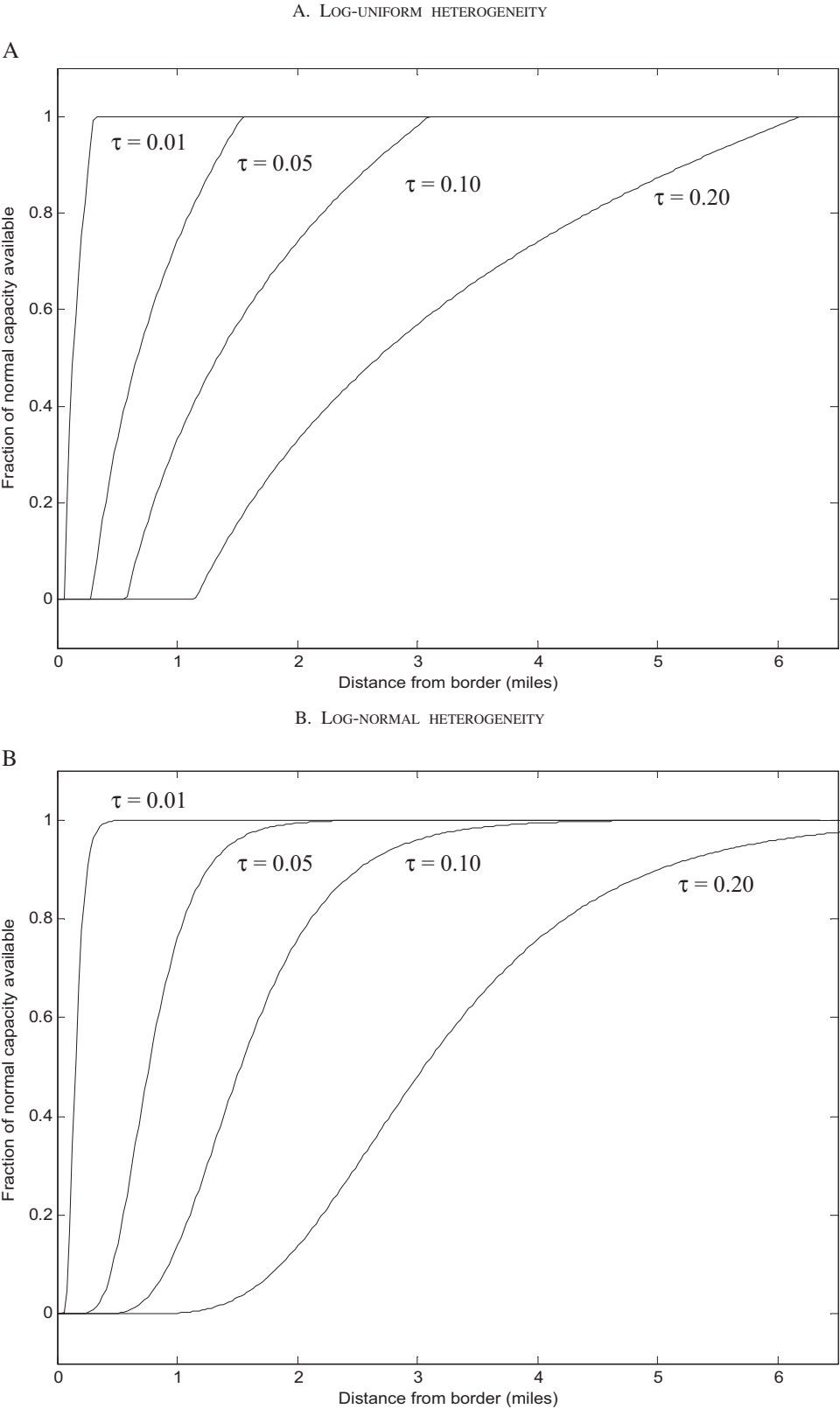
Regarding the heterogeneity itself, we typically estimate statistically significant values for consumer heterogeneity reflected in σ , although the precision of these estimates is somewhat less than that of our estimates of the mean trade-off. This heterogeneity reflects gradual, rather than abrupt, transitions to normal capacity in high-tax regimes. Across all specifications, the log-normal distribution yields notably smaller estimates of σ_v than the log-uniform case. This finding is not surprising since the unbounded tail of the log-normal distribution can yield similar heterogeneity with a lower σ_v when compared to the bounded log-uniform distribution. Notably, the inclusion of smokers tends to lower the estimated magnitude of the heterogeneity in the price-distance trade-off. With smokers, less heterogeneity is necessary since savings associated with cigarettes, and not heterogeneity in willingness to travel, can account for some of the variation in observed capacity levels.

To illustrate the implications of our estimates in more detail, figure 3 presents a two-dimensional depiction of the implied capacity outcomes in a high-tax region. To construct this figure, we use the preference parameters from the last two columns of table 9 and normalize the mass of consumers in each location to 1. For simplicity, we also assume that cigarette taxes are not higher in this region. We then consider the impact of various unit tax differences, ranging from \$0.01 to \$0.20 per gallon, on the fraction of gasoline capacity available at different distances from the border with a low-tax region.⁴⁰

³⁹ Our estimates of $\frac{\delta}{\beta}$ in columns 2 and 3 of both tables are precise since, in those columns, that coefficient is the primary way for the model to match the sharp changes in capacity around the borders. As a result, small changes in $\frac{\delta}{\beta}$ have substantial implications for the model's fit.

⁴⁰ The horizontal axis in figure 3 represents actual distance from the border. Hence, a consumer located at a distance of five miles in figure 3 would incur a round-trip distance of ten miles when patronizing the low-tax region.

FIGURE 3.—DIFFERENT UNIT TAX RATES FOR GASOLINE AND IMPLIED AVAILABLE CAPACITY IN A HIGH-TAX REGION AT DIFFERENT DISTANCES FROM A TAX BORDER



Note: Solid lines correspond to fraction of available capacity given relevant per gallon tax difference. Estimates correspond to preference parameters from columns 6 and 7 of table 9.

Figure 3 indicates that our estimates imply the same pattern depicted in the first panel of figure 1. Heterogeneity among consumers implies a gradual decline in activity in the high-tax region where the slope of this decline depends on the magnitude of the price difference. All of these consumers would induce a spike in capacity for the low-tax region at the border. At distances sufficiently far from the border, activity in the high-tax region is restored to its normal level, although in the log-normal case, the cut-off point for an effect can be quite high. Notably, the overall patterns predicted by both distributional assumptions are fairly similar except for the tails in the log-normal case. At a per gallon tax of \$0.10, which corresponds to a total savings of \$1.30 per trip in our framework, the log-uniform figure indicates that approximately 50% of normal capacity will exist around 1.5 miles from the border, while the border effect essentially dissipates by 3 miles. Conversely, a \$0.20 per gallon tax yields an effect on capacity up to slightly more than 6 miles.

Our model also allows us to calculate the deadweight loss associated with travel in the Chicagoland retail gasoline market. This deadweight loss reflects the presence of consumers who would prefer to purchase at home were prices equalized across regions, but instead choose to travel some distance to obtain a lower price. One interesting calculation involves the per fill-up deadweight loss associated with travel. This deadweight loss in dollars associated with a purchase by the average consumer in location i with respect to tax region t is

$$DWL_{it} = d_{ik_{it}} \phi_{ik_{it}} E \left[\frac{\delta}{\beta} v_i \middle| U_{ik_{it}} > \max_s U_{ik_{is}} \right], \quad (11)$$

where $d_{ik_{it}}$ is the distance from a consumer in location i to the closest alternative in tax region t , $\phi_{ik_{it}}$ is the probability that a consumer would travel from location i to tax region t , and $E \left[\frac{\delta}{\beta} v_i \middle| U_{ik_{it}} > \max_s U_{ik_{is}} \right]$ is the average monetary valuation of the disutility experienced by those who actually make that trip. Weighting for the smoker population share, the product of these terms yields the average deadweight loss across all consumers at location i from the price difference and travel associated with tax region t . The total per fill-up deadweight loss would then be the sum of DWL_{it} across all tax regions t .

This calculation is potentially problematic for a number of reasons. First, our model does not include the possibility of an aggregate elasticity of demand for gasoline. Thus, our calculation will not reflect the conventional deadweight loss associated with a tax that arises when some consumers cease purchasing altogether. Moreover, that deadweight loss would arise in both tax regions, not just the one with the higher tax. As such, our deadweight loss should provide a lower bound on the potential deadweight loss and should be interpreted as the deadweight loss solely related to the possibility that some consumers travel.

A related problem arises due to the fact that price differences induce travel and, hence, any tax difference would yield deadweight loss in our framework. This observation has an unattractive implication. We could envision taxes being equalized across Chicago and Indiana. Deadweight loss due to travel would then arise if a higher tax were imposed in Chicago. Conversely, we could equivalently view a tax cut in Indiana as the source of this deadweight loss. The natural way to remedy this deadweight loss would be to mandate equal taxes everywhere regardless of the tax level, a conclusion that largely stems from the inelasticity of aggregate demand in our model. In light of this symmetry, our calculation should be viewed as the result of the simple presence of a difference in taxes rather than relating that difference to increases or cuts in taxes in a particular region.

Figure 4 depicts the average deadweight loss per fill-up for consumers in Chicago. These calculations are based on the log-uniform point estimates from column 6 of table 9.

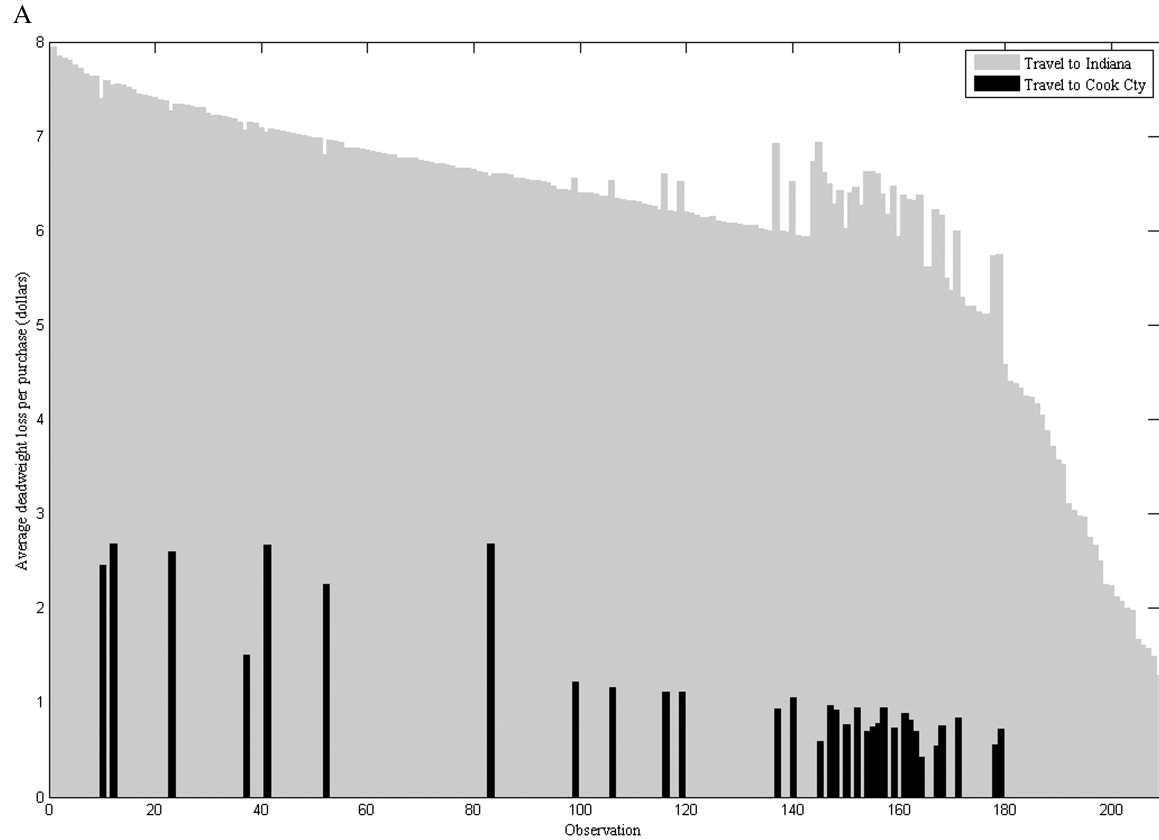
Figure 4A depicts estimates of $d_{ik_{it}} E \left[\frac{\delta}{\beta} v_i \middle| U_{ik_{it}} > \max_s U_{ik_{is}} \right]$, the average deadweight loss experienced by consumers who travel to either Cook County or Indiana. Figure 4B then weights these deadweight losses by the probability of travel to those different regions, $\phi_{ik_{it}}$, in order to compute the total average deadweight loss per purchase across all consumers in location i . In both figures, observations are ordered according to the total average deadweight loss from figure 4B. This ordering imperfectly corresponds to proximity to the Indiana border, with higher indices indicating locations that are closer to Indiana.⁴¹

Figure 4A indicates that at some locations, distance erodes three times the entire gasoline price savings for the average consumer who travels to Indiana. These observations are reflecting the fact that smokers benefit from an additional source of tax savings. The total gasoline savings per purchase from a trip to Indiana is approximately \$2.60 for Chicago consumers, but the tax savings on a carton of cigarettes is almost \$9.00. In tracts that are sufficiently far from the border, only smokers travel, and average travelers at some locations would be willing to pay \$8.00 in order to obtain the Indiana price without travel. Figure 4B, however, reveals that these smoking consumers have a less dramatic impact on the aggregate deadweight loss. Instead, most consumers would not make such a trip, so that while those who do travel experience substantial deadweight loss, the vast majority of consumers purchase locally. They do not receive the price savings associated with Indiana but also do not incur substantial travel.

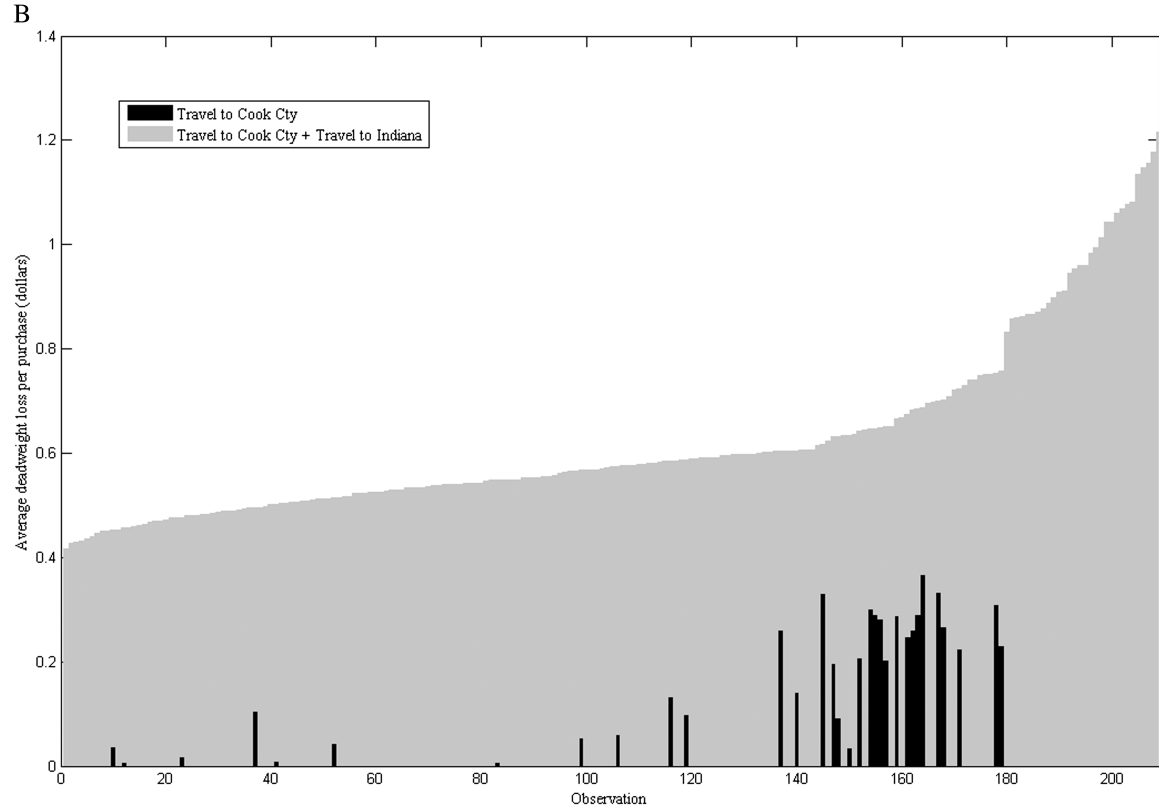
Moving rightward through figure 4A, the average deadweight loss for travelers to Indiana decreases at locations that are nearer to Indiana. While the composition of travelers is changing so that nonsmokers and consumers with low willingness to travel (i.e., high $\frac{\delta}{\beta}$) are induced to make trips,

⁴¹ This ordering is an imperfect reflection of distance to Indiana due to the presence of Cook County alternatives as a source of deadweight loss.

FIGURE 4.—AVERAGE DEADWEIGHT LOSS PER PURCHASE IN CHICAGO
A. AVERAGE DEADWEIGHT LOSS AMONG THOSE WHO TRAVEL TO SPECIFIC TAX REGIONS



B. AVERAGE DEADWEIGHT LOSS AMONG ALL CONSUMERS



Note: Observations are ordered on the horizontal axes according to total average DWL per purchase depicted in panel B. The black bar in each graph depicts the average DWL arising from trips to Cook County. In panel A, the gray bar is the average DWL experienced by those who travel to Indiana. The gray bar in panel B is total average deadweight loss reflecting trips to both Cook County and Indiana.

the travel distances become sufficiently small that their total average deadweight loss is decreasing, with the right-most close to \$1.25 per purchase. In other words, consumers who are closest to Indiana are also those who, on average, experience the least deadweight loss associated with a trip. These deadweight losses are still fairly substantial, as travel dissipates about 50% of the dollar savings from a fill-up in Indiana. Moreover, because virtually all of those consumers do travel, the average deadweight loss among all consumers tends to be high at locations close to Indiana, as can be seen in the right-most part of figure 4B.

A less straightforward calculation involves the aggregate deadweight loss in the different regions. To compute the aggregate deadweight loss, we have to weight the average deadweight losses per purchase by the number of consumers at each location. In principle, the weighting in the deadweight loss for location k would appear to involve M_k , the consumer base at that location. This weighting, however, can provide only a snapshot of deadweight loss because our model has no temporal component and, as a result, does not explicitly reflect the possibility of multiple purchases over time, each of which generates unnecessary travel. Moreover, our estimates do not actually identify M_k . Instead, we really estimate $\tilde{M}_k = \rho q^G M_k$, where $q^G = 13$ is the assumed fill-up size and ρ reflects the transformation of consumer base into capacity.⁴² If we were to use $\frac{\tilde{M}_k}{q^G}$ as our weight, we could view the resulting deadweight loss as an unknown scalar multiple of the actual immediate deadweight loss. Alternatively, we can use \tilde{M}_k , in which case the deadweight loss is measured in terms of absent pumps.

If we perform this calculation for Chicago using the values of \tilde{M}_k implied by the point estimates in column 6 of table 9, we obtain deadweight losses of 651 and 60 pumps arising due to Indiana and Cook County, respectively, with a total deadweight loss of 711 pumps. Analogous calculations for Cook County yield deadweight losses of 430 from Indiana and 26 from Will County, yielding a total deadweight loss of 456 pumps. One interpretation is that our surveyed area of Chicago, which currently has 966 fueling positions, is missing slightly more than 40% of the capacity that it would have were taxes equalized. Moreover, the assumed linearity of the transformation of demand into capacity implies that we can examine the ratio of deadweight loss across these regions. This comparison implies that regardless of the actual dollar values of the deadweight losses, the welfare loss in Chicago is approximately 1.6 times the loss in Cook County.

V. Conclusion

In this paper, we look to answer the question posed in the title: How far are consumers willing to travel in order to save money on purchases like gasoline fill-ups? In an ideal

application, detailed consumer-level information about travel, purchase patterns, and market conditions would provide information about willingness to travel. Such disaggregate data on consumers are difficult to obtain. We must instead often rely on market-level data in which we relate aggregate activity in an industry to aggregate information about the locations of consumers. In such settings, one rarely has information on exogenous factors that influence patterns of economic activity in a way that is clearly related to the price-distance trade-off faced by consumers.

In our analysis, we exploit exactly this type of exogenous variation by examining patterns of activity in the retail gasoline industry in a region where political boundaries and associated tax differences provide exogenous variation in the prices that consumers would pay at different locations. Arguably, consumers have little motivation to travel to lower tax regions beyond the price savings that they obtain. Moreover, the spikes that arise at the immediate borders of lower tax regions in our data are difficult to rationalize without the arrival of traveling consumers from high tax areas. The asymmetries in gasoline capacity around the border, particularly in comparison to interior areas of the different tax regions, allow us to infer the size of the consumer base that must be traveling. We can then relate the magnitudes of traveling consumers to their distances from the borders to learn about the distribution of consumer preferences related to willingness to travel.

Of interest to future work, the state of Indiana raised its sales tax to 6% in 2002 and its per gallon gasoline tax to \$0.18 in 2003, largely eliminating the gasoline tax difference between Indiana and Will County, Illinois. The difference in cigarette taxes, however, has exploded with Cook County's recent tenfold increase in its unit tax. The steady-state impact of these taxes is therefore ambiguous. While introduction and closure of gasoline stations may take some time, these changes in tax rates nevertheless provide an opportunity to revisit our data at some point in the future to evaluate the robustness of our results.

More generally, this paper is an example of the way in which variation in market outcomes caused by government policies can be exploited to estimate economic primitives that govern the behavior of consumers or firms. Key features of our application are the relative ease with which government policy variation can be linked to monetary implications of different actions for consumers and our ability to translate different consumer actions at a disaggregate level into different observed aggregate outcomes. In other settings, the ability to draw the link between government policies, agent behavior, and observed outcomes may be less straightforward but still possible.

In particular, policy variations across adjacent political jurisdictions that provide the foundation of our study are common. In some instances, like the one that we consider or other situations involving tax differences, the monetary implications of policy differences for consumers are rela-

⁴² Recall that at the end of section III, we absorb q^G and ρ into the demand parameters α_1 .

tively easy to calculate. The remaining task involves linking those price differences to consumer behavior and observed outcomes. In other cases, neither of these tasks is simple. For example, blue laws that limit commercial activity on Sundays as exist in various parts of the country or laws that restrict certain types of activity such as the sale of alcoholic beverages in Pennsylvania clearly could affect outcomes in those and surrounding locales. As noted in section I, a large literature attempts to measure and quantify the effects of such policies. The next logical step involves attempting, as we do in this study, to link those policies to underlying economic primitives.

REFERENCES

- Alm, James, Edward Sennoga, and Mark L. Skidmore, "Perfect Competition, Urbanization, and Tax Incidence in the Retail Gasoline Market," *Economic Inquiry* 47 (2009), 118–134.
- Berry, Steven T., "Estimation of a Model of Entry in the Airline Industry," *Econometrica* 60:4 (1991), 889–917.
- Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica* 63:4 (1995), 841–890.
- Borenstein, Severin, and Andrea Shepard, "Dynamic Pricing in Retail Gasoline Markets," *Rand Journal of Economics* 27:3 (1996), 429–451.
- Bresnahan, Timothy F., and Peter C. Reiss, "Entry and Competition in Concentrated Markets," *Journal of Political Economy* 99:5 (1991), 977–1009.
- Campbell, Jeffrey, and Beverly Lapham, "Real Exchange Rate Fluctuations and the Dynamics of Retail Trade Industries on the U.S.-Canada Border," *American Economic Review* 94:4 (2004), 1194–1206.
- Capeheart, Thomas C., "Trends in the Cigarette Industry after the Master Settlement Agreement" (Washington, DC: Department of Agriculture, 2001), <http://usda.mannlib.cornell.edu/reports/erssor/specialty/tbsbb/2001/tbs250-01.pdf>. Accessed August 15, 2006.
- Card, David, and Alan B. Krueger, "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review* 84:4 (1994), 772–793.
- Centers for Disease Control, "Cigarette Smoking among Adults—United States, 2004," *Morbidity and Mortality Weekly Report* (2005). http://www.cdc.gov/tobacco/research_data/adults_prev/prevali.htm. Accessed August 15, 2006.
- Chouinard, Hayley, and Jeffrey M. Perloff, "The Incidence of Federal and State Gasoline Taxes," *Economic Letters* 83:1 (2004), 55–60.
- Conley, Timothy G., "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics* 92:1 (1999), 1–45.
- Davis, Peter J., "Spatial Competition in Retail Markets: Movie Theaters," *Rand Journal of Economics* 37:4 (2006), 964–982.
- Doyle, Joseph J., Jr., and Krislert Samphantharak, "\$2.00 Gas! Studying the Effects of a Gas Tax Moratorium," *Journal of Public Finance* 92 (2008), 869–884.
- Holmes, Thomas J., "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders," *Journal of Political Economy* 106:4 (1998), 667–705.
- Hotelling, Harold, "Stability in Competition," *Economic Journal* 39:153 (1929), 41–57.
- Illinois Department of Employment Security, "Occupational Wages, Illinois Department of Employment Security, Wage Data as of 2004, 2nd Quarter, Region: Cook County" (2005). <http://lmi.ides.state.il.us/wagedata/countyfiles/CookCountyWages.pdf>. Accessed October 30, 2005.
- Jin, Ginger, and Phillip Leslie, "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards," *Quarterly Journal of Economics* 118:2 (2003), 409–451.
- Manuszak, Mark D., "Predicting the Impact of Upstream Mergers on Downstream Markets with an Application to the Retail Gasoline Industry," *International Journal of Industrial Organization*, forthcoming.
- Milyo, Jeffrey, and Joel Waldfogel, "The Effect of Price Advertising on Prices: Evidence in the Wake of 44 Liquormart," *American Economic Review* 89:5 (1999), 1081–1096.
- NPD Group, "The NPD Group Reports Escalating Gasoline Prices Not Likely to Discourage Summer Driving" (2000). http://www.npd.com/press/releases/press_000710.htm. Accessed September 3, 2003.
- Thomadsen, Raphael, "The Effect of Ownership Structure on Prices in Geographically Differentiated Industries," *Rand Journal of Economics* 36:4 (2005), 908–929.