Competition among Spatially Differentiated Firms: An Estimator with an Application to Cement[*]

Nathan H. Miller
Department of Justice

Matthew Osborne
Bureau of Economic Analysis

April 2011

Abstract

We develop an estimator for models of competition among spatially differentiated firms. In contrast to existing methods (e.g., Houde (2009)), the estimator has flexible data requirements and is implementable with data that are observed at any level of aggregation. Further, the estimator is the first to be applicable to models in which firms price discriminate among consumers based on location. We apply the estimator to the portland cement industry in the U.S. Southwest over 1983-2003. We estimate transportation costs to be $0.30 per tonne-mile and show that, given the topology of the U.S. Southwest, these transportation costs permit more geographically isolated plants to discriminate among consumers. We conduct a counterfactual experiment and determine that disallowing this spatial price discrimination would increase consumer surplus by $12 million annually, relative to a volume of commerce of $1.3 billion. Heretofore it has not been possible examine the surplus implications of spatial price discrimination in specific, real-world settings; these implications have been known to be ambiguous theoretically since at least Gronberg and Meyer (1982) and Katz (1984). Additionally, our methodology can be used to construct transportation margins, which are an important component of input-output tables.

Keywords: spatial differentiation; price discrimination; transportation costs; cement
JEL classification: C51; L11; L40; L61

1 Introduction

In many industries, firms are differentiated in geographic space and transportation is costly. Seminal theoretical contributions demonstrate that these conditions can soften the intensity of competition, facilitate markups above marginal cost, and induce firms to discriminate among consumers based on location (Hotelling (1929), Salop (1979), Anderson and de Palma (1988), Vogel (2008)).[1] The empirical literature of industrial organization, however, only recently has grappled with the structural estimation of models of spatial differentiation; existing estimation strategies have strict data requirements and cannot incorporate spatial price discrimination (e.g., Thomadsen (2005), Davis (2006), McManus (2007), Houde (2009)).

In this paper, we develop an estimator for models of spatial differentiation that is implementable with data observed at any level of aggregation (e.g., firm prices or regional production) and applicable to models that incorporate spatial price discrimination. Relative to existing estimation strategies, the estimator we introduce more fully leverages the information contained in the structure of the model. This leads to more flexible data requirements and enables the estimation of models that previously would have been too complicated or too demanding. Nonetheless, the estimator uses familiar minimum distance techniques and we provide conditions under which estimates are consistent and asymptotically normal. We also discuss how estimation can be conducted efficiently using recently developed numerical techniques (e.g., La Cruz, Martínez, and Raydan (2006)).

We apply the estimator to the portland cement industry in the U.S. Southwest over the period 1983-2003. This is a good match for the estimator because the industry features high transportation costs and a homogenous product (aside from geographic considerations). Producers negotiate private contracts with their customers and can engage in spatial price discrimination. We exploit variation in publicly available state-level data to identify the structural supply and demand parameters. The estimation results indicate that consumers pay $0.30 per tonne mile, given diesel prices at the 2000 level.[2] We show that, given the topology of the U.S. Southwest, these transportation costs permit the more geographically isolated plants to discriminate substantially among consumers. By contrast, plants located nearby many other plants have less localized market power and do not discriminate among

---

[1]There is limited evidence that such spatial price discrimination is common in business-to-business industries. Greenhut, Greenhut, and Li (1980) report that 32 percent of surveyed firms employ some form of spatial price discrimination, though the sample is small and not clearly representative. To our knowledge, more systematic efforts to identify the extent of employ spatial price discrimination have not been made.

[2]The 1974 edition of the Minerals Yearbook, an annual publication of the U.S. Geological Survey, indicates transportation costs of $0.35 per tonne mile (when adjusted to real 2000 dollars), which partially corroborates the estimation results. Subsequent editions do not provide the magnitude of transportation costs.

consumers. We note that our approach could also be used to back out transportation margins when observed prices include transport costs. Transportation margins are an important part of input-output tables, which are used in national accounting. Input-output tables present a matrix detailing how the production of goods and services is allocated between component inputs, and how those goods and services flow to outputs.[3] As outlined in the System of National Accounts 2008 (SNA 2008), when a product's measured price includes transport costs, those transport costs need to be subtracted from the price to avoid double counting (see SNA 2008, pg. 508). Our approach facilitates this exercise because the estimator provides equilibrium prices net of transport costs. One would simply subtract equilibrium prices from observed prices to arrive at the transport costs faced by consumers.

We conduct two counter-factual experiments to demonstrate the power of the estimation results. First, we consider the consumer surplus implications of spatial price discrimination. The theoretical literature has long recognized that spatial price discrimination can increase or decrease social welfare (e.g., Gronberg and Meyer (1982), Katz (1984), Hobbs (1986), Anderson, de Palma, and Thisse (1989)).[4] Although this suggests an important role for empirical research, heretofore it has not been possible to examine spatial price discrimination in specific, real-world settings. The results of the counter-factual experiment indicate that disallowing spatial price discrimination would increase consumer surplus by $12 million annually, relative to a volume of commerce of $1.3 billion. We quantify and provide support for the standard intuition that discrimination harms consumers located nearby cement plants (these consumers tend to be inframarginal) but benefits more distant consumers.

Second, we consider a hypothetical merger between the two largest portland cement manufacturers in the U.S. Southwest in 1986, and find that the merger would have increased prices in southern California and Arizona by three and five percent, respectively. By way of contrast, one standard antitrust model that exploits state-level data yields price effects of one percent in southern California and 25 percent in Arizona. This highlights the fact that our approach allows one to estimate more realistic economics models with more limited data – i.e., to do more with less. This may be particularly useful to antitrust authorities, which often must conduct merger investigations under tight deadlines and without access to comprehensive industry data.

We now detail more explicitly the methodological contribution. To frame the contribution, we first discuss a data availability problem that hinders the estimation of spatial

---

[3]For more detailed definitions, the reader may refer to the System of National Accounts, 2008.

[4]When discrimination is legal, firms are incentivized to charge high prices to nearby consumers and low prices to distant consumers; it follows that discrimination harms some consumers and benefits others.

models. The easiest way to identify spatial differentiation (equivocally, transportation costs) is to measure how firms' market shares differ between nearby and distant consumers. For instance, if one observes that each firm captures greater market shares among nearby consumers then one can infer that transportation costs are large and that firms are spatially differentiated. Implementation is difficult, however, because data on the geographic distributions of market shares are seldom available. Indeed, the extant empirical literature has yet to utilize such data.[5] The problem is only exacerbated when firms employ spatial price discrimination because one must account not only for the geographic distributions of market shares but also for the geographic distributions of prices.

Our central insight is that one can solve the data availability problem by relying on numerical approximations to equilibrium. That is, one can *compute* the geographic distributions of markets shares and prices that characterize the equilibrium of the economic model, given a set of candidate parameters. If these parameters imply high transportation costs, for example, then one would compute that firms capture greater market shares among nearby consumers. These distributions can be used to construct aggregated equilibrium predictions at the level of the available data. This process is repeatable for any set of parameters, so one can search for parameters that minimize the "distance" between the aggregated equilibrium predictions and the data. Further, since one can match data that are observed at any level of aggregation (e.g., plant-level, firm-level, or region-level data), the data requirements of the are completely flexible, provided there is sufficient variation to identify the parameters of the model.[6]

We make the identifying assumption that discrepancies between data and the aggregated equilibrium predictions, when evaluated at the population parameters, are orthogonal to firm locations, cost shifters, and demand shifters. This yields a multiple-equation non-linear least squares estimator (e.g., as analyzed in Greene (2003, p. 369)).[7] Each equation

---

[5]Instead, much of the empirical literature uses market delineation to sidestep the data availability problem (e.g., Pesendorfer (2003), Salvo (2008), Collard-Wexler (2009), Ryan (2010)). This facilitates estimation with market-level data but sacrifices realism in the underlying economic model. In Section 6.4, we discuss market delineation in greater detail and compare some of our results to those obtained in one recent study that employs market delineation to estimate a model of the portland cement industry (Ryan (2010)).

[6]The key empirical patterns that drive parameter estimates can be transparent despite the complicated nonlinear relationships involved. In our application, the transportation cost estimate is driven by differences between consumption and production within specific geographic regions. Suppose, for instance, that one observes that consumption is greater than production in one region but less than production in another. This implies that inter-regional trade flows exist, and an estimate of transportation costs can be selected to rationalize these trade flows within the structure of the model.

[7]Equivalently, the estimator can be interpreted as method-of-moments with optimal instruments based on firm locations, cost shifters, and demand shifters.

matches observations on a single time-series of data to the corresponding aggregated equilibrium predictions. Thus, if the observed data are total industry output and the average industry price, then the estimator would exploit time-series observations on two nonlinear equations. The twist is that one must compute the model predictions (i.e., the "right-hand-side") using numerical techniques for every candidate parameter vector.

Although we focus on the estimation of models in which firms are differentiated in geographic space, some broad parallels can be drawn between our estimator and existing estimators for models in which firms are differentiated in product space (e.g. Berry, Levinsohn, and Pakes (1995), Nevo (2001)). When differentiation is in product space, the central challenge is to recover structural parameters when prices and quantities are fully observed but non-price product characteristics are imperfectly observed. By contrast, when differentiation is in geographic space, the challenge is that prices and quantities are imperfectly observed. Yet, in both settings, one can use numerical techniques to recover the unobserved metrics: the contraction mapping of Berry (1994) obtains the unobserved product characteristic when differentiation is in product space, and equilibrium computations obtain the relevant prices and quantities when differentiation is in geographic space.

Finally, we note that the estimator could be employed to define stage-game payoffs inside dynamic estimation routines, such as that of Bajari, Benkard, and Levin (2007).[8] This may extend the reach of researchers dramatically. For instance, the theoretical literature makes it clear that firms select locations to secure a base of profitable customers, provide separation from efficient competitors, and deter nearby entry; our estimator may make permit researchers to explore the practical effects of these incentives.

The paper proceeds as follows. We close the introduction below with an overview of existing estimation strategies for models of spatial differentiation. In Section 2, we outline the institutional details of the portland cement industry and describe the relevant data. We then formalize a model of spatial price discrimination in Section 3. The model is general but can be tailored to the salient features of the cement industry, including capacity constraints and the constraining influence of foreign imports. In Section 4, we derive the estimator, prove consistency and asymptotically normality under specified conditions, and discuss how estimation can be conducted efficiently with recently developed numerical techniques. We discuss identification and related topics in Section 5. We present the results of estimation in Section 6 and the results of the two counterfactual experiments in Section 7. Section 8 concludes.

---

[8]Seim (2006) estimates a dynamic game of spatial competition but uses reduced-form payoffs that circumvent the data availability problem.

1.1 Review of the empirical literature

We build upon the recent contributions of Thomadsen (2005), Davis (2006), McManus (2007) and Houde (2009), each of which provides an estimator for models of spatial differentiation that exploits variation in firm-level data on prices and quantities.[9] The strategy we introduce has two advantages relative to this literature: First, it incorporates spatial price discrimination for the first time. Second, it has more flexible data requirements and is implementable with data that are observed at any level of aggregation.

To help build this intuition, we outline the estimation strategy of McManus (2007) in some detail; the estimation strategies of Thomadsen (2005), Davis (2006) and Houde (2009) are analogous. McManus estimates the structural demand parameters for coffee shops on the grounds of the University of Virginia, exploiting data on prices and sales for each shop. These data are assumed (reasonably) to be observed without error. McManus also observes the distance between each coffee shop and number of "location points," as well as a measure of the student population at each location point. Estimation proceeds by calculating the sales of each shop to consumers at each location point, given the observed prices, the structure of the demand model, and a candidate parameter vector. These sales predictions are then aggregated to the shop level and compared against the data. McManus estimates using maximum likelihood; alternatively, the generalized method of moments technique of Berry, Levinsohn, and Pakes (1995) could be employed.

This approach requires the econometrician to observe all the relevant prices. Demand cannot be estimated separately from supply if some prices are unobserved because demand is a function of prices. Our estimator overcomes this limitation by leveraging supply-side information to compute prices numerically. These computed prices can then be used in the place of data to construct demand. Unobserved prices can occur in many situations. A prime example is when firms employ spatial price discrimination: this generates a distribution of prices across geographic space that is difficult to capture with data. Measurement error can create a similar problem – the use of imprecise prices in the demand equation should yield inconsistent estimates. With our approach, one could compute the true prices numerically and match moments on observed prices and quantities. Finally, firm-level data are sometimes unavailable in policy situations. Antitrust regulators might have access to price data for the merging parties but more aggregated data for non-merging parties. Our estimator could be

---

[9]An alternative approach for non-discriminatory spatial models is developed in Pinske, Slade, and Brett (2002). The paper introduces a reduced-form estimator that can be applied usefully to evaluate the extent to which competition is localized. However, the estimator does not recover the underlying structural parameters of the model, including the transportation cost, and does not enable counter-factual policy experimentation.

used to compute the unobserved prices, estimate the demand and supply parameters, and conduct merger simulations.

2 Portland Cement

2.1 Industry details

Portland cement is a finely ground dust that forms concrete when mixed with water and coarse aggregates such as sand and stone.[10] Concrete, in turn, is an essential input to many construction and transportation projects because its local availability and lower maintenance costs make it more economical than substitutes such as steel, asphalt, and lumber (Van Oss and Padovani (2002)). The producers of portland cement adhere to strict industry standards that govern the production process. Aside from geographic considerations, product differentiation in the industry is minimal.[11]

Producers negotiate private contracts with their customers, predominately ready-mix concrete firms and large construction firms. Most contracts specify a mill price (or a "free-on-board" price) for portland cement at the location of production. Customers are responsible for door-to-door transportation, which is an important consideration because portland cement is inexpensive relative to its weight.[12] This is well understood in the academic literature. For example, Scherer et al (1975) calculates that transportation would account for roughly one-third of total customer expenditures on a hypothetical 350-mile route between Chicago and Cleveland, and a Census Bureau study (1977) reports that more than 80 percent is transported within 200 miles.[13] More recently, Salvo (2010) presents evidence consistent with the importance of transportation costs in the Brazilian portland cement industry.

The production process can be summarized as follows. Limestone, typically mined from a quarry adjacent to the cement plant, is fed into coal-fired rotary kilns that reach peak temperatures of 1400-1450° Celsius. The output of the kilns, known as "clinker," is

---

[10]We draw heavily from the publicly available documents and publications of the United States Geological Survey and the Portland Cement Association to support the analysis in this section. We defer detailed discussion of these sources for expositional convenience.

[11]Standards are maintained by the the American Society for Testing and Materials Specification for Portland Cement, in order to protect the quality and reliability of construction materials.

[12]The bulk of portland cement is moved by truck, though some is sent by train or barge to distribution terminals and only then trucked to customers. Barge transport is not feasible in the U.S. Southwest due to the lack of navigable rivers.

[13]Scherer et al (1975) examined more than 100 commodities and determined that the transportation costs of portland cement were second only to those of industrial gases. Other commodities identified as having particularly high transportation costs include concrete, petroleum refining, alkalies/chlorine, and gypsum.

cooled, mixed with a small amount of gypsum, and ground in electric mills to form portland cement. Kilns operate at peak capacity with the exception of an annual maintenance period. The duration of this maintenance period can be adjusted to meet demand conditions. When demand is particularly strong, managers sometimes forego maintenance at the risk of breakdowns and kiln damage.[14]

In our application, we focus on California, Arizona, and Nevada over the period 1983-2003. We refer to these three states as the "U.S. Southwest" for expositional convenience. Figure 1 maps the geographic configuration of the industry in the U.S. Southwest circa 2003. Most plants are located along an interstate highway, nearby one or more population centers. Some firms own multiple plants but ownership is not particularly concentrated – the capacity Herfindahl-Hirschman index (HHI) of 1260 is well below the threshold that defines highly concentrated markets in the 2010 Merger Guidelines. Foreign imports enter through four customs offices, located in San Francisco, Los Angeles, San Diego, and Nogales. Foreign imports are mostly produced by large, efficient plants located in Southeast Asia. Exports are negligible because domestic plants are not competitive in the international market.

It is worth noting that trade flows between the U.S. Southwest and other domestic areas are negligible. To demonstrate, in Figure 2 we plot foreign imports ("observed imports") and consumption less production in the U.S. Southwest ("apparent imports").[15] The nearly exact correlation between these two measures reveals no net trade flows between the U.S. Southwest and other domestic regions. Other statistics published by the USGS imply that gross trade flows are also negligible, as well.[16] The fact that cement can be shipped economically into the U.S. Southwest from foreign countries (e.g., Thailand) but not from nearby domestic areas is due to a number of factors, including the cost discrepancy between freighter and truck transportation and the relative efficiency of the large foreign plants.

We observe only two plant closures, one plant entry, and three substantive kiln upgrades over 1983-2003. This degree of capital persistence is consistent with the substantial

---

[14]A recent report prepared for the Environmental Protection Agency identifies five main variable input costs of production: raw materials, coal, electricity, labor, and kiln maintenance (EPA (2009)).

[15]The figure also plots consumption and production. Both are highly cyclical, consistent with the role of cement as an input to construction projects. However, consumption is more cyclical due to the costliness of capacity adjustments (e.g., as documented in Ryan (2010)) and the gap between consumption and production increases in overall activity. Thus, while imported cement generally represents a small fraction of total consumption, it plays an important role when demand outstrips domestic capacity.

[16]More than 98 percent of cement produced in southern California was shipped within the U.S. Southwest over 1990-1999, and more than 99 percent of cement produced in California was shipped within the region over 2000-2003. Outflows from Arizona and Nevada are unlikely because consumption routinely exceeds production in those states. And since net trade-flows between the U.S. Southwest and other domestic regions are insubstantial, these data points imply that gross domestic inflows must also be insubstantial.
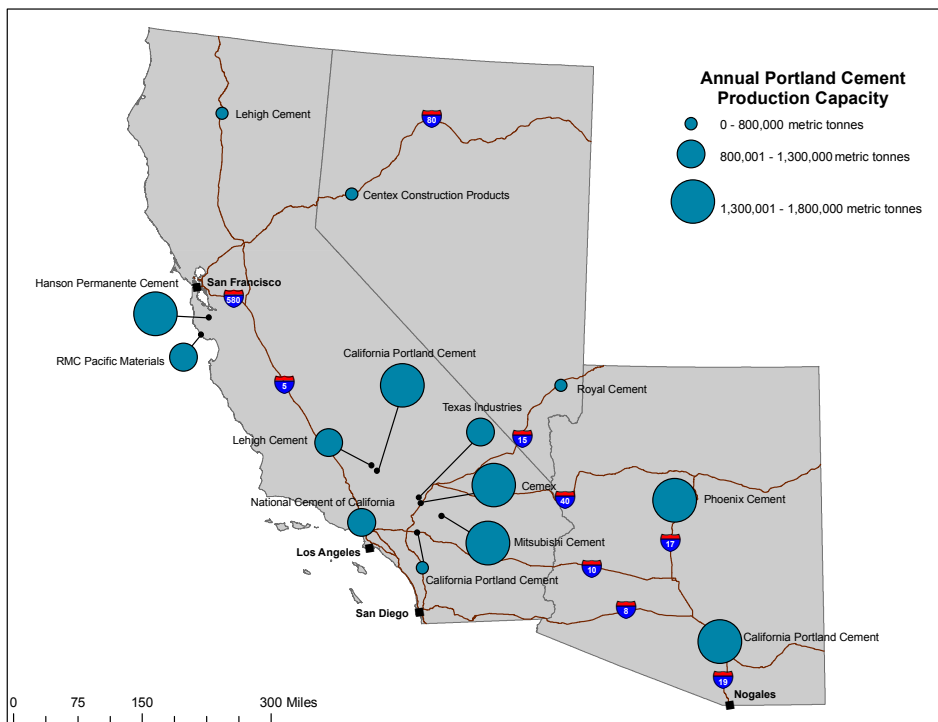
Figure 1: Portland Cement Production Capacity in the U.S. Southwest circa 2003.

sunk costs of kiln construction (e.g., Ryan (2010)). Our treatment of plant locations as pre-determined may be reasonable given the specific geographic area and time period examined.

2.2 The data

Our primary data source is the U.S. Geological Survey (USGS). The USGS conducts an annual establishment-level census of portland cement producers, aggregates responses to protect the confidential information, and publishes the results in its Minerals Yearbook. Data are available over the full sample period, 1983-2003. Of particular interest are average (production weighted) mill prices, regional production, and regional consumption. The consumption data are available separately for northern California, southern California, Arizona, and Nevada. The USGS aggregates the Arizona and Nevada data on prices and production data for 1983-1991.[17] We also make use of data on cross-region shipments that appear in the California Letter, another annual publication of the USGS that is available for 1990-2003.

[17]In later years, the USGS instead aggregates Nevada with states outside the U.S. Southwest due to confidentiality concerns – the USGS must adjust its aggregation scheme as the number and location of plants in operation changes. We do not make use of USGS data for Nevada after 1991.
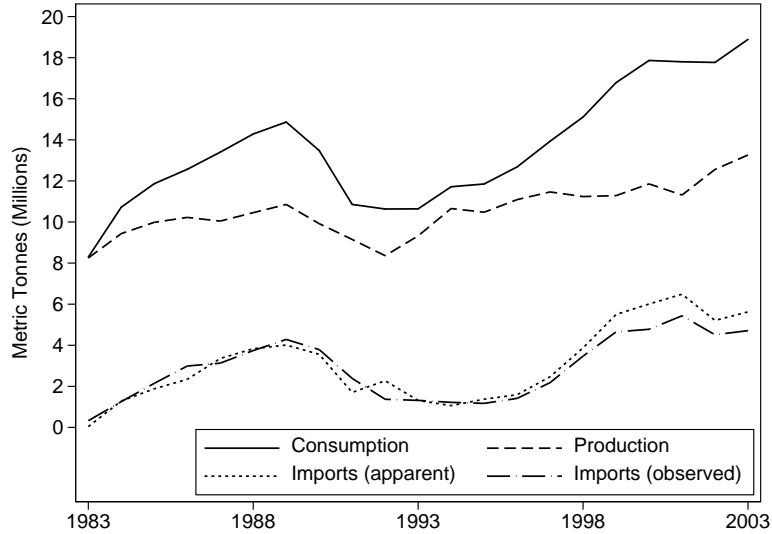
8

Figure 2: Consumption, Production, and Imports of Portland Cement. Apparent imports are defined as consumption minus production. Observed imports are total foreign imports shipped into San Francisco, Los Angeles, San Diego, and Nogales.

Plant locations are available from the Plant Information Survey (PIS), an annual publication of the Portland Cement Association. The PIS also provides the annual capacity of the kilns. To model input prices, we collect data on coal and electricity prices from the Energy Information Agency (EIA), data on the average wages of durable good manufacturing employees from the BEA, and data on crushed stone prices from the USGS. These data are observed at the year-state level. As we discuss below, we allow transportation costs to fluctuate with diesel prices; we use the diesel price index published by the EIA. We use import prices obtained from the USGS Minerals Yearbook to help model the role of import competition. Finally, we use county-level data from the Census Bureau on construction employment and residential construction permits to normalize the potential demand of each county. We refer the reader to the appendix for summary statistics and details on the data collection process.

3 The Model Of Spatial Price Discrimination

3.1 The geographic space

We define the relevant *geographic space* to be a compact, connected set $\mathbb{C}$ in the Euclidean space $\mathbb{R}^2$. The geographic space is the U.S. Southwest in our application. We take as given

9

that J plants compete in the space, and assume that each plant is endowed with a fixed location defined by the geographic coordinates $\{z_1, z_2, \ldots, z_J\}$, where $z_j \in \mathbb{C}$. We further take as given that a continuum of consumers spans the space, and assume that each consumer has unit demand and a fixed location $w \in \mathbb{C}$. The absolute measure $\phi(w)$ characterizes the geographic distribution of consumers and we define $M = \int_{\mathbb{C}} \phi(w) dw$ to be the potential demand of the space. We denote the distance between any two points in the geographic space, say a and b, as the Euclidean distance $\|a - b\|$.

We partition this geographic space into mutually exclusive *consumer areas*. As we formalize below, we permit plants to set different mill prices in each consumer area. The partition is best interpreted as determining the extent which firms engage in spatial price discrimination. Finer partitions of the geographic space imply more sophisticated discrimination, and if only a single area exists then firms do not discriminate. In our application, we use the 90 counties of the U.S. Southwest to define the consumer areas. Within the context of the model, these areas have no economic significance aside from their implications for price discrimination. Since every plant competes in every area, the partition of the geographic space does not artificially limit competition and is not analogous to a "market delineation" assumption under which plants compete only within prescribed geographic boundaries. Each area $\mathbb{C}_n$ (for $n = 1, \ldots, N$) has the potential demand $M_n = \int_{\mathbb{C}_n} \phi(w) dw$.

We sketch one possible geographic space in Figure 3. The dashed lines delineate three consumer areas, $\mathbb{C}_1$, $\mathbb{C}_2$, and $\mathbb{C}_3$. Two plants operate in the space and are characterized by the locations z_1 and z_2. A distribution of consumers span the space, and both plants compete for every consumer. The plants imperfectly price discriminate by setting different prices in each consumers area. Thus, there are six prices in the space, which we represent with the arrows labeled $\{p_{11}, p_{12}, p_{13}, p_{21}, p_{22}, p_{23}\}$. Finally, we plot the location of a single consumer characterized by location w. The dashed line labeled $\|w - z_1\|$ is the Euclidean distance between the consumer and the first plant.

3.2 Supply

We take as given that F firms and J plants exist in the geographic space. Each firm operates some subset $\mathbb{J}_f$ of the plants and can ship from any plant $j \in \mathbb{J}_f$ to any consumer. We assume that firms employ imperfect spatial price discrimination by setting different mill prices to different areas. This mill price does not include the transport cost: a consumer's total payment for the product include the mill price and the door-to-door transportation cost.

Equilibrium is the result of a Bertrand-Nash pricing game: each firm chooses a vector
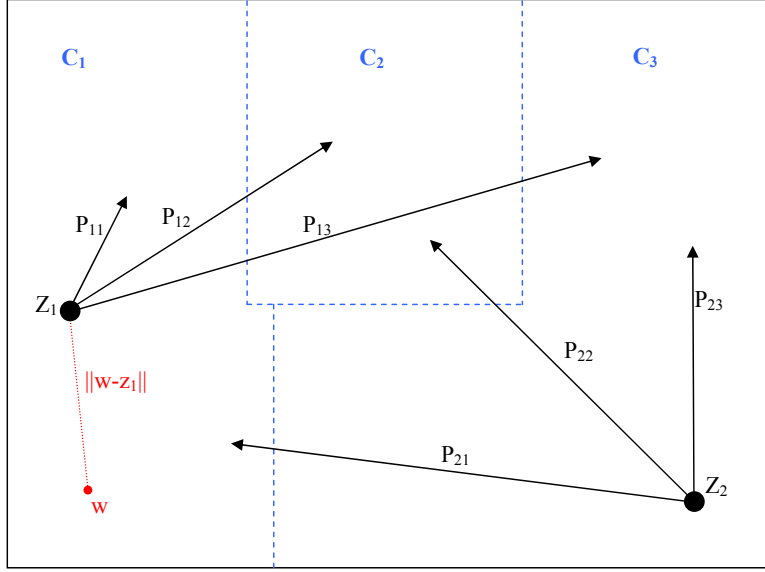
Figure 3: A Geographic Space.

of prices, $\mathbf{p}_f = (p_{jn}; j \in \mathbb{J}_f, n = 1, \cdots, N)$, to maximize its short run profits conditional on the prices chosen by all other firms. Formally, the equilibrium prices $\boldsymbol{p}^*$ solve the following maximization problem:

$$\boldsymbol{p}_f^* = \arg\max_{\boldsymbol{p}_f} \pi_f(\boldsymbol{p}_f, \boldsymbol{p}_{-f}^*; \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \quad \forall f = 1, ..., F, \tag{1}$$

where the firm profit function is

$$\pi_f(\boldsymbol{p}_f, \boldsymbol{p}_{-f}; \cdot) = \sum_{j \in \mathbb{J}_f} \sum_n p_{jn} q_{jn}(\boldsymbol{p}_n; \boldsymbol{x}_n, \boldsymbol{\beta}) - \sum_{j \in \mathbb{J}_f} \int_0^{Q_j(\boldsymbol{p}; \, \boldsymbol{x}, \boldsymbol{\beta})} c(Q; \boldsymbol{w}_j, \boldsymbol{\alpha}) dQ. \tag{2}$$

Here, the quantity demanded from plant j by consumers in area n, denoted $q_{jn}(\cdot)$, is a function of all the prices in the area (denoted $\boldsymbol{p}_n$). Total production at plant j is $Q_j(\cdot) = \sum_n q_{jn}(\cdot)$. The vectors $\boldsymbol{x}$ and $\boldsymbol{w}$ include demand shifters and cost shifters, respectively, and $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ include the corresponding parameters. Finally, $c(\cdot)$ is a marginal cost function that is convex in Q_j and differentiable in all its arguments.[18]

---

[18]Spatial price discrimination is at the core of the firm's pricing problem: in equilibrium, firms will charge higher mill prices to nearby consumers and to consumers for whom the firm's competitors are more distant. However, aside from price discrimination, the firm's pricing problem follows standard intuition. A firm that contemplates a higher mill price from one of its plants to a given area must evaluate (1) the tradeoff between lost sales to marginal consumers and greater revenue from inframarginal consumers; and (2) whether the firm would recapture lost sales with its other plants. If marginal costs are not constant, then the firm must

We assume that equilibrium exists and is unique, for reasons that we clarify below. Recent theoretical contributions demonstrate that existence and uniqueness hold for two special cases of our model: nested logit demand with single-plant firms (Mizuno 2003), and logit demand with increasing marginal costs and multi-plant firms (Konovalov and Sándor 2010). Although the uniqueness property is not satisfied generally, numerical methods can be used to evaluate the property over a large portion of the parameter space when uniqueness is difficult or impossible to demonstrate on theoretical grounds. We provide guidance on how to do this in Section 5.1.

The convexity of the marginal cost curve allows one to incorporate of nonlinear production factors, such as capacity constraints, that are are common in many industrial settings. In our application, we do so by specifying a marginal cost function that depends non-linearly on the level of capacity utilization:

$$c(Q_j(\cdot); \boldsymbol{w_j}, \boldsymbol{\alpha}, \gamma, \nu, \mu) = \boldsymbol{w'_j}\boldsymbol{\alpha} + \gamma \, 1\left\{\frac{Q_j(\cdot)}{CAP_j} > \nu\right\}\left(\frac{Q_j(\cdot)}{CAP_j} - \nu\right)^{\mu}, \qquad (3)$$

where CAP_j is total plant capacity. This treatment of capacity constraints, an innovation of Ryan (2010), imbeds the intuition that production near capacity creates shadow costs due to foregone kiln maintenance. Thus, marginal costs increase in production once utilization exceeds ν, and the combination $\gamma(1 - \nu)^{\mu}$ represents the penalty associated with production at capacity. The function is continuously differentiable for $\mu > 1$. In practice, we find that it is difficult to estimate both γ and μ and we normalize the latter to 1.5.[19]

In our application, we augment the formal structure above to account for foreign import competition. Specifically, we assume that domestic plants compete against a competitive fringe of foreign importers. We denote this fringe as "plant" $J + 1$, and assign the fringe to four locations in the U.S. Southwest based on the customs offices through which cement can enter (San Francisco, Los Angeles, San Diego, and Nogales). Consumers pay the door-to-door cost of transportation from these customs offices. We rule out spatial price discrimination on the part of the fringe, consistent with perfect competition among importers, and assume that the import price is set exogenously based on the marginal costs of the importers or other considerations. Thus, the supply specification is capable of generating the stylized fact that foreign importers provide substantial quantities of portland cement to the U.S. Southwest

---

also evaluate how the lost sales would affect the plant's competitiveness in other areas.

[19]The cost shifters we incorporate are input prices for coal, electricity, labor, and limestone. This constant portion of marginal costs can be derived from a Leontief production function (i.e., the factors of production are used in fixed proportions) and is consistent with the economics of portland cement production.

when demand is strong (see Table 2).

3.3 Demand

We model consumer behavior using a conventional discrete-choice demand system. Each consumer observes the plant locations and the available mill prices, and either purchases from one of the $J+1$ plants or foregoes a purchase altogether (i.e., selects the outside good). The indirect utility that consumer i receives from plant j is:

$$u_{ij} = \beta^c + \beta^p p_{jn} + \beta^d d_{jn} + \boldsymbol{x}'_{jn}\boldsymbol{\beta}^x + \epsilon_{ij}, \tag{4}$$

where d_{jn} is the average distance between plant j and consumers in area n, the vector $\boldsymbol{x}_{jn}$ includes demand shifters (e.g., product characteristics), and ϵ_{ij} is a preference shock that is i.i.d. across consumers. Following standard practice, we normalize the mean utility of the outside option to zero. Finally, $\boldsymbol{\beta} = (\beta^c, \beta^p, \beta^d, \boldsymbol{\beta}^x)$ are the demand parameters and the ratio β^d/β^p is the unit transportation cost incurred by consumers.

> *Example:* We provide motivation for this indirect utility function based on our application to the portland cement industry. The end-users of cement are construction firms that use the cement as an input to various construction projects. Suppose that project i requires a certain quantity of cement and that the unit cost of purchasing this cement from plant j is given by
>
> $$b_{ij} = F + p_{jn} - \beta^d\|w_i - z_j\|,$$
>
> where F includes fixed costs, $\|w_i - z_j\|$ measures the distance between customer i and plant j, and $\beta^d < 0$.[20] We have normalized the price coefficient such that $\beta^p = -1$, for tractability. Further suppose that a substitute for cement, such as lumber or steel, can be used in the project at the unit cost $b_{i,0}$. Then the construction firm's cost minimization problem can be rewritten as a utility maximization problem in which indirect utility is given by
>
> $$u_{ij} = \begin{cases} \beta^c - p_{jn} + \beta^d d_{jn} + \epsilon_{ij} & \text{for } j = 1, \dots, J, \\ \epsilon_{ij} & \text{for } j = 0, \end{cases}$$

---

[20]Portland cement is often purchased by firms that mix the cement with water and aggregates to form ready-mix concrete, which subsequently is shipped to construction sites. The fixed costs F would include any markups charged by these firms.

where

$$\epsilon_{ij} = \begin{cases} \beta^d(\|w_i - z_j\| - d_{jn}) & \text{for } j = 1, \ldots, J, \\ \int b_{i,j} f_b(b_{i,j}) - b_{i,j} & \text{for } j = 0, \end{cases}$$

the utility constant is defined such that $\beta^c = \int b_{i,0} f_b(b_{i,0}) - F$, and $f_b(\cdot)$ is a continuously distributed probability density function for the cost of the substitute good over the population of projects.[21]

The preference shock in this example arises due to two sources of heterogeneity: consumers have different valuations of the outside good and differ in their proximity to plants. This second source of heterogeneity provides an errors-in-variables motivation for the preference shock, as the distance between counties and plants is an imperfect proxy for the distance between consumers and plants. The dual sources of heterogeneity also provide a theoretical basis for preferring distributional assumptions that divorce the aggregate elasticity of demand from the plant-level elasticities (i.e., that accommodate inelastic aggregate demand and elastic plant-level demand). The nested logit demand system is based on one such distributional assumption.

In our application, we proxy distance using the miles between the plant and the centroid of the consumer's area, multiplied by a diesel price index. For the foreign import option, distance is specified using the miles between the consumer's area and the nearest customs office, again adjusted for diesel prices. Thus, transportation costs increase linearly in Euclidian distance and fuel costs. Since we measure miles in thousands and mill prices are per metric tonne, the ratio β^d/β^p gives the transportation cost per thousand tonne-miles when the diesel price index equals one. The domestic mill prices that appear in the indirect utility function are not observed in the data and must be computed as the solution to equation (1). This procedure takes as given import prices that are exogenously-determined, non-discriminatory, and observed in the data.

Estimation requires that demand be twice continuously differentiable and downward sloping in price. In our application, we assume a distribution of preference shocks that yields the nested logit demand system. We place the inside options (i.e., the domestic plants and foreign imports) in a different nest than the outside option. Following Cardell (1997), we let the parameter λ characterize the degree to which valuations of the inside options are

---

[21]The fact that d_{jn} appears in the indirect utility function and in the error term is not problematic because d_{jn} is orthogonal to the residual deviation $\|w_i - z_j\| - d_{jn}$. We provide a proof in the appendix. The constant term of utility, β^c, can be interpreted as the average unit cost of cement relative to the average unit cost of the alternative product. This constant could also account for the possibility that construction firms require different amounts of cement versus the substitute good to complete the project.

correlated across consumers. Valuations are perfectly correlated if $\lambda = 0$ and uncorrelated if $\lambda = 1$; the model collapses to a standard logit in the latter case. The nested logit structure makes available analytical expressions for the quantity of cement that each plant sells to each area (i.e., $q_{jn}(\boldsymbol{p_n}; \boldsymbol{x}_n, \boldsymbol{\beta})$) and makes estimation computationally feasible.[22]

To close the model, one must specify the potential demand in each consumer area. This is typically done by normalizing potential demand based on some set of plausibly exogenous demand factors (e.g., Berry, Levinsohn, and Pakes (1995), Nevo (2001)). In our application, we define consumer areas as counties and normalize potential demand based on the number of construction employees and the number of new residential building permits. Thus, we implicitly assume that total construction spending is unaffected by cement prices. This seems reasonable because cement accounts for a small fraction of total construction expenditures (e.g., see Syverson (2004)). The procedure indicates that potential demand is concentrated in a small number of counties. In 2003, the largest 20 counties account for 90 percent of potential demand, the largest ten counties account for 65 percent of potential demand, and the largest two counties – Maricopa County and Los Angeles County – together account for nearly 25 percent of potential demand.[23]

4   Estimation

4.1   Overview

In this section, we describe the estimator formally, provide assumptions under which the estimator is consistent and asymptotically normal, and then discuss numerical techniques for the computation of equilibrium. Some additional notation is useful. We denote the vector

---

[22]The substitution patterns between cement plants are characterized by the independence of irrelevant alternatives (IIA) within the inside good nest. This is reasonable for our application. Portland cement is purchased nearly exclusively by ready-mix concrete plants and other construction companies. These firms employ similar production technologies and compete under comparable demand conditions. Thus, we are skeptical that meaningful heterogeneity exists in consumer preferences for plant observables (e.g., price and distance). Without such heterogeneity, the IIA property arises quite naturally – for example, the random coefficient logit demand model collapses to standard logit when the distribution of consumer preferences is degenerate.

[23]To perform the normalization, we regress regional portland cement consumption on the demand predictors (aggregated to the regional level), impute predicted consumption at the county level based on the estimated relationships, and then scale predicted consumption by a constant of proportionality to obtain potential demand. The regression of regional portland cement consumption on the demand predictors yields an R^2 of 0.9786. Additional predictors, such as land area, population, and percent change in gross domestic product, contribute little additional explanatory power. We use a constant of proportionality of 1.4, which is sufficient to ensure that potential demand exceeds observed consumption in each region-year observation.

of endogenous data available to the econometrician for period t as $\boldsymbol{y}_t$. In our application, this vector contains production, consumption, and average prices for various geographic regions, as well as trade flows between some of those regions. For notational brevity, we stack the distances, demand shifters, and cost shifters into a single matrix $\boldsymbol{X}_t$. We let the K-dimensional vector $\boldsymbol{\theta}_0$ contain the model parameters. Finally, the vector of aggregated equilibrium predictions, denoted $\widetilde{\boldsymbol{y}}_t(\boldsymbol{\theta}; \boldsymbol{X}_t)$, is a function of the candidate parameter vector and the exogenous data.

4.2 The estimator

The estimator minimizes the weighted sum of squared deviations between the endogenous data and the aggregated equilibrium predictions. It takes the following form:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{T} \sum_{t=1}^{T} [\boldsymbol{y}_t - \widetilde{\boldsymbol{y}}_t(\boldsymbol{\theta}; \boldsymbol{X}_t)]' \boldsymbol{C}_T^{-1} [\boldsymbol{y}_t - \widetilde{\boldsymbol{y}}_t(\boldsymbol{\theta}; \boldsymbol{X}_t)], \tag{5}$$

where $\boldsymbol{\Theta}$ is some compact parameter space. The estimation procedure is multiple-equation nonlinear least squares. Each element of the vector $[\boldsymbol{y}_t - \widetilde{\boldsymbol{y}}_t(\boldsymbol{\theta}; \boldsymbol{X}_t)]$ defines a single nonlinear equation and $\boldsymbol{C}_T$ is a positive definite matrix that weights the equations.

In our application, we use eleven aggregated equilibrium predictions for which empirical analogs are available: average mill prices (production weighted) charged by plants in northern California, in southern California, and in Arizona and Nevada; total production by plants in the same three geographic regions; total consumption by consumers in northern California, in southern California, in Arizona, and in Nevada; and shipments from plants in California to consumers in northern California.[24] The empirical analogs are available annually over 1983-2003 for the first ten predictions (prices, production and consumption) and over 1990-2003 for the eleventh prediction (cross-region shipments).[25] Thus, estimation exploits variation in 21 time-series observations on ten nonlinear equations and 14 time-series observations on one nonlinear equation. We use methods developed in Srivastava and Zaatar (1973) and

---

[24]For reasons of data availability, we combine plants from Arizona and Nevada when constructing prices and production but not when constructing consumption. The ability to handle such data mismatches is one of the strengths of our estimation approach. Other data on cross-region shipments are available but there are fewer data points – for instance, shipments from California to Nevada are available only over 2000-2003. We find that, in practice, the inclusion of additional shipping data undermines the invertibility of the weighting matrix. Still, the withheld data provide natural checks on the model predictions that we examine in detail after estimation.

[25]The shipments data are necessary to pin down the coal price coefficient in the marginal cost specification. This is unexpected but could be attributable to the high degree of correlation between coal prices and diesel prices over the sample period.

Hwang (1990) to account for the unequal numbers of observations across equations.

The estimator can be derived from the assumption that discrepancies between the endogenous data and the equilibrium predictions, when evaluated at the population parameters, are independent of the distances, demand shifters, and cost shifters. The moment conditions are

$$E[\boldsymbol{\omega}_t | \boldsymbol{X}_t] = 0, \tag{6}$$

where $\boldsymbol{\omega}_t = \boldsymbol{y}_t - \widetilde{\boldsymbol{y}}_t(\boldsymbol{\theta}_0; \boldsymbol{X}_t)$. This seems sensible for many applications. As an example, discrepancies between the endogenous data and the equilibrium predictions could be attributed in part to measurement error in the endogenous data. In our application, it is an open question whether plants respond to the USGS census with complete accuracy, given the costs of creating, modifying, and verifying internal company data. Further, the USGS imputes its data when plants are non-responsive, which could introduce additional noise.[26]

If there is reason to suspect that discrepancies between the endogenous data and the equilibrium predictions *are* correlated with the distances, demand shifters, or cost shifters, then estimation remains feasible provided instruments are available. Suitable instruments should be correlated with the equilibrium prediction but uncorrelated with the discrepancies, when evaluated at the population parameters. Such instruments may be readily available. For instance, if discrepancies are correlated with the cost shifters then the distances and demand shifters would provide valid instruments. One could then construct the generalized method of moments analog to equation 5 and proceed with estimation.

Finally, efficiency is improved when a consistent estimate of the cross-equation variance matrix (i.e., $E[\boldsymbol{\omega}_t | \boldsymbol{X}_t] E[\boldsymbol{\omega}_t | \boldsymbol{X}_t]'$) is used to weight the nonlinear equations. The two-step procedure of Hansen (1982) is applicable. In the first stage, we find that using a diagonal matrix in which each element is the sample variance of the relevant endogenous series (e.g., $\frac{1}{T} \sum_{t=1}^{T} (y_t - \overline{y})^2$) improves performance relative to an identity matrix. We use the methods of Hansen (1982) and Newey and McFadden (1994) to calculate standard errors that are robust to heteroscedasticity and arbitrary correlations among the equations of each period.

---

[26]Under these moment conditions, the estimator can be interpreted as method of moments with optimal instruments. Despite the multiple equations, the model is exactly identified because there are K parameters and K moments. The optimal instruments and the corresponding sample moment conditions are

$$\boldsymbol{Z}_t^* = -\frac{\partial \widetilde{\boldsymbol{y}}(\boldsymbol{\theta}_0; \boldsymbol{X}_t)}{\partial \boldsymbol{\theta}_0} \boldsymbol{\Lambda}_0(\boldsymbol{\theta}_0)^{-1} \quad \text{and} \quad \frac{1}{T} \sum_{t=1}^{T} -\frac{\partial \widetilde{\boldsymbol{y}}(\boldsymbol{\theta}; \boldsymbol{X}_t)}{\partial \boldsymbol{\theta}} \boldsymbol{C}_T^{-1} (\boldsymbol{y}_t - \widetilde{\boldsymbol{y}}_t(\boldsymbol{\theta}; \boldsymbol{X}_t)),$$

respectively, where $\boldsymbol{\Lambda}_0(\boldsymbol{\theta}_0) \equiv E[\boldsymbol{\omega}_t | \boldsymbol{X}_t] E[\boldsymbol{\omega}_t | \boldsymbol{X}_t]'$. We refer the reader to Greene (2003) for details.

4.3 Obtaining aggregate equilibrium predictions

To evaluate the objective function, one must obtain the aggregated equilibrium predictions of the economic model (i.e., $\widetilde{\boldsymbol{y}}_t(\boldsymbol{\theta}; \boldsymbol{X}_t)$) for each candidate parameter vector. The key ingredient is the equilibrium price vector, which can be computed from the first order conditions of the firms' profit maximization problem.

There are $J \times N$ first-order conditions, reflecting the modeling assumption that each plant can discriminate between the consumers of different areas. For notational convenience, we define the block-diagonal matrix $\boldsymbol{\Omega}(\boldsymbol{p}; \boldsymbol{X}_t, \boldsymbol{\theta})$ as the combination of $n = 1, \ldots, N$ sub-matrices, each of dimension $J \times J$. The elements of the sub-matrices are defined as follows:

$$\Omega_{jk}^n(\boldsymbol{p_n}; \boldsymbol{X}_t, \boldsymbol{\theta}) = \begin{cases} \frac{\partial q_{jn}(\boldsymbol{p_n}; \boldsymbol{X}_t, \boldsymbol{\theta})}{\partial p_{kn}} & \text{if } j \text{ and } k \text{ have the same owner} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The elements of each sub-matrix characterize substitution patterns within area $\mathbb{C}_n$, and $\boldsymbol{\Omega}$ has a block diagonal structure because $q_{jn}(\boldsymbol{p_n}; \boldsymbol{X}_t, \boldsymbol{\theta})$ is free of $\boldsymbol{p}_{-n}$. Thus, the construction of $\boldsymbol{\Omega}$ builds on the premises that (1) consumers in each area $\mathbb{C}_n$ select among all J plants, and (2) demand in area $\mathbb{C}_n$ is unaffected by mill prices in area $\mathbb{C}_m$ for $n \neq m$. With this notation in hand, the first-order conditions take the form

$$\boldsymbol{f}(\boldsymbol{p}; \boldsymbol{X}_t, \boldsymbol{\theta}) \equiv \boldsymbol{p} - \boldsymbol{c}(\boldsymbol{Q}(\boldsymbol{p}; \boldsymbol{X}_t, \boldsymbol{\theta}); \boldsymbol{X}_t, \boldsymbol{\theta}) + \boldsymbol{\Omega}^{-1}(\boldsymbol{p}; \boldsymbol{X}_t, \boldsymbol{\theta})\boldsymbol{q}(\boldsymbol{p}; \boldsymbol{X}_t, \boldsymbol{\theta}) = \boldsymbol{0}. \tag{8}$$

A vector of prices that solves this system of equations is a Bertrand-Nash equilibrium. In most applications, however, analytic solutions are unobtainable. Rather, one must solve equation (8) numerically using a nonlinear equation solver to produce a vector of *computed equilibrium prices*, which we denote $\widetilde{\boldsymbol{p}}^*(\boldsymbol{\theta}; \boldsymbol{X}_t)$. Specifically, the nonlinear equation solver selects the vector $\widetilde{\boldsymbol{p}}^*$ to satisfy

$$\frac{1}{JN} \parallel \boldsymbol{f}(\widetilde{\boldsymbol{p}}^*; \boldsymbol{X}_t, \boldsymbol{\theta}) \parallel < \delta,$$

where δ is a user specified tolerance. A tolerance of 1e-13 performs well in numerical experiments based on our application. Numerical error can propagate into the objective function when the tolerance is substantially looser (e.g., 1e-7), which slows overall estimation time and can produce poor estimates. (These thresholds are specific to our application because tolerance is not unit free and must be evaluated relative to the price level.)

Once the equilibrium price vector is obtained, it can be manipulated into the aggre-

gated equilibrium predictions. To formalize this process, we define a function $S : \mathbb{R}^{JN} \to \mathbb{R}^L$ that maps from the equilibrium price vector to the aggregate equilibrium predictions; L is the number of predictions that must be calculated (i.e., the length of $\boldsymbol{y}_t$). The aggregate equilibrium predictions that enter the objective function are given by $\widetilde{\boldsymbol{y}}_t(\boldsymbol{\theta}; \boldsymbol{X}_t) = S(\widetilde{\boldsymbol{p}}^*(\boldsymbol{\theta}; \boldsymbol{X}_t))$. We assume that $S(\cdot)$ is continuously differentiable, which holds for applications based on averaged or summed endogenous data.

> *Example:* In our application, the estimator makes use of 11 nonlinear equations in most time periods. Three of these relate to the average mill prices (production weighted) charged by plants in specific geographic regions. Thus, denoting the set of plants in region r as A_r, these aggregate equilibrium predictions can be calculated as.
>
> $$\widetilde{P}_{rt}(\boldsymbol{\theta}, \boldsymbol{X}_t) = \sum_{j \in A_r} \sum_n \frac{q_{jn}(\widetilde{\boldsymbol{p}}_n^*; \boldsymbol{X}_t, \boldsymbol{\theta})}{\sum_{j \in A_r} \sum_n q_{jn}(\widetilde{\boldsymbol{p}}_n^*; \boldsymbol{X}_t, \boldsymbol{\theta})} \widetilde{p}_{jn}^*.$$
>
> The aggregate equilibrium predictions for production, consumption, and cross-region shipments can be written analogously. These predictions can be stacked into the vector $\widetilde{\boldsymbol{y}}_t(\boldsymbol{\theta}; \boldsymbol{X}_t)$ and compared to the data.

The estimator has a nested structure in which a numerical optimizer finds the parameter vector that minimizes the objective function and a nonlinear equation solver computes equilibrium prices conditional on the parameter vector. This structure complicates implementation because the dimensionality of the equilibrium price vector that must be computed can be quite large. In our application, there are 90 consumer-areas and 14 plants (in a typical year), resulting in a price vector with 1,260 elements. In many standard numerical packages, solving for such a large price vector is computationally intensive. One way to reduce computational complexity is to assume that the firm's marginal cost function is constant. Under this assumption, one can solve for the equilibrium prices in each consumer-area individually, substantially saving computational time.

In many applications, however, marginal costs are unlikely to be constant and the prices that characterize equilibrium in different consumer areas are not independent. If marginal costs increase with production (e.g., due to capacity constraints), then lowering price in one consumer area will increase overall quantity sold by a plant, raising its cost, and hence its equilibrium price, in other areas. In general, one may need to solve for the entire vector of prices jointly. We use a large-scale nonlinear equation solver developed in La Cruz, Martínez, and Raydan (2006) to compute equilibrium in our application. The equation solver employs

a quasi-Newton method and exploits simple derivative-free approximations to the Jacobian matrix; it converges more quickly than other algorithms and does not sacrifice precision. This algorithm is available as part of the BB package in the statistical programming language R. Our application uses a Fortran version of the nonlinear equation solver, which significantly increases computational speed.[27]

4.4 Consistency and Asymptotic Normality

The demonstration of consistency and asymptotic normality is complicated by the fact that the objective function is constructed using equilibrium predictions. These predictions are functions of the implicit solution to the firms' first-order conditions. In addition to an identification assumption and standard regularity conditions, consistency and asymptotic normality require (1) the continuity of the implicit solution in its arguments; (2) the differentiability of the implicit solution when evaluated at the population parameter vector, for almost all realizations of the exogenous data; and (3) the objective function to satisfy a Lipschitz condition. In this subsection, we specify properties of the first-order conditions that guarantee that these properties hold. All proofs appear in Appendix A. Readers who are only interested in the application of our method may skip this section.

It is useful to write the objective function as

$$\frac{1}{T} \sum_{t=1}^{T} m(\boldsymbol{\theta}, \boldsymbol{y}_t, \boldsymbol{X}_t) \equiv$$

$$\frac{1}{T} \sum_{t=1}^{T} (\boldsymbol{y}_t - S(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t))' \boldsymbol{W}_t (\boldsymbol{y}_t - S(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t)), \quad (9)$$

where $S(\cdot)$ is the function maps from the equilibrium price vector to the aggregate equilibrium predictions, as defined in Section 4.3, and $\boldsymbol{W}_t \equiv \boldsymbol{C}_t^{-1}$. We assume that $\boldsymbol{W} = \lim_{t \to \infty} \boldsymbol{W}_t$ exists and is positive definite. Following the notation in Section 4.2, the data generating process for the endogenous data is:

$$\boldsymbol{y}_t = S(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t) + \boldsymbol{\omega}_t. \quad (10)$$

---

[27]The function that implements the solver is titled dfsane. Our experience is that Fortran reduces the computational time of the inner loop by a factor of 30 or more, relative to the dfsane function in R. The numerical computation of equilibrium takes between 2 and 12 seconds for most candidate parameter vectors when run on a 2.40GHz dual core processor with 4.00GB of RAM.

We denote the distribution and support of $\boldsymbol{X}$ as F_x and U, respectively. Since $\boldsymbol{X}$ includes the distance between plants and counties, it is continuous in at least one of its elements. We denote the distribution of $\boldsymbol{\omega}$ as F_ω.

Assumption A1 (Global Identification): *The parameter vector $\boldsymbol{\theta}_0$ is globally identified in $\boldsymbol{\Theta}$. Formally, $E[\boldsymbol{y}_t - S(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}), \boldsymbol{\theta}, \boldsymbol{X})|\boldsymbol{X}] = \boldsymbol{0} \leftrightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0$.*

Identification assumptions such as A1 are standard in empirical industrial organization because the basic conditions for identification in nonlinear models are difficult to formulate and verify (e.g., Ruud (2000)). In our case, however, A1 could be violated even if the parameters of the model are identifiable with disaggregate data (i.e., with individual prices and quantities). This is more likely when aggregation is particularly coarse. Empirically, one can evaluate the potential for aggregation problems using artificial data experiments, and we develop one such evaluation in our application.

Assumption A2 (Existence and Uniqueness): *A unique Bertrand-Nash equilibrium exists, and the prices that support it are strictly positive. Formally, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ there exists a vector $\boldsymbol{p}_1 \in \mathbb{R}_+^{JN}$ such that $\boldsymbol{f}(\boldsymbol{p}_1; \boldsymbol{X}, \boldsymbol{\theta}) = \boldsymbol{0}$. Further, $\boldsymbol{f}(\boldsymbol{p}_1; \boldsymbol{X}, \boldsymbol{\theta}) = \boldsymbol{f}(\boldsymbol{p}_2; \boldsymbol{X}_t, \boldsymbol{\theta}) = \boldsymbol{0} \leftrightarrow \boldsymbol{p}_1 = \boldsymbol{p}_2$.*

Recent theoretical contributions demonstrate that A2 holds for two special cases of our model (see Section 3.2 for discussion). We recommend that researchers evaluate whether uniqueness holds using numerical techniques when violations cannot be dismissed on theoretical grounds. It is worth noting that A2 may be overly strong – existence may suffice if, for instance, the econometrician can compute the universe of equilibria and select the equilibrium closest to the data (e.g., as in Bisin, Moro, and Topa (2010)). We defer the evaluation of such possibilities to future research. One result of A2 is the following Lemma:

Lemma 1 (Continuity): *Under A2, the mapping $\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X})$ is continuous in $\boldsymbol{\theta}$ and $\boldsymbol{X}$.*

The corollary that $S(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}), \boldsymbol{\theta}, \boldsymbol{X})$ is continuous in $\boldsymbol{\theta}$ and $\boldsymbol{X}$ follows from the properties of the aggregating function $S(\cdot)$.

Next, since the Jacobian matrix of the first-order conditions need not be nonsingular over all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we cannot rely on the Implicit Function Theorem (IFT) to guarantee that $\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X})$ is continuously differentiable in a neighborhood of $(\boldsymbol{\theta}, \boldsymbol{X})$. We proceed with the weaker requirement that the first order conditions are well-behaved, in the sense that if the Jacobian is singular at $(\boldsymbol{\theta}_0, \boldsymbol{X}_0)$ then a perturbation to $\boldsymbol{X}_0 + \epsilon$ yields nonsingularity:
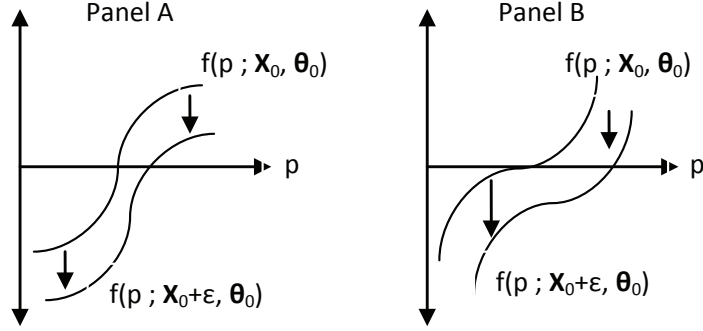
Figure 4: The Implicit Function Theorem is inapplicable at $(\boldsymbol{\theta}_0, \boldsymbol{X}_0)$.

Assumption A3: *Denote the Jacobian matrix of the first order condition $\boldsymbol{f}$ with respect to $\boldsymbol{p}$ at some point $(\boldsymbol{\theta}, \boldsymbol{X})$ as $J\boldsymbol{f}_{\boldsymbol{p}}(\boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{X})$. Consider the true parameter value $\boldsymbol{\theta}_0$ and the set of points*

$$B(\boldsymbol{\theta}_0) = \{\boldsymbol{X} : J\boldsymbol{f}_{\boldsymbol{p}}(\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X}), \boldsymbol{\theta}_0, \boldsymbol{X}) \text{ is singular}\}.$$

For each point $\boldsymbol{X}_0$ in $B(\boldsymbol{\theta_0})$, there exists a neighborhood $N(\boldsymbol{X}_0, \boldsymbol{\theta}_0)$ around $\boldsymbol{X}_0$ such that the Jacobian matrix $J\boldsymbol{f}_{\boldsymbol{p}}(\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X}), \boldsymbol{\theta}_0, \boldsymbol{X})$ is nonsingular for all $\boldsymbol{X} \in N(\boldsymbol{X}_0, \boldsymbol{\theta}_0)$ and $\boldsymbol{X} \neq \boldsymbol{X}_0$.

Under A3, if the differentiability of $\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X}_0)$ in $\boldsymbol{\theta}$ fails then the IFT guarantees continuous differentiability at the new equilibrium price $\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X}_0 + \epsilon)$. Provided that one can apply the IFT in an open ball around $\boldsymbol{X}_0$, and at least one element in $\boldsymbol{X}$ is continuously distributed, the set of points at which the IFT fails occurs with zero probability. Figure 4 provides a graphical illustration for the case of a one-dimensional price vector. Equilibrium is characterized by the intersection of $f(p, \boldsymbol{\theta}_0, \boldsymbol{X}_0)$ and the horizontal axis.[28] The IFT cannot be applied when the slope of the first-order condition is undefined (panel A) or zero (panel B) at equilibrium. In these cases, A3 guarantees that a perturbation of the exogenous data yields nonzero derivatives at equilibrium. Lemma 2 follows:

Lemma 2 (Differentiability): *Under A2-A3, $S(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}), \boldsymbol{\theta}, \boldsymbol{X})$ is differentiable in $\boldsymbol{\theta}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for almost all $\boldsymbol{X}$ in U.*

Finally, consistency and asymptotic normality require the objective function to satisfy a Lipschitz condition. The condition holds provided that partial derivatives of $\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ almost always exist and can be bounded (when they exist):

---

[28]Assumption A2 guarantees a single crossing.

Panel A:

θ

Potential non-differentiability of p*

(x_0, θ_0)

$B(\theta_0)$

x

Panel B:

θ

x-nbd of x_0

Potential non-differentiability of p*

(x_0, θ_0)

$B(\theta_0)$

x

Panel C:

θ

Potential non-differentiability of p*
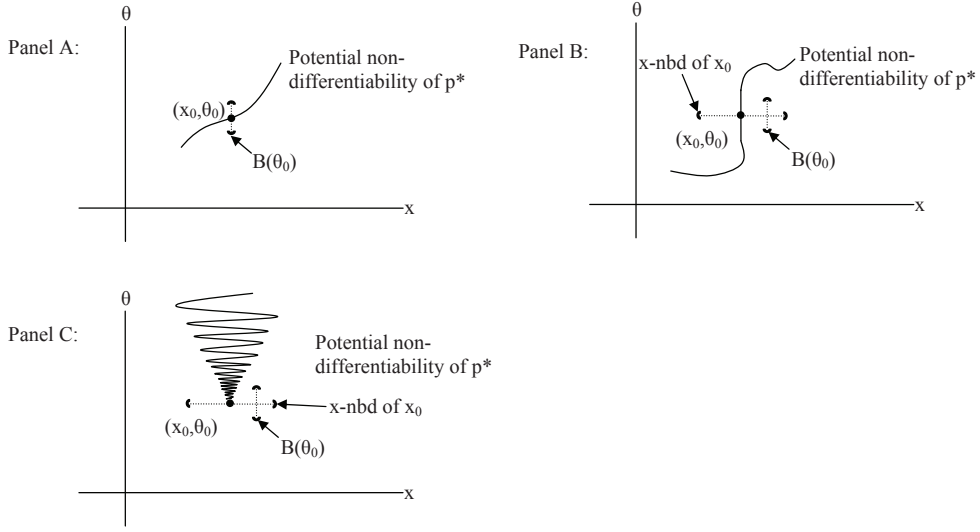
x-nbd of x_0

(x_0, θ_0)

$B(\theta_0)$

x

Figure 5: Points at which the partial derivatives of $\boldsymbol{p}^*$ do not exist.

Assumption A4: *The partial derivatives of $\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ are bounded by a measurable function $M(\boldsymbol{X})$ at all points $\boldsymbol{X} \in U$ for which $J\boldsymbol{f}_{\boldsymbol{p}}(\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X}), \boldsymbol{\theta}_0, \boldsymbol{X})$ is nonsingular. Further, for any point $(\boldsymbol{\theta}_0, \boldsymbol{X}_0)$ at which $J\boldsymbol{f}_{\boldsymbol{p}}(\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X}_0), \boldsymbol{\theta}_0, \boldsymbol{X}_0)$ is singular, there exists a neighborhood $B(\boldsymbol{\theta}_0)$ in $\boldsymbol{\theta}$-space in which either:*

(i) *For all $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0), \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, the partial derivatives of $\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_0)$ with respect to the elements of $\boldsymbol{\theta}$ exist and are bounded by a measurable function $M(\boldsymbol{X})$.*

(ii) *For all $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0)$, the partial derivatives of $\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X})$ with respect to the elements of $\boldsymbol{\theta}$ exist in a neighborhood of $\boldsymbol{X}$ around $\boldsymbol{X}_0$, with $\boldsymbol{X} \neq \boldsymbol{X}_0$. These partial derivatives are bounded by a measurable function $M(\boldsymbol{X}) \leq M < \infty$.*

Figure 5 provides a graphical illustration of A4 for one-dimensional $\boldsymbol{X}$ and $\boldsymbol{\theta}$. The line in each panel represents the combinations of x and θ for which the partial derivatives of $\boldsymbol{p}^*(\theta, x)$ do not exist. If the line passes through (θ_0, x_0), A4 guarantees that either (1) the partial derivatives exist for $\theta \in B(\theta_0)$ and $\theta \neq \theta_0$, as in panel A; or (2) the partial derivatives exist for all $\theta \in B(\theta_0)$ in a neighborhood of x around x_0 that excludes x_0, as in panel B. The assumption rules out problematic cases such as when the points of non-differentiability are given by the function $\sin(1/x)$, as shown in panel C. In that case, for a small enough neighborhood around x_0, there are always points of non-differentiability for $\theta \in B(\theta_0)$. A4 yields the Lipschitz condition:

Lemma 3 (Lipschitz Continuity of m in $\boldsymbol{\theta}$-space): *Under A2-A4, there is a measurable function $\dot{m}(\boldsymbol{y}, \boldsymbol{X})$ such that*

$$|m(\boldsymbol{\theta}_1, \boldsymbol{y}, \boldsymbol{X}) - m(\boldsymbol{\theta}_2, \boldsymbol{y}, \boldsymbol{X})| \leq \dot{m}(\boldsymbol{y}, \boldsymbol{X}) \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$$

for every $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in some open neighborhood of $\boldsymbol{\theta}_0$

The asymptotic properties of the estimator are now obtainable:

Proposition 1 (Consistency and Asymptotic Normality): *Under A1-A5 and certain regularity conditions enumerated in the appendix,*

(i) $\text{plim}\, \widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$

(ii) $\sqrt{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to^d N\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\theta}_0}^{-1}\left[\int_U \int_{\boldsymbol{\omega}} \nabla m(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{X}) \nabla m(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{X})' F_x(\boldsymbol{X}) F_\omega(\boldsymbol{\omega})\right] \boldsymbol{V}_{\boldsymbol{\theta}_0}^{-1}\right)$

where $\boldsymbol{V}_{\boldsymbol{\theta}_0}$ is a symmetric matrix that contains the second derivatives of $m(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{X})$ with respect to $\boldsymbol{\theta}$, evaluated at $\boldsymbol{\theta}_0$.

5 Identification and Related Topics

5.1 Aggregation problems and multiple equilibria

The asymptotic properties of the estimator rest on a number of assumptions, some of which can be evaluated numerically. First, point identification can fail if multiple candidate parameters produce equilibrium predictions that are identical once aggregated to the level of the available data. This is more likely when the data are relatively coarse so that aggregation entails a substantial loss of information. We conduct an artificial data experiment to check for this sort of aggregation problem in our empirical application. We pair a vector of "true" parameters with 40 randomly-drawn sets of exogenous data. Both the parameters and the data are chosen to mimic the application. For each of set of exogenous data, we compute equilibrium, generate the relevant aggregated data, and estimate the model. We argue that the parameters are reasonably identified if the estimates are close to the true parameters.[29]

---

[29]The exogenous data includes the plant capacities, the potential demand of counties, the diesel price, the import price, and two cost shifters. We randomly draw capacity and potential demand from the data (with replacement), and we draw the remaining data from normal distributions. Specifically, we use the following distributions: diesel price $\sim N(1, 0.28)$, import price $\sim N(50, 9)$, cost shifter $1 \sim N(60, 15)$, and cost shifter $2 \sim N(9, 2)$. We redraw data that are below zero and data that lead the estimator to nonsensical areas of

Table 1: Artificial Data Test for Identification

Variable	Parameter	Truth (θ)	Transformed ($\widetilde{\theta}$)	Mean Est	RMSE
Demand					
Cement Price	β^p	-0.07	-2.66	-2.51	0.66
Miles×Diesel Price	β^d	-25.00	3.22	2.86	0.59
Import Dummy	β^i	-4.00	-4.00	-6.07	1.23
Intercept	β^c	2.00	2.00	1.11	0.51
Inclusive Value	λ	0.09	-2.31	-1.73	0.54
Marginal Costs					
Cost Shifter 1	α_1	0.70	-0.36	-0.88	0.51
Cost Shifter 2	α_2	3.00	1.10	0.54	0.45
Utilization Threshold	ν	0.90	2.19	1.71	0.59
Over-Utilization Cost	γ	300.00	5.70	6.14	1.05

Results of estimation on 40 data sets that are randomly drawn based on the "true" parameters listed. The parameters are transformed prior to estimation to place constraints on the parameter signs/magnitudes (see Appendix C). Mean Est and RMSE are the mean of the estimated (transformed) parameters and the root mean-squared error, respectively.

Table 1 shows the results of the artificial data experiment. Interpretation is complicated somewhat because we use non-linear transformations to constrain the some of coefficients (e.g., $\beta^p < 0$), and we defer details on these transformations to Appendix C. Nonetheless, it is clear that the means of the estimated coefficients are close to transformed true parameters. The means of the price and distance coefficients are within 6 percent and 11 percent of the truth, respectively. This precision is notable because the ration of price and distance coefficients determines the unit transportation cost and thereby the degree of spatial differentiation. The other means of the estimated demand coefficients are somewhat farther from the truth. Among the marginal cost parameters, the mean estimated coefficients are accurate for the utilization threshold and the over-utilization cost but less accurate for the constant cost shifters. We conclude that the primary coefficients of interest (for spatial considerations) are likely well-identified but that some skepticism of the other coefficients may be appropriate, especially with regard to the constant marginal cost shifters.

Second, the continuity and differentiability of the implicit solution to the firms' first

---

parameter space. Throughout, we hold plant and county locations fixed to maintain tractability, and rely on the random draws of capacity, potential demand, and diesel prices to create variation in the distances between production capacity and consumers. Each artificial data set includes 21 draws on the exogenous data, with each draw representing a single time-series observation.

order condition fails if multiple equilibria are present. We search for only a single equilibria in the inner loop in our application. For robustness, we conduct a Monte Carlo experiment and search for the existence of multiple equilibria. In particular, we compute equilibrium at eleven different starting points for thousands of randomly-drawn candidate parameter vectors. We then evaluate whether, for each given candidate parameter vector, the computed equilibrium prices are sensitive to the starting points.[30] More precisely, for each candidate parameter vector, we calculate the standard deviation of each equilibrium price across the eleven starting points. (So there are 1,260 standard deviations for a typical equilibrium price vector of 1,260 plant-area elements.) The results indicate that the maximum standard deviation, over all candidate parameter vectors and all plant-area prices, is zero to computer precision. Thus, the Monte Carlo experiment finds no evidence of multiple equilibria. This may be unsurprising because, theoretically, uniqueness is ensured for two close cousins of our model: nested logit demand with single-plant firms (Mizuno 2003) and logit demand with increasing marginal costs and multi-plant firms (Konovalov and Sándor 2010).

5.2   Key empirical relationships

The empirical relationships that drive parameter estimates can be transparent despite the complicated nonlinear relationships involved. We plot some of these relationships in Figure 6. On the demand side, the price coefficient is primarily determined by the relationship between consumption and price. In panel A, we plot cement prices and the ratio of consumption to potential demand ("market coverage") over the sample period. There is weak negative correlation, consistent with downward-sloping but inelastic aggregate demand. Next, the distance coefficient is primarily determined by (1) the cross-region shipments, and (2) the relationship between consumption and production in different regions. To illustrate this second source of identification, we plot the gap between production and consumption ("excess production") for each region in panel B. In many years, excess production is positive in Southern California and negative elsewhere, consistent with inter-regional trade flows. The magnitude of these implied trade flows drives the distance coefficient. Interestingly, the

---

[30]We consider 300 parameter vectors for each of the 21 years in the sample, for a total of 6,300 candidate parameter vectors. For each $\theta_i \in \boldsymbol{\theta}$, we draw from the distribution $N(\widehat{\mu}_i, \widehat{\sigma}_i^2)$, where $\widehat{\mu}_i$ and $\widehat{\sigma}_i$ are the coefficient and standard error, respectively, reported in Table 2. We then compute the numerical equilibrium for each parameter vector, using eleven different starting vectors. We define the elements of the starting vectors to be $p_{jnt} = \phi \overline{p}_t$, where $\overline{p}_t$ is the average price of portland cement and $\phi = 0.5, 0.6, \ldots, 1.4, 1.5$. Thus, we start the equation solver at initial prices that are sometimes larger and sometimes smaller than the average prices observed in the data. The equal-solver computes numerical equilibria for 90 percent of the candidate vectors. See appendix C for a discussion of non-convergence in the inner-loop.
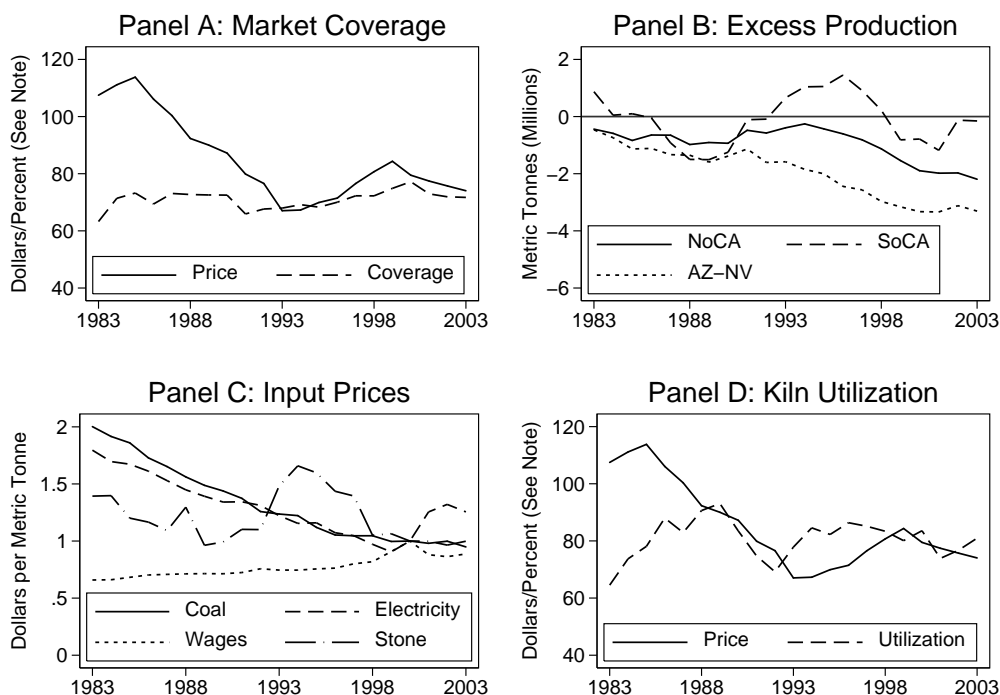
Figure 6: Empirical Relationships in the U.S. Southwest. Panel A plots average cement prices and market coverage. Prices are in dollars per metric tonne and market coverage is defined as the ratio of consumption to potential demand (times 100). Panel B plots excess production in each region, which we define as the gap between between production and consumption. Excess production is in millions of metric tonnes. Panel C plots average coal prices, electricity prices, durable-goods manufacturing wages, and crushed stone prices in California. For comparability, each time-series is converted to an index that equals one in 2000. Panel D plots the average cement price and industry-wide utilization (times 100).

implied trade flows are higher later in the sample, when the diesel fuel is less expensive.

On the supply side, the parameters are determined by the relationships between prices and the marginal cost shifters. In panel C, we plot the coal price, the electricity price, the durable-goods manufacturing wage, and the crushed stone price for California. Coal and electricity prices are highly correlated with the cement price, consistent with a strong influence on marginal costs; inter-regional variation in input prices helps disentangle the two effects. It is less clear that wages and crushed stone prices are positively correlated with cement prices. Finally, the utilization parameters are primarily determined by (1) the relative pro-cyclicality of production and consumption, and (2) the relationship between utilization and prices. We explore the second source of identification in panel D, which shows cement prices and industry-wide utilization over the sample period. The two metrics are negatively correlated over 1983-1987 and positively correlated over 1988-2003.

6 Empirical Results

6.1 Transportation costs and price discrimination

Table 2 presents the parameter estimates. The price and distance coefficients are the two primary objects of interest on the demand side; both are negative and precisely estimated.[31] Together, the coefficients imply that consumers pay roughly $0.30 per tonne mile, given diesel prices at the 2000 level.[32] The mean shipping distance that arises in equilibrium is 92 miles, and only 10 percent of shipments are more than 175 miles. The other demand parameters take reasonable values and are precisely identified. The coefficient on the import dummy is negative because observed import prices do not reflect the full price of imported cement. The inclusive value coefficient suggests that consumer tastes for the different cement providers are highly correlated; the standard (non-nested) logit model is easily rejected.

We find that transportation costs facilitate the exercise of localized market power in some counties but not others.[33] Table 3 contrasts Maricopa County (i.e., Phoenix) and Los Angeles County in 2003, based on the equilibrium computed with the parameter estimates presented above. As shown, fully 89 percent of the cement consumed in Maricopa County is supplied by two plants – operated by Phoenix Cement and California Cement – that are approximately 100 miles north and south of the county, respectively. The mill prices of these plants are around $80 per metric tonne and consumers must spend around $30 on transportation. The mill prices of the southern California plants to Maricopa County are lower but transportation costs are much higher (e.g., the mill price of the Cemex plant is $63 but transportation is $87). As a result, the two plants in Arizona can support mill prices to Maricopa County that are well above the cost of production.[34] By contrast, the leading

---

[31]The aggregate elasticity implied by the price coefficient is -0.16 in the median year, consistent with the conventional wisdom that materials such as steel, asphalt, and lumber are poor substitutes for portland cement in most construction projects. The median firm-level elasticity of -5.70 is indicative of substantial price competition among the firms.

[32]The ratio of the distance and price coefficients is the willingness-to-pay for proximity, incorporating transportation costs and any other distance-related costs (e.g., reduced reliability). We refer to the willingness-to-pay as the transportation cost although the two concepts may not be strictly equivalent. The calculation is $\frac{26.42}{0.087} \frac{i}{1000} = 0.3037$, where $i = 1$ in 2000.

[33]An interesting implication of the specification – one that we have not fully explored – is that transportation costs and spatial differentiation fluctuate with diesel prices. The extent to which carbon or gasoline taxes would have unintended consequences on the intensity of competition in industries such as portland cement remains an open question.

[34]The margin shown is based on the mill price and the constant portion of marginal costs, and approximates a variable cost margin. In the notation established, $m = (p_{jn} - \boldsymbol{w}'_j \widehat{\boldsymbol{\alpha}})/p_{jn}$. Incorporating utilization costs would yield the Lerner index. We find that plants with localized market power typically operate at higher utilization rates, presumably due to the economic profits available.

Table 2: Estimation Results

Variable	Parameter	Estimate	St. Error
Demand			
Cement Price	β^p	-0.087	0.002
Miles×Diesel Price	β^d	-26.42	1.78
Import Dummy	β^i	-3.80	0.06
Intercept	β^c	1.88	0.08
Inclusive Value	λ	0.10	0.004
Marginal Costs			
Coal Price	α_1	0.64	0.05
Electricity Price	α_2	2.28	0.47
Hourly Wages	α_3	0.01	0.04
Crushed Stone Price	α_4	0.29	0.31
Utilization Threshold	ν	0.86	0.01
Over-Utilization Cost	γ	233.91	38.16

Estimation exploits variation in regional consumption, production, and average prices over the period 1983-2003, as well as variation in shipments from California to Northern California over the period 1990-2003. The prices of cement, coal, and crushed stone are in dollars per metric tonne. Miles are in thousands. The diesel price is an index that equals one in 2000. The price of electricity is in cents per kilowatt-hour, and hourly wages are in dollars per hour. The marginal cost parameter ϕ is normalized to 1.5, which ensures the theoretical existence of equilibrium. Standard errors are robust to heteroscedasticity and contemporaneous correlations between moments.

Table 3: Leading Plants in Maricopa County and Los Angeles County in 2003

Plant Owner	Plant Location	Distance	Mill Price	Trans. Cost	Margin	Share
Maricopa County (Phoenix)						
Phoenix Cement	Clarkdale, AZ	101	$80	$31	0.49	47%
California Cement	Rillito, AZ	104	$81	$31	0.50	42%
Cemex	Victorville, CA	290	$63	$88	0.21	2%
Los Angeles County						
Cemex	Victorville, CA	55	$66	$17	0.25	22%
National Cement	Encino, CA	19	$77	$6	0.36	21%
California Cement	Mojave, CA	50	$71	$15	0.30	16%

Based on estimation results. Distance is the miles between the plant and the county centroid. Mill Price and Transportation Cost are per metric tonne. Mill Price is computed based on the estimation results. Margin is based on the mill price and the constant portion of marginal costs (it ignores utilization costs). Share is the proportion of domestic cement consumed in the county that is produced by the plant.

suppliers of Los Angeles County are less differentiated spatially and thus have less localized market power – the top three plants set mill prices closer to the cost of production yet supply only 59% percent of consumption.

The geographic configuration of the U.S. Southwest permits some, but not all, plants to discriminate among consumers. In Figure 7, we plot the "total cost of purchase" (i.e., the mill price plus the transportation cost) for counties within 400 miles of the Cemex plant in southern California and the Phoenix Cement plant in Arizona. In the absence of price discrimination, one would expect the total cost of purchase to increase linearly in distance. This is precisely what one observes for the Cemex plant. The line of best fit is produced from a regression of total purchase cost on distance, using *only* counties farther than 200 miles from the plant. Yet it predicts total purchase costs for closer plants equally well. Further, since the slope of the line is 0.2953, total purchase costs increase at the same rate as transportation costs (which we estimate at $0.30 per tonne-mile). By contrast, the Phoenix Cement plant price discriminates among consumers. The total costs of purchase for consumers in counties within 200 miles exceed the line of best fit based on counties farther than 200 miles from the plant by $10.83 on average; this is due to higher mill prices for consumers in nearby counties.[35] That the slope of the best fit line is 0.3023 indicates that

[35]The gap between equilibrium prices and the line of best fit can be interpreted as a back-of-the-envelop calculation of how much localized market power increases prices. This calculation excludes competitive interactions, however. If Phoenix Cement were to change its price schedule then, presumably, so would its competitors. We account for these competitive interactions in a counter-factual policy experiment presented
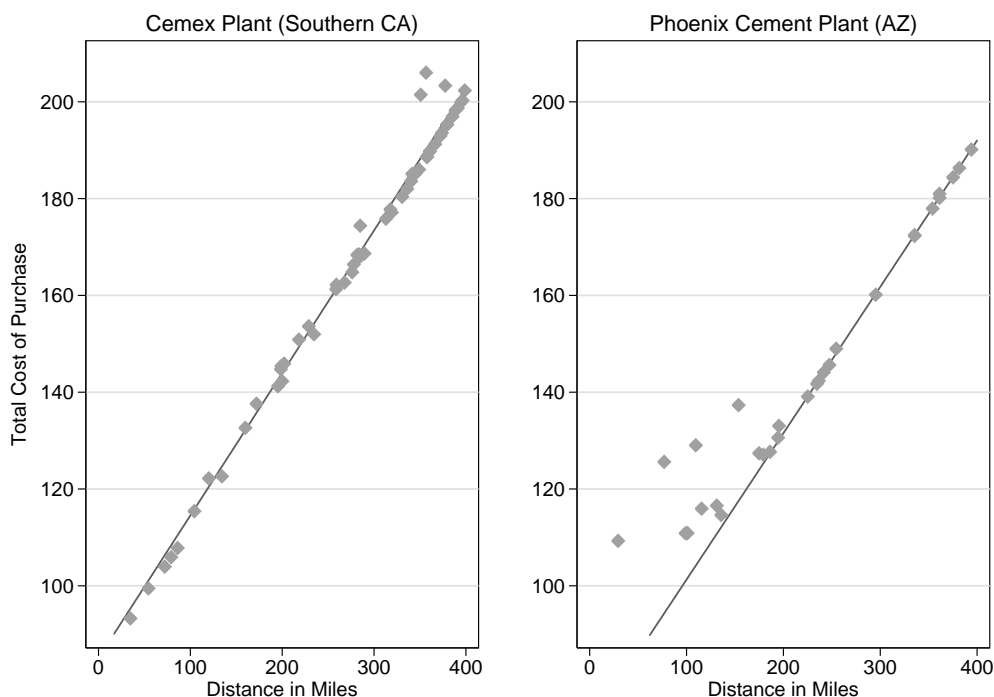
Figure 7: Price Discrimination at Two Plants in 2003. The vertical axis is the total cost of purchase, i.e. the mill price plus the transportation cost incurred by the consumer. The mill price is computed based on the estimation results. The horizontal axis is the distance in miles between the plant and the county centroid. Each dot represents the total cost of purchase for a plant-county pair. The line of best fit is from a regression of total cost of purchase on distance, using pairs with distance greater than 200 miles.

spatial price discrimination is a local phenomenon – the plant does not discriminate between "distant" and "very distant" consumers.

The critical difference between the Cemex plant in southern California and the Phoenix Cement plant in Arizona is location. The presence of nearby competitors constrains price discrimination on the part of Cemex plant, whereas the Phoenix Cement plant is more differentiated spatially (e.g., see Figure 1 and Table 3). To generalize this somewhat, we plot the plant-county specific margins in Figure 8.[36] Plants should earn higher margins from sales to nearby counties only to the extent they price discriminate. It is apparent that the most pronounced discrimination occurs at plants that are differentiated from competitors – the Phoenix Cement and California Cement plants in Arizona, the Centex plant in Nevada,

in Section 7.

[36]Again, the margin shown is based on the mill price and the constant portion of marginal costs, and approximates a variable cost margin.
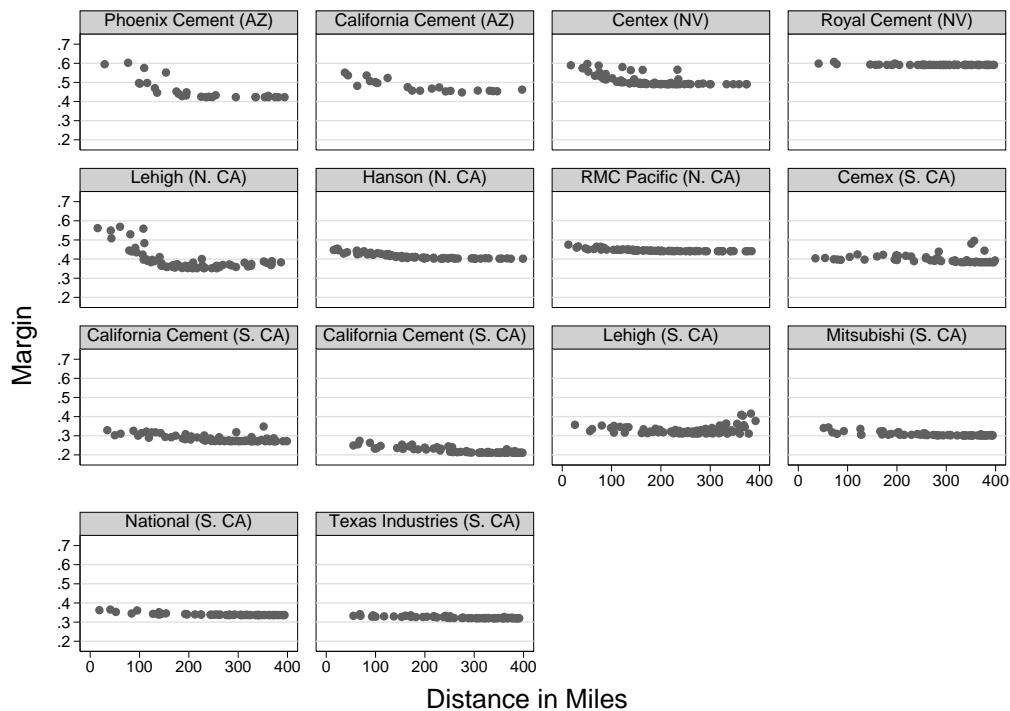
31

Figure 8: Price Discrimination and Margins in 2003, by plant. The vertical axis is a margin based on mill prices and the constant portion of marginal costs (i.e., it excludes utilization costs). Margins are computed based on the estimation results. The horizontal axis is the distance in miles between the plant and the county centroid. Each dot represents the margin for a plant-county pair.

and the Lehigh Cement plant in northern California.[37] By contrast, price discrimination is more subdued at the plants in southern California and near the San Francisco Bay Area.

6.2 Marginal costs

We estimate marginal costs to be \$69.40 in the mean plant-year (weighted by production). Of these marginal costs, \$60.50 is attributable to costs related to coal, electricity, labor and raw materials, and the remaining \$8.90 is attributable to high utilization rates. Integrating the marginal cost function over the levels of production that arise in numerical equilibrium yields an average variable cost of \$51 million. Virtually all of these variable costs – 98.5 percent – are due to coal, electricity, labor and raw materials, rather than due to high utilization. Thus, although capacity constraints may have substantial affects on marginal costs, the results

---

[37]The exception is the low-capacity Royal Cement plant in southern Nevada. The plant ships more than 90% of its output to consumers in Clark County (i.e., Las Vegas), and it incurs substantial utilization costs that prevent the plant from profitably lowering its price to consumers in more distant counties.

suggest that their cumulative contribution to variable costs can be minimal. Taking the accounting statistics further, we calculate that the average plant-year has variable revenues of $73 million and that the average gross margin (variable profits over variable revenues) is 0.32. As argued in Ryan (2010), margins of this magnitude may be needed to rationalize entry given the sunk costs associated with plant construction.[38,39]

Finally, we discuss the individual parameter estimates shown in Table 2, each of which deviates somewhat from production data available from the Minerals Yearbooks and EPA (2009). To start, the coal parameter implies that plants burn 0.64 tonnes of coal to produce one tonne of cement, whereas in fact plants burn roughly 0.09 tonnes of coal to produce each tonne of cement. The electricity parameter implies that plants use 228 kilowatt-hours per tonne of cement, whereas the true number is closer to 150. Each tonne of cement requires approximately 0.34 employee-hours yet the parameter on wages is essentially zero. Lastly, the crushed stone coefficient of 0.29 is too small, given that roughly 1.67 tonnes of raw materials are used per tonne of cement. We suspect that these discrepancies are due to measurement error in the data.[40] Alternatively, they may be due to a failure of identification (e.g., see Section 5.1) or due to the implicit assumption that plant productivity is fixed over the sample period – it seems clear that the renegotiation of onerous labor contracts improved productivity in the 1980s (e.g., Northrup (1989), Dunne, Klimek, and Schmitz (2009)).

6.3 Regression fits

One measure of an econometric model's viability is its ability to fit the data. In Figure 9, we plot observed consumption against predicted consumption (panel A), observed production against predicted production (panel B), and observed prices against predicted prices (panel C). Univariate regressions of the data on the predictions indicate that the model explains

---

[38]Lafarge North America, one of the largest domestic producers, reports an average gross margin of 0.33 over 2002-2004 in its public accounting records.

[39]Fixed costs are well understood to be important for production, as well. The trade journal *Rock Products* reports that high capacity portland cement plants incurred averaged $6.96 in maintenance costs per production tonne in 1993 (Rock-Products (1994)). Evaluated at the production levels that correspond to numerical equilibrium in 1993, this number implies that the average plant would have incurred $5.7 million in maintenance costs relative to variable profits of $17.7 million. Our results suggest that the bulk of these maintenance costs are best considered fixed rather than due to high utilization rates. Of course, the static nature of the model precludes more direct inferences about fixed costs.

[40]In particular, the coal prices in the data are free-on-board and do not reflect any transportation costs paid by cement plants; cement plants may negotiate individual contracts with electrical utilities that are not reflected in the data; the wages of cement workers need not track the average wages of durable-goods manufacturing employees; and cement plants typically use limestone from a quarry adjacent to the plant, so the crushed stone price may not proxy the cost of limestone acquisition (i.e., the quarry production costs).
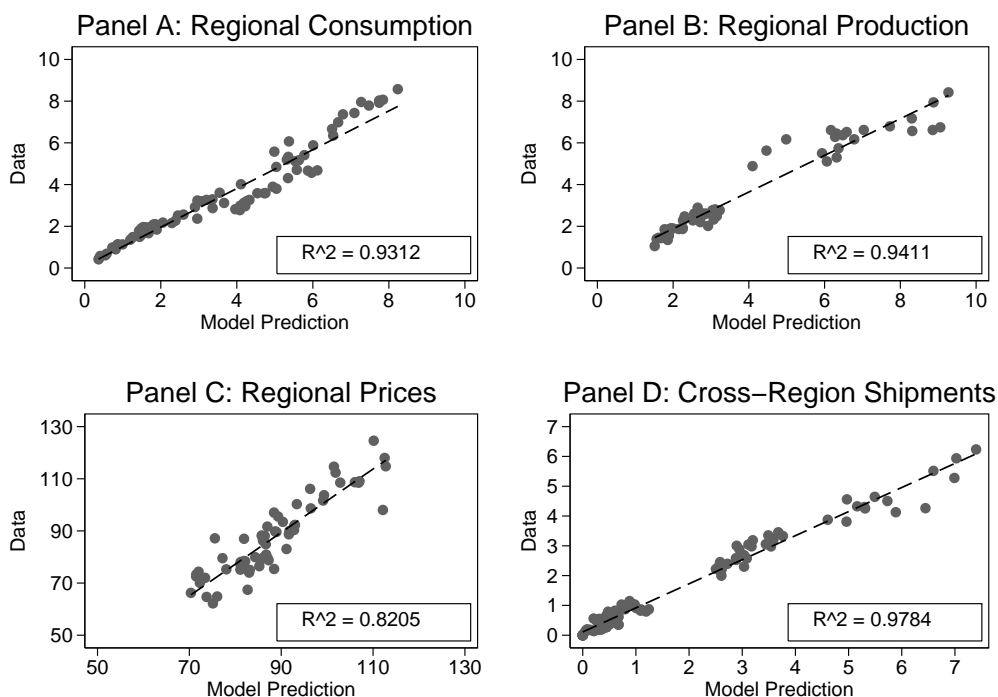
Figure 9: Estimation Fits for Regional Metrics. Consumption, production, and cross-region shipments are in millions of metric tonnes. Prices are constructed as a weighted-average of plants in the region, and are reported as dollars per metric tonne. The lines of best fit and the reported R^2 values are based on univariate OLS regressions.

93 percent of the variation in regional consumption, 94 percent of the variation in regional production, and 82 percent of the variation in regional prices. Thus, the model performs reasonably well in accounting for the variation in the endogenous data. It is also telling to examine the model's out-of-sample predictions. We also plot observations on cross-region shipments against the corresponding model predictions (panel D). We use 14 of these observations in the estimation routine – the shipments from plants in California to consumers in northern California over 1990-2003 – but the remaining 82 data points are withheld from the estimation procedure and do not influence the estimated parameters. Even so, the model explains 98 percent of the variation in these data.

6.4   Comparison to market delineation

A sizeable empirical literature uses market delineation to sidestep the complications of spatial differentiation (e.g., Pesendorfer (2003), Salvo (2008), Collard-Wexler (2009), Ryan (2010)). Such models assume the existence of precisely bounded markets; each market includes firms

that compete (typically in quantities) to supply the consumers of the market. This enables estimation with market-level data but sacrifices realism in the underlying economic model. For instance, it is well understood that market delineation imposes the mutually incompatible assumptions that (1) transportation costs are large enough to preclude competition across market boundaries, and (2) transportation costs are small enough that spatial differentiation is negligible within markets. Syverson (2004) and others discuss how this tension compels researchers to seek compromise between markets that are too broad or too narrow. Further, market delineation precludes inferences about spatial differentiation because the effects of transportation costs are assumed rather than estimated.

Our approach facilitates the estimation of more realistic models without imposing overly cumbersome data requirements. To illustrate, we compare our elasticities to those estimated in Ryan (2010). Ryan conducts a careful empirical study of the portland cement industry, based on the same data sources we employ, and delineates markets based on the USGS reporting regions. Regressions of annual market-level production on annual market-level average prices yield estimates of the aggregate demand elasticity. Ryan's preferred specification suggests an aggregate elasticity of -2.96. When controls for residential construction permits are added, however, aggregate elasticity falls to -0.15.[41] This is much closer to our estimate of -0.16. Housing permits are an important predictor of cement demand (see Section 3.3); we account for permits through the specification of county-level potential demand. Yet Ryan discards the less elastic estimate because, given Cournot competition, it implies firm elasticities that are below one in magnitude and inconsistent with profit maximization.[42] By contrast, our approach divorces firm elasticities from the aggregate elasticity and is suitable for industries with elastic firm demand and inelastic overall demand.[43]

---

[41]See Table 3 in Ryan (2010).

[42]Cournot competition links firm elasticities to the aggregate elasticity according to $e_j = e/s_j$, where e_j, e, and s_j denote firm elasticity, aggregate elasticity, and market share, respectively.

[43]We do not think this discrepancy diminishes the contribution of Ryan (2010), which estimates an innovative dynamic discrete choice game and focuses primarily on the dynamic parameters; market delineation is used simply to determine the payoffs at different realizations of the state space.

7 Counter-factual experiments

7.1 Spatial discrimination and consumer surplus

The consumer surplus implications of spatial price discrimination have long been recognized as ambiguous (e.g., Gronberg and Meyer (1982), Katz (1984), Hobbs (1986), Anderson, de Palma, and Thisse (1989)). We conduct a counter-factual policy experiment to evaluate the implications of spatial price discrimination in the portland cement industry. We solve numerically for equilibrium, given the estimated parameters and the topology of the industry in the year 2003, under the restriction that each plant must charge the same price (net of transport costs) to all consumers.[44]

Figure 10 characterizes the consumer surplus implications of disallowing spatial price discrimination, on a county-by-county basis. Counties that are shaded in dark gray or black are harmed by the ban whereas counties shaded in light gray or white are benefited. The net effect of the ban, aggregating across all counties, is a $12 million gain in consumer surplus. This can be calibrated against a volume of commerce in the U.S. Southwest of $1.3 billion.[45] However, the effects of disallowing discrimination vary widely across counties and are consistent with the generalization that the ban benefits consumers located nearby cement plants and harms more distant consumers. Since nearby consumers tend to be infra-marginal whereas distant consumers tend to be marginal, this follows the economics of the model – price discrimination enables plants to extract surplus accruing to inframarginal consumers without sacrificing sales to marginal consumers.

The heterogeneous effects of spatial price discrimination are starkest in Maricopa County and the two counties immediately to the north and south (Yavapai County and Pima County, respectively). The predominate domestic suppliers of cement in these counties are the Phoenix Cement plant in Clarkdale, Arizona and the California Cement plant in Rillito, CA. Table 4 shows the mill prices set by these plants in the discriminatory regime ("Pre-Price") and the non-discriminatory regime ("Post-Price"). The price discrimination ban leads the closest supplier to reduce prices in Yavapai and Pima. Thus, the mill price of the Phoenix Cement plant to consumers in Yavapai falls from $100 per metric tonne to $83, and the mill price of the California Cement plant to consumers in Pima falls from $88 to $85. Due to these price effects, disallowing price discrimination creates $2.2 million and $1.7 million in consumer surplus in these counties, respectively. By contrast, the prices that

---

[44] Although we focus on 2003, the results obtained from other years are similar.

[45] Volume of commerce is calculated as price times quantity for all sales by plants in the U.S. Southwest.
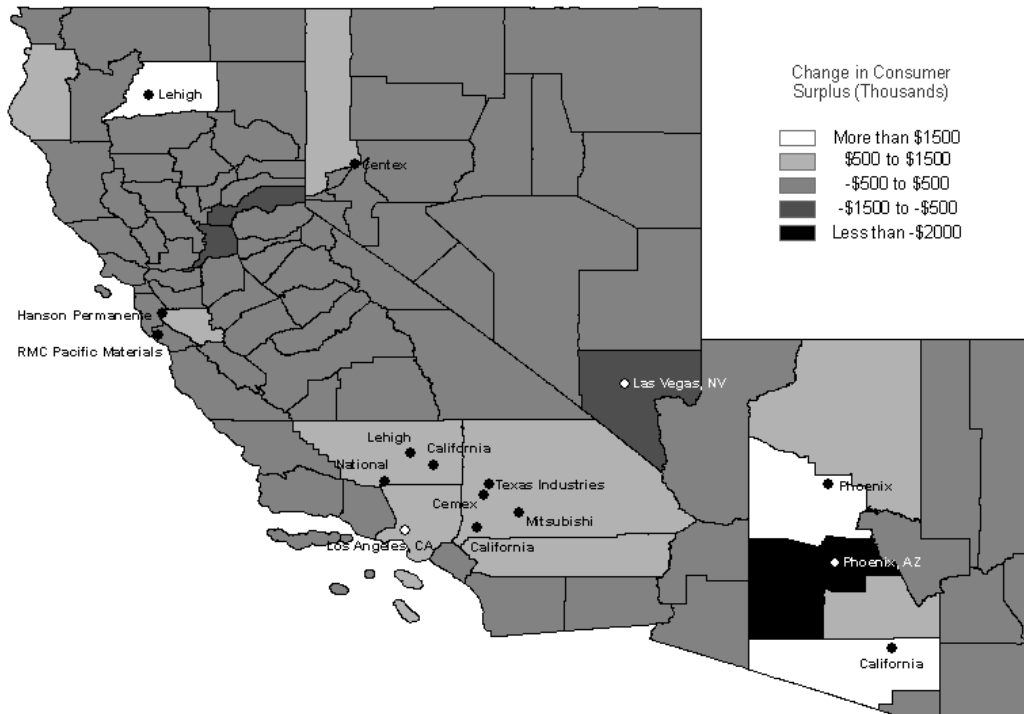
Figure 10: Effects of Disallowing Price Discrimination on Consumer Surplus

these plants charge to the consumers in Maricopa (who tend to be more marginal) increase due to the price discrimination ban, and these price increases leads to $2.3 million in lost consumer surplus.

7.2 Merger simulation

Antitrust authorities routinely support merger investigations with coarse or incomplete data due to tight statutory deadlines. For instance, the full complement of firm-level data needed to estimate the models of Thomadsen (2005), Davis (2006), McManus (2007) and Houde (2009) is rarely available. In these cases, the flexible data requirements of our estimator are particularly valuable. To illustrate, we use counter-factual simulations to evaluate a hypothetical merger between Calmat and Gifford-Hill in 1986.[46] Together, these two firms operated four of the eight plants in southern California and both of the plants in Arizona.

---

[46]We follow standard practice to perform the counterfactuals. We use the equations of McFadden (1981) and Small and Rosen (1981) to calculate consumer surplus.

Table 4: Effects of Disallowing Price Discrimination on Prices in Selected Counties

Plant Owner	Plant Location	Distance	Trans. Cost	Pre-Price	Post-Price
Maricopa County (Phoenix)					
Phoenix Cement	Clarkdale, AZ	101	$31	$80	$83
California Cement	Rillito, AZ	104	$31	$81	$85
Yavapai County (Clarkdale)					
Phoenix Cement	Clarkdale, AZ	29	$9	$100	$83
California Cement	Rillito, AZ	174	$53	$74	$85
Pima County (Rillito)					
Phoenix Cement	Clarkdale, AZ	186	$56	$71	$83
California Cement	Rillito, AZ	44	$13	$88	$85

Results of the counter-factual experiment. Distance is the miles between the plant and the county centroid. Transportation Cost is per metric tonne. Pre-Price is the mill price in the discriminatory regime and Post-Price is the mill price in the non-discriminatory regime; both are per metric tonne.

The simulation results suggest the merger leads to prices at the Calmat and Gifford-Hill plants that are three percent higher in southern California and five percent higher in Arizona, on average. This induces consumer switching; and consumers that do switch split evenly between other domestic plants (48 percent) and foreign importers (52 percent). Prices at other domestic plants increases by only 0.5 percent. Overall, consumer surplus falls by more than $20 million, relative to a total volume of commerce in southern California and Arizona of $801 million.[47]

Absent our estimation strategy, aggregate data could still support merger simulation provided one imposes market delineation assumptions. This can yield quite different merger predictions, however. We calculate the percentage price increases one would predict in southern California and Arizona under the assumptions that these two regions are independent markets, competition is Cournot, demand has constant elasticity, plants share a constant marginal cost, and there are no foreign imports. Aside from the marginal cost assumption, this mimics the modeling framework of Ryan (2010). Post-merger prices are

$$p^{post} = \frac{N(N-1)e - (N-1)}{N(N-1)e - N} p^{pre},$$ (11)

where N is the number of firms and e is the aggregate elasticity of demand.[48] The merger

---

[47]Volume of commerce is calculated as price times quantity for all sales by plants in southern California and Arizona.

[48]In obtaining this expression it is useful to keep in mind the relationship between firm elasticities and

has the effect of reducing the number of firms from N to $N-1$; six firms operate plants in southern California during 1986, and two firms operated plants in Arizona. We calibrate using the Ryan (2010) aggregate elasticity estimate of 2.96. This yields price increases of one percent in southern California and 25 percent in Arizona. Thus, the application of market delineation assumptions would appear to well understate harm in southern California and well overstate harm in Arizona.

Another advantage of our estimation strategy is that it better informs divestiture negotiations.[49] Since the challenge faced by antitrust authorities is to identify the plant best located to provide competition post-merger, a serious treatment of spatial differentiation is central. Figure 11 maps the distribution of consumer harm that arises from the hypothetical merger of Calmat and Gifford Hill. Panel A focuses on effects of the merger absent any divestitures. Harm is concentrated in the counties surrounding Los Angeles and Phoenix. Panel B plots harm under the most powerful single-plant divestiture, that of Gifford Hill's Oro Grande plant ("Gifford-Hill 2" in the figure). This divestiture eliminates 55% of total harm. Panel B shows this relief occurs mainly in southern California; the divestiture does little to reduce harm in Arizona. Additional counterfactual exercises indicate that another divestiture is needed to mitigate this harm as well.

8   Conclusion

The literature of the "new empirical industrial organization" focuses largely on the structural estimation of competition models and the recovery of the underlying parameters that guide firm and consumer decisions. Econometric innovations and greater computer power have improved our ability to link empirical correlations with sensible theoretical models of behavior. One area of particular interest has been the estimation of product differentiation models, as in Berry, Levinsohn, and Pakes (1995) and Nevo (2001). Yet geographic considerations – often critical drivers of differentiation – have received relatively little attention. In this paper, we have developed an estimator for models of competition among spatially differentiated firms that has flexible data requirements and is implementable with data at

the aggregate elasticity, i.e., that $e_j = Ne$ where e_j is the firm elasticity. Then manipulation of the Lerner index yields a familiar expression for post-merger prices:

$$p^{post} = \frac{(N-1)e}{(N-1)e-1}c \quad \text{where} \quad c = \frac{Ne-1}{Ne}p^{pre}.$$

[49]Models based on standard market delineation assumptions are less informative about divestures in this context because spatial considerations are ignored – within a market, all plants are assumed to be equal.
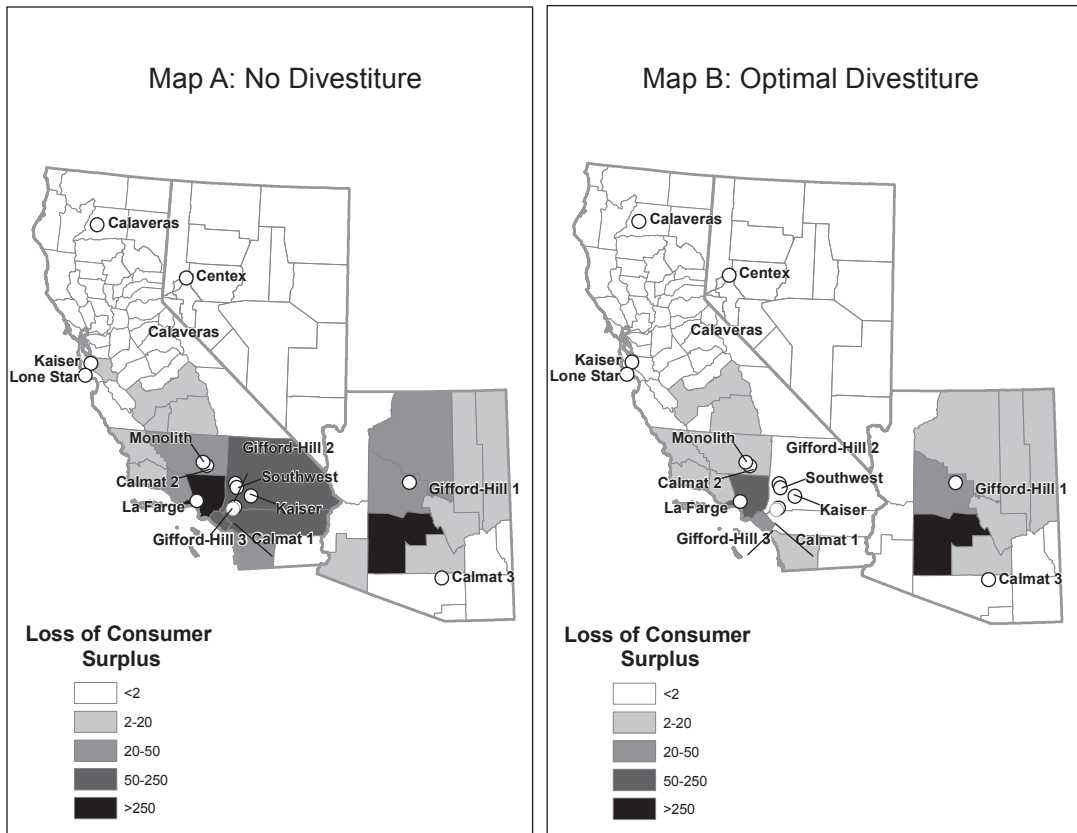
Figure 11: Loss of Consumer Surplus Due to a Hypothetical Merger

any level of aggregation. Further, the estimator is the first to be applicable to models in which firms price discriminate among consumers based on location.

Our hope is that the estimator extends the reach of empirical researchers. Our application to the portland cement industry provides an example. In a counter-factual policy experiment, we determine that disallowing spatial price discrimination would increase consumer surplus by a modest $12 million, relative to a volume of commerce of $1.3 billion. Heretofore, it has not been possible examine the surplus implications of spatial price discrimination in specific, real-world settings; these implications have been known to be ambiguous theoretically since at least Gronberg and Meyer (1982) and Katz (1984). Other applications have equal promise. Researchers could study the relationship between transportation costs and the intensity of competition or the proper construction of antitrust markets. And, though our application is static, the estimator could be used to define payoffs in strategic dynamic games. Such extensions could examine an array of interesting topics including entry deterrence, optimal location choice, and the effects of various government policies (e.g., carbon

taxes or import duties) on welfare and the long-run location of production.

Methodological extensions would have value, as well. The estimator as developed does not accommodate unobserved plant-level heterogeneity. This limitation helped motivate our selection of the portland cement industry: due to concerns regarding construction quality and reliability, the production of cement is subject to strictly enforced standards that minimize heterogeneity. However, we suspect that the estimator could be extended to accommodate unobserved heterogeneity using the simulated method of moments of McFadden (1989). The estimator could then be applied more broadly to industries in which spatial differentiation is important. We speculate that such simulation methods would prove feasible because the computation burden of estimation should increase only linearly in the number of simulation draws used to model the unobserved heterogeneity.

References

Anderson, S. P. and A. de Palma (1988). Spatial price discrimination with heterogeneous products. *Review of Economic Studies 55*, 573–592.

Anderson, S. P., A. de Palma, and J.-F. Thisse (1989). Spatial price policies reconsidered. *Journal of Industrial Economics 38*, 1–18.

Bajari, P., L. Benkard, and J. Levin (2007). Estimating dynamic models of imperfect competition. *Econometrica*, 1331–1370.

Berry, S. (1994). Estimating discrete choice models of product differentiation. *RAND Journal of Economics*, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995, July). Automobile prices in market equilibrium. *Econometrica*, 841–890.

Bisin, A., A. Moro, and G. Topa (2010). The empirical content of models with multiple equilibria in economies with social interactions. Mimeo.

Cardell, S. N. (1997). Variance components structures for the extreme value and logistic distributions with applications to models of heterogeneity. *Journal of Economic Theory 13*, 185–213.

Collard-Wexler, A. (2009). Productivity dispersion and plant selection in the ready-mix concrete industry. *Mimeo*.

Davis, P. (2006). Spatial competition in retail markets: Movie theaters. *The RAND Journal of Economics 37*, 964–982.

Dunne, T., S. Klimek, and J. A. Schmitz (2009). Does foreign competition spur productivity? Evidence from post-WWII U.S. cement manufacturing. *Federal Reserve Bank of Minneapolis Staff Report*.

EPA (2009). *Regulatory Impact Analysis: National Emission Standards for Hazardous Air Pollutants from the Portland Cement Manufacturing Industry*. Prepared by RTI International.

Greene, W. H. (2003). *Econometric Analysis 5th Ed.* Prentice-Hall Upper Saddle River NJ.

Greenhut, J., M. Greenhut, and S.-Y. Li (1980). Spatial pricing patterns in the united states. *Quarterly Journal of Economics 94*, 329–350.

Gronberg, T. J. and J. Meyer (1982). Spatial pricing, spatial rents, and spatial welfare. *Quarterly Journal of Economics 97*, 633–644.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

Hobbs, B. F. (1986). Mill pricing versus spatial price discrimination under bertrand and cournot spatial competition. *Journal of Industrial Organization 35*, 173–191.

Hotelling, H. (1929). Stability in competition. *Economic Journal 39*, 41–57.

Houde, J.-F. (2009). Spatial differentiation and vertical contracts in retail markets for gasoline. Mimeo.

Hwang, H.-s. (1990). Estimation of a linear SUR model with unequal numbers of observations. *Review of Economics and Statistics 72*.

Katz, M. J. (1984). Price discrimination and monopolistic competition. *Econometrica 52*, 1453–1471.

Konovalov, A. and Z. Sándor (2010). On price equilibrium with multi-product firms. *Economic Theory 44*, 271–292.

La Cruz, W., J. Martínez, and M. Raydan (2006). Spectral residual method without gradient information for solving large-scale nonlinear systems of equations. *Mathematics of Computation 75*, 1429–1448.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Mathematics 2*, 164–168.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics 11*, 431–331.

McFadden, D. (1981). *Econometric Models of Probabilistic Choice*. MIT Press. Structural Analysis of Discrete Data, C.F. Manski and D. McFadden (eds.).

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*, 995–1026.

McManus, B. (2007). Nonlinear pricing in an oligopoly market: The case of specialty coffee. *RAND Journal of Economics 38*, 512–532.

Mizuno, T. (2003). On the existence of a unique price equilibrium for modles of product differentiation. *International Journal of Industrial Organization 21*(6), 761–793.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*, 307–342.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*.

Northrup, H. R. (1989). From union hegemony to union disintegration: Collective bargaining in cement and related industries. *Journal of Labor Research 10*, 337–376.

Pesendorfer, M. (2003). Horizontal mergers in the paper industry. *RAND Journal of Economics 34*, 495–515.

Pinske, J., M. E. Slade, and C. Brett (2002). Spatial price competition: A semiparametric approach. *Econometrica 70*(3), 1111–1153.

Rock-Products (1994). Cement plant operating cost study. Overland Park, KS: Intertic Publishing.

Ruud, P. A. (2000). *An Introduction to Classical Econometric Theory*. Oxford University Press.

Ryan, S. (2010). The costs of environmental regulation in a concentrated industry. Mimeo.

Salop, S. (1979). Monopolistic competition with outside goods. *Bell Journal of Economics 10*, 141–156.

Salvo, A. (2008). Inferring market power under the threat of entry: The case of the Brazilian cement industry. *Mimeo*.

Salvo, A. (2010). Trade flows in a spatial oligolopy: Gravity fits well, but what does it explain? *Canadian Journal of Economics 43*, 63–96.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics 37*, 619–640.

Small, K. and H. Rosen (1981). Applied welfare economics with discrete choice methods. *Econometrica 49*, 105–130.

Srivastava, J. and M. K. Zaatar (1973). A Monte Carlo comparison of four estimators of the dispersion matrix of a bivariate normal population, using incomplete data. *Journal of the American Statistical Association 68*, 180–183.

Syverson, C. (2004, December). Market structure and productivity: A concrete example. *Journal of Political Economy 112*, 1181–1222.

Thomadsen, R. (2005). The effect of ownership structure on prices in geographically differentiated industries. *RAND Journal of Economics 36*, 908–929.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Van Oss, H. G. and A. Padovani (2002). Cement manufacture and the environment, Part I: Chemistry and technology. *Journal of Industrial Ecology 6*, 89–105.

Vogel, J. (2008, June). Spatial competition with heterogeneous firms. *Journal of Political Economy 116*(3), 423–466.

A   Proofs

Proof of Lemma 1: We demonstrate that the implicit solution $\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X})$ is continuous in $\boldsymbol{\theta}$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, taking as given the existence and uniqueness of equilibrium. The proof is by contradiction. We note that $\boldsymbol{f}$ is continuous for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{p}$ in $\mathbb{R}^{JN}$. We suppress $\boldsymbol{X}$ for notational simplicity and note that the arguments we apply to $\boldsymbol{\theta}$ apply to $\boldsymbol{X}$ as well. To start, suppose by way of contradiction that $\boldsymbol{p}^*(\boldsymbol{\theta})$ is not continuous at some point $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}$. Then there exists an $\epsilon > 0$ such that for all $\delta > 0$ there exists a $\boldsymbol{\theta}_2$ such that

$$0 < \|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1\| < \delta$$

and

$$\|\boldsymbol{p}^*(\boldsymbol{\theta}_2) - \boldsymbol{p}^*(\boldsymbol{\theta}_1)\| \geq \epsilon.$$

Uniqueness of the equilibrium price $\boldsymbol{p}^*$ implies that if $\|\boldsymbol{p}^*(\boldsymbol{\theta}_2) - \boldsymbol{p}^*(\boldsymbol{\theta}_1)\| \geq \epsilon > 0$, then

$$\|\boldsymbol{f}(\boldsymbol{p}^*(\boldsymbol{\theta}_2), \boldsymbol{\theta}_1)\| > b > 0.^{[50]}$$

Continuity of $\boldsymbol{f}$ in $\boldsymbol{\theta}$ implies that for all $\tilde{\epsilon}$ there exists a $\tilde{\delta} > 0$ such that if

$$0 < \|\boldsymbol{\theta} - \boldsymbol{\theta}_1\| < \tilde{\delta}$$

then

$$\|\boldsymbol{f}(\boldsymbol{p}^*(\boldsymbol{\theta}), \boldsymbol{\theta}) - \boldsymbol{f}(\boldsymbol{p}^*(\boldsymbol{\theta}), \boldsymbol{\theta}_1)\| = \|\boldsymbol{f}(\boldsymbol{p}^*(\boldsymbol{\theta}), \boldsymbol{\theta}_1)\| < \tilde{\epsilon}.$$

A contradiction immediately follows from this if we choose $\tilde{\epsilon} = b$. Our initial assertion would imply that for $\tilde{\delta}(b)$ we could find a $\boldsymbol{\theta}_2(\tilde{\delta}(b))$ where

$$0 < \|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1\| < \tilde{\delta}$$

and

$$\|\boldsymbol{f}(\boldsymbol{p}^*(\boldsymbol{\theta}_2), \boldsymbol{\theta}_1)\| \geq b = \tilde{\epsilon}.$$

$\square$

Proof of Lemma 2: We prove that under Assumption A3, $S(\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X}), \boldsymbol{\theta}_0, \boldsymbol{X})$ is differentiable at $\boldsymbol{\theta}_0$ for almost all $\boldsymbol{X}$. The modifier "almost all" is understood to mean that

---

[50]This is because $\|\boldsymbol{p}^*(\boldsymbol{\theta}) - \boldsymbol{p}^*(\boldsymbol{\theta}_1)\| > \epsilon$ implies that $\boldsymbol{p}^*(\boldsymbol{\theta}) \neq \boldsymbol{p}^*(\boldsymbol{\theta}_1)$. Our definition of $\boldsymbol{p}^*$ and the assumption of a unique equilibrium implies $\boldsymbol{f}(\boldsymbol{p}, \boldsymbol{\theta}_1) = \boldsymbol{0}$ at $\boldsymbol{p}^*(\boldsymbol{\theta}_1)$, and no other price.

the set of $\boldsymbol{X}$ points for which differentiability fails occurs with measure zero, where the measure being used is the probability measure generated by the probability distribution of $\boldsymbol{X}$. First, the function $S(\boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{X})$ is continuously differentiable in its arguments by assumption. Thus it remains to show that the equilibrium price function $\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ is almost everywhere differentiable in $\boldsymbol{\theta}_0$. Assumption A3 guarantees that for every $\boldsymbol{X}_0$ in $B(\boldsymbol{\theta}_0)$, there is an $\boldsymbol{X}$-neighborhood around $\boldsymbol{X}_0$ where the Jacobian of $\boldsymbol{f}$ with respect to $\boldsymbol{p}$ is nonsingular. The Implicit Function Theorem guarantees that $\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ is continuously differentiable for the $\boldsymbol{X}$ points in this neighborhood. Because each point of possible nondifferentiability $\boldsymbol{X}_0$ is surrounded by an open neighborhood of differentiable points, and at least one element of $\boldsymbol{X}_0$ has a continuous distribution, under the probability measure for $\boldsymbol{X}$ points of nondifferentiability occur with measure zero.

$\square$

Proof of Lemma 3: Here we prove the Lipschitz condition on the objective function $m(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{X})$ under Assumptions A2-A4. Recall that we want to prove that there exists a measurable function $\dot{m}(\boldsymbol{y}, \boldsymbol{X})$ such that

$$|m(\boldsymbol{\theta}_1, \boldsymbol{y}, \boldsymbol{X}) - m(\boldsymbol{\theta}_2, \boldsymbol{y}, \boldsymbol{X})| \leq \dot{m}(\boldsymbol{y}, \boldsymbol{X})\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$$

for every $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in some open neighborhood of $\boldsymbol{\theta}_0$.

First, consider the points $(\boldsymbol{\theta}_0, \boldsymbol{X})$ at which the Jacobian of $\boldsymbol{f}$ with respect to $\boldsymbol{p}$ is nonsingular. At these points, the Implicit Function Theorem guarantees that the implicit solution $\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ is continuously differentiable in a $\boldsymbol{\theta}$-neighborhood around $\boldsymbol{\theta}_0$ because $\boldsymbol{f}$ is continuously differentiable in $\boldsymbol{\theta}$. It follows that the partial derivatives of $\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X})$ with respect to $\boldsymbol{\theta}$ exist in this neighborhood, and Assumption A4 guarantees that the partial derivatives are bounded by $M(\boldsymbol{X})$. Now we turn to $m(\boldsymbol{\theta}, \boldsymbol{y}_t, \boldsymbol{X}_t)$. Since the matrix $\boldsymbol{W}_t$ has a finite limit, each of its elements $w_{ij,t}$ can be bounded by $\overline{w}_{ij}$.[51] We can write

$$
\begin{aligned}
m(\boldsymbol{\theta}, \boldsymbol{y}_t, \boldsymbol{X}_t) &= \sum_{i,j} m_{ij,t}(\boldsymbol{\theta}, \boldsymbol{y}_t, \boldsymbol{X}_t) \\
&= \sum_{i,j} (y_{it} - S_i(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t)) w_{ij,t}(y_{jt} - S_j(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t)).
\end{aligned}
$$

---

[51]By the definition of the limit, for all ϵ there is some T for which $|\lim_{t \to \infty} w_{ij,t} - w_{ij,t}| < \epsilon$ for $t > T$. So for all $t > T$, $w_{ij,t}$ can be bounded. $\max_{t \leq T}\{w_{ij,t}\}$ must also exist and be finite, since there are finitely many $w_{ij,t}$'s prior to T. We have implicitly assumed that all the elements of $\boldsymbol{W}_t$ are finite; violations would make numerical maximization of the objective function impossible for some values of t.

Consider the partial derivative of $m_{ij,t}(\boldsymbol{\theta}, \boldsymbol{y}_t, \boldsymbol{X}_t)$ with respect to some θ_k. We know that

$$
\begin{aligned}
\frac{\partial m_{ij,t}(\boldsymbol{\theta}, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k} = {} & -w_{ij,t}(y_{jt} - S_j(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t)) \cdot \\
& \left[\sum_{n,l} \frac{\partial S_i(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t)}{\partial p_{nl}} \frac{\partial p_{nl}^*(\boldsymbol{\theta}, \boldsymbol{X}_t)}{\partial \theta_k} + \frac{\partial S_i(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t)}{\partial \theta_k} \right] \\
& -w_{ij,t}(y_{it} - S_i(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t)) \cdot \\
& \left[\sum_{n,l} \frac{\partial S_j(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t)}{\partial p_{nl}} \frac{\partial p_{nl}^*(\boldsymbol{\theta}, \boldsymbol{X}_t)}{\partial \theta_k} + \frac{\partial S_j(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t)}{\partial \theta_k} \right].
\end{aligned}
$$

Our assumption that S is continuously differentiable in its arguments means that there is some $\boldsymbol{\theta}$ neighborhood around $\boldsymbol{\theta}_0$ where $S(\boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{X}_t)$ and the partial derivatives of $S(\boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{X}_t)$ with respect to the elements of $\boldsymbol{\theta}$ are bounded (this follows from the definition of continuity). Moreover, because $\boldsymbol{p}^*$ is continuous in its arguments, it is also bounded in some neighborhood of $\boldsymbol{\theta}_0$. This means that $S(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}_t), \boldsymbol{\theta}, \boldsymbol{X}_t)$ and its partial derivatives with respect to both $\boldsymbol{\theta}$ and $\boldsymbol{p}$ can be bounded in a neighborhood of $\boldsymbol{\theta}_0$. We denote the lower bound on S as $\underline{S}$ and the upper bound on the partial derivatives as $\overline{S}'$. Recalling that Assumption A4 guarantees that all the partial derivatives of $\boldsymbol{p}^*$ with respect to θ_k are bounded by $|M(\boldsymbol{X}_t)|$, through repeated applications of the triangle inequality we can put a bound on $\frac{\partial m_{ij,t}(\boldsymbol{\theta}, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k}$:

$$
\begin{aligned}
\left| \frac{\partial m_{ij,t}(\boldsymbol{\theta}, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k} \right| & \leq \left| \overline{w}_{ij} \overline{S}' \right| \left(\sum_{n,l} |M(\boldsymbol{X}_t)| + 1 \right) (|y_{it} - \underline{S}| + |y_{jt} - \underline{S}|) \\
& = \dot{m}_{ij}(\boldsymbol{y}_t, \boldsymbol{X}_t).
\end{aligned}
$$

Recalling that $\boldsymbol{\theta}$ is L dimensional, we can write:

$$
\begin{aligned}
m(\boldsymbol{\theta}_1, \boldsymbol{y}_t, \boldsymbol{X}_t) - m(\boldsymbol{\theta}_2, \boldsymbol{y}_t, \boldsymbol{X}_t) & = \sum_{k=1}^{L} m(\theta_{11}, ..., \theta_{1k}, \theta_{2,k+1}, ..., \tilde{\theta}_{2L}, \boldsymbol{y}_t, \boldsymbol{X}_t) \\
& \quad - m(\theta_{11}, ...\theta_{1,k-1}, \theta_{2k}, ..., \theta_{2L}, \boldsymbol{y}_t, \boldsymbol{X}_t) \\
& = \sum_{k=1}^{L} \frac{\partial m(\tilde{\theta}_k, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k} (\theta_{1k} - \theta_{2k}).
\end{aligned}
$$

The second step follows from the Mean Value Theorem for $\tilde{\theta}_k = (\theta_{11}, ... \theta_{1,k-1}, \gamma, \theta_{2,k+1}, ..., \theta_{2L})$, where γ is between θ_{1k} and θ_{2k}. It follows that:

$$
\begin{aligned}
|m(\boldsymbol{\theta}_1, \boldsymbol{y}_t, \boldsymbol{X}_t) - m(\boldsymbol{\theta}_2, \boldsymbol{y}_t, \boldsymbol{X}_t)| &\leq \sum_{k=1}^{L} \left| \frac{\partial m(\tilde{\theta}_k, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k} \right| |\theta_{1k} - \theta_{2k}| \\
&\leq \sum_{k=1}^{L} \left| \frac{\partial m(\tilde{\theta}_k, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k} \right| \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \\
&\leq L \max_{i,j} \{\dot{m}_{ij}(\boldsymbol{y}_t, \boldsymbol{X}_t)\} \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.
\end{aligned}
$$

Hence, $\dot{m}(\boldsymbol{y}_t, \boldsymbol{X}_t) = L \max\{\dot{m}_{ij}(\boldsymbol{y}_t, \boldsymbol{X}_t)\}$ and Lemma 3 holds for the points $(\boldsymbol{\theta}_0, \boldsymbol{X})$ at which the Jacobian of $\boldsymbol{f}$ with respect to $\boldsymbol{p}$ is nonsingular.

Second, we prove the lemma at points of nondifferentiability, i.e., points $(\boldsymbol{\theta}_0, \boldsymbol{X})$ at which the Jacobian of $\boldsymbol{f}$ with respect to $\boldsymbol{p}$ is singular. We first consider Case (ii) of Assumption A4 and then return to Case (i). For any $\boldsymbol{X} \neq \boldsymbol{X}_0$, we can argue that

$$
\left| \frac{\partial m(\boldsymbol{\theta}, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k} \right| \leq \left| \overline{w}_{ij} \overline{\overline{S}}' \right| (NJ |M| + 1) \left(|y_{it} - \underline{S}| + |y_{jt} - \underline{S}| \right).
$$

This follows from arguments similar to those presented above. Assumption A4 (Case (ii)) guarantees that the partial derivatives of $\boldsymbol{p}^*$ are bounded by a constant M. Additionally, since $S(\cdot)$ is continuously differentiable, and since $\boldsymbol{p}^*$ is continuous in our $\boldsymbol{X}$-neighborhood of $\boldsymbol{X}_0$, $S(\cdot)$ and its derivative are bounded by $\underline{S}$ and $\overline{\overline{S}}'$, respectively.[52] This implies that the upper bound $\{\dot{m}_{ij}(\boldsymbol{y}_t, \boldsymbol{X}_t)\}$ is not a function of $\boldsymbol{X}$. It follows that:

$$
|m(\boldsymbol{\theta}_1, \boldsymbol{y}_t, \boldsymbol{X}_t) - m(\boldsymbol{\theta}_2, \boldsymbol{y}_t, \boldsymbol{X}_t)| \leq L \max_{i,j} \{\dot{m}_{ij}(\boldsymbol{y}_t)\} \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.
$$

Taking limits of both sides of this inequality, we see that

$$
\begin{aligned}
\lim_{\boldsymbol{X} \to \boldsymbol{X}_0} |m(\boldsymbol{\theta}_1, \boldsymbol{y}_t, \boldsymbol{X}_t) - m(\boldsymbol{\theta}_2, \boldsymbol{y}_t, \boldsymbol{X}_t)| &= |m(\boldsymbol{\theta}_1, \boldsymbol{y}_t, \boldsymbol{X}_0) - m(\boldsymbol{\theta}_2, \boldsymbol{y}_t, \boldsymbol{X}_0)| \\
&\leq L \max_{i,j} \{\dot{m}_{ij}(\boldsymbol{y}_t)\} \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.
\end{aligned}
$$

The first line is due to continuity of $\boldsymbol{p}^*$ and $S(\cdot)$. The last line is where the requirement that $M(\boldsymbol{X}_t)$ not be a function of $\boldsymbol{X}_t$ binds. To build intuition, suppose that, at the points of nondifferentiability graphed in Panel B of Figure 5, the partial derivative of $\boldsymbol{p}^*$ with respect

---

[52]If the $\boldsymbol{X}$ neighborhood is large enough that they are not bounded, we can simply shrink the neighborhood until they are.

to $\boldsymbol{\theta}$ approaches infinity as $\boldsymbol{X}$ approaches $\boldsymbol{X}_0$. It is still possible for the partial derivatives to be bounded by a function of $\boldsymbol{X}$ which also approaches infinity as $\boldsymbol{X}$ approaches $\boldsymbol{X}_0$. Making that function bounded solves this problem.

To finish, we turn to Case (i) of Assumption A4. We fix $\boldsymbol{X}$ at $\boldsymbol{X}_0$, and again consider applying the mean value theorem to each component of

$$ m(\boldsymbol{\theta}_1, \boldsymbol{y}_t, \boldsymbol{X}_t) - m(\boldsymbol{\theta}_2, \boldsymbol{y}_t, \boldsymbol{X}_t). $$

Consider some component

$$ m(\theta_{11}, ..., \theta_{1k}, \theta_{2,k+1}, ..., \tilde{\theta}_{2L}, \boldsymbol{y}_t, \boldsymbol{X}_t) - m(\theta_{11}, ... \theta_{1,k-1}, \theta_{2k}, ..., \theta_{2L}, \boldsymbol{y}_t, \boldsymbol{X}_t). $$

There are two possibilities to consider. First, suppose that the vector $(\theta_{11}, ... \theta_{1,k-1})$ is different from $(\theta_{01}, ... \theta_{0,k-1})$ in at least one element, or $(\theta_{2,k+1}, ..., \theta_{2L})$ is different from $(\theta_{0,k+1}, ..., \theta_{0L})$ in at least one element. In this case, the vector $\tilde{\theta}_k = (\theta_{11}, ... \theta_{1,k-1}, \gamma, \theta_{2,k+1}, ..., \theta_{2L})$ can never be equal to θ_0. Assumption A4 guarantees that the partial derivatives of $\boldsymbol{p}^*$, and hence m, exist for all possible $\tilde{\theta}_k$ so we can apply the single variable Mean Value Theorem as above. The second possibility is that $(\theta_{11}, ... \theta_{1,k-1})$ equals $(\theta_{01}, ... \theta_{0,k-1})$ and $(\theta_{2,k+1}, ..., \theta_{2L})$ equals $(\theta_{0,k+1}, ..., \theta_{0L})$. If $\theta_{1k} = \theta_{2k} = \theta_{0k}$ then the difference above is simply zero. If not, we can prove the following inequality:

$$ \frac{|m(\theta_{11}, ..., \theta_{1k}, \theta_{2,k+1}, ..., \tilde{\theta}_{2L}, \boldsymbol{y}_t, \boldsymbol{X}_t) - m(\theta_{11}, ... \theta_{1,k-1}, \theta_{2k}, ..., \theta_{2L}, \boldsymbol{y}_t, \boldsymbol{X}_t)|}{|\theta_{1k} - \theta_{2k}|} \leq $$
$$ \max \left\{ \left| \frac{\partial m(\tilde{\theta}_{1k}, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k} \right|, \left| \frac{\partial m(\tilde{\theta}_{2k}, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k} \right| \right\}. $$

To prove this, define $g(\gamma) = m(\theta_{11}, ..., \gamma, \theta_{2,k+1}, ..., \tilde{\theta}_{2L}, \boldsymbol{y}_t, \boldsymbol{X}_t)$. Assuming without loss of generality that $\theta_{1k} < \theta_{2k}$, from A4 we know that $g(\gamma)$ is differentiable on the open intervals $(\theta_{1k}, \theta_{0k})$ and $(\theta_{0k}, \theta_{2k})$ and it is continuous on the interval $[\theta_{1k}, \theta_{2k}]$ due to continuity of S and $\boldsymbol{p}^*$. Hence we can apply the Mean Value Theorem on the interval $(\theta_{1k}, \theta_{0k})$ and $(\theta_{0k}, \theta_{2k})$ to show that

$$ \frac{|g(\theta_{0k}) - g(\theta_{1k})|}{|\theta_{0k} - \theta_{1k}|} \leq \left| \frac{\partial m(\tilde{\theta}_{1k}, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k} \right|, \text{ and } \frac{|g(\theta_{2k}) - g(\theta_{0k})|}{|\theta_{2k} - \theta_{0k}|} \leq \left| \frac{\partial m(\tilde{\theta}_{2k}, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k} \right|, $$

for some $\tilde{\theta_{1k}} \in (\theta_{1k}, \theta_{0k})$ and $\tilde{\theta_{2k}} \in (\theta_{0k}, \theta_{2k})$. We next show that

$$ \frac{|g(\theta_{2k}) - g(\theta_{1k})|}{|\theta_{2k} - \theta_{1k}|} \leq \max \left\{ \frac{|g(\theta_{0k}) - g(\theta_{1k})|}{|\theta_{0k} - \theta_{1k}|}, \frac{|g(\theta_{2k}) - g(\theta_{0k})|}{|\theta_{2k} - \theta_{0k}|} \right\}. $$

To show this inequality, we first make the following definitions:

$$m_1 = \frac{g(\theta_{2k}) - g(\theta_{1k})}{\theta_{2k} - \theta_{1k}}$$

$$m_2 = \frac{g(\theta_{0k}) - g(\theta_{1k})}{\theta_{0k} - \theta_{1k}}$$

$$m_3 = \frac{g(\theta_{2k}) - g(\theta_{0k})}{\theta_{2k} - \theta_{0k}}.$$

Then define three lines on the interval $[\theta_{1k}, \theta_{2k}]$:

$$L_1(\theta) = m_1\theta + b_1$$
$$L_2(\theta) = m_2\theta + b_2$$
$$L_3(\theta) = m_3\theta + b_3,$$

where we define

$$b_1 = g(\theta_1) - m_1\theta_1$$
$$b_2 = g(\theta_1) - m_2\theta_1$$
$$b_3 = g(\theta_2) - m_3\theta_2.$$

Because of the way we have defined these lines, and because of the continuity of g, it must be the case that $L_2(\theta_0) = L_3(\theta_0)$, $L_1(\theta_1) = L_2(\theta_1)$, and $L_1(\theta_2) = L_3(\theta_2)$. Let us suppose by way of contradiction that

$$|m_1| > \max\{|m_2|, |m_3|\}.$$

There are a number of cases that we have to consider. First, suppose that m_1, m_2, and m_3 are all positive. Then it must be the case that for $\theta > \theta_1$, $L_1(\theta) > L_2(\theta)$ since $L_1(\theta_1) = L_2(\theta_1)$ and L_1 has a steeper slope than L_2. It must also be the case that for $\theta < \theta_2$, $L_1(\theta) < L_3(\theta)$ since L_1 is more steep than L_3 and $L_1(\theta_2) = L_3(\theta_2)$. Since $\theta_1 < \theta_0 < \theta_2$, this implies that $L_3(\theta_0) > L_1(\theta_0) > L_2(\theta_0)$. This contradicts $L_2(\theta_0) = L_3(\theta_0)$. Next suppose that $m_1 > 0$, $m_2 < 0$, and $m_3 > 0$. It is easy to show that it must be the case that $L_2(\theta_0) < L_1(\theta_0)$ (because L_2 slopes down from θ_1, while L_1 slopes upward), and $L_3(\theta_0) > L_1(\theta_0)$ (by the assumption that $m_1 > m_3$), again leading to a contradiction. Then suppose that $m_1 > 0$, $m_2 > 0$ and $m_3 < 0$. The assumption that $m_1 > m_2$ implies that $L_2(\theta_0) < L_1(\theta_0)$. Since we assumed that m_3 is negative, L_3 slopes up from $\theta < \theta_2$ and L_1 slopes down, implying that $L_3(\theta_0) > L_1(\theta_0)$. This again is a contradiction of $L_2(\theta_0) = L_3(\theta_0)$. The cases where $m_1 < 0$

can be shown with similar logic. The fact that $|m_1| \leq \max\{|m_2|, |m_3|\}$ implies that

$$\frac{|g(\theta_{1k}) - g(\theta_{2k})|}{|\theta_{1k} - \theta_{2k}|} \leq \max\left\{\left|\frac{\partial m(\tilde{\theta}_{1k}, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k}\right|, \left|\frac{\partial m(\tilde{\theta}_{2k}, \boldsymbol{y}_t, \boldsymbol{X}_t)}{\partial \theta_k}\right|\right\}.$$

From this point on, similar logic to what was used to prove the last two cases can be used to show

$$|m(\boldsymbol{\theta}_1, \boldsymbol{y}_t, \boldsymbol{X}_t) - m(\boldsymbol{\theta}_2, \boldsymbol{y}_t, \boldsymbol{X}_t)| \leq L \max_{i,j}\{\dot{m}_{ij}(\boldsymbol{y}_t, \boldsymbol{X}_t)\}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.$$

$\square$

Proof of Proposition 1: With Lemmas 1-3 in hand, the proof of Proposition 1 follows directly from Theorem 5.23 in van der Vaart (1998), pages. 53-54. Two additional normalcy conditions are required:

(i) $E_{\boldsymbol{\omega}} E_{\boldsymbol{X}} \dot{m}(\boldsymbol{y}, \boldsymbol{X})^2 < \infty$.

(ii) The mapping

$$\begin{aligned}\boldsymbol{\theta} \to Pm(\boldsymbol{\theta}) &= \int_U \left[(S(\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X}), \boldsymbol{\theta}_0, \boldsymbol{X}) - S(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}), \boldsymbol{\theta}, \boldsymbol{X}))' \cdot \right. \\ &\quad \left. \boldsymbol{W}\left(S(\boldsymbol{p}^*(\boldsymbol{\theta}_0, \boldsymbol{X}_t), \boldsymbol{\theta}_0, \boldsymbol{X}) - S(\boldsymbol{p}^*(\boldsymbol{\theta}, \boldsymbol{X}), \boldsymbol{\theta}, \boldsymbol{X}))\right)\right] F_x(\boldsymbol{X}) + E\boldsymbol{\omega}'\boldsymbol{W}\boldsymbol{\omega}\end{aligned}$$

admits a second-order Taylor expansion at $\boldsymbol{\theta}_0$.

$\square$

Additional proof: In footnote 20 of Section 3.3, we make the claim that the mean distance between plant j and consumers in area n (denoted d_{jn}) is orthogonal to the consumer-specific deviation ($\|w_i - z_j\|_d - d_{jn}$). We prove that claim here, using a continuous version of Ruud's (2000, p.31) proof that the predicted vector of a regression is orthogonal to the residual vector. For simplicity, we define $d_i = \|w_i - z_j\|_d$, where $\|\|_d$ denotes Euclidean distance and where we have dropped the j subscript. Let d_i have a continuous density function $f_d(d_i)$. Note that the space of univariate functions of d_i under the norm

$$\int g(x)^2 f_d(x)$$

with the usual vector addition and subtraction form a Hilbert space, which we denote as $\mathbb{F}$. Therefore, by the Hilbert projection theorem, if $\mathbb{S}$ is a closed subspace of $\mathbb{F}$, then a necessary

and sufficient condition for minimizing the distance $\|y - x\|$ where $y \in \mathbb{F}$ and $x \in \mathbb{S}$ is that $y - x$ is orthogonal to x. So let's consider the identity function for y, and the subspace

$$\mathbb{S} = \{g : g(d) = \kappa_n \text{ if } d \in \mathbb{C}_n\},$$

where κ_n is some constant. We simply need to show that the solution to the minimization problem

$$\min_{g \in \mathbb{S}} E(d_i - g(d_i))^2$$

is

$$g^*(d_i) = \begin{cases} \frac{\int_{\mathbb{C}_n} df_d(d)}{\int_{\mathbb{C}_n} f_d(d)} = E(d_i | d_i \in \mathbb{C}_n) & \text{if } d_i \in \mathbb{C}_n \\ 0 & \text{otherwise.} \end{cases}$$

Note that we can write:

$$
\begin{aligned}
E(d_i - g(d_i))^2 &= E\left([d_i - g^*(d_i)] + [g^*(d_i) - g(d_i)]\right)^2 \qquad (12) \\
&= E\left([d_i - g^*(d_i)]\right)^2 + E\left([g^*(d_i) - g(d_i)]\right)^2,
\end{aligned}
$$

where the second equality follows from the fact that

$$E\left([d_i - g^*(d_i)][g^*(d_i) - g(d_i)]\right) = 0.$$

The proof of this line is

$$
\begin{aligned}
&\sum_{n=1}^{N} \int_{\mathbb{C}_n} \left([d_i - g^*(d_i)][g^*(d_i) - g(d_i)]\right) f_d(d_i) \\
&= \sum_{n=1}^{N} \int_{\mathbb{C}_n} \left(\left[d_i - \frac{\int_{\mathbb{C}_n} df_d(d)}{\int_{\mathbb{C}_n} f_d(d)}\right] \left[\frac{\int_{\mathbb{C}_n} df_d(d)}{\int_{\mathbb{C}_n} f_d(d)} - \kappa_n\right]\right) f_d(d_i) \\
&= \sum_{n=1}^{N} \left[\frac{\int_{\mathbb{C}_n} df_d(d)}{\int_{\mathbb{C}_n} f_d(d)} - \kappa_n\right] \left(\int_{\mathbb{C}_n} \left[d_i - \frac{\int_{\mathbb{C}_n} df_d(d)}{\int_{\mathbb{C}_n} f_d(d)}\right]\right) f_d(d_i) \\
&= \sum_{n=1}^{N} \left[\frac{\int_{\mathbb{C}_n} df_d(d)}{\int_{\mathbb{C}_n} f_d(d)} - \kappa_n\right] \left(\int_{\mathbb{C}_n} d_i f_d(d_i) - \int_{\mathbb{C}_n} df_d(d)\right) \\
&= 0
\end{aligned}
$$

where the second equality follows from the fact that g is in the subspace $\mathbb{S}$, and the third from the fact that the expectations and the κ_n's don't depend on d_i, so they can be factored

53

out of the integral. It is then obvious from inspection of equation (12) that $E(d_i - g(d_i))^2$ is minimized if and only if $g(d_i) = g^*(d_i)$, and by the Hilbert projection theorem it must be the case that $d_i - g^*(d_i)$ is orthogonal to $g^*(d_i)$.

$\square$

B Summary Statistics

We provide selected summary statistics in Table 5. Some patterns stand out: First, substantial variation in each metric is available, both inter-temporally and across regions, to support estimation. Second, Southern California is larger than the other regions, whether measured by consumption or production. Third, consumption exceeds production in Northern California, Arizona, and Nevada; these shortfalls must be countered by cross-region shipments and/or imports. The observation that plants in these regions charge higher prices is consistent with transportation costs providing some degree of local market power. Finally, imports are less expensive than domestically produced portland cement. This discrepancy exists for two reasons: First, imports typically come in the form of clinker, which observes water from the air more slowly than cement. The clinker is ground into cement only after it clears customs. The import price does not include the grinding cost. Second, the import price does not include tariffs and duties, which are substantial. We include the import dummy in the demand specification to adjust for these factors.

C Estimation details

We minimize the objective function using the Levenberg-Marquardt algorithm (Levenberg (1944), Marquardt (1963)), which interpolates between the Gauss-Newton algorithm and the method of gradient descent. We find that the Levenberg-Marquardt algorithm outperforms simplex methods such as simulated annealing and the Nelder-Mead algorithm, as well as quasi-Newton methods such as BFGS. We implement the minimization procedure using the nls.lm function in R, which is downloadable as part of the minpack.lm package.

We use observed prices to form the basis of the initial vector in the inner loop computations, which limits the distance that the nonlinear equation solver must walk to compute numerical equilibrium. In practice, the equation solver occasionally fails to compute a numerical equilibrium at the specified tolerance level (1e-13) within the specified maximum number of iterations (600). The candidate parameter vectors that generate non-convergence

Table 5: Consumption, Production, and Prices

Description	Mean	Std	Min	Max
Consumption				
Northern California	3,513	718	2,366	4,706
Southern California	6,464	1,324	4,016	8,574
Arizona	2,353	650	1,492	3,608
Nevada	1,289	563	416	2,206
Production				
Northern California	2,548	230	1,927	2,894
Southern California	6,316	860	4,886	8,437
Arizona-Nevada	1,669	287	1050	2,337
Domestic Prices				
Northern California	85.81	11.71	67.43	108.68
Southern California	82.81	16.39	62.21	114.64
Arizona-Nevada	92.92	14.24	75.06	124.60
Import Prices [excludes duties and grinding costs]				
U.S. Southwest	50.78	9.30	39.39	79.32

Statistics are based on observations at the region-year level over the period 1983-2003. Production and consumption are in thousands of metric tonnes. Prices are per metric tonne, in real 2000 dollars. Import prices exclude duties. The region labeled "Arizona-Nevada" incorporates information from Nevada plants only over 1983-1991.

in the inner loop tend to be less economically reasonable, and may be consistent with equilibria that are simply too distant from observed prices. When this occurs, we construct regional-level metrics based on the price vector that comes closest to satisfying our definition of numerical equilibrium.

To further speed the inner loop computations, we re-express the first-order condition of 8 such that inversion of $\mathbf{\Omega}(\boldsymbol{p}; \boldsymbol{X}, \boldsymbol{\theta})$ is avoided. The computation of equilibrium for each time period can be parallelized, which further speeds the inner loop calculations. We also note that were production characterized by constant marginal costs, then one could further ease the computational burden of the inner loop by solving for equilibrium prices in each consumer area separately.

We constrain the signs and/or magnitudes of some parameters based on our understanding of economic theory and the economics of the portland cement industry, because some parameter vectors hinder the computation of numerical equilibrium in the inner loop. For instance, a positive price coefficient would preclude the existence of Bertrand-Nash equilibrium. We use the following constraints: the price and distance coefficients (β_1 and β_2) must be negative; the coefficients on the marginal cost shifters ($\boldsymbol{\alpha}$) and the over-utilization cost (γ) must be positive; and the coefficients on the inclusive value (λ) and the utilization threshold (ν) must be between zero and one. We use nonlinear transformations to implement the constraints. As examples, we estimate the price coefficient using $\widetilde{\beta_1} = \log(-\beta_1)$ in the GMM procedure, and we estimate the inclusive value coefficient using $\widetilde{\lambda} = \log\left(\frac{\lambda}{1-\lambda}\right)$. We calculate standard errors with the delta method.

D   Data Details

We make various adjustments to the data in order to improve consistency over time and across different sources. We discuss some of these adjustments here, in an attempt to build transparency and aid replication. To start, we note that the California Letter is based on a monthly survey rather than on the annual USGS census, which creates minor discrepancies. We normalize the California Letter data prior to estimation so that total shipments equal total production in each year. The 96 cross-region data points include:

- CA to N. CA over 1990-2003
- CA to S. CA over 2000-2003
- CA to AZ over 1990-2003
- CA to NV over 2000-2003
- N. CA to N. CA over 1990-1999

- S. CA to N. CA over 1990-1999
- S. CA to S. CA over 1990-1999
- S. CA to AZ over 1990-1999
- S. CA to NV over 1990-1999
- N. CA to AZ over 1990-1999.

The (single) Arizona-Nevada region includes Nevada data only over 1983-1991. Starting in 1992, the USGS combined Nevada with Idaho, Montana and Utah to form a new reporting region. We tailor the estimator accordingly. Additionally, this region also includes information from a small plant located in New Mexico. We scale the USGS production data downward, proportional to plant capacity, to remove for the influence of this plant. Since the two plants in Arizona account for 89 percent of kiln capacity in Arizona and New Mexico in 2003, we scale production by 0.89. We do not adjust prices.

The portland cement plant in Riverside closed its kiln permanently in 1988 but continued operating its grinding mill with purchased clinker. We include the plant in the analysis over 1983-1987, and we adjust the USGS production data to remove the influence of the plant over 1988-2003 by scaling the data downward, proportional to plant grinding capacities. Since the Riverside plant accounts for 7 percent of grinding capacity in Southern California in 1988, so we scale the production data for that region by 0.93.

We exclude one plant in Riverside that produces white portland cement. White cement takes the color of dyes and is used for decorative structures. Production requires kiln temperatures that are roughly 50°C hotter than would be needed for the production of grey cement. The resulting cost differential makes white cement a poor substitute for grey cement.

The PCA reports that the California Cement Company idled one of two kilns at its Colton plant over 1992-1993 and three of four kilns at its Rillito plant over 1992-1995, and that the Calaveras Cement Company idled all kilns at the San Andreas plant following the plant's acquisition from Genstar Cement in 1986. We adjust plant capacity accordingly.

We multiply kiln capacity by 1.05 to approximate cement capacity, consistent with the industry practice of mixing clinker with a small amount of gypsum (typically 3 to 7 percent) in the grinding mills.

The data on coal and electricity prices from the Energy Information Agency are avail-

able at the state level starting in 1990. Only national-level data are available in earlier years. We impute state-level data over 1983-1989 by (1) calculating the average discrepancy between each state's price and the national price over 1990-2000, and (2) adjusting the national-level data upward or downward, in line with the relevant average discrepancy.