Growth Policy, Agglomeration and (the Lack of) Competition*

Wyatt J. Brooks† Notre Dame

Joseph P. Kaboski‡ Notre Dame and NBER

Yao Amber Li§ HKUST

This version: January 25, 2018

Abstract

Industrial clusters policies are common growth policies in both developing and advanced economies. This paper evaluates whether clusters and cluster policies limit the extent of competition, however. We develop, validate, and apply a novel approach for screening for a lack of competition among a subset of firms. When a firm behaves independently, its markup depends on its own market share, but under cooperative pricing amongst a group of firms, markups converge and depend instead on the total market share of the group. Empirically, we validate the screen using plants with common owners, and then measure the extent of competition using data from Chinese manufacturing firms. We find strong evidence for a lack of competition within a subset of industrial clusters, and we find the level of non-competitive pricing is almost three times as high in Chinese special economic zones as outside those zones.

*We are thankful for comments received David Atkin, Timo Boppart, Hugo Hopenhayn, Oleg Itskhoki, Gaurav Khanna, Daniel Xu and participants in presentations at: Bank of Portugal Conference on Monetary Economics; BREAD; CEMFI; Cornell University; Dartmouth College; Duke University; the Federal Reserve Bank of Chicago; HKUST IEMS/World Bank Conference on Urbanization, Structural Change, and Employment; the London School of Economics/University College of London; London Macro Development Conference; NEUDC; Northwestern University; Princeton University; SED; SITE; the Taipei International Conference on Growth, Trade and Development; the University of Cambridge; the World Bank; and Yale University. We are thankful to the International Growth Centre and the HKUST IEMS (Grant No. IEMS16BM02) for financial support.

†Brooks: Department of Economics, University of Notre Dame. Email: wbrooks@nd.edu.

‡Kaboski: Corresponding author. Department of Economics, University of Notre Dame. Email: jkaboski@nd.edu. Research Associate at National Bureau of Economic Research.

§Li: Department of Economics and Faculty Associate of the Institute for Emerging Market Studies (IEMS), Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR-PRC. Email: yaoli@ust.hk. Research Affiliate of the China Research and Policy Group at University of Western Ontario.

1 Introduction

Both rich and poor countries generally regard industrial clusters, i.e., geographic concentrations of firms in the same industry, as good for productivity, growth, and development. The conventional economic wisdom dates back to the causes of agglomeration identified by Marshall (1890). Marshall's first two explanations, resource and demand concentration, imply that agglomeration occurs naturally without any need for policy intervention. His third cause, positive external spillovers on nearby firms, could lead to too little agglomeration or coordination concerns, and this is used to justify cluster-fostering industrial policies. Many studies find support for Marshall's hypotheses.[1] Influential work, including Marshall, has also viewed industrial clusters as productivity-enhancing through the pro-competitive pressures they may foster (e.g., Porter (1990)). Therefore, perhaps we should not be surprised that both advanced and developing economies adopt policies that promote clusters.[2]

Industrial clusters may indeed be cost reducing and productivity enhancing, but there is an even older concern – dating back to at least Adam Smith – that gathering competitors in the same locale and fostering cooperation among firms could instead lead to non-competitive behavior.[3] Close proximity and frequent interaction facilitate easy communication and observation that can enable cooperative behavior among firms. This cooperation can be beneficial. For example, firm associations have been shown to foster cooperation and information sharing, and increase the level of trust among CEOs while also increasing profits (Cai and Szeidl (2017)). However, in other cases, these impacts may also reduce the extent of competition between firms.[4]

There are certainly important cases of non-competitive behavior within industrial clusters. Historically, the most famous industrial clusters in the United States have all been accused of explicit collusion.[5] In China, our empirical focus, our own interviews with firm owners and administrators of industrial clusters uncovered explicit cooperation on sales and pricing, as we discuss. Such smoking guns for particular cases exist, but what we lack is a sense of the overall prevalence of such non-competitive behavior in the economy, and the extent to which

[1]See, for example, Greenstone, Hornbeck and Moretti (2010), Ellison, Glaeser and Kerr (2010), and Guiso and Schivardi (2007), for recent evidence. In contrast, Cabral, Wang and Xu (2015) finds little evidence of agglomeration economies in Detroit's Motor City.

[2]There are currently an estimated 1400 global initiatives fostering industrial clusters.

[3]Smith (1776)'s famous quote: "People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices. It is impossible indeed to prevent such meetings, by any law which either could be executed, or would be consistent with liberty and justice. But though the law cannot hinder people of the same trade from sometimes assembling together, it ought to do nothing to facilitate such assemblies; much less to render them necessary. (Book I, Chapter X)."

[4]See, for example, Green and Porter (1984), a theoretical case where easy observation helps support tacit collusion, or Marshall and Marx (2012) and Genesove and Mullin (1998), who document the behavior of actual cartels.

[5]See Bresnahan (1987) for evidence of Detroit's Big 3 automakers in the 1950s, and Christie, Harris and Schultz (1994) for Wall Street in the 1990s. The major Hollywood production studios were convicted of anti-competitive agreements in the theaters that they owned in the Paramount anti-trust case of the 1940s. Ongoing litigation alleges non-compete agreements for workers among Silicon Valley firms.

it may be linked to development policy.

This paper examines the hypothesis that geographic concentration and cluster policies are associated with non-competitive behavior from a much broader perspective. We define non-competitive behavior as behavior in either firm sales, hiring, or input purchasing that internalizes pecuniary externalities on other firms. We make three major contributions toward this end. First, we derive a novel, intuitive screen for measuring the extent of internalization among firms competing in the same industry. Independent firms consider their own market share but not the market shares of other firms when setting markups. When firms cooperate by internalizing the impacts of their behavior on other firms, however, their markups tend to depend on the aggregate market share of the cooperating firms. Second, using panel data on Chinese manufacturing firms, we validate our screen by confirming that affiliate plants of the same parent company are not behaving independently, which we would expect from firms with the same owner. Third, we show evidence of non-competitive behavior at the level of organized industrial clusters in the Chinese economy. Although we find limited levels of non-competitive behavior in the economy overall, it is almost three times as high in China's "special economic zones" (SEZs) than outside of them. Furthermore, we find that the levels of non-competitive behavior are also high in a set of industry-geography pairs that we pre-identified using the theory.

We derive our screen from a standard nested, constant-elasticity-of-substitution (CES) demand system with a finite number of competing firms and with a higher elasticity of substitution within an industry than across industries. As is well known in this setup and empirically confirmed (e.g., Atkeson and Burstein (2008), Edmond, Midrigan and Xu (2015)), the gross markup that a firm charges is increasing in its own market share. Our theoretical contribution is to show that when a subset of firms internalize their impact on the profits of the other firms, it leads to convergence in markups across these firms, and each firm's markup depends on the total market share of the cooperating firms rather than its own firm-specific market share.

Following this, our screen is to regress the reciprocal of the firm's markup on the firm's own market share and the total market share of its potential set of fellow syndicate members.[6] An index of the lack of competition is a simple function of the relative size of the coefficients on group market share vs. own market share. The screen is similar in spirit to the standard risk-sharing regression of Townsend (1994), focusing on a syndicate of local (cooperating) firms rather than a syndicate of local (risk-sharing) households.[7] It has similar strengths,

[6] Throughout the paper we consider several different possibilities for sets of firms that are cooperating, such as firms with a common owner, firms in the same geographic region, and firms in the same special economic zone.

[7] An important difference between our context and that of Townsend (1994) is the potential confounding effect of measurement error. Ravallion and Chaudhuri (1997) argue that idiosyncratic measurement error potentially biases the measure of risk-sharing upward when measurement error in the dependent variable is unrelated to that in the independent variable. In our context, measurement error in revenue affects both our measure of market shares and of markups. As discussed in Section 3.2, that implies that idiosyncratic measurement error in sales actually biases our measure of internalization *downward*. Hence, idiosyncratic

in that it allows for the two extreme cases of independent decision-making and perfect joint maximization, but it also allows intermediate cases. As in Townsend, we can be somewhat agnostic about the actual details of how non-competitive behavior occurs; we instead focus on the outcomes, i.e., whether increased concentration among a set of firms (measured by market share) is associated with increased market power of those firms (measured by markups). The screen is also robust along other avenues. Importantly, our theoretical results, and so the validity of the screen, depend only on the constant elasticity demand system. They are therefore robust to arbitrary assumptions on the cost functions and geographical locations of the individual firms. Moreover, we use simulations to show that our screen performs well for plausible levels of firm uncertainty, including correlated demand or cost shocks, and when we relax our strong assumptions on the demand system. Indeed, simulations calibrated to our empirical exercise show only small biases when departures from our assumptions are in the empirically plausible range.

Empirically, we use the screen to assess the lack of independent competition in Chinese industrial clusters and SEZs. SEZs are generally considered to have played a key role in China's growth miracle, and we have a high quality panel of firms with a great deal of spatial and industrial variation. The panel structure of the Annual Survey of Chinese Industrial Enterprises (CIE) allows us to estimate markups using the cost-minimization methods of De Loecker and Warzynski (2012) and implement our screen using within-firm variation.

Our screen both identifies non-competitive pricing in simple validation exercises and rejects it in simple placebo tests. Specifically, we test for joint profit maximization among groups of affiliates with the same parent company and in the same industry. Similarly, we test for joint profit maximization among state-owned firms in the same industry. Consistent with the theory, in our validation tests we estimate a highly significant relationship between markups and combined market share, but an insignificant relationship with the individual firms' own market share. This is exactly what the theory predicts for firms that maximize their joint profits. In our placebo tests, we find no response in markups to industrial cluster market shares and no influence of SEZs on markup behavior among these sets of firms.

In the broader sample of Chinese firms, the level of competitive behavior appears high, but as we move to smaller geographic definitions of a cluster the level of independent competition falls. Moreover, we find stronger evidence in subsets of clusters: SEZs and clusters pre-screened as having low initial cross-sectional variation in markups. SEZs target firms in specific industries and locations, giving them benefits such as special tax treatment or favorable regulation.[8] They also attempt to foster cooperation through industry associations, trade fairs, and coordinated marketing, but such venues can be used to reduce competition. We find that the intensity of cooperative pricing is nearly three times higher for clusters in SEZs than for those not in SEZs. Moreover, we apply our pre-screening criteria, focusing

measurement error cannot explain our results.

[8] We use SEZ in the broad sense of the term. See Alder, Shao and Zilibotti (2013) for a summary of SEZs, their history, and their policies.

on clusters in the lowest three deciles of cross-sectional markup variation, and find that only the cluster market share is a significant predictor of the panel variation in markups. That is, this subsample appears to be dominated by jointly cooperative, syndicate-like behavior. These clusters are characterized by disproportionately higher concentration industries, have lower export intensities, and contain a greater proportion of private domestic enterprises (as opposed to foreign or state-owned ventures).

The lack of competition in some industrial clusters, and in SEZs, in particular, may not negate the overall value of these policies, but they are nonetheless of normative significance. Spillovers across firms, such as externalities on productivity, could be positive. In this paper, we do not attempt to weigh these effects against the effects we find on competition. However, both effects are important for understanding the role and significance of cluster policy.

Our paper contributes and complements the literatures on both industrial clusters and competition.

First, we contribute to an emerging literature examining the role of firm competition – markups in particular – on macro development, including Asturias, Garcia-Santana and Ramos (2015), Edmond, Midrigan and Xu (2015), Galle (2016), and Peters (2015). Aghion et al. (2015) study pro-competitive industrial policy in China. Similarly, a recent literature has looked at firm networks and firm cooperation and the productive benefits they may foster (e.g., Cai and Szeidl (2017), Brooks, Donovan and Johnson (2017)).

The local growth impact of Chinese SEZs has been studied in Alder, Shao and Zilibotti (2013), Wang (2013), and Cheng (2014), and they have been found to have sizable positive effects using panel level data at the local administrative units. Our firm-level evidence of non-competitive behavior suggests that the growth from these policies may at least partially reflect important, unintended consequences.[9] Measured value added may be higher among firms in SEZs in part because cooperation allowed them to achieve higher markups, which is an important caveat when interpreting the previous results.

Finally, several papers have examined explicit collusion in cooperative industry associations, industrial clusters or agglomerations. The 19th century railroad associations in the U.S., originally formed to cooperate on technical (e.g., track width) and safety standards to link the various rails, soon turned to an explicit cartel designed to manage competition (see, e.g., Chandler (1977)). Colluding clusters in the 20th century have also been studied. Bresnahan (1987) studied collusion of the Big 3 automakers in Detroit, and Christie, Harris and Schultz (1994) examine NASDAQ collusion on Wall Street. More recently, Gan and Hernandez (2013) shows that hotels near one another effectively collude.

Methodologically, the recent industrial organization literature has tended toward "smok-

[9]While a lack of competition is likely an unintended consequence of agglomeration it is not obvious that the effect is negative. In a second best world, reduced competition may be welfare improving over high levels of competition. See, for example, Galle (2016) or Itskhoki and Moll (2015) for the case where financial frictions are present. In this paper we do not need to take any stand on whether the welfare consequences of cooperation are negative or positive.

ing gun" analysis of explicit collusion: detailed case studies of particular industries, making less stringent assumptions on demand or basing them on deep institutional knowledge of the industry.[10] Our approach is different but complementary, developing a screen of effective competition and applying the entire economy of a developing country that has actively promoted industrial clusters. Thus, our screen can be used to guide broad industrial policy *ex ante* and considering competition more broadly, rather than focusing on a case study of a the extreme case of a cartel *ex post*.

The rest of this paper is organized as follows. Section 2 presents the model and derives the key theoretical results. Section 3 lays out are empirical screen and reviews our empirical application. Section 4 discusses our data and methods for identifying markups. We discuss direct evidence on firm cooperation in Section 5, while Section 6 discusses the broader empirical results. Section 7 concludes.

2 Model

We develop a simple static model of a finite number of differentiated firms that yields different relationships between firm markups and market shares under independent competition and under syndicate behavior, and we show the robustness of these results to various assumptions. We assume a nested CES demand system of industries and varieties within the industry, which we assume is independent of location. Whereas the structure of demand is critical, we make minimal assumptions on the production side, allowing for a wide variety of determinants of firms costs, such as location choice, arbitrary productivity spillovers, and productivity growth for firms.[11]

2.1 Firm Demand

A finite number of firms operate in an industry i. The demand function of firm n in industry i is:

$$y_{ni} = D_i \left(\frac{p_{ni}}{P_i} \right)^{-\sigma} \left(\frac{P_i}{P} \right)^{-\gamma}, \tag{1}$$

where p_{ni} is the firm's price, and P_i and P are the price indexes for industry i and the economy overall, respectively. Thus, $\sigma > 1$ is the own price elasticity of any variety within industry i, while $\gamma > 1$ is the elasticity of industry demand to changes in the relative price index of the industry.[12] Typically, $\sigma > \gamma$, so that products are more substitutable within industries than industries are with one another. The parameters D_i captures the overall demand at the

[10] Einav and Levin (2010) give an excellent review of the rationale for moving away from cross-industry identification. Our screen also relies on within-industry (indeed, within-firm) identification.

[11] Our assumption that demand is independent of location implicitly assumes negligible trade costs in output, which is important for allowing for agglomeration based on externalities rather than local demand. Empirically, we will focus on manufactured goods.

[12] We analyze highly disaggregated industries, so the assumption $\gamma > 1$ is natural.

industry level. For exposition, we define units so that demand is symmetric across firms in the same industry, but this is without loss of generality. As each firm in the industry faces symmetric demand, the industry price index within industry i is:

$$P_i = \left(\sum_{m \in \Omega_i} p_{mi}^{1-\sigma} \right)^{1/(1-\sigma)}, \tag{2}$$

where Ω_i is the set of all firms operating in industry i.

As we show in the online appendix, this demand system can be derived as the solution to a household's problem that has nested CES utility.

One can invert the demand function to get the following inverse demand:

$$p_{ni} = P \left(\frac{y_{ni}}{Y_i} \right)^{-1/\sigma} \left(\frac{Y_i}{D_i} \right)^{-1/\gamma}, \tag{3}$$

where:

$$Y_i = \left(\sum_{m \in \Omega_i} y_{mi}^{1-1/\sigma} \right)^{\frac{\sigma}{\sigma-1}}. \tag{4}$$

To establish notation that will be used throughout this paper, we define market shares as:

$$s_{ni} = \frac{p_{ni} y_{ni}}{\sum_{m \in \Omega_i} p_{mi} y_{mi}} = \frac{y_{ni}^{1-1/\sigma}}{\sum_{m \in \Omega_i} y_{mi}^{1-1/\sigma}}, \tag{5}$$

where the second equality follows from substituting in (1) for prices and simplifying.

This demand system implies that the cross-price elasticity is given by a simple expression:

$$\forall m \neq n, \frac{\partial \log(y_{in})}{\partial \log(p_{im})} = (\sigma - \gamma)\, s_{im}. \tag{6}$$

which allows for simple aggregation in the results that follow. Our structure of demand, which implies a this cross-price elasticity restriction and a constant elasticity of demand, allows us to be very general in our specification of firm costs. The cost to firm n of producing y_{ni} units of output is $C(y_{ni}; X_{ni})$, where X_{ni} represents a general vector of characteristics such as capital, technology, firm productivity, location, externalities operating through the production levels of other firms, and any other characteristics that are taken as given by the producer when making production choices. For example, a special case of our model would be one in which an initial stage involves a firm placement game in which each firms' productivity is determined by the placement of each other firm through external spillovers, local input prices, or other channels. Then the results from that first stage determine X_{ni} that firms take as given when production choices are made, which is a special case of our framework.[13]

[13] However, note that the fact that firms maximize static profits below implicitly limits the way the vector

We now separately consider two extreme cases: firms operating totally independently and firms acting as a perfect syndicate. We then consider intermediate cases.

2.2 Firms Operating Independently

First, we consider the case of all firms operate independently of one another. The static profit maximization problem of a firm n in industry i is:

$$\pi_{ni} = \max_{y_{ni}} p_{ni} y_{ni} - C(y_{ni}; X_{ni}). \tag{7}$$

Using (3), the firm's optimal pricing condition equates marginal revenue with marginal cost:

$$p_{ni} \left(\frac{\sigma - 1}{\sigma} + \left[\frac{1}{\sigma} - \frac{1}{\gamma} \right] \frac{y_{ni}^{1-1/\sigma}}{\sum_{m \in \Omega_i} y_{mi}^{1-1/\sigma}} \right) = C'(y_{ni}; X_{ni}). \tag{8}$$

Using the definition of market shares, s_{ni}, given above, rearranging (8), and defining the firm's gross markup, μ_n, as the ratio of price to marginal cost yields the well-known result:[14]

$$\frac{1}{\mu_{ni}} = \frac{\sigma - 1}{\sigma} + \left(\frac{1}{\sigma} - \frac{1}{\gamma} \right) s_{ni}. \tag{9}$$

When a firm is operating independently, and given values of the elasticity parameters, this equation implies that the only information that is needed to predict a firm's markup is that firm's market share. In particular, while factor prices, productivity, and local externalities captured by X_{ni} would certainly affect quantities, prices, costs, and profits, markups are only affected by X_{ni} through their impact on market shares. For $\sigma > \gamma$, the empirically relevant case, additional sales that accompany lower markups come more from substitution within the industry than from growing the relative size of the industry itself. Firms with larger market shares have more to lose by lowering their prices, so they charge higher markups.

2.3 Perfect Syndicate

We contrast the case of independent firms with the opposite extreme: a subset of firms within an industry forms a syndicate to maximize the sum of their profits. Within an industry i, consider a set $S \subseteq \Omega_i$ of firms that solve the following joint maximization problem:

$$\sum_{m \in S} \pi_{mi} = \max_{\{y_{mi}\}_{m \in S}} \sum_{m \in S} p_{mi} y_{mi} - C(y_{mi}; X_{mi}). \tag{10}$$

X_{ni} can relate to past production decisions, such as dynamic learning-by-doing, sticky market shares, or dynamic contracts.

[14]See, for example, Edmond, Midrigan and Xu (2015) or Atkeson and Burstein (2008).

Using our definition of market shares again, we can express the first-order condition as:

$$\forall n \in S, \qquad C'(y_{ni}; X_{ni}) = p_{ni}\frac{\sigma - 1}{\sigma} + p_{ni}\sum_{m\in S}\left(\frac{1}{\sigma} - \frac{1}{\gamma}\right)s_{mi}. \tag{11}$$

Then rearranging (11) gives the relationship between markups and market shares:

$$\frac{1}{\mu_{ni}} = \frac{\sigma - 1}{\sigma} + \left(\frac{1}{\sigma} - \frac{1}{\gamma}\right)\sum_{m\in S} s_{mi}. \tag{12}$$

The markup of a firm within the set S depends only on the total market share of all firms within the group. While the independent firm considered only its own market share, the syndicate internalizes the costs to its own members of any one firm selling more goods, and these cost depends on the total market shares of the member firms. In this extreme case of a perfect syndicate, the firm's own market share influences its markup only to the extent that it affects the syndicate's share.

A number of corollary results follow from equations (9) and (12). First, clearly $\sigma > \gamma > 1$ implies that an independent firm's markup is increasing in its own market share. Second, for a firm in a syndicate, the firm's markup is increasing in the total market share of the syndicate. That is, the firm's own market share plays no role except to the extent that it affects the syndicate market share. Third, syndicate members all charge the same markup, since their markup is based on the sum of their market shares. In our empirical work later we interpret this to mean that there is less variation in markups when firms behave cooperatively than they would have if they operated independently. Fourth, if any member of a syndicate were instead operating independently, that firm's markup would be lower and its market share would be higher. Finally, the market shares of any set of cooperating firms exhibit more variation than if the same set of firms was operating independently.

We summarize the above characterization in the following proposition.

Proposition 1. *Given $\sigma > \gamma > 1$:*

1. *When operating independently, firm markups are increasing in the firm's own market share.*
2. *When maximizing joint profits, firm markups are increasing in total syndicate market share, with the firm's own market share playing no additional role.*
3. *syndicate firm markups are more similar under perfect syndicate than independent decisions.*
4. *Firm markups are higher under perfect syndicate decisions than independent decisions.*
5. *Firm market shares are less similar under perfect syndicate decisions than independent decisions.*

Each of these claims will be addressed in our empirical work that follows. We will use the first two claims to derive our screen in Section 3, while the third and fourth claims will be used to pre-identify potential collusive clusters. Finally, we will use the fifth claim as additional testable implication. We have intentionally written Proposition 1 in general language. In the subsection below, we will show that, while the precise formulas vary, these more general claims are robust to several alternative specifications.

2.4 Alternative Models

We present related results below for the cases of firm-specific price elasticities, Bertrand competition rather than Cournot, an imperfect syndicate, a more general demand structure, and monopsonistic internalization.

2.4.1 Firm-specific price elasticities

To allow for markups to vary among competitive firms with the same market share, we allow for a firm-specific elasticity of demand. In particular, suppose that inverse demand takes the form:

$$p_{in} = D_i^{1/\gamma} P y_{in}^{-1/\sigma+\delta_{in}} Y_i^{1/\gamma-1/\sigma}. \tag{13}$$

Here δ_{in} captures the firm-specific component of demand, and we think of these as deviations from the average elasticity σ: $\sum_{n\in\Omega_i} \delta_{in} = 0$. Proceeding as before to derive markup equations, the first order conditions for an independent firm imply:

$$\frac{1}{\mu_{ni}} = \delta_{ni} + \frac{\sigma-1}{\sigma} + \left(\frac{1}{\gamma} - \frac{1}{\sigma}\right) s_{ni}, \tag{14}$$

and for a syndicate, the analogous equation is:

$$\frac{1}{\mu_{ni}} = \delta_{ni} + \frac{\sigma-1}{\sigma} + \left(\frac{1}{\gamma} - \frac{1}{\sigma}\right) \cdot \sum_{m\in S} s_{mi} \tag{15}$$

Firm markups are again increasing in either the firm or syndicate's market share and the magnitude of this relationship is governed by the difference between the within- and across-industry elasticities. In addition, however, the presence of δ_{ni} in both equations shows the level of markups may be firm-specific, even when market share is arbitrarily small or firms are members of the same syndicate. This could explain why firms in the same syndicate have differing markups.

2.4.2 Bertrand competition

Now we consider the case where firms take competitors' prices as given instead of quantities when making production choices. From the demand function (1), we can write the problem of a firm operating independently as:

$$\max_{\{p_{ni},y_{ni}\}} p_{ni}y_{ni} - C(y_{ni}; X_{ni})$$

$$\text{subject to:} \quad y_{ni} = D_i \left(\frac{p_{ni}}{P_i}\right)^{-\sigma} \left(\frac{P_i}{P}\right)^{-\gamma}.$$

Taking first-order conditions with respect to both choice variables and dividing them yields the following equations, which are analogous to (9) and (12), respectively:

$$\frac{\mu_{in}}{\mu_{in} - 1} = \sigma - (\sigma - \gamma)s_{in} \tag{16}$$

and

$$\frac{\mu_{in}}{\mu_{in} - 1} = \sigma - (\sigma - \gamma)\sum_{m\in S} s_{im}. \tag{17}$$

Equation (16) corresponds to the case where firms operate independently, and equation (17) to the case where firms are in a perfect syndicate. Again, given elasticity parameters we see that firms' market shares (in the case of independent firms) or syndicates' market shares (in the case of perfect syndicates) are sufficient to solve for the firms' markups. As before, higher markups coincide with higher market shares, and the magnitude of this increasing relationship depends on the gap between the two elasticity parameters.

2.4.3 Imperfect Syndicate

Purely independent pricing and pure syndicate represent two extreme cases. Here we consider an imperfect syndicate, in which firms place a positive weight $\kappa \in (0, 1)$ on other firms' profits relative to its own, so that each firm maximizes:

$$\pi_{in} + \kappa \sum_{m\in S/\{n\}} \pi_{im}.$$

It is easy to show that the markup now depends on both the firm and syndicate market shares. For the Cournot case, we have:

$$\frac{1}{\mu_{ni}} = \frac{\sigma - 1}{\sigma} + (1 - \kappa)\left(\frac{1}{\sigma} - \frac{1}{\gamma}\right)s_{ni} + \kappa\left(\frac{1}{\sigma} - \frac{1}{\gamma}\right)\sum_{m\in S} s_{mi}. \tag{18}$$

2.4.4 General Demand

Our result is not true for all demand systems, but it is useful to consider the extent to which it may hold for other demand systems, and what are the chief characteristics of demand driving this relationship. To examine this, we start with a very general demand system $p_{in}(y_{in}; y_{im})$. Denoting the inverse price elasticity $\frac{y_{im}\partial p_{in}}{p_{in}\partial y_{im}}$ as ε_{nm}, we can solve the Cournot problem to derive the following general relationship for the perfect syndicate:

$$\frac{1}{\mu_{ni}} = 1 - \varepsilon_{nn} - \sum_{m \in S} \varepsilon_{nm} \frac{s_{mi}}{s_{ni}}. \tag{19}$$

In order for this to approximate equation (12) above, we need to assume $\varepsilon_{nn} = \varepsilon^*_{1,nn} + \varepsilon^*_{2,nn} s_{ni}$ and $\varepsilon_{nm} = \varepsilon^*_{nm} s_{ni}$, where the starred elasticities are (approximately) constant. That leads to

$$\frac{1}{\mu_{ni}} = \frac{\varepsilon^*_{1,nn} - 1}{\varepsilon^*_{1,nn}} - \varepsilon^*_{2,nn} \left(s_{ni} + \sum_{m \in S} \frac{\varepsilon^*_{nm}}{\varepsilon^*_{2,nn}} s_{mi} \right). \tag{20}$$

In this expression, inverse markups involve a constant and an elasticity weighted sum of own and syndicate market shares.

Interpreting the above assumption, the inverse own price elasticity has both a component that is independent of market share and a component that increases in market share, while the inverse cross price elasticity is inversely related to market share. The components of these elasticities that are increasing in market share capture the idea that the impact on a price of a percentage output increase of a firm depends positively on the relative size of that firm in the market overall. The precise summation result depends on the inverse cross-price elasticities being equal to the second component of the inverse own price elasticity.

2.4.5 Monopsony Behavior

Instead of cooperating to increase output prices, firms may instead cooperate to reduce input costs. As a simple case to evaluate this possibility, suppose each firm n in location j uses a single factor to produce its output by a production function $y_{nj} = F(l_{nj}; X_{nj})$. To fix ideas, we refer to this as labor. The aggregate supply of labor L_j depends on the market wage w_j, which is common across firms in a given location. For simplicity, we assume the function for the market wage takes the following form:

$$w_j(L_j) = A_j L_j^{\phi}. \tag{21}$$

Firms take the labor demand decisions of other firms (or those outside their own syndicate) as given. To isolate the effect of monopsony power, suppose that firms take the price of their output as given. Then the problem of an independent firm n in location j is:

$$\max_{y_{nj}, l_{nj}} p_{nj} y_{nj} - w_j(L_j) l_{nj}$$

$$\text{subject to:} \quad y_{nj} \leq F(l_{nj}; X_{nj})$$

$$L_j = \sum_m l_{mj}.$$

Since there are a finite number of firms purchasing labor, firm optimality implies a markup because firms restrict their purchases of labor to keep wages low. A firm n in location j has labor market share:

$$s_{nj}^{L} = \frac{l_{nj}}{L_j}. \tag{22}$$

Optimality for the independent firm implies that the markup is given by:

$$\mu_{nj} = 1 + \phi s_{nj}^{L} \tag{23}$$

and the analog for the syndicate imply a result similar to (12):

$$\mu_{nj} = 1 + \phi \sum_{m \in S} s_{mj}^{L} \tag{24}$$

Three things are important to note. First, the expressions above define marginal cost as the cost of producing an additional unit *at market prices*. Therefore the markup is:

$$\mu_{nj} = \frac{p_{nj}}{w_j(L_j)/F'(l_{nj}; X_{nj})} \tag{25}$$

Second, the shares in the expressions depend critically on the view of labor markets and the definition of relevant labor supply, L_j. If labor is mobile across industries but not across locations, it would be the total local labor force. If labor is specialized by industry but mobile across locations, it would be the total industry labor force. If immobile along both dimensions, it would be the total local industry-specific labor, while if mobile in both dimensions, it would be the economy-wide total labor force. Finally, note that it would be trivial to replace labor with any other input in the analysis.

3 Empirical Approach

In this section, we present our empirical screen for non-competitive pricing; assess the robustness of the screen on Monte Carlo simulations; and discuss our application to China, including the data and methods of acquiring markups.

3.1 Screen for Non-Competitive Pricing

The model of the previous section yielded the result that the markups of competitive firms depend on the within-industry elasticity of demand and their own market share, while the markups of fully internalizing firms depend on the total market share of the firms in the syndicate. This motivates the following single empirical regression equation for inverse markups:

$$\frac{1}{\mu_{nit}} = \theta_t + \alpha_{ni} + \beta_1 s_{nit} + \beta_2 \sum_{m \in S} s_{mit} + \varepsilon_{nit} \tag{26}$$

for firm n, a member of (potential) syndicate S, in industry i at time t.

In the case of purely independent pricing, the hypothesis is $\beta_2 = 0$ and $\beta_1 < 0$. For the case of a pure syndicate, we have the inverted hypothesis of $\beta_2 < 0$ and $\beta_1 = 0$. The relationships in equations (9) and (12) hold deterministically. The error term ε_{nit} could stem from (classical) measurement error in the estimation of markups, which we discuss in Section 4.2, or from uncertainty or other model specification error as discussed in Section 3.2.

Moreover, for the case of intermediate levels of internalization, κ in (18) can be easily estimated from equation (26) as:

$$\hat{\kappa} = \frac{\hat{\beta}_2}{\hat{\beta}_1 + \hat{\beta}_2} \tag{27}$$

Furthermore, equation (18) implies that we can use the regression in equation (26) to estimate the elasticity parameters. These equations imply that:

$$\frac{1}{\hat{\sigma}} - \frac{1}{\hat{\gamma}} = \hat{\beta}_1 + \hat{\beta}_2 \tag{28}$$

$$\frac{\hat{\sigma} - 1}{\hat{\sigma}} = \frac{1}{N} \sum_i \sum_{n \in \Omega_i} \left(\frac{1}{\mu_{ni}} - \hat{\beta}_1 s_{ni} - \hat{\beta}_2 \sum_{m \in S_{ni}} s_{mi} \right)$$

where N is the number of firms. It is then immediate to solve these equations simultaneously to generate estimates of the elasticity parameters.

An alternative interpretation of $\hat{\kappa}$ as a measure of the intensity of cooperation can be derived from considering the case of a subset of $\tilde{S} \subset S$ firms who fully internalize, while the others compete independently. This also leads to intermediate estimates in both coefficients, with β_1 larger and β_2 smaller for $\tilde{S}$ than for S. Under somewhat stronger assumptions that the distribution of market shares is the same for syndicate and non-syndicate firms, we can show that κ equals the fraction of firms in a perfect syndicate.[15]

Equation (26) has strong parallels with the risk-sharing test developed by Townsend (1994). In that family of risk-sharing regressions, household consumption is regressed on household income and total (village) consumption in the risk-sharing syndicate. Townsend solves the problem of a syndicate of households jointly maximizing utility and perfectly risk-sharing, and contrasts that with households in financial autarky; We solve the problem of a syndicate of firms jointly maximizing profits in a perfect syndicate and contrast with those independently maximizing profits. Townsend posited that households in proximity are likely to be able to more easily cooperate, defining villages as the appropriate risk-sharing network.

[15] Details of this claim are provided in the appendix.

We posit the same is true for firms and examine local cooperation of firms. Our screen also shares another key strength of risk-sharing tests: we do not need to be explicit about the details of how this cooperation arises.[16] Instead, we directly address the effects of less competition that are of most concern: an ability of firms to use their collective market power to raise markups. Finally, as discussed in Section 2.4, firms could compete as in Cournot or Bertrand, and the essential elements of the screen hold in each.

We also note the presence of time and firm dummies in our screening equation. The time dummies, θ_t, capture time-specific variation, which is important since markups have increased over time, as we show in the next section. In principle, firm-specific fixed effects are not explicitly required in the case of symmetric demand elasticities.[17] Nevertheless, we add α_{ni} to capture fixed firm-specific variation in the markup, stemming perhaps from firm-specific variation in demand elasticities, as discussed in Section 2.4. Together, these time and firm controls assure that the identification in the regression stems from within-cluster and within-firm variation over time in markups and market shares.

3.2 Simulation Results of Robustness

We derived our screen from the model in Section 2, which assumed that (i) all relevant information is known to the firm before it makes its production or pricing decisions, (ii) demand is nested-CES, and (iii) there is no measurement error. In reality firms face unanticipated shocks to production costs and demand, and they take this uncertainty into account when making decisions. Indeed we require such unanticipated shocks in order to identify our production functions used in our empirical implementation. Moreover, demand may not be CES, and there may be measurement error with specific levels of correlation. Here we examine the robustness of our screen to relaxing these assumptions by running our regression on simulated data from an augmented model.

We augment demand and technologies for firm n in industry i located in region k in year t according to the following equations:

$$y_{nikt} = \varepsilon_{nikt} D_{nikt} \left(\frac{p_{nikt} + \bar{p}}{P_i} \right)^{-\sigma} \left(\frac{P_i}{P} \right)^{-\gamma}, \tag{29}$$

$$y_{nikt} = \rho_{nikt} z_{nikt} l_{nikt}^{\eta}$$

The parameter η allows for curvature in the cost function, while the parameter $\bar{p}$ allows for decreasing ($\bar{p} < 0$) and increasing ($\bar{p} > 0$) demand elasticities. Here D_{nikt} and z_{nikt} are the known component of (firm-specific) demand and productivity, respectively, while ε_{nikt}

[16] For example, we do not need to distinguish between implicit or explicit cooperation.

[17] Here the parallel with Townsend breaks, since risk-sharing regression require household fixed effects, or differencing, in order to account for household-specific Pareto weights. In contrast, syndicates maximize profits rather than Pareto-weighted utility, and as long as profits can be freely transferred – an assumption needed for a perfect syndicate – all profits are weighted equally.

and ρ_{nikt} are the unanticipated shocks to demand and productivity, respectively. Note that demand and productivity shocks are not equivalent in this model, since productivity shocks affect marginal cost, while demand shocks do not.

We then augment the firm's problem to allow for partial internalization captured by κ and take into account firm uncertainty:

$$\max_{l_{nikt}} \int_{\varepsilon} \int_{\rho} \left[(1-\kappa)\pi_{nikt}(l,\varepsilon,\rho) + \kappa \sum_{m \in S_{ikt}} \pi_{mikt}(l,\varepsilon,\rho) \right] dF(\varepsilon) dG(\rho) \tag{30}$$

where the unsubscripted ε, ρ, l are *vectors* of demand shocks, cost shocks, and labor input choices. We assume that each firm belongs to a cluster S_{ikt}, and they jointly solve (30). In later sections we consider different cases for the sets of firms that may be operating as a syndicate, but in this section we refer to them generally as clusters. Notice that F and G are probability distributions over vectors. We will consider covariation of these shocks across firms at the firm, cluster, region-industry, industry, and year levels.

We simulate this model for various parameter values, run our screening regression on the simulated data, and evaluate the bias in κ as measured by equation (27). We overview the results here, and full details are given in the online appendix.

Our first exercise is to measure the bias to our estimates from unanticipated shocks. When shocks are at the level of the individual firm or are correlated at the level of the cluster, we find that they can bias our results, but these work in opposite directions. Unanticipated shocks at the individual level push our estimate of κ toward zero, while those at the cluster level push κ toward one. This is because individual shocks cause comovement in markups and individual shares independent of the cluster shares, which causes the coefficient on the individual share to increase in magnitude. The opposite is true for the cluster shock, which causes the coefficient on cluster share to increase in magnitude relative to that on the individual share. These effects can bias our $\hat{\beta}_1$ and $\hat{\beta}_2$ estimates. These estimates can lead to bias in $\hat{\kappa}$ for two reasons. First, biases in $\hat{\beta}_1$ and $\hat{\beta}_2$ feed directly into $\hat{\kappa}$. Second, since $\hat{\kappa}$ is a nonlinear function of $\hat{\beta}_1$ and $\hat{\beta}_2$, variance in the estimates of those coefficients leads to bias in $\hat{\kappa}$.

In all of these results, we stress that this bias only results from unanticipated shocks, and any shocks to cost or demand that are anticipated will not bias our results no matter how those shocks are correlated across firms as discussed in Section 2. In particular, if changes in the price of inputs are spatially or industrially correlated it only biases our results to the extent to which they are unanticipated.

In our second exercise, we study how large these unanticipated shocks would have to be to generate economically significant bias in our results. We parameterize the simulation to match the regression output from our baseline exercise, which is discussed in Section 6.2. We select the variance of individual shocks, the variance of cluster shocks as well as values of σ, κ and γ in order to match the point estimates and standard errors on the coefficients on own and cluster shares, the average markup, the estimated value of κ and the adjusted R^2 (when averaged across all simulations) to their counterparts in the Chinese analysis. We find that magnitudes of these shocks are not large enough to substantially bias our estimates of κ. In our parameterized simulation, the true value of κ is 0.29 while the estimated value is 0.28. In general, the quantitative importance of these depend on the magnitude of shocks relative to predictable variation in the data. Hence, large biases in estimates of κ would require extremely low R^2, substantially less than even the small R^2 levels we observe in the data.

In our third exercise, we simulate a non-CES demand system. Applying the form of non-CES demand given in equation (29), we find that, as the CES-deviating parameter, $\bar{p}$, moves away from zero, our estimated coefficient on firms' own shares can be biased. In the case of $\bar{p} > 0$ (implying a decreasing elasticity, as in linear demand, for example), the estimate would be downward biased, since a firm's markups would increase with its output (and firm's market share) simply from the decreasing elasticity. That is, this additional force would increase the absolute magnitude of $\hat{\beta}_1$, decreasing $\hat{\kappa}$. The converse is true for $\bar{p} < 0$. Nevertheless, the coefficient estimate on the cluster shares are unbiased. This is important because, if we wished to screen for the presence of any cooperation, our model implies that we should test if the coefficient on the cluster share is positive. Thus, the fact that our coefficient on cluster share is unbiased with non-CES demand implies that our screen for the presence of cooperation is unaffected by non-CES demand. However, the fact that the coefficient on firms' own shares could be biased implies that our estimate of the *magnitude* of cooperation/internalization, $\hat{\kappa}$, is biased if demand were non-CES, and the direction of bias would depend on the direction of the deviation from CES demand.

Our final exercise is to consider measurement error in revenues and costs in the model to see how that affects our estimate of κ.[18] One might suspect that idiosyncratic measurement error would lead to overestimation of cooperation in a way that it can lead to overestimates of risk-sharing.[19] However, we find that idiosyncratic measurement error actually leads to a *downward* bias in our estimate of κ.

This bias may seem surprising, but it has a simple explanation. Measurement error in regressors typically biases their coefficient estimates toward zero, so measurement error in a firm's own market shares alone should shrink that regressor's coefficient and push the estimate of κ toward one. However, this intuition relies on market share measurement error being independent of markups, but measurement error in revenue affects both measured market shares and measured markups. If the measured value of revenue is higher than its true value, both measured markups *and* measured market shares are by construction higher than their true values, and therefore idiosyncratic measurement error causes them to positively comove.

[18] Measurement error is distinguished from the case of model misspecification described above in that unanticipated shocks are taken into account when firms make choices, while measurement error has no effect on firm choices.

[19] See, for example, Ravallion and Chaudhuri (1997)'s critique of Townsend (1994).

We therefore *overestimate* the strength of the relationship between the two, increasing our estimate of $\hat{\beta}_1$ and causing a downward bias in $\hat{\kappa}$. Hence, if measurement error is idiosyncratic, we would tend to *underestimate* the extent of cooperation.[20]

Although quantitative magnitudes of the biases are small in the empirically relevant parameter space, we nevertheless consider shocks correlated at the level of the cluster to be the most serious threat to the interpretation of our results. We address this possibility in multiple ways, as discussed in detail in Section 6. We examine variation across different sets of firms, where we have stronger or weaker *a priori* reasons to suspect cooperation. First, we examine affiliated of the same parent company as a validation. Second, using similar reasoning, we evaluate firms that are state-owned enterprises within an industry, and we also run a placebo test for local cooperation in the sample of state-owned firms. Third, we examine our estimated patterns of cooperation among groups of firms more likely to form industry associations in the United States. Fourth, we utilize the result in Proposition 1 that internalization makes markups more similar (Result 3) to motivate separately examining clusters with low coefficients of variation in markups over the cross-section of firms in the cluster. To limit potential endogeneity, we identify these clusters using the cross-sectional variation of firms in the initial year of our data (1999). Within the model, these clusters could have low markup variation because (i) they are internalizing the impact of their behavior on others profits, or (ii) they have lower variation in market shares (because of similarity in firm-specific demand or technology, for example). We assume the former in our *ex ante* identification strategy, but then we evaluate the latter *ex post*. Finally, as a robustness check, we add region-time specific fixed effects to control for any region-time specific cost shocks, such as unanticipated shocks to factor prices. All of these exercises alleviate our concerns.

4 Application to Chinese Data

For our empirical analysis, we examine manufacturing firms in China. Manufacturing firms have the advantage of being highly tradable, as is consistent with the assumption in our model that demand does not depend on location or local markets. Our measurement methods are standard and closely follow the existing literature.

4.1 Why China?

China has several advantages. First, it has the world's largest population and second largest economy. The size of the Chinese country and economy give us wide industrial and geographic heterogeneity. Second, China is a well-known development miracle, and its success is often attributed, at least in part, to its policies fostering special economic zones and industrial

[20] By the same argument, however, measurement error that is perfectly correlated at the level of the cluster biases the estimate of κ upward, and the overall bias for a mix of idiosyncratic and cluster-specific measurement error depends on the relative strength of each.

clusters.[21] Third, both agglomeration and markups have increased over time as shown in Figure 1, which plots the average level of industrial agglomeration (as defined below) and average markups.

Finally, we have a high quality panel of firms for China: the Annual Survey of Chinese Industrial Enterprises (CIE), which was conducted by the National Bureau of Statistics of China (NBSC). The database covers all state-owned enterprises (SOEs), and non-state-owned enterprises with annual sales of at least 5 million RMB (about $750,000 in 2008).[22] It contains the most comprehensive information on firms in China. These data have been previously used in many influential development studies (e.g., Hsieh and Klenow (2009), Song, Storesletten and Zilibotti (2011)).

4.2 Measurement

Between 1999 and 2009, the approximate number of firms covered in the NBSC database varied from 162,000 to 411,000. The number of firms increased over time, mainly because manufacturing firms in China have been growing rapidly, and over the sample period, more firms reached the threshold for inclusion in the survey. Since there is a great variation in the number of firms contained in the database, we used an unbalanced panel to conduct our empirical analysis.[23] This NBSC database contains 29 2-digit manufacturing industries and 425 4-digit industries.[24]

The data also contain detailed data on revenue, fixed assets, labor, and, importantly, firm location at the province, city, and county location. Of the three designations, provinces are largest, and counties are smallest. We construct real capital stocks by deflating fixed assets using investment deflators from China's National Bureau of Statistics and a 1998 base year. The "parent id code", which we use to identify affiliated firms, is only available for the year 2004, but we assume that ownership is time invariant. We construct market shares using sales data and following the definition in Equation (5). We also use firms' registered designation to distinguish state-owned enterprises (SOEs) from domestic private enterprises (DPEs), multinational firms (MNFs), and joint ventures (JVs).

We do not have direct measures of prices and marginal cost, so we cannot directly measure markups. Instead, we must estimate firm markups using structural assumptions and structural methods, the method of De Loecker and Warzynski (2012), referred to as DW hereafter,

[21] For example, a World Bank volume (Zeng, 2011) cites industrial clusters as an "undoubtedly important engine [in China's] meteoric economic rise."

[22] We drop firms with less than ten employees, and firms with incomplete data or unusual patterns/discrepancies (e.g., negative input usage). The omission of smaller firms precludes us from speaking to their behavior, but the impact on our proposed screen would only operate through our estimates of market share and should therefore be minimal.

[23] The Chinese growth experience necessitates that we use the unbalanced panel. Using a balanced panel would require dropping the bulk of our firms (from 1,470,892 to 60,291 observations), or shortening the panel length substantially.

[24] We use the adjusted 4-digit industrial classification from Brandt, Van Biesebroeck and Zhang (2012).

in particular. DW extend Hall (1987) to show that one can use the first-order condition for any input that is flexibly chosen to derive the firm-specific markup as the ratio of the factor's output elasticities to its firm-specific factor payment shares:

$$\mu_{i,t} = \frac{\theta^{v}_{i,t}}{\alpha^{x}_{i,t}}. \tag{31}$$

This structural approach has the advantage of yielding a plant-specific, rather than a product-specific, markup. The result follows from cost-minimization and holds for any flexibly chosen input where factor price equals the value of marginal product. Importantly, we use materials as the relevant flexibly chosen factor. The denominator $\alpha^{x}_{i,t}$ is therefore easily measured.

The more difficult aspect is calculating the firm-specific output elasticity with respect to materials, $\theta^{v}_{i,t}$, which requires estimating firm-specific production functions. The issue is that inputs are generally chosen endogenously to productivity (or profitability). We address this by applying Ackerberg, Caves and Frazer (2006)'s methodology, presuming a 3rd-order translog gross output production function in capital, labor, and materials that is:

$$\begin{aligned} q_{nit} = {} & \beta_{k,i} k_{nit} + \beta_{l,i} l_{nit} + \beta_{m,i} m_{nit} + \\ & \beta_{k2,i} k^{2}_{nit} + \beta_{l2,i} l^{2}_{nit} + \beta_{m2,i} m^{2}_{nit} + \beta_{kl,i} k_{nit} l_{nit} + \beta_{km,i} k_{nit} m_{nit} + \\ & \beta_{lm,i} l_{nit} m_{nit} + \beta_{k3,i} k^{3}_{nit} + \ldots + \omega_{nit} + \epsilon_{nit}. \end{aligned} \tag{32}$$

Note that the coefficients vary across industry i, but only the level of productivity is firm-specific. This firm-specific productivity has two stochastic components. ϵ_{nit} is a shock that was unobserved/anticipated by the firm (and could reflect measurement error, as mentioned above) and is therefore exogenous to the firm's input choices. However, ω_{nit} is a component of TFP that is observed/anticipated, and so it is potentially correlated with $k_{i,t}$, l_{nit}, and m_{nit} because the inputs are chosen endogenously based on knowledge of the former. They assume that ω_{nit} is Markovian and linear in $\omega_{ni(t-1)}$. Identification comes from orthogonality moment conditions that stem from the timing of decisions, namely lagged labor and materials and current capital (and their lags) are all decided before observing the innovation to the TFP shock, and a two-step procedure is used to first estimate ϵ_{nit} and then the production function.

Production functions are estimated at the industry-level (although the estimation allows for firm-specific factor-neutral levels of productivity). The precision of the production function estimates – and hence the measurement error in markups – therefore depends on the number of firms in an industry. For this reason, we follow DW and weight the data in our regressions using the total number of firms in the industry. Moreover, estimation of markups is noisy in practice, and within each industry we windsorize the 3 percent of observations in the tails.

Finally, we use information on the geographic industries and clusters that we study. Namely, we merge our geographic and industry data together with detailed data from the China SEZs Approval Catalog (2006) on whether or not a firm's address falls within the geographic boundaries of targeted SEZ policies, and, if so, when the SEZ started. We use the broad understanding of SEZs, including both the traditional SEZs but also the more local zones such as High-tech Industry Development Zones (HIDZ), Economic and Technological Development Zones (ETDZ), Bonded Zones (BZ), Export Processing Zones (EPZ), and Border Economic Cooperation Zones (BECZ). Since no SEZs were added after 2006, these data are complete. Since our data start in 1999, the broad, well-known SEZs that were established earlier offer us no time variation. We also measure agglomeration at the industry level using the Ellison and Glaeser (1997) measure, where 0 indicates no geographic agglomeration (beyond that expected by industrial concentration), 1 is complete agglomeration, and negative would indicate "excess diffusion" relative to a random balls-and-bins approach.[25]

Table 1 presents the relevant summary statistics for our sample of firms.

5 Direct Evidence of Cooperation in Chinese Industrial Clusters

This sectional details direct evidence on cooperation in Chinese industrial clusters, some of which may be productivity enhancing and some of which may reduce competition. We have direct evidence on the operation of industrial clusters and firm behavior from a small number of field visits to industrial clusters involving qualitative interviews with firm owners, government officials, and other support services in Chinese industrial clusters. Comparison with narrative reports from the field visits of other researchers indicate that the observed cluster behavior appears representative (Zhang and Mu, 2017).

The clusters we visited were in different regions of the country and different industries. Each of the clusters focused on a unique consumer good industry with products involving a measure of standard automation but differentiated by quality, style and fashion rather than process technology.[26] Each cluster involved production for both the domestic and export markets – typically each firm had some mix – but some clusters focus on the domestic market disproportionately, while others focus on the export market. Indeed, by government

[25]Specifically, start by defining a measure of geographic concentration, G:

$$G \equiv \sum_i (s_i - x_i)^2$$

, where s_i is the share of industry employment in area i and x_i is the share of total manufacturing employment in area i. This therefore captures disproportionate concentration in industry i relative to total manufacturing. Using the Herfindahl index $H = \sum_{j=1}^{N} z_j^2$, where z_j is plant j's share in total industry employment, we have the following formula for the agglomeration index g:

$$g \equiv \frac{G - \left(1 - \sum_i x_i^2\right) H}{\left(1 - \sum_i x_i^2\right)(1 - H)}.$$

[26]Clusters tend to be highly specialized, at a finer level than our industry codes, such as cups, woolen sweaters, or hardware tools, for example.

design, China has multiple industrial clusters in the same industry that are located in different regions. Some focus on the domestic markets, while others on the foreign market, thus partially segmenting the total market across clusters. These field visits uncover several avenues of firm cooperation, including government-firm relationships, industrial associations, coordinated marketing activities, and order sharing. The last reflects an explicit form of anti-competitve firm behavior.

Government cooperation is a common element of industrial clusters, and this government leadership can lead to coordination among firms. Many industrial clusters – though not all – have an official designation as a SEZ (or HIDZ, ETDZ, etc.). In some cases, these official designations and the policies associated with them were implemented at the foundation of the cluster, but typically they have been given to existing clusters to encourage their growth. Special economic zones assist in many ways, including streamlined export processing, preferential regulations, and tax benefits. Much of this is directed by local government officials.

Government cooperation also plays an important role in land markets and pollution permitting. In some clusters, the local officials allocate land within the special economic zone to certain firms. In another cluster we visited, the land was owned by a private developer, but the land was purchased by the real estate development company in conjunction with an influential member of parliament who assisted in getting proper regulatory access. In some polluting industries, pollution rights also come from local governments with the influence of more influential government leaders at the national level.

Often, local governments organize business associations within SEZs that also foster cooperation. In the clusters we visited, the industry associations met weekly, biweekly, or monthly. The business leaders insisted that one of the key advantages of being in the clusters, in addition to access to specialized suppliers, was sharing information in order to have a pulse on market trends. They were able to differentiate their products from the competition (one way of segmenting the market), coordinate the mass of purchasers in the area (the scale of the market), as well as gather information about prevailing prices.

In many of our interviews, members of clusters discussed order sharing, which can take multiple forms.[27] In some cases, a large firm receives a large order, then breaks up the order to be fulfilled by smaller firms. In another case, an industry association would coordinate the bids of its members to allocate orders among firms. Since member firms do not usually compete against one another, this eliminates competitive pressure among members of the association. As the president of an industry association explained, "We do not allow internal competition on pricing. If a firm tried price cutting, we would kick them out." This president acted as a planner among the firms, allocating orders to member firms.

Other forms of cooperation within SEZs, such as information sharing, discussion of best practices, and entrepreneurship training, are consistent with previous studies showing positive productivity spillovers from firm-to-firm cooperation. In China, Cai and Szeidl (2017)

[27] See Zhang and Mu (2017) for more discussion of order sharing among firms in industrial clusters in China.

find that business associations, exogenously organized among medium-sized manufacturing firms, improved revenues and growth among firms by enhancing supplier-client matching and learning from peers. Similarly, Brooks, Donovan and Johnson (2017) find that exogenously introducing business owners with more experienced mentor-entrepreneurs in Kenya improved profitability of firms by helping young firms find low cost suppliers. Anecdotally, many entrepreneurs in our interviews reported similar effects. For example, entrepreneurs in one textile SEZ that we visited reported that membership in the SEZ has improved their business, which we can observe directly in our data. Firms inside SEZs enjoy, on average, higher labor productivity (value added per worker 15.4% higher), larger gross output value (6% larger) and sales (8% larger) relative to their counterpart firms in the same industry located outside of SEZs. The differences in means are highly statistically significant.[28] Therefore, there are other potentially important forms of cooperation among firms that are not captured by our screen, and may have positive effects on firm productivity.

Thus, we have direct evidence of both anti-competitive practices and productivity enhancing behaviors from firm cooperation. However, it is *a priori* unclear how quantitatively important these coordinated activities are, how representative these firm patterns are, and the extent to which the higher sales and revenue reflect the internalization of technological or pecuniary externalities. The normative implications are potentially important, which motivates are empirical work below.

6 Empirical Results

We start by presenting the results validating our screen using firms with common ownership. We then present the results for the overall sample (which are mixed), the results for those pre-identified clusters with low variation in markups across firms (which strongly indicate internalization), and some important characteristics of these collusive clusters. Throughout our regression analysis, we report robust standard errors, clustered at the firm level.[29]

6.1 Validation and Placebo Exercises

We start by running our screen on the sample of affiliated firms. That is, we define our potential syndicates in equation (26) as groups of affiliated firms in the same industry who all have the same parent, and we construct the relevant market shares of these syndicates. We know from existing empirical work (e.g., Edmond, Midrigan and Xu (2015)) that markups tend to be positively correlated with market share. Our hypothesis is $\beta_1 = 0$ and $\beta_2 < 0$,

[28] To be clear, this is only a comparison of means, and we cannot claim this statement is causal.

[29] We cluster at the firm level, since the identification involves within-firm variation, and we can maintain the same clustering for all our analysis. The significance of our main results are robust to clustering at the "cluster" level as well, but such clustering varies from analysis to analysis, while clustering at the firm level allows us to remain consistent throughout, which allows for clearer comparison across results.

however, so that own market share will not impact markups after controlling for total market share of the syndicate firms. We estimate (26) for various definition of industries: 2-digit, 3-digit, and 4-digit industries. Note that the definition of industry affects not only the market share of the firm and syndicate, but the set of affiliates in the syndicate. The broader industry classification incorporates potential vertical cooperation, but it also makes market shares themselves likely less informative.

Table 2 present the estimates, $\hat{\beta}_1$ and $\hat{\beta}_2$. (We omit the firm and time fixed effects from the tables.) The first column shows the estimates, where we assume perfectly independent behavior and constrain the coefficient on the internalized share to be zero. In the next three columns, we assume perfect internalization at the cluster level (constraining the coefficient on firm share to be zero), and define clusters at the 2-digit, 3-digit, and 4-digit levels, respectively. The last three columns are analogous in their cluster definitions, but we do not constrain either coefficient. The sample of observations is a very small subset (less than two percent) of our full sample both because we only include affiliates, and because we only have parent/affiliate information for firms present in the 2004 subsample.

Focusing on the last three columns, we see that our hypothesis is confirmed for all three industry classifications with the coefficients on syndicate share being larger and statistically significant, while the coefficients on own share are smaller and not significant. The coefficients are larger for the broader classifications, implying very low elasticities of substitution between broadly defined markets. Since our model is one of horizontal competition, *a priori* we view the 4-digit classification as most appropriate. Applying (28) to the results that constrain $\hat{\beta}_1$ to zero (i.e., column (4)) yields estimates of $\sigma = 4.5$ and $\gamma = 3.3$. The corresponding values implied by column (7) are very similar at 4.5 and 4.1. At this 4-digit level, the implied demand elasticities in all of our results are consistent with those found using other methods, e.g., elasticities based on international trade patterns in Simonovska and Waugh (2014), which is encouraging given the potential biases discussed in Section 3.2.

Our second validation exercise is analogous. Instead of examining private affiliates owned by the same parent, however, we examine state-owned enterprises (SOEs), which are all owned by the government. The variation in the data naturally reflect the privatization process occurring in China over the period (declining market share of SOEs), and the corresponding decrease in markups, but we hypothesize that competition amongst SOEs is weaker than competition between SOEs and private firms.

Indeed, the results in Table 3 verify this hypothesis. Columns 2-4 examine cooperation at different industry aggregations, and our screen is consistent with perfect internalization at the disaggregate industry level. In column 4, we find the coefficient on own share to be insignificant at the 4-digit level, while the coefficient on cluster's share is negative and significant. While our screen uncovers negative and statistically significant coefficients on cluster's share at the broader industry level (e.g., the 2-digit industry classification) too, own share is also significant and the implied κ values are tiny. Again, our model is one of horizontal competition, so it is natural that the results are most consistent when using the most disaggregate industries. For this reason, we focus on the 4-digit industry classification, our narrowest, for the remainder of our analyses.

Columns 5-7 consider variants where SOEs only collude with other SOEs (in their 4-digit industry) that are in geographic proximity, i.e., at more local levels of province, city, or county, respectively. Our results are strongest at the province level. We view this in some sense as a placebo test, and indeed the evidence for cooperation disappears at the more local levels. We take this as evidence that the presence of any correlated local shocks are not enough to erroneously lead to an assumption of only *local* internalization in the case of SOEs.

We also run placebo tests that replicate our screens for industrial cluster-based cooperation, but use these subsets of firms. We use the identical measure of industrial cluster market share that we use below, but consider only the markup response for these sets of firms. The results are quite strong: we find no significant responses of markups to the total market share of industrial clusters in either the SOE or affiliated firm samples, and no effect of being in an SEZ (see Tables A.1 and A.2 in the online appendix for full results). These negative results are an important counter-example to the idea that something about the construction of our screen (e.g., biases due to spurious local correlations) or our data automatically lead to false positives in detecting internalization at cluster levels.

In sum, both validation exercises are consistent with firms cooperation within ownership structures at the disaggregate industry level, and our screen is able to reject cluster-based cooperation in placebo tests.

6.2 Non-Competitive Behavior in Industrial Clusters

We now turn to industrial clusters more generally by defining our potential syndicates as sets of firms in the same industry and geographic location. Table 4 presents the results. The first column shows the estimates, where we assume perfectly independent behavior and constrain the coefficient on syndicate share to be zero. In the next three columns, we assume perfect internalization at the cluster level (constraining the coefficient on firm share to be zero), and define clusters at the province, city, and county level, respectively. The next three columns allow for both shares to influence inverse markups, while the final three interact firm market share and cluster market share with an indicator variable for whether the firm is in a SEZ.

Focusing on columns 1 through 7, we note several strong results. First, all of the estimates are highly significant indicating that both firm share and market share are strongly related to markups. Because all estimates are statistically different from zero, we can rule out either perfectly independent behavior or perfect internalization at the cluster level. Second, all the coefficients on market shares are negative, as we would predict if output within an industry are more substitutable than output between industries. Third, the magnitudes are substantially larger for own firm share. Fourth, as we define clusters at a more local level, the coefficient on cluster share increases in magnitude, while the the coefficient on own share decreases. This suggests that cooperation is indeed more prevalent among firms that are in proximity to one another.

The $\beta_2 < 0$ estimates indicate some level of cluster-level collusion in the overall sample.[30] Again, applying equation (28), we can interpret the magnitude of the implied elasticities and the extent of internalization. At the county level, we estimate $\hat{\kappa} = 0.29$, while we estimate just $\hat{\kappa} = 0.08$ at the province level. This indicates a relatively low level of non-competitive behavior overall, especially when examining firms only located within the same province. The implied elasticity estimates are $\sigma = 4.8$ and $\gamma = 2.9$. These implied elasticities are quite similar to those implied in the smaller sample of affiliated firms, even though the level of internalization is greater.

Finally, we examine the role of SEZs examined in columns 8-10 of Table 4. The coefficients on the interaction of the SEZ dummy with firm market share are positive and significant but smaller in absolute value than the coefficient on firm market share itself. Adding the two coefficients, own market share is therefore a less important a predictor of (inverse markups) in SEZs. Similarly, the coefficients on cluster market share are negative, so that overall cluster market share is a more important predictor in SEZs. Indeed, using the county-level estimates in the last column, we estimate an internalization index $\hat{\kappa} = 0.42$ for firms within SEZs, nearly three times as high as that of firms not in SEZs, where $\hat{\kappa} = 0.16$. Again, the results for SEZs are strongest, the more local the definition of clusters. Recall, that SEZs are essentially pro-business zones, combining tax breaks, infrastructure investment, and government cooperation in order to attract investment. A common goal with industry-specific zones or clusters is to foster technical coordination in order to internalize productive externalities. The evidence suggests that such zones may also facilitate marketing coordination and internalizing pecuniary externalities.

We have estimated similar regressions where we differentiate across industries using the Rauch (1999) classification. Rauch classifies industries depending on whether they sell homogeneous goods (e.g., goods sold on exchanges), referenced priced goods, and differentiated goods. Without agriculture and raw materials, our sample of homogeneous goods is limited, but we can distinguish between industries that produce differentiated goods, and those that produce homogenous/reference priced goods. Our estimates of κ range from 0.15 to 0.28 for the former and range from 0.31 to 0.68 for the latter (depending on which Rauch specification is used as shown in Table A.3 in the online appendix), indicating stronger cooperation among firms producing fmore homogeneous goods, consistent with existing arguments and evidence that collusion is less beneficial and common in industries with differentiated products Dick (1996). Equally interesting, the coefficients themselves are much larger for these goods, consistent with a larger ρ, which would be expected, since goods should be highly substitutable

[30]We verify that this is not driven by the affiliated firms in two ways: (i) dropping the affiliated firms from the sample, and (ii) assigning the parent group share within the cluster to firm share. Neither changes affect our results substantially.

within these industries. Again, we view this latter consistency as further evidence that our results are driven by the markup-market share mechanism we highlight rather than some other statistical phenomenon.

We have also examined robustness of the (county-level, unrestricted) results in Table A.4 to various alternative specifications. Although the theory motivates weighting our regressions, neither the significance nor magnitudes of our results are dependent on the weighting in our regressions. We can also use the Bertrand specification rather than Cournot, by replacing the dependent variable with $\mu_{nit}/(\mu_{nit} - 1)$. This Bertrand formulation require us to Windsorize the data, however, because for very low markups the dependent variable explodes. These observations take on huge weight, and very low markups are inconsistent with the model for reasonable values of γ. If we drop all observations below 1.06, a lower bound on markups for a conservative estimate of $\gamma = 10$ (much larger than implied by the Cournot estimates, for example), we get similar results, with implied elasticities $\sigma = 5.4$ and $\gamma = 2.3$ and and intensity of internalization, $\kappa = 0.36$. Finally, we can use log markup, rather than inverse markup, as our dependent variable. The log function may make these regressions more robust to very large outlier markups. Naturally, the predicted signs are reversed, but they are both statistically significant, indicating partial internalization, and the implied semi-elasticities with respect to own and cluster share are 11.8 and 5.2 percent, respectively. The details of these robustness studies are in our online appendix.

We next turn to clusters which appear *a priori* likely to be potentially behaving as a synicate because they have low cross-sectional variation in markups. We do this by sorting clusters into deciles according to their coefficient of variation of the markup. Table 5 presents the coefficient of variation of these deciles, along with other cluster decile characteristics, when clusters are defined at the county level. Note that the average markup increases with coefficient of variation of markups over the top seven deciles, but that this pattern inverts for the lowest three deciles, where the average markup is actually higher as the coefficient of variation decreases. Higher markups and lower coefficients of variation may indicate co-operating behavior, given claims 3 and 4 in Proposition 1. We therefore focus on firms in the these bottom three clusters, and the lowest thirty percent is also consistent with the $\hat{\kappa}$ interpretation that 29 percent of firms collude.[31]

The other key characteristics of these lowest deciles of clusters are also of interest. First, although they have lower variation in markups, this does not appear to be connected to lower variation in market shares, as the coefficients of variations in market shares are similar, showing no clear patterns across the deciles. They have fewer firms per cluster, and are in industries with higher geographic concentration (measured by the Ellison-Glaeser agglomeration index) and higher industry concentration (as measured by the Hirschman-Herfindahl index). The firms themselves are somewhat smaller in terms of fewer employees per firm.

[31] These low markup variation deciles contain fewer firms on average, however, and so they constitute only 16 percent of firms.

Fewer firms in these clusters export, and overall exports are a lower fraction of sales. Finally, although there are not sharp differences in the ownership distribution, they are disproportionately domestic private enterprises and somewhat less likely to be multi-national enterprises or joint ventures.[32]

Table 6 presents the results for this restricted sample of the lower three deciles. The columns follow a parallel structure as in Table 4, but there are three columns even for the regressions that only include firm market share because the set of firms here varies depending on whether we define our clusters at the province, city, or county level. In the results that assume perfectly independent behavior we again find negative and significant estimates at the province and county level.[33] In the results that assume perfectly cooperative behavior, we again find negative significant estimates on cluster market share, and the results are again stronger, the more locally the cluster is defined. The most interesting results in the table, however, are those where we do not constrain either coefficient. In this restricted sample, we again find evidence of partially internalizing behavior at the province level.

What is striking, however, is that the internalization appears complete at local levels within these restricted samples: only the $\hat{\beta}_2$ estimates are negative and significant. The *positive* $\hat{\beta}_1$ at the city level is admittedly at odds with the theory, but the coefficient is not statistically significant. The county-level estimate in column (9) implies a within-industry elasticity σ that compares well with that in the full sample (5.0 vs. 4.8), but the between-industry elasticity is somewhat higher than in the full sample (6.6 vs. 2.9).

Once again, we find significant impacts of SEZs when interacted with market share. For counties, the region's share is nearly twice as large for firms in SEZs.

6.3 Robustness

We now examine the robustness of our results to various alternatives. In particular, we attempt to address the issue that the correlation between markups and cluster share may simply be driven by spatially correlated shocks to costs or demand across firms. (Although our Monte Carlo simulations indicated this was unlikely to be problematic quantitatively.) We address this concern in two ways.

First, we add region-time specific fixed effects as controls into our regressions. Our Monte Carlo simulations showed that these effectively control for any general shocks or trends to production or costs at the region level, e.g., rising costs of land or (non-industry-specific) labor from agglomeration economies. Controlling for these, our regressions will only be identified by cross-industry variation in market shares within a geographic location. Table 7 shows these results for the sample of clusters with low initial variation in markups. The patterns are quite

[32]Moreover, the single most disproportionately overrepresented industry in these clusters is petroleum refining, a classic syndicate in U.S. history.

[33]The city estimates have fewer observations, since there are fewer firms in the low markup variation deciles of city clusters.

similar to those in Table 6, for example, the magnitudes of the coefficients on cluster share are -0.065 vs. -0.064 in column 9. The results are significant at a one percent level. We find very similar results for the overall sample, but since our SEZs show very little variation with counties, we cannot separately run our SEZ regression using these fixed effects. Nonetheless, we view the robustness of our results as evidence that spatially correlated shocks (or trends) do not drive our inference, although in principle, industry-specific spatially correlated shocks could still play a role.

Second, we attempt an instrumental variable approach, since shares themselves are endogenous. Identifying general instruments may be difficult, but in the context of the model and our Ackerberg, Caves and Frazer (2006) estimation, exogenous productivity shocks affect costs and therefore exogenously drive both market share and markups. We motivate our instrument using an approximation, the case of known productivity z_{in} and monopolistic competition. This set up yields the following relationship between shares and the distribution of productivity:

$$s_{in} = \frac{p_{in}y_{in}}{\sum\limits_{m\in\Omega_i} p_{im}y_{im}} \approx \frac{z_{in}^{1-1/\sigma}}{\sum\limits_{m\in\Omega_i} z_{im}^{1-1/\sigma}} \tag{33}$$

We construct instruments for own market share (I_1) and cluster market share (I_2) using variants of the above formula that exclude the firm's own productivity and the productivities of all firms in the firm's cluster (S_n), respectively:

$$I_1 = \frac{1}{\sum\limits_{m\in\Omega_i/n} z_{im}^{1-1/\sigma}}, \qquad I_2 = \frac{1}{\sum\limits_{m\in\Omega_i/S_n} z_{im}^{1-1/\sigma}} \tag{34}$$

This two-stage estimation yields very similar results (see Table A.5). For example, the coefficient on cluster share in the analog to column (9) is -0.048 and is significant at the one percent level. Again, the patterns we develop are broadly robust.

In sum, we have shown for China that: the screen detects internalization among firms owned by the same parents in the affiliated and SOE samples; the markups of SOEs in a placebo test do not respond to their market share in a local cluster; the estimates are consistent with the model's mechanism based on the Rauch classification; our internalization patterns are stronger in SEZs; the internalization patterns are very strong in clusters that the model pre-identifies as likely syndicate clusters; these patterns are robust to inclusion of time-region specific fixed effects and instrumenting for market share.

7 Conclusion

We have developed a simple, intuitive and robust screen for identifying non-competitive behavior for subsets of firms competing in the same industry. Using this screen we have found evidence of a lack of competition in Chinese industrial clusters. These results are strongest within narrowly-defined clusters in terms of narrow industries and narrow geographic units. A small but non-negligible share of firms and clusters appear to exhibit from non-competitive behavior. This behavior is disproportionately strong – nearly three times greater – in special economic zones.

The results open several avenues for future research. In this paper we have focused exclusively on China. However, the fact that it satisfied our validation exercises means it could easily applied more generally to other countries and contexts where firm panel data are available. Furthermore, the potential normative importance of our results are compelling with respect to evaluating industrial policies that promote clustering, such as local tax breaks, subsidized credit, or targeted infrastructure investments. They motivate more rigorous evaluation of various normative considerations, including weighing the extent to which syndicates hurt (or perhaps even help) consumers, productivity gains from external economies of scale vs. monopoly pricing losses from syndicates, and local vs. global welfare implications and incentives. Precisely these issues are the subjects of our continuing research.

References

Ackerberg, Daniel, Kevin Caves, and Garth Frazer. 2006. "Structural identification of production functions." University Library of Munich, Germany MPRA Paper 38349.

Aghion, Phillipe, Jing Cai, Mathias Dewatripont, Luosha Du, Ann Harrison, and Patrick Legros. 2015. "Industrial Policy and Competition." *American Economic Journal: Macroeconomics*, 7(4): 1–32.

Alder, Simon, Lin Shao, and Fabrizio Zilibotti. 2013. "The Effect of Economic Reform and Industrial Policy in a Panel of Chinese Cities." Working Paper.

Asturias, Jose, Manuel Garcia-Santana, and Roberto Ramos. 2015. "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral in India." Working Paper.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review*, 98: 1998–2031.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang. 2012. "Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing." *Journal of Development Economics*, 97(2): 339–351.

Bresnahan, Timothy. 1987. "Competition and Collusion in the American Automobile Industry: The 1955 Price War." *Journal of Industrial Economics*, 35(4): 457–482.

Brooks, Wyatt, Kevin Donovan, and Terence Johnson. 2017. "Mentors or Teachers? Microenterprise Training in Kenya." *American Economic Journal: Applied Economics*. Forthcoming.

Cabral, Luis, Zhu Wang, and Daniel Yi Xu. 2015. "Competitors, Complementors, Parents and Places: Explaining Regional Agglomeration in the U.S. Auto Industry." Working Paper.

Cai, Jing, and Adam Szeidl. 2017. "Interfirm Relationships and Business Performance." *Quarterly Journal of Economics*. Forthcoming.

Chandler, Alfred. 1977. *The Visible Hand: The Managerial Revolution in American Business.* Harvard University Press.

Cheng, Yiwen. 2014. "Place-Based Policies in a Development Context - Evidence from China." Working Paper.

Christie, William, Jeffrey Harris, and Paul Schultz. 1994. "Why did NASDAQ Market Makers Stop Avoiding Odd-Eighth Quotes?" *Journal of Finance*, 49(5): 1841–1860.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status." *American Economic Review*, 102(6): 2437–71.

Dick, Andrew. 1996. "When are Cartels Stable Contracts?" *Journal of Law and Economics*, 39(1): 241–283.

Edmond, Chris, Virgiliu Midrigan, and Daniel Xu. 2015. "Competition, Markups and the Gains from International Trade." *American Economic Review*, 105(10): 3183–3221.

Einav, Liran, and Jonathan Levin. 2010. "Empirical Industrial Organization: A Progress Report." *Journal of Economic Perspectives*, 24(2): 145–162.

Ellison, Glenn, and Edward L Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy*, 105(5): 889–927.

Ellison, Glenn, Edward Glaeser, and William Kerr. 2010. "What Causes Industrial Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review*, 100(3): 1195–1213.

Galle, Simon. 2016. "Competition, Financial Constraints and Misallocation: Plant-Level Evidence from Indian Manufacturing." Working Paper.

Gan, Li, and Manuel Hernandez. 2013. "Making Friends with Your Neighbors? Agglomeration and Tacit Collusion in the Lodging Industry." *Review of Economics and Statistics*, 95(3): 1002–1017.

Genesove, David, and Wallace Mullin. 1998. "Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890-1914." *RAND Journal of Economics*, 29(2): 355–377.

Green, Edward, and Robert Porter. 1984. "Noncooperative Collusion under Imperfect Price Information." *Econometrica*, 52(1): 87–100.

Greenstone, Michael, Rick Hornbeck, and Enrico Moretti. 2010. "Identifying Productivity Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy*, 118(3): 536–598.

Guiso, Luigi, and Fabiano Schivardi. 2007. "Spillovers in Industrial Districtrs." *Economic Journal*, 117(516): 68–93.

Hall, Robert. 1987. "Productivity and the Business Cycle." *Carnegie-Rochester Conference Series on Public Policy*, 27: 421–444.

Hsieh, Chang-Tai, and Peter Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403–1448.

Itskhoki, Oleg, and Benjamin Moll. 2015. "Optimal Development Policies with Financial Frictions." Working Paper.

Marshall, Alfred. 1890. *Principles of Economics.* London: MacMillan.

Marshall, Robert, and Leslie Marx. 2012. *The Economics of Collusion.* MIT Press.

Peters, Michael. 2015. "Heterogeneous Mark-Ups, Growth and Endogenous Misallocation." Working Paper.

Porter, Michael. 1990. *The Competitive Advantage of Nations.* Free Press, New York.

Rauch, James E. 1999. "Networks Versus Markets in International Trade." *Journal International of Economics*, 48(1): 7–35.

Ravallion, Martin, and Shubham Chaudhuri. 1997. "Risk and Insurance in Village India: Comment." *Econometrica*, 65(1): 171–184.

Simonovska, Ina, and Michael Waugh. 2014. "Elasticity of Trade: Estimates and Evidence." *Journal of International Economics*, 92(1): 34–50.

Smith, Adam. 1776. *The Wealth of Nations.* W. Strahan and T. Cadell, London.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti. 2011. "Growing Like China." *American Economic Review*, 101(1): 202–241.

Townsend, Robert. 1994. "Risk and Insurance in Village India." *Econometrica*, 62(3): 539–591.

Wang, Jin. 2013. "The economic impact of Special Economic Zones: Evidence from Chinese municipalities." *Journal of Development Economics*, 101: 133–147.

Zeng, Douglas. 2011. "How Do Special Economic Zones and Industrial Clusters Drive China's Rapid Development?" In *Building Engines for Growth and Competitiveness in China.* , ed. Douglas Zeng. Washington, D.C.

Zhang, Xiaobo, and Tianshi Mu. 2017. "Survey Summary (in Chinese, "Zhong Guo Qi Ye Chuang Xin Chuang Ye Diao Cha: Diao Yan Zong Jie")." In *Entrepreneur Survey for Innovation and Entrepreneurship in China (ESIEC).* by National School of Development, Peking University, and Institute of Social Science Survey, Peking University.

Figure 1: Increasing Agglomeration and Markups over Time in China

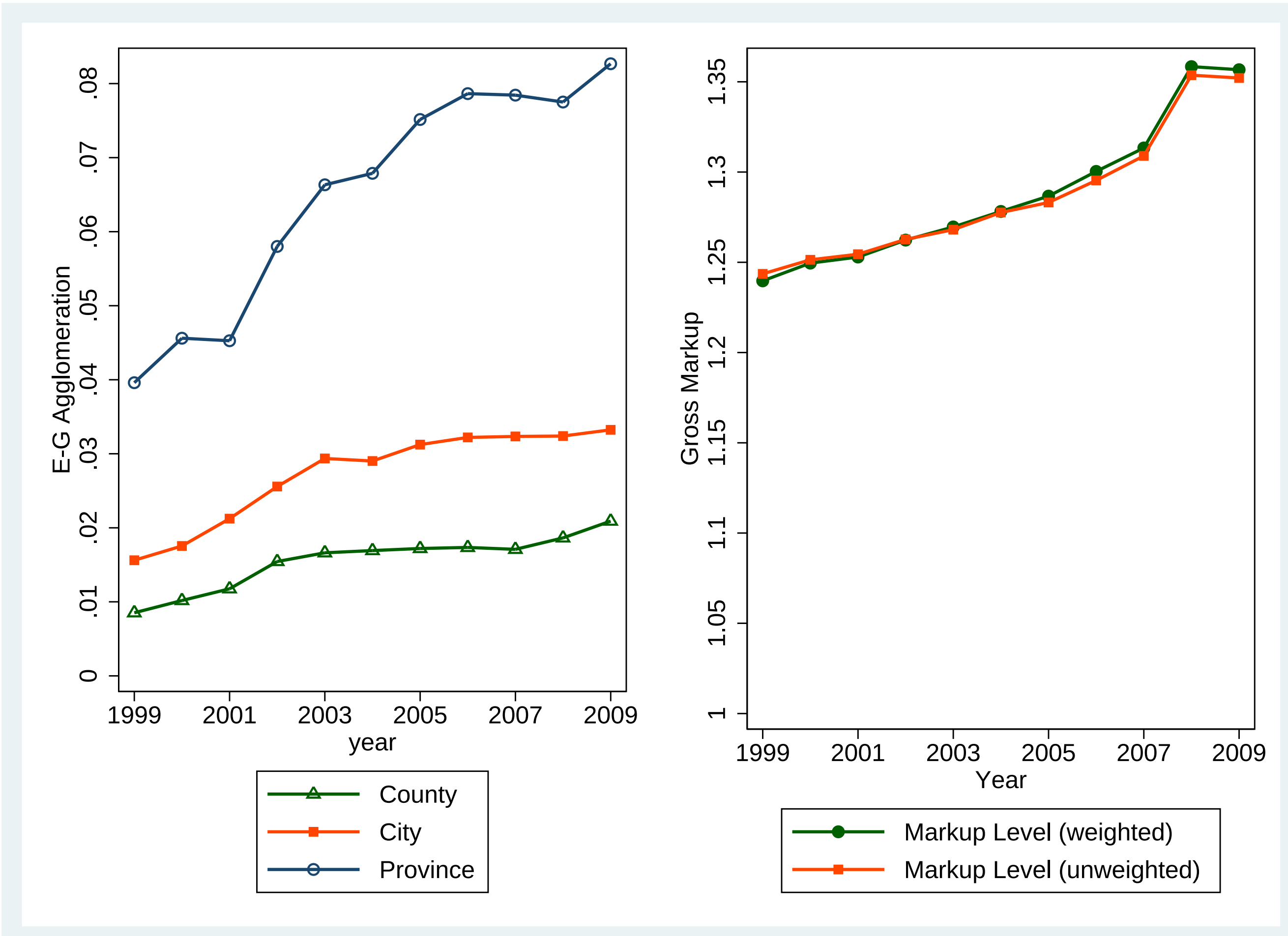


Table 1: Key Summary Statistics of Data

Variable	Mean	Median	S.D.	Min	Max
Markup	1.29	1.26	0.21	0.61	4.76
Firm Share	0.00	0.00	0.01	0	1
Cluster Share (Province)	0.14	0.10	0.14	0	1
Cluster Share (City)	0.04	0.02	0.06	0	1
Cluster Share (County)	0.02	0.00	0.04	0	1
Capital per Firm	320	50	3720	0.01	1,035,380
Materials per Firm	720	170	5940	0.05	860,550
Real Output per Firm	1000	240	7970	0.08	1,434,840
Workers per Firm	290	120	1010	10	166,860
No. of Firms			408,848		

Notes: Market shares are computed using 4-digit industries. Capital, output and materials are in thousand RMB (in real value).

Table 2: Baseline Results Using Affiliated Firms

	Dependent Variable: $\frac{1}{\mu_{nit}}$						
	(1) 4-digit	(2) 2-digit	(3) 3-digit	(4) 4-digit	(5) 2-digit	(6) 3-digit	(7) 4-digit
Firm's share	-0.009				0.108	0.264	0.109
	(0.060)				(0.850)	(0.209)	(0.083)
Cluster's share		-0.465***	-0.270***	-0.077	-0.470***	-0.320***	-0.133*
		(0.149)	(0.091)	(0.054)	(0.150)	(0.092)	(0.070)
Year FEs	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES
Observations	24780	24780	24780	24780	24780	24780	24780
Overall R^2	0.010	0.005	0.009	0.012	0.005	0.009	0.013

Notes: Robust standard errors clustered at firm level in parentheses. Significance: ***: 1%, **: 5%, *: 10%. Various industry aggregation levels are employed, including 4-digit industry (in specifications 1, 4 and 7), 3-digit industry (in specifications 3 and 6), and 2-digit industry (in specifications 2 and 5). All specifications are regressions weighted by the number of observations for each two-digit CIC sector production function estimation reported (following De Loecker et al. 2014). All regressions include a constant term.

Table 3: Baseline Results Using SOEs as Cluster

	Dependent Variable: $\frac{1}{\mu_{nit}}$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	all SOEs in the industry (nation-wise)				province	city	county
	4-digit	2-digit	3-digit	4-digit	4-digit	4-digit	4-digit
Firm's Share	-0.047	-3.141***	-0.682***	-0.035	0.014	-0.002	-0.023
	(0.055)	(0.929)	(0.145)	(0.055)	(0.061)	(0.068)	(0.117)
Cluster's Share		-0.028*	-0.010	-0.016**	-0.061**	-0.045	-0.024
		(0.017)	(0.009)	(0.006)	(0.027)	(0.041)	(0.103)
Year FEs	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES
Observations	106434	106434	106434	106434	106434	106434	106434
Overall R^2	0.034	0.052	0.041	0.050	0.036	0.035	0.034

Notes: Robust standard errors clustered at firm level in parentheses. Significance: ***: 1%, **: 5%, *: 10%. Various industry aggregation levels are employed, including 4-digit industry (in specifications 1 and 4-7), 3-digit industry in specifications, and 2-digit industry in specifications 2. All specifications are regressions weighted by the number of observations for each two-digit CIC sector production function estimation reported (following De Loecker et al. 2014). All regressions include a constant term.

Table 4: Baseline Results Using Overall Sample

	Dependent Variable: $\frac{1}{\mu_{nit}}$									
	(1)	(2) Province	(3) City	(4) County	(5) Province	(6) City	(7) County	(8) Province	(9) City	(10) County
Firm's Share	-0.143***				-0.132***	-0.111***	-0.099***	-0.189***	-0.170***	-0.163***
	(0.021)				(0.021)	(0.021)	(0.022)	(0.026)	(0.026)	(0.027)
Region's Share		-0.014***	-0.037***	-0.053***	-0.012***	-0.029***	-0.040***	-0.007***	-0.024***	-0.031***
		(0.002)	(0.004)	(0.006)	(0.002)	(0.004)	(0.006)	(0.002)	(0.005)	(0.007)
SEZ*Firm's Share								0.083**	0.091**	0.095**
								(0.037)	(0.038)	(0.039)
SEZ*Region's Share								-0.008**	-0.014**	-0.018*
								(0.003)	(0.006)	(0.010)
SEZ Dummy								0.001	0.000	0.000
								(0.001)	(0.001)	(0.001)
Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	1346860	1346860	1346860	1346860	1346860	1346860	1346860	1105162	1105162	1105162
Overall R^2	0.028	0.023	0.026	0.027	0.025	0.027	0.028	0.025	0.026	0.026

Notes: Robust standard errors clustered at firm level in parentheses. Significance: ***: 1%, **: 5%, *: 10%. Regions are defined at various aggregation levels, including province (in specifications 2, 5, and 8), city (in specifications 3, 6, and 9), and county (in specifications 4, 7, and 10). All specifications are regressions weighted by the number of observations for each two-digit CIC sector production function estimation reported (following De Loecker et al. 2014). All regressions include a constant term.

Table 5: Cluster Characteristics by Cluster Decile of Coefficient of Variation of Markup

Cluster Decile of Coefficient of Variation of Markup	Avg. Coefficient of Variation of Markup	Local Competition Indicator			Industry Concentration Characteristics		Firm Exporting	Firm Ownership Distribution		
		Avg. Mean Markup	Avg. Coefficient of Variation of Market Share	Avg. Firms Per Cluster	Avg. EG Agglomeration Index	Hirschman-Herfindahl Index	Avg. Mean Export Share	Avg. Percent DPE	Avg. Percent FIE	Avg. Percent SOE
1	0.01	1.224	4.3	1.69	0.0102	0.023	0.16	79	14	7
2	0.03	1.223	4.3	2.22	0.0100	0.020	0.14	81	14	5
3	0.04	1.223	4.5	2.92	0.0099	0.018	0.16	81	14	4
4	0.06	1.219	4.4	3.72	0.0100	0.017	0.19	79	17	4
5	0.08	1.225	4.4	4.35	0.0098	0.017	0.22	79	18	4
6	0.09	1.232	4.9	4.70	0.0099	0.017	0.24	75	21	4
7	0.11	1.234	4.4	4.99	0.0100	0.016	0.23	70	25	4
8	0.14	1.252	5.5	5.03	0.0096	0.018	0.28	62	33	5
9	0.17	1.280	5.0	4.57	0.0095	0.018	0.29	57	37	6
10	0.26	1.345	4.5	3.27	0.0095	0.022	0.20	54	34	12

Table 6: Baseline Results Using Low CV Deciles

	Dependent Variable: $\frac{1}{\mu_{nit}}$								
	(1) Province	(2) City	(3) County	(4) Province	(5) City	(6) County	(7) Province	(8) City	(9) County
Firm's share	-0.071*	-0.001	-0.080***				-0.060	0.015	-0.016
	(0.038)	(0.061)	(0.029)				(0.038)	(0.061)	(0.035)
Region's share				-0.013**	-0.013	-0.071***	-0.012**	-0.016	-0.064***
				(0.005)	(0.014)	(0.017)	(0.005)	(0.012)	(0.020)
Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	263053	154108	187120	263053	154108	187120	263053	154108	187120
Overall R^2	0.033	0.016	0.024	0.028	0.016	0.021	0.029	0.016	0.022

Notes: Robust standard errors clustered at firm level in parentheses. Significance: ***: 1%, **: 5%, *: 10%. Regions are defined at various aggregation levels, including province (in specifications 1, 4, and 7), city (in specifications 2, 5, and 8), and county (in specifications 3, 6 and 9). All specifications are regressions weighted by the number of observations for each two-digit CIC sector production function estimation reported (following De Loecker et al. 2014). All regressions include a constant term.

Table 7: Low CV Deciles with Region-Year Fixed Effects

	(1) Province	(2) City	(3) County	(4) Province	(5) City	(6) County	(7) Province	(8) City	(9) County
Firm's Share	-0.075*	-0.007	-0.079***				-0.066*	0.005	-0.014
	(0.039)	(0.057)	(0.028)				(0.039)	(0.058)	(0.034)
Region's Share				-0.011**	-0.011	-0.071***	-0.010*	-0.012	-0.065***
				(0.006)	(0.014)	(0.017)	(0.006)	(0.013)	(0.020)
Province-Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	263053	154108	187120	263053	154108	187120	263053	154108	187120
Overall R^2	0.032	0.031	0.017	0.022	0.032	0.017	0.023	0.032	0.017

Notes: Robust standard errors clustered at firm level in parentheses. Significance: ***: 1%, **: 5%, *: 10%. Regions are defined at various aggregation levels, including province (in specifications 1, 4, and 7), city (in specifications 2, 5, and 8), and county (in specifications 3, 6 and 9). All specifications include province-year fixed effects. Results are also robust to adding city-year or county-year fixed effects in specifications where regions are defined at city or county level. All specifications are regressions weighted by the number of observations for each two-digit CIC sector production function estimation reported (following De Loecker et al. 2014). All regressions include a constant term.

Online Appendix

A1. Derivation of the Demand Function

Suppose the household solves the following problem:

$$\max_{\{Y_i\}} \left(\sum_i D_i^{1/\gamma} Y_i^{\frac{\gamma-1}{\gamma}} \right)^{\frac{\gamma}{\gamma-1}} \tag{A1}$$

subject to:

$$\sum_i P_i Y_i \leq P$$

We take the budget of the household P to be exogenous. Cost minimization on the part of the representative household implies the demand function:

$$Y_i = D_i \left(\frac{P_i}{P} \right)^{-\gamma} \tag{A2}$$

The final product in each industry is assembled by competitive firms in each industry that solves:

$$P_i Y_i = \min_{\{y_{ni}\}} \sum_{n \in \Omega_i} p_{ni} y_{ni} \tag{A3}$$

subject to:

$$Y_i = \left(\sum_i y_{ni}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

Cost minimization on the part of these competitive firms implies:

$$y_{ni} = Y_i \left(\frac{p_{ni}}{P_i} \right)^{-\sigma} \tag{A4}$$

Combining equations (A2) and (A4) implies:

$$y_{ni} = D_i \left(\frac{P_i}{P} \right)^{-\gamma} \left(\frac{p_{ni}}{P_i} \right)^{-\sigma} \tag{A5}$$

A2. Proof of Proposition 1

Suppose marginal costs of all firms are bounded and non-decreasing. Proposition 1 has the following five parts:

1) If operating independently, firm markups are increasing in a firm's own market share,

2) If operating as a cartel, cartel markups are increasing in total cartel market share with each firm's own market share playing no additional role,

3) Firm markups are higher under cartel decisions than when operating independently,

4) Firm markups are more similar when operating as a cartel than when operating independently,

5) Firm market shares are more similar when operating independently than when operating as a cartel

PROOF:

Suppose any firm n in industry i weights the profits of the set of firms $S \subset \Omega_i$ with constant $\kappa \in [0,1]$. Then their objective is:

$$\max_{y_{ni}} p(y_{ni})y_{ni} - C(y_{ni}; X_{ni}) + \kappa \sum_{m \in S} [p(y_{mi})y_{mi} - C(y_{mi}; X_{mi})] \tag{A6}$$

Then for μ_{ni} defined as price divided by marginal cost and share defined as the firm's revenue divided by the sum of firm revenues in the industry, the firm's first order condition can be rewritten as:

$$\frac{1}{\mu_{ni}} = 1 + (1-\kappa)\frac{\partial \log(p_{ni})}{\partial \log(y_{ni})} + \kappa \sum_{m \in S} \frac{s_{mi}}{s_{ni}} \frac{\partial \log(p_{mi})}{\partial \log(y_{ni})} \tag{A7}$$

If inverse demand is given by:

$$p_{ni} = D_i y_{ni}^{-1/\sigma} \left(\sum_{m \in \Omega_i} y_{mi}^{1-1/\sigma} \right)^{\frac{\sigma}{\gamma}\frac{\gamma-1}{\sigma-1}-1} \tag{A8}$$

Then the cross-price elasticities are:

$$\frac{\partial \log(p_{mi})}{\partial \log(y_{ni})} = \left(\frac{1}{\sigma} - \frac{1}{\gamma} \right) s_{ni} \tag{A9}$$

The own-price elasticity is:

$$\frac{\partial \log(p_{ni})}{\partial \log(y_{ni})} = -\frac{1}{\sigma} + \left(\frac{1}{\sigma} - \frac{1}{\gamma} \right) s_{ni} \tag{A10}$$

Together these imply that:

$$\frac{1}{\mu_{ni}} = 1 - \frac{1}{\sigma} + \left(\frac{1}{\sigma} - \frac{1}{\gamma} \right) \left((1-\kappa)s_{ni} + \kappa \sum_{m \in S} s_{mi} \right) \tag{A11}$$

Firms operating independently is the case where $\kappa = 0$, so then:

$$\frac{1}{\mu_{ni}} = 1 - \frac{1}{\sigma} + \left(\frac{1}{\sigma} - \frac{1}{\gamma}\right) s_{ni} \tag{A12}$$

This implies result 1, when $\sigma > \gamma$. Likewise, if firms are operating as a perfect cartel, then $\kappa = 1$:

$$\frac{1}{\mu_{ni}} = 1 - \frac{1}{\sigma} + \left(\frac{1}{\sigma} - \frac{1}{\gamma}\right) \sum_{m \in S} s_{mi} \tag{A13}$$

This immediately implies the second result. Moreover, equations (A12) and (A13) together imply the fourth result, as cartels have no variation in markups (even if they have variation in market shares) while independent firms have markups that vary with their shares.

To compare firms in a cartel to those operating independently, we construct an artificial single firm that is equivalent to the cartel. That is, suppose $\kappa = 1$ so that the cartel solves:

$$\max_{\{y_{mi}\}} \sum_{m \in S} \left(p_{mi} y_{mi} - C(y_{mi}; X_{mi})\right) \tag{A14}$$

where p_{mi} is given by (A8). Now define a cartel aggregate of production:

$$Y = \left(\sum_{m \in S} y_{mi}^{1-1/\sigma}\right)^{\frac{\sigma}{\sigma-1}} \tag{A15}$$

Let $\tilde{C}(Y)$ be the cost function of the cartel defined as:

$$\tilde{C}(Y) = \min_{\{y_{mi}\}} \sum_{m \in S} C(y_{mi}; X_{mi}) \tag{A16}$$

$$\text{subject to:} \quad Y = \left(\sum_{m \in S} y_{mi}^{1-1/\sigma}\right)^{\frac{\sigma}{\sigma-1}}$$

Then the following problem is equivalent to (A14):

$$\max_{Y} D_i Y^{1-1/\sigma} \left(Y^{1-1/\sigma} + \sum_{n \notin S} y_{ni}^{1-1/\sigma}\right)^{\frac{\sigma}{\gamma}\frac{\gamma-1}{\sigma-1}-1} - \tilde{C}(Y) \tag{A17}$$

First notice that the Envelope Theorem applied to the problem in (A16):

$$\text{(A18)} \qquad \forall m \in S, \qquad \tilde{C}'(Y) = \lambda = \frac{C'(y_{mi}; X_{mi})}{y_{mi}^{-1/\sigma} Y^{1/\sigma}}$$

Then we can relate the size of the cartel to the cost of the cartel's production.

LEMMA 1: *Consider a cartel made up of in $T \subset S$. Then for every level of production Y, the marginal cost in the cartel composed of T is strictly higher than in the cartel composed of S.*

To prove this lemma, suppose y_{mi}^T is how much firm m produces when part of the cartel composed of T and y_{mi}^S is how much the same firm produces when part of the cartel composed of S. Then for any given Y it must be the case that:

$$y_{mi}^S < y_{mi}^T \implies \frac{C'(y_{mi}^S; X_{mi})}{{y_{mi}^S}^{-1/\sigma} Y^{1/\sigma}} < \frac{C'(y_{mi}^T; X_{mi})}{{y_{mi}^T}^{-1/\sigma} Y^{1/\sigma}} \implies \tilde{C^S}'(Y) < \tilde{C^T}'(Y)$$

where the second implication follows from the fact that all firms have non-decreasing marginal costs. The first inequality follows from bounded marginal costs and Inada conditions in the aggregation of individual firm production to cartel-level production. Therefore, if more firms are added to a cartel, marginal costs for the cartel are reduced for every level of output.

Given this lemma, notice that as a cartel grows, the markup that the cartel charges strictly increases. This follows immediately from that fact that, given the lemma, marginal costs decline so cartel production increases, and as another firm from within the same industry is brought into the cartel, that firm's production is no longer counted in the denominator when computing the cartel's market share. Therefore, the cartel's market share strictly increases as more firms are added. Hence, by (A13), the markup charged by the cartel increases.

A special case of this result is part 3 of Proposition 1. If a firm is operating outside of an existing cartel then is brought into it, the new cartel would have strictly higher markups than either the original cartel or the formerly independent firm.

To demonstrate the last result, consider any two firms n and m within the same cartel. Manipulating (A18) gives:

$$\text{(A19)} \qquad \frac{C'(y_{mi}; X_{mi})}{C'(y_{ni}; X_{ni})} = \left(\frac{y_{mi}}{y_{ni}}\right)^{-\frac{1}{\sigma}} = \left(\frac{s_{mi}}{s_{ni}}\right)^{\frac{1}{1-\sigma}}$$

Then consider two other firms v and w that are operating independently. Then the relationship between marginal cost and market share is:

$$\text{(A20)} \qquad \frac{C'(y_{vi}; X_{vi})}{C'(y_{wi}; X_{wi})} = \left(\frac{s_{vi}}{s_{wi}}\right)^{\frac{1}{1-\sigma}} \frac{1 - 1/\sigma + (1/\sigma - 1/\gamma)s_{vi}}{1 - 1/\sigma + (1/\sigma - 1/\gamma)s_{wi}}$$

Suppose these two pairs of firms have the same relative marginal costs. Then:

$$\text{(A21)} \qquad \frac{C'(y_{mi}; X_{mi})}{C'(y_{ni}; X_{ni})} = \frac{C'(y_{vi}; X_{vi})}{C'(y_{wi}; X_{wi})} \implies$$

$$\left(\frac{s_{mi}}{s_{ni}}\right)^{\frac{1}{1-\sigma}} = \left(\frac{s_{vi}}{s_{wi}}\right)^{\frac{1}{1-\sigma}} \frac{1 - 1/\sigma + (1/\sigma - 1/\gamma)s_{vi}}{1 - 1/\sigma + (1/\sigma - 1/\gamma)s_{wi}}$$

Without loss, if firms v and m have relatively high costs, then:

$$\text{(A22)} \qquad \frac{C'(y_{mi}; X_{mi})}{C'(y_{ni}; X_{ni})} = \frac{C'(y_{vi}; X_{vi})}{C'(y_{wi}; X_{wi})} > 1 \implies$$

$$\frac{1 - 1/\sigma + (1/\sigma - 1/\gamma)s_{vi}}{1 - 1/\sigma + (1/\sigma - 1/\gamma)s_{wi}} > 1 \implies \frac{s_{ni}}{s_{mi}} > \frac{s_{wi}}{s_{vi}}$$

Therefore, independently operating firms have wider variation in market shares conditional on marginal cost than do firms operating as a cartel. This completes the proof.

A3. Simulation of Model with Shocks to Demand and Costs

We now consider a version of the model where some uncertainty in costs or demand is realized after production choices are made. Firm i in industry j located in region k in year t solves the following problem:

$$\max_{l_{ijkt}} \int_{S_\varepsilon} \int_{S_\rho} \left[(1-\kappa)\pi_{ijkt}(l, \varepsilon, \rho) + \kappa \sum_{m \in \omega_{jkt}} \pi_{mjkt}(l, \varepsilon, \rho) \right] dF(\varepsilon) dG(\rho)$$

where:

$$\pi_{ijkt}(l, \varepsilon, \rho) = D_j (\varepsilon_{ijkt} l_{ijkt}^{1/\eta})^{1-1/\sigma} \left(\sum_{m \in \Omega_{jt}} (\varepsilon_{mjkt} l_{mjkt}^{1/\eta})^{1-1/\sigma} \right)^{\frac{\sigma}{\gamma}\frac{\gamma-1}{\sigma-1} - 1} - \rho_{ijkt} \frac{l_{ijkt}}{z_{ijkt}}$$

Here ε is the vector of demand shocks, ρ is the vector of cost shocks, and l is the vector of production choices. The set of firms operating in industry j at time t is Ω_{jt}, and its subset of firms operating within region k is ω_{jkt}. For any given firm, z_{ijkt} is the component of their costs that is known before production decisions are made. Without heterogeneity in this, there would be no heterogeneity in l_{ijkt}. The parameter η allows for curvature in the cost function.

Notice that F and G are probability distributions over vectors, and we will consider covariance at the cluster, industry and year levels.

The first order condition implies:

$$\int_{S_\rho} \frac{\eta \rho_{ijkt} l_{ijkt}^{1-1/\eta}}{z_{ijkt}} dG(\rho) =$$

$$= \int_{S_\varepsilon} p_{ijkt}(l,\varepsilon) \left[\frac{\sigma - 1}{\sigma} + \left(\frac{1}{\sigma} - \frac{1}{\gamma} \right) \frac{\kappa(\varepsilon_{ijkt} l_{ijkt}^{1/\eta})^{1-1/\sigma} + (1-\kappa) \sum_{n \in \omega_{jkt}} (\varepsilon_{njkt} l_{njkt}^{1/\eta})^{1-1/\sigma}}{\sum_{m \in \Omega_{jt}} (\varepsilon_{mjkt} l_{mjkt}^{1/\eta})^{1-1/\sigma}} \right] dF(\varepsilon)$$

where:

$$p_{ijkt}(l,\varepsilon) = D_j \varepsilon_{ijkt}^{1-1/\sigma} l_{ijkt}^{-1/\eta\sigma} \left(\sum_{m \in \Omega_{jt}} (\varepsilon_{mjkt} l_{mjkt})^{1/\eta(1-1/\sigma)} \right)^{\frac{\sigma}{\gamma}\frac{\gamma-1}{\sigma-1}-1}$$

Firms face a variety of shocks at different levels:

$$\varepsilon_{ijkt} = \nu_1 \varepsilon_t^1 + \nu_2 \varepsilon_{jt}^2 + \nu_3 \varepsilon_{ijkt}^3 + \nu_4 \varepsilon_{jkt}^4 + \nu_5 \varepsilon_{kt}^5$$

$$\rho_{ijkt} = \mu_1 \rho_t^1 + \mu_2 \rho_{jt}^2 + \mu_3 \rho_{ijkt}^3 + \mu_4 \rho_{jkt}^4 + \mu_5 \rho_{kt}^5$$

Therefore, we can separately analyze shocks at different levels.

Computational Implementation

The simulated dataset has T years, J industries and K regions. Every industry-region-year has I firms within it. The vectors ε and ρ are therefore of length $I \times J \times K \times T$. First, both ε and ρ are simulated M times. Then a vector L is drawn. Then L is input as the vector of production choices of firms. Using the first order condition, we then solve for the vector Z of anticipated costs that rationalizes the vector L. Together, Z, L, and the realization of shocks implies markups (using the method of De Loecker and Warzynski) and market shares for each firm. Then, for each realization, the regression described in the paper is run on the simulated data. This is done M times.

For these results we choose $\sigma = 5$, $\gamma = 3$, and $\kappa = 0.3$. We set $T = 11$, $J = 5$, $K = 8$, $I = 10$ and $M = 1000$. We assume that the log of each shock is a standard normal random variable.

Effects of Shocks: Comparative Statics

First we look at the effects of all twelve types of shocks individually. The table below presents the results of setting $\mu_1 = ... = \mu_5 = \nu_1 = ... = \nu_5 = 0$, then individually setting each to 1.

In each iteration of the simulation we run the following regression:

$$\frac{1}{\text{markup}_{ijkt}} = \alpha + \beta_1 s_{ijkt} + \beta_2 c_{jkt} + \delta_{ijkt}$$

where:

$$s_{ijkt} = \frac{(\varepsilon_{ijkt} y_{ijkt})^{1-1/\sigma}}{\sum_{m \in \Omega_{jt}} (\varepsilon_{mjkt} y_{mjkt})^{1-1/\sigma}}$$

$$c_{jkt} = \sum_{l \in \omega_{jkt}} s_{ljkt}$$

Here we present the simulated moments of $\hat{\kappa}$ defined by:

TABLE A1—SIMULATION RESULTS: EX POST SHOCKS

	No Fixed Effects			Region-Year and Firm FEs		
Cost Shocks:	Avg. $\hat{\kappa}$	St. Dev. $\hat{\kappa}$	Adj. R^2	Avg. $\hat{\kappa}$	St. Dev. $\hat{\kappa}$	Adj. R^2
Year	0.3000	0.0075	-0.0006	0.2995	0.0171	0.9999
Industry-Year	0.3000	0.0015	0.0030	0.3000	0.0024	0.1719
Firm-Year	0.0120	0.0562	0.0093	0.0044	0.0577	0.0130
Cluster-Year	0.9759	0.0059	0.0805	0.9982	0.0059	0.2628
Region-Year	0.2227	0.0332	0.0175	0.3178	0.0253	0.6229
Demand Shocks:	Avg. $\hat{\kappa}$	St. Dev. $\hat{\kappa}$	Adj. R^2	Avg. $\hat{\kappa}$	St. Dev. $\hat{\kappa}$	Adj. R^2
Year	0.3000	0.0065	-0.0006	0.3000	0.0145	0.9999
Industry-Year	0.2999	0.0086	-0.0006	0.2998	0.0149	0.1650
Firm-Year	0.0652	6.7790	-0.0001	0.0199	5.3935	-0.0003
Cluster-Year	1.1084	24.2151	0.0016	0.8891	5.2687	0.1846
Region-Year	0.0179	23.5850	0.0021	0.3018	0.0215	1.0000
	Firm FEs			Region-Year FEs		
Cost Shocks:	Avg. $\hat{\kappa}$	St. Dev. $\hat{\kappa}$	Adj. R^2	Avg. $\hat{\kappa}$	St. Dev. $\hat{\kappa}$	Adj. R^2
Year	0.3003	0.0124	-0.1004	0.3004	0.0098	1.000
Industry-Year	0.3000	0.0020	-0.0147	0.3000	0.0016	0.1694
Firm-Year	0.0035	0.0499	0.0125	0.0074	0.0605	0.0094
Cluster-Year	0.9986	0.0053	0.0984	0.9752	0.0066	0.2483
Region-Year	0.2253	0.0306	-0.0113	0.3101	0.0222	0.5584
Demand Shocks:	Avg. $\hat{\kappa}$	St. Dev. $\hat{\kappa}$	Adj. R^2	Avg. $\hat{\kappa}$	St. Dev. $\hat{\kappa}$	Adj. R^2
Year	0.2995	0.0124	-0.1004	0.2999	0.0085	1.0000
Industry-Year	0.2998	0.0126	-0.0187	0.2998	0.0111	0.1711
Firm-Year	-0.2294	15.1836	0.0002	-0.1389	6.0075	-0.0002
Cluster-Year	0.7296	8.6585	-0.0002	1.1787	2.6894	0.1893
Region-Year	-1.3071	53.0598	0.0092	0.3004	0.0110	0.9999

$$\hat{\kappa} \equiv \frac{\beta_2}{\beta_1 + \beta_2}$$

The results from these experiments are given in Table A.A3. We provide four sets of results based on the set of fixed effects considered, and for each case we provide the average and standard deviation of κ across the 1000 simulations. We also provide the adjusted R^2 averaged across the 1000 simulations.

These results demonstrate two important things to help understand how our estimates of κ could be biased. Firm-year shocks bias estimates of κ downward, and cluster-year and region-year shocks bias estimates upward. The region-year shocks can be mitigated with region-year fixed effects: the bias is almost eliminated for cost shocks and is less severe for demand shocks. In the other cases, the adjusted R^2 of the model can fall considerably, but we see little evidence of bias in estimates of κ.

Calibrated Example

The previous subsection demonstrates that the most serious bias arises when ex post shocks are at the firm-year and cluster-year level. We now repeat the numerical exercise from the previous section but now we parameterize the model to replicate the results of our baseline results in column 7 of Table 4. As in that regression, we include firm and year fixed effects and cluster standard errors at the firm level. We consider ex post shocks to productivity at the firm-year level and the cluster-year level, and we include measurement error at the firm-level. We also have idiosyncratic firm-year ex ante shocks. Each shock is assumed to be log-normal.

We calibrate six parameters: the variance of the three shocks, γ, σ, and κ. We match six moments: the coefficient estimate on the firm's own share and on the cluster's share, the standard errors on the firm's own share and on the cluster share, the average markup, and the regression's within-R^2.

The calibrated value of κ is 0.29, while the value in the model, as in the data, is 0.28. This demonstrates that, in this case, we actually underestimate the degree of collusion with our procedure relative to its true value. The calibrated standard deviation of the firm-year productivity shock is 0.011 while that of the cluster-year shock is 0.009. Our estimate of σ is 4.57 and γ is 2.74. The standard deviation of the measurement error is 0.044.

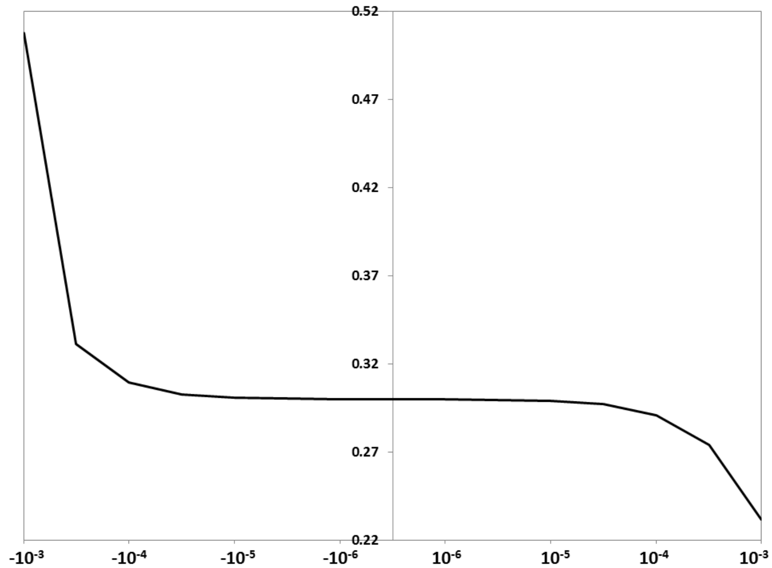
Departure from CES Demand

Next we consider the case where the demand system is instead given by:

$$q_{ijkt} = \left(\frac{p_{ijkt} + \bar{p}}{P_{jt}}\right)^{-\sigma} \left(\frac{P_{jt}}{P_t}\right)^{-\gamma} \tag{A23}$$

Proceeding with the same simulation technique as above, we consider the case where there are no ex post shocks and vary the magnitude of $\bar{p}$.

Figure A1. Varying Non-Homotheticity: Estimated κ



The results are summarized below in Figures A1 and A2. As the value of $\bar{p}$ varies, as shown on the horizontal axis in both figures, on average our measure of κ will be affected monotonically as shown in Figure A1. As before, the true value of κ in this simulation is equal to 0.3. Figure A2 shows that this bias is entirely due to bias in the coefficient on firms' own shares. In fact, the coefficient on cluster shares is unbiased by $\bar{p}$.

This supports our conclusion that a non-CES demand system of this type affects our estimate of the magnitude of collusion. However, if we interpret the t-test of whether or not the coefficient on the cluster share is positive to be a test of collusion, that test is unaffected by non-CES demand systems of this form.

Measurement Error

Finally, we consider the case where revenues are measured with error. We proceed as before, but now instead of unanticipated shocks, we study the effect of increases in the variances of the measurement error.

Following the parameterization in the first simulation exercise, A.A3 shows the effects of measurement error. In the "Idiosyncratic" columns, we assume that measurement error has no correlation across firms. In the "Cluster" columns, we consider the extreme case of correlation within clusters where measurement errors are equal in all firms of the same cluster.

Figure A2. Varying Non-Homotheticity: Coefficient Estimates

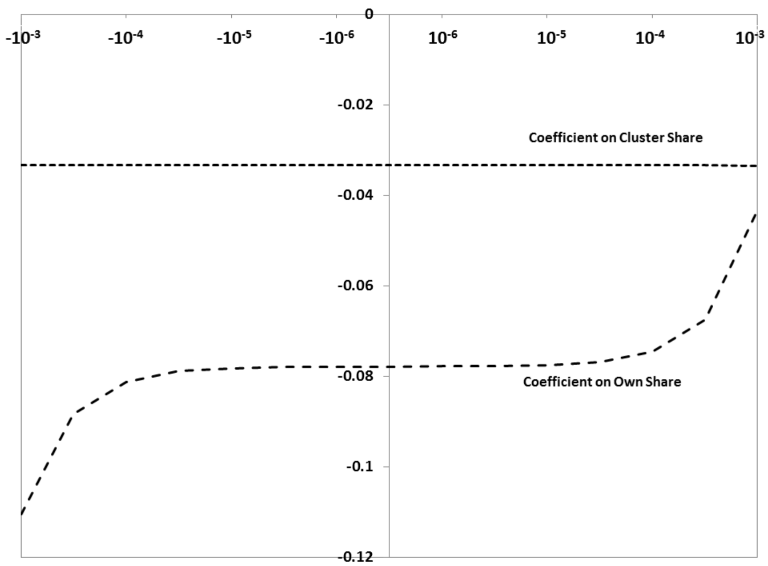


Table A2—Effects of Measurement Error

	Measurement Error, Idiosyncratic			
Var. of Error	Avg. $\hat{\kappa}$	St. Dev. $\hat{\kappa}$	Avg. $\hat{\beta}_1$	$Avg.\hat{\beta}_2$
0.1	0.2763	0.0520	-0.0899	-0.0337
0.2	0.2135	0.0781	-0.1280	-0.0333
0.3	0.1563	0.0812	-0.1926	-0.0335
0.4	0.1170	0.0746	-0.2736	-0.0341
0.5	0.0840	0.0672	-0.3846	-0.0328
	Measurement Error, Cluster			
Var. of Error	Avg. $\hat{\kappa}$	St. Dev. $\hat{\kappa}$	Avg. $\hat{\beta}_1$	$Avg.\hat{\beta}_2$
0.1	0.3602	0.0855	-0.0777	-0.0457
0.2	0.4954	0.1102	-0.0775	-0.0826
0.3	0.6322	0.1116	-0.0771	-0.1469
0.4	0.7253	0.0869	-0.0769	-0.2236
0.5	0.8033	0.0599	-0.0761	-0.3391

Table A.1: Appendix Table–Placebo Test Using Affiliate Sample

	Dependent Variable: $\frac{1}{\mu_{nit}}$									
	(1)	(2) Province	(3) City	(4) County	(5) Province	(6) City	(7) County	(8) Province	(9) City	(10) County
Firm's share	-0.009				-0.029	-0.016	-0.038	-0.066	-0.048	-0.064
	(0.060)				(0.063)	(0.067)	(0.080)	(0.065)	(0.070)	(0.082)
Region's share		0.019	0.004	0.013	0.021	0.007	0.030	0.028	0.009	0.026
		(0.021)	(0.030)	(0.040)	(0.022)	(0.033)	(0.054)	(0.023)	(0.035)	(0.056)
SEZ*Firm's share								0.094	0.100	0.100
								(0.137)	(0.137)	(0.137)
SEZ*Region's share								0.019	0.022	0.023
								(0.039)	(0.039)	(0.039)
SEZ Dummy								0.004	0.004	0.004
								(0.005)	(0.005)	(0.005)
Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	24780	24780	24780	24780	24780	24780	24780	20686	20686	20686
Overall R^2	0.010	0.012	0.010	0.010	0.012	0.010	0.010	0.010	0.009	0.009

Notes: Robust standard errors clustered at firm level in parentheses. Significance: ***: 1%, **: 5%, *: 10%. Regions are defined at various aggregation levels, including province (in specifications 2, 5, and 8), city (in specifications 3, 6, and 9), and county (in specifications 4, 7, and 10). All specifications are regressions weighted by the number of observations for each two-digit CIC sector production function estimation reported (following De Loecker et al. 2014). All regressions include a constant term.

Table A.2: Appendix Table–Placebo Test Using SOE Sample

	Dependent Variable: $\frac{1}{\mu_{nit}}$									
	(1)	(2) Province	(3) City	(4) County	(5) Province	(6) City	(7) County	(8) Province	(9) City	(10) County
Firm's share	-0.047				-0.029	-0.066	-0.061	-0.078	-0.101*	-0.147*
	(0.055)				(0.056)	(0.058)	(0.070)	(0.054)	(0.060)	(0.076)
Region's share		-0.024*	0.005	-0.020	-0.021	0.020	0.013	-0.010	0.016	0.060
		(0.013)	(0.023)	(0.037)	(0.014)	(0.025)	(0.048)	(0.016)	(0.028)	(0.055)
SEZ*Firm's share								0.052	0.050	0.050
								(0.120)	(0.120)	(0.119)
SEZ*Region's share								-0.014	-0.016	-0.016
								(0.039)	(0.039)	(0.039)
SEZ dummy								-0.000	-0.000	-0.000
								(0.004)	(0.004)	(0.004)
Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	106434	106434	106434	106434	106434	106434	106434	69608	69608	69608
Overall R^2	0.034	0.030	0.032	0.033	0.031	0.033	0.034	0.034	0.035	0.035

Notes: Robust standard errors clustered at firm level in parentheses. Significance: ***: 1%, **: 5%, *: 10%. Regions are defined at various aggregation levels, including province (in specifications 2, 5, and 8), city (in specifications 3, 6, and 9), and county (in specifications 4, 7, and 10). All specifications are regressions weighted by the number of observations for each two-digit CIC sector production function estimation reported (following De Loecker et al. 2014). All regressions include a constant term.

Table A.3: Appendix Table–Rauch Product Classification Results

	Dependent Variable: $\frac{1}{\mu_{nit}}$					
	(1) homo/ref	(2) diff.	(3) overall	(4) homo/ref	(5) diff.	(6) overall
Firm's Share	-0.167**	-0.074***	-0.127**	-0.141	-0.046*	-0.179***
	(0.080)	(0.023)	(0.051)	(0.202)	(0.024)	(0.040)
Region's Share	-0.076***	-0.026***	-0.067***	-0.296***	-0.016*	-0.070***
	(0.011)	(0.008)	(0.010)	(0.089)	(0.009)	(0.009)
Differentiated X Firm's Share			0.042			0.126***
			(0.056)			(0.044)
Differentiated X Region's Share			0.045***			0.056***
			(0.013)			(0.012)
Differentiated Dummy			-0.002			-0.001
			(0.001)			(0.001)
Year FEs	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES
Observations	273327	935702	1279149	75692	642132	1279149
Overall R^2	0.036	0.024	0.027	0.015	0.019	0.027

Notes: Robust standard errors clustered at firm level in parentheses. Significance: ***: 1%, **: 5%, *: 10%. Specifications 1-3 refer to product classification using "most frequent" principle; specifications 4-6 refer to product classification using "pure" principle. All specifications are regressions weighted by the number of observations for each two-digit CIC sector production function estimation reported (following De Loecker et al. 2014). All regressions include a constant term.

Table A.4: Appendix Table–Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Dependent Variable = $1/\mu_{nit}$								
Firm's Share	-0.143***		-0.099***	-0.163***	-0.083***		-0.046**	-0.089***
	(0.021)		(0.022)	(0.027)	(0.017)		(0.018)	(0.023)
Region's Share		-0.053***	-0.040***	-0.031***		-0.044***	-0.036***	-0.024***
		(0.006)	(0.006)	(0.007)		(0.006)	(0.006)	(0.007)
SEZ*Firm's Share				0.095**				0.073**
				(0.039)				(0.035)
SEZ*Region's Share				-0.018*				-0.021**
				(0.010)				(0.010)
Observations	1346860	1346860	1346860	1105162	1346860	1346860	1346860	1105162
Overall R^2	0.028	0.027	0.028	0.026	0.028	0.027	0.027	0.026
Panel B: Dependent Variable = $\mu_{nit}/(\mu_{nit}-1)$ *(full sample)*								
Firm's Share	225.521		195.421	388.275	469.783		439.703	800.015
	(200.154)		(204.416)	(318.785)	(474.848)		(480.277)	(772.947)
Region's Share		53.610	27.789	18.889		109.975	28.807	15.500
		(33.033)	(25.908)	(31.765)		(88.601)	(33.949)	(41.299)
SEZ*Firm's Share				-369.619				-694.560
				(251.815)				(525.775)
SEZ*Region's Share				28.451				38.929
				(26.538)				(35.645)
Observations	1346860	1346860	1346860	1105162	1346860	1346860	1346860	1105162
Overall R^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Panel C: Dependent Variable = $\mu_{nit}/(\mu_{nit}-1)$ *(drop* $\mu_{nit} < 1.06$*)*								
Firm's Share	-3.229***		-1.985***	-3.783***	-2.471***		-1.311***	-2.488***
	(0.561)		(0.594)	(0.745)	(0.464)		(0.500)	(0.692)
Region's Share		-1.404***	-1.140***	-1.051***		-1.339***	-1.103***	-0.932***
		(0.170)	(0.183)	(0.220)		(0.161)	(0.175)	(0.208)
SEZ*Firm's Share				3.248***				2.545***
				(0.964)				(0.905)
SEZ*Region's Share				-0.272				-0.372
				(0.259)				(0.251)
Observations	1228255	1228255	1228255	1006748	1228255	1228255	1228255	1006748
Overall R^2	0.010	0.009	0.009	0.009	0.010	0.009	0.009	0.009
Panel D: Dependent Variable = $log(mu)_{nit}$								
Firm's Share	0.174***		0.118***	0.193***	0.099***		0.051**	0.102***
	(0.027)		(0.028)	(0.035)	(0.022)		(0.024)	(0.030)
Region's Share		0.067***	0.052***	0.035***		0.055***	0.046***	0.028***
		(0.008)	(0.008)	(0.009)		(0.007)	(0.008)	(0.009)
SEZ*Firm's Share				-0.103*				-0.085*
				(0.053)				(0.047)
SEZ*Region's Share				0.034**				0.035***
				(0.014)				(0.013)
Observations	1346860	1346860	1346860	1105162	1346860	1346860	1346860	1105162
Overall R^2	0.028	0.027	0.028	0.027	0.028	0.027	0.027	0.026
All Panels								
Year FEs	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Robust standard errors clustered at firm level in parentheses. Significance: ***: 1%, **: 5%, *: 10%. Regions are defined at county level. Specifications 1-4 are weighted regressions; specifications 5-8 are unweighted regressions. All regressions include a constant term and SEZ dummy.

Table A.5: Appendix Table–Instrumental Variable Estimation Results Using Low CV Deciles

	Dependent Variable: $\frac{1}{\mu_{nit}}$								
	(1) Province	(2) City	(3) County	(4) Province	(5) City	(6) County	(7) Province	(8) City	(9) County
Firm's share	0.007	-0.383***	-0.224***				0.100	-0.229*	-0.113***
	(0.130)	(0.114)	(0.052)				(0.178)	(0.121)	(0.037)
Region's share				0.010	-0.042***	-0.057***	0.018	-0.030***	-0.048***
				(0.016)	(0.011)	(0.016)	(0.022)	(0.012)	(0.017)
Year FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	258745	154012	187050	258745	154012	187050	258745	154012	187050
Overall R^2	0.031	0.014	0.023	0.031	0.015	0.023	0.031	0.015	0.023
First-Stage Instruments:	(Sum of other firms' productivity; Sum of outside-cluster firms' productivity)								
Weak Instrument (Prob > F)	0.0000								

Notes: Robust standard errors clustered at firm level in parentheses. Significance: ***: 1%, **: 5%, *: 10%. Regions are defined at various aggregation levels, including province (in specifications 1, 4, and 7), city (in specifications 2, 5, and 8), and county (in specifications 3, 6 and 9). All specifications are regressions weighted by the number of observations for each two-digit CIC sector production function estimation reported (following De Loecker et al. 2014). All regressions include a constant term.