

Market structure, oligopsony power and productivity

Michael Rubens*

April 15, 2020

[Latest version here](#)

Abstract

This paper examines the effects of ownership consolidation among Chinese cigarette manufacturers on monopsony power, product market power and productive efficiency. I combine a structural model of production and input market competition with a natural experiment in which firms under certain production quantity thresholds were forced to exit. This consolidation led to an increase in tobacco leaf price markdowns of 37%, while cigarette price markups fell by 23%. Although the objective of the consolidation policy was to increase efficiency, I do not find strong evidence for such gains. Rising monopsony power has important distributional consequences. The policy of forcing small manufacturers to exit explains 42% of the increase in income inequality between farmers and manufacturing workers in the tobacco industry between 2003 and 2006.

Keywords: Market Power, Monopsony, Concentration, Productivity, Inequality

JEL Codes: L10, J42, O25, D33

*KU Leuven and Research Foundation Flanders. E-mail: michael.rubens@kuleuven.be.

This paper previously circulated under the title ‘Ownership consolidation, monopsony power and efficiency: Evidence from the Chinese tobacco industry’. I am grateful to Jan De Loecker and Jo Van Biesebroeck for their mentoring and advice. I also thank Frank Verboven, Chad Syverson, Otto Toivanen, Dan Akerberg, Myrto Kalouptsi, and Benjamin Vatter for their helpful suggestions, and participants at EARIE 2019, IIOC 2019, the Warwick Economics PhD Conference and KU Leuven IO and Development workshops. I thank Yale’s Economics department for hosting me as a visiting PhD student. This is the first chapter of my Ph.D. dissertation at KU Leuven. Financial support from the Research Foundation Flanders (Project 11A4618N), Fulbright Belgium and Belgian American Educational Foundation are gratefully acknowledged. Any remaining errors are my own.

1 Introduction

There is an ongoing debate about the economy-wide evolution of market concentration (Rossi-Hansberg, Sarte, & Trachter, 2018; Autor, Dorn, Katz, Patterson, & Van Reenen, 2017; Covarrubias, Gutiérrez, & Philippon, 2019), and of markups (De Loecker, Eeckhout, & Unger, 2018; De Loecker & Eeckhout, 2018). In theory, the degree of market concentration can affect both competition and productive efficiency. Ownership consolidation may induce firms to change their markup of prices over marginal costs, but they could also adjust the extent to which input prices are marked down below marginal input products.¹ Moreover, production could become more efficient, for instance due to increasing returns to scale or scale economies. Empirical work on ownership consolidation tends to focus on one or two of these channels while ruling out the others. In order to fully understand the welfare consequences of ownership consolidation, however, all three potential effects need to be jointly examined.

This paper fills this gap by examining the effects of ownership consolidation in the Chinese cigarette manufacturing industry on both cigarette price markups, tobacco leaf price markdowns, *and* productive efficiency. Besides the fact that this is a large industry,² it also provides the ideal setting to study consolidation due to quasi-experimental variation in market structure. In 2003, the Chinese government initiated a large consolidation wave during which cigarette manufacturers under specific output thresholds were forced to close down. This variation is useful because other sources of market structure variation, such as mergers and acquisitions and exit and entry, tend to be endogenous to productivity and market power.

Another interesting feature of the Chinese tobacco industry is that market concentration increases along the value chain. Around 20 million farmers sell leaves to manufacturers, with the number of manufacturers decreasing from 350 to 150 during the consolidation. In turn, manufacturers sell cigarettes domestically to a monopsonic government-controlled wholesaler.³ We can thus expect monopsony power to be present along the chain.⁴

The key driver of monopsony power on leaf markets is the legal obligation of tobacco farmers to sell their entire output at local purchasing stations, where only local manufacturers compete for leaves. Transporting leaves outside these small isolated markets is illegal.⁵ Such localized markets

¹The relationship between concentration, markups and markdowns is theoretically ambiguous, see Syverson (2019).

²Annual industry revenue exceeds \$7 billion, and 40% of the world's cigarettes are made in China.

³The tobacco industry remained largely domestic even after China's WTO accession, as exports make up for less than 1% of industry revenue. This eliminates various potentially confounding factors which relate to international trade.

⁴Public health externalities are, finally, an idiosyncratic aspect of the tobacco industry. I will abstract from these health concerns in this paper.

⁵Counterfeiting and leaf trade on black markets does happen in China. In the absence of data on such illegal transactions, I abstract from illegal trade flows.

are widespread across the rural developing world, and are, for instance, a driver of monopsony power on Indian agricultural markets as well (Chatterjee, 2019). Although farmers can switch crops, this is costly: an experiment with crop substitution from 2008 found that switching to other crops increased the annual revenue per acre of Chinese tobacco farmers by 21% to 110% (Li, Wang, Xia, Tang, & Wang, 2012). The fact that farmers forgo these large revenue gains suggests large crop switching costs.⁶ Leaving agriculture altogether is another possibility, but mostly involves migration. Both the Hukou permit system and land tenure uncertainty increase rural-urban migration costs (Minale, 2018). Finally, county governments in tobacco-rich provinces are heavily dependent on fiscal revenue from the tobacco industry, and there are reports of tobacco farmers being coerced not to switch towards other crops by local authorities (Wang, 2013).

I start by providing reduced-form evidence for the effects of the consolidation on both leaf and cigarette prices. I compare manufacturers which competed with firms that produced below the exit threshold at the onset of the consolidation in 2003 (the treatment group) with manufacturers without such competitors (the control group). I find that for firms in the treatment group, leaf prices fell by 40% to 64% more compared to firms in the control group between 2003 and 2006. Labor wages did, in contrast, fall by 6% to 15%, but this drop was not statistically significant. Finally, factory-gate cigarette prices fell between 24-34% more for the treatment group. Although this reduced-form evidence is arguably causal, it is not sufficient to draw conclusions about the underlying mechanism. Prices could have changed due to changes in market power on product or input markets, but also due to changes in productive efficiency. In order to identify the exact mechanisms through which the consolidation affected prices, more structure is needed.

I therefore construct and estimate a structural model in which cigarette price markups, leaf price markdowns and manufacturing efficiency are jointly identified. I combine firm-level data on costs, prices and production quantities and build on the cost-side approach of Hall (1986) and De Loecker and Warzynski (2012), with the key distinction that I allow for a subset of inputs to be non-substitutable.⁷ Substitutability between tobacco leaf and manufacturing workers is low, so I use a Leontief model in leaf and labor.⁸ I combine this approach with a model of leaf supply.⁹ I start by estimating the production function for cigarettes, and discuss the conditions under which prior control function approaches identify the production function when input prices are endogenous. Next, I use the productivity resid-

⁶These switching costs are a key feature of the agricultural economics literature (Song, Zhao, & Swinton, 2011; Scott, 2013).

⁷In their analysis of market power in the beer industry, De Loecker and Scott (2016) also allowed for a non-substitutable input, but not for input market power.

⁸I estimate the elasticity of substitution between leaf and labor and find that a Leontief cannot be rejected, while a Cobb-Douglas production function can be rejected.

⁹More structure on input market competition was also placed in Tortarolo and Zarate (2018), but with a very different identification strategy, and while still relying on substitutability between all inputs.

uals as input demand shifters to identify the tobacco leaf supply function and infer markdowns. The condition for this exclusion restriction to hold is that, conditional on the leaf price, manufacturing productivity shocks are excluded from farmer utility. Finally, I bring the production model and input supply model together to infer markups without imposing any additional assumptions on how manufacturers compete on cigarette markets.

I find that markups of cigarette producers were not significantly different from one, meaning that prices were equal to marginal costs. Leaf price markdowns were, in contrast, large: the average cigarette manufacturer paid its tobacco farmers 38% of their marginal revenue product. This is much less compared to most prior monopsony studies such as Goolsbee and Syverson (2019) for US tenure-track college professors (83%) and Oaxaca and Ransom (2010) for US grocery clerks (70%). The combination of low markups and high markdowns of manufacturers is consistent with the fact that they bought from many small farmers, but sold to a single large buyer.¹⁰

Having estimated both markups, markdowns and productive efficiency, I examine how the consolidation policy affected these three outcomes. I find that markdowns increased on average by 37% more in the treatment group relative to the control group when defining input markets at the province-level, and by even more when defining input markets more narrowly. The markdown increase was the largest in markets with high numbers of immigrants, non-registered inhabitants and high unemployment rates. Markups of manufacturers fell, in contrast, by 23% as the wholesaler used its own buying power to decrease factory-gate prices.

Finally, I examine the effects of the consolidation on redistribution and scale economies. By increasing markdowns on tobacco leaf markets, but not on manufacturing labor markets, the consolidation led to an increase of 42% in income inequality between farmers and manufacturing workers. This surge in rural-urban inequality was not in line with official policy objectives, as laid out in President Hu Jintao's *Harmonious Society* program during the mid-2000s. In 2017, the 13th five-year plan introduced targeted subsidies to alleviate poverty among tobacco farmers.¹¹ Such transfer schemes may not have been necessary in the absence of a consolidation. Although the policy's official objective was to increase efficiency by exploiting scale economies, I do not find strong evidence for either increasing returns to scale or other scale economies.

The principal contribution of this paper is to study the effects of ownership consolidation on both markups, markdowns and productivity together. There is a large empirical literature on market power and ownership consolidation, such as Miller and Weinberg (2017); Nevo (2001), which mostly builds on the 'demand-side' approach of S. Berry, Levinsohn, and Pakes (1995). In principle, this approach

¹⁰High markdowns are also consistent with widespread poverty among Chinese tobacco farmers, in contrast to most other tobacco-growing countries where tobacco ranks high among crops in terms of profitability (FAO, 2003).

¹¹<https://www.tobaccoasia.com/features/china-aims-to-increase-tobacco-farmers-income/>

can be used to identify the effects of consolidation on both markups, markdowns and productive efficiency. Doing so requires imposing a model on how firms compete on product markets and input markets, and identifying both the product demand and input supply curves. Moreover, observable and exogenous variation in market structure is necessary to identify efficiency gains. In practice, applications using this approach have tended to assume either input prices or efficiency gains to be exogenous to ownership consolidation.

Alternatively, the production and cost approach to markups of Hall (1986); De Loecker and Warzynski (2012) can be used as well. If all inputs are substitutable, then this approach can be used to identify both markups, markdowns and productive efficiency without imposing a model of competition on either input or product markets, but at the cost of having to identify the production function.¹² I show, however, that joint identification of markups and markdowns using this approach alone fails whenever firms have buying power over non-substitutable inputs. In many other industries, a subset of intermediate inputs are not substitutable, or substitutable to a limited extent. Examples include beer brewing (hop), coffee roasting (beans) and consumer electronics (rare earth metals).¹³

I solve this identification challenge by combining the production and cost approach with a model of input supply. This allows identification of markups, markdowns and productive efficiency without imposing a model of how firms compete on cigarette markets. This modeling choice is driven by the institutional setting of the Chinese tobacco industry: it is far easier to model leaf supply than cigarette demand. Tobacco leaf is sold on static spot markets that are isolated by law. In contrast, cigarette markets are not well-delineated, conduct is unknown due to the government-owned wholesaler, retail prices are unobserved, and cigarette demand is dynamic due to addiction. In other settings, however, it may be preferred to combine a model of product demand and production, or a model of product demand and input supply.

I find that ignoring non-substitutability in the production and cost approach leads to mis-inference of both markups and markdowns. Markups are biased upwards when inputs are erroneously assumed to be substitutable. In the substitutable inputs production model, marginal costs can be expressed for input separately, and depend on the input price and the markdown. In the Leontief production model, however, the prices and markdowns of all inputs enter marginal cost, because all inputs need to be moved together to change output. Markdowns are, conversely, underestimated. The reason for this is that an output elasticity below one will be imposed for the Leontief input. I illustrate these sources of bias in my application: assuming substitutable tobacco leaf delivers markup and markdown averages

¹²This was shown in De Loecker, Goldberg, Khandelwal, and Pavcnik (2016) and applied in, among others, Morlacco (2017); Brooks, Kaboski, Yao, and Qian (2019)

¹³Even if intermediate inputs are partially substitutable, but to a lower degree than implied by a Cobb-Douglas production function, the implications of this paper still matter. Markdowns and markups would then be weakly identified, rather than non-identified. One could, for instance, think about settings in which firms can substitute in-house production of intermediate inputs with outsourcing.

of 5.63 and 0.77, while these are 1.06 and 2.87 when using the Leontief production function.

I make three additional contributions to the literature. First, I contribute to the literatures about efficiency gains from consolidation (Braguinsky, Ohyama, Okazaki, & Syverson, 2015; Blonigen & Pierce, 2016; Grieco, Pinkse, & Slade, 2017) and State Owned Enterprise (SOE) reform and privatization (Gupta, 2005; Hsieh & Song, 2015; Chen, Igami, Sawada, & Xiao, 2018). These papers usually find evidence for large TFP gains from SOE privatization. Hsieh and Song (2015) finds that consolidation policies similar to the one studied in this paper led to an increase in aggregate TFP of 20% across all Chinese manufacturing industries. My contribution to this literature is to allow for endogenous input prices, which changes the interpretation of the productivity residual. Using a production function with exogenous leaf prices leads to the conclusion that average TFP increased by 30% due to the consolidation. In reality, leaf prices fell as monopsony power increased on leaf markets. This has important policy implications: large-scale consolidation policies, such as the one studied in this paper, are increasingly implemented. China consolidated many of its SOEs into industrial giants in various important industries such as energy, transport utilities, telecommunication and defense industries.¹⁴ Other countries, such as Indonesia, have recently adopted similar policies as well. If these reforms lead to rising monopsony power rather than productivity growth, this changes their welfare effects.

Second, I contribute to the literature on monopsony power, such as Card and Krueger (1994); Manning (2003).¹⁵ Recent papers on this topic includes Matsudaira (2014) and Wollmann (2019) for nurses, Naidu, Nyarko, and Wang (2016) for migrant workers in the UAE, and Goolsbee and Syverson (2019) for US academics. These papers usually estimate an input supply function. Valid input prices instruments are hence crucial, but hard to find.¹⁶ I propose using estimated productivity shocks from the production model as input demand shifters to identify the input supply curve, and discuss the conditions under which the underlying exclusion restriction holds.

Third, I contribute to the literature on rural-urban income inequality in developing countries. Many papers have been devoted to this margin of inequality in China, as it has increased rapidly since the early 1990s (Yang, 1999; Benjamin, Brandt, & Giles, 2005; Ravallion & Chen, 2009). My contribution consists of showing that SOE consolidation and the ensuing rise in monopsony power on agricultural markets was an important driver of this surge in inequality, at least in the tobacco industry.¹⁷ In a closely related paper, Chatterjee (2019) examines the consequences of interstate agricultural trade restrictions on farm income. In contrast to that paper, I focus on a different policy measure, SOE

¹⁴These policies are known as “*Grasping the large and letting the small go*” (Naughton, 2007).

¹⁵A review with empirical work is in Ashenfelter, Farber, and Ransom (2010).

¹⁶S. Berry, Azar, and Marinescu (2019) gives an overview of potential instruments that can be used.

¹⁷While the relationship between market power and income inequality has been studied before, e.g. in De Loecker and Eeckhout (2018); De Loecker et al. (2018), their focus has been mainly on product market power rather than on monopsony power.

consolidation reforms.

The remainder of this paper is structured as follows. In the next section, I briefly summarize the industry background and the data, and provide reduced-form evidence on how prices were affected by the consolidation. In section 3, I present the model and discuss identification of markups, markdowns and productivity. Section 4 contains the estimation and the main results, while the distributional consequences and scale benefits are discussed in section 5. I revisit the main modeling assumptions and present some extensions in section 6, after which I conclude.

2 The Chinese tobacco industry

2.1 Industry background

This paper focuses on the production of cigarettes in China, the value chain of which is visualized in figure 1. Cigarette manufacturing firms, which are the entity observed in the data, buy tobacco leaf from farmers at ‘purchasing stations’. In 1997, there were around 20 million tobacco farmers in China, which were mostly organized at the household level and operated small plots of around 0.3-0.4 ha (FAO, 2003). Tobacco leaf needs to be ‘cured’ before it can be consumed, and various curing processes are possible.¹⁸ Farmers must sell their leaf at their nearest purchasing station. Only official cigarette manufacturing firms are allowed to purchase tobacco leaf at these stations, and transporting tobacco leaf beyond these isolated markets without official approval is strictly forbidden by law (State Council of the People’s Republic of China, 1997). Cigarette manufacturers, of which there were 150-350 during the time period studied, can buy at more than one station, but only within the boundaries of their own provinces. Tobacco leaves are sorted into quality ‘grades’, each of which sells at a different price.

[Figure 1 here]

Four out of five manufacturers in the dataset are owned by county, prefectural or provincial governments. All manufacturing firms are formally part of a state-owned holding, the *Chinese National Tobacco Corporation*. In practice, they are autonomous in how they operate and set input prices and how they compete against each other (Peng, 1996; Wang, 2013). SOEs are also controlled by local or provincial governments, rather than the central government. They sell their cigarettes to wholesalers which are controlled by the State Tobacco Monopoly Administration (STMA) through its commer-

¹⁸Examples include air curing, fire curing and flue curing.

cial counterpart, the *Chinese National Tobacco Trade Corporation* (CNTTC).¹⁹ This organization is centrally controlled and operates a monopoly on the cigarette market. In contrast to tobacco leaf, cigarettes are transported within and sold throughout China (State Council of the People's Republic of China, 1997). The distinction between centrally controlled wholesaling and decentralized manufacturing has been at the core of the STMA system since its inception in the early 1980s.

Even after China acceded to the WTO in 2001, the Chinese tobacco industry has been shielded from international competition. Industry-wide exports and imports were merely 1.0% and 0.2% of total industry revenue between 1998 and 2007 (United Nations, 2019). The fiscal importance of the tobacco industry may be an important reason for this protection: in 1997, tobacco taxes and monopoly profits made up for 10.4% of central government revenue. In 2015, tax revenues from the cigarettes industry amounted to ¥840 B, which is 6.2% of China's total tax revenue (Goodchild & Zheng, 2018; State Administration of Taxation, 2015).

Cigarette production process

Cigarette factories turn 'cured' tobacco leaf, paper and filters into cigarettes using labor and capital. Intermediate inputs make up for 90% of variable input expenditure, and tobacco leaf accounts for two thirds of intermediate input expenditure.²⁰ Manufacturers treat, shred and compress the leaf, and then insert it into cigarettes, cigars or other tobacco products. While the focus of the analysis will be on cigarette manufacturers, I also include other tobacco users such as cigar and chewing tobacco producers in the market definitions, as these compete for leaf as well. They account for less than 5% of industry revenue, however. A picture of the consecutive steps in the cigarette production process are in panels (b)-(d) of figure 1. The focus will be on firms, rather than individual factories, as cigarette factories are usually located in urban areas, often in the prefectural seat.

A map of tobacco manufacturing locations in China is in figure 2, with dots indicating counties with at least one manufacturing firm. The tobacco industry is scattered across the country, even if the most important tobacco provinces, such as Yunnan, Guizhou, Sichuan and Henan, are in the south. Comparing the map of locations in 1999 and 2006 reveals that the number of counties with at least one firm decreased drastically over this period of time. In 1999, there were 339 firms in 268 counties. By 2006, 160 firms were left in 121 counties.

[Figure 2 here]

¹⁹STMA and CNTTC share most of their leadership (Wang, 2013)

²⁰The Chinese data do not break down intermediate inputs into more detailed categories, but US census data from 1997 show that tobacco leaves make up for for 60% of all intermediate input costs in tobacco manufacturing firms (U.S. Census Bureau, 1997b)

Farmers

Tobacco farming is not a very profitable activity in China. In 1997, just before the sample period of this paper, tobacco was the median cash crop in terms of profitability (FAO, 2003). By 2004, however, it had become the least profitable cash crop (Hu et al., 2006). Although tobacco farmers can switch to other crops, this entails large switching costs. A policy intervention in which tobacco farmers were helped to substitute crops in 2008 found that substituting increased annual revenue per acre by 21% to 110% (Li et al., 2012). The fact that farmers do not substitute despite these potential gains implies that crop switching costs are large. Farmers can also exit agriculture altogether, but rural mobility is constrained due to the Hukou registration system. Some sources also make mention of tobacco farmers being coerced not to switch crops by local politicians, due to the importance of tobacco for local fiscal revenue (Peng, 1996). Land tenure insecurity does, finally, also make migration more costly. Rural land is the property of villages or collectives, and if households move they lose their exclusive use rights (Minale, 2018). This makes migration more costly and risky.

2.2 Data

For the main analysis in this paper, I combine two datasets. First, I use firm-level production and cost data between 1999 and 2006 from the *Annual Survey of Industrial Firms*, which is conducted by the National Bureau for Statistics (NBS). I retain all firms with CIC codes 1610, 1620 and 1690, which together correspond to HS code 24, “Tobacco and Manufactured Tobacco Substitutes”. This results in 478 firms and 2139 observations. The above-scale survey includes non-SOEs with sales exceeding 5 million RMB and all SOEs irrespective of their size. I refer to Brandt, Van Biesebroeck, and Zhang (2012) for a comprehensive discussion of this dataset. Second, I obtain product-firm-month level production quantities during the same time period, again from the NBS. Quantities are observed for only 1,260 observations and 274 firms.²¹ In addition to these data, I match county-level population census data and brand-level cigarette characteristics to the dataset as well. I use this data for the robustness checks. More details about all datasets used and data cleaning is in appendix A.

²¹Some sample selection may be going on due to missing quantities. Firms for which quantities are unobserved have on average less employees. The labor and material shares of revenue are, however, not significantly different between firms with and without observed quantities. Whether quantities are observable explains barely any variation in revenue shares.

Summary statistics

Summary statistics of selected variables are in table 1. The average firm earned a revenue of \$105 million (in 2006 US dollars) and sold 340,000 cases per year. The average factory-gate price for a case of 50,000 cigarettes was \$1623, so the price for a pack of 20 cigarettes was on average \$0.65. The 5th and 95th percentiles of pack prices were \$0.02 and \$1.16. Using retail price data from Nargis et al. (2019), this means that factory-gate prices were on average around 20% to 30% of retail prices, and the difference between both includes wholesale margins, retail margins, transport costs and taxes.²²

[Table 1 here]

2.3 Reduced-form evidence: consolidation and prices

The consolidation

The number of tobacco manufacturing firms fell from 351 in 1998 to 148 in 2007. In its annual report from 2000, the STMA decided that “competitive large enterprise groups” had to be formed, without specifying a concrete timing (Wang, 2013).²³ The main motivation for this policy was to “enable China’s cigarette industry to achieve scale and efficiency” (STMA, 2002). The top chart in figure 3 shows that the number of manufacturers indeed started to decrease from that time onwards. In May 2002, the STMA published a concrete implementation plan which ordered firms producing less than 100,000 cigarette cases per year to be closed down in 2003,²⁴ while firms with an annual production below 300,000 cases were encouraged to merge with larger firms.²⁵

The second graph in figure 3 compares the number of firms which produce less and more than 100,000 cases per year. As quantities are observed for only a subset of firms, the annual number of firms reported is lower compared to the previous graph. The number of firms under the exit threshold fell sharply between 2002 and 2004, from 98 to 25, while the number of firms above the thresholds fell from 98 to 66. These smaller firms were economically meaningful: 36% and 46% of firms produced less than 100,000 and 300,000 cases respectively in 2002, generating 6% and 11% of industry

²²Harris (1998) reports that US wholesale prices were 59% of retail prices in 1998. Assuming a similar retail margin for China would leave 29% to 39% of the retail price as profit for the wholesaler. This is a large margin, which is consistent with the STMA’s monopoly power in wholesaling.

²³I test for announcement effects in appendix C.1

²⁴One case contains 50,000 sticks of cigarettes (Fang, Lee, & Sejpal, 2017)

²⁵The thresholds were calculated based on production in 2002. In appendix C.1, I test for bunching just above the exit threshold, and find no evidence for this.

revenue. As figure 3 shows, average market shares of the provincial market leaders increased from below 50% to 70% between 2002 and 2007. For the two largest firms in each province, joint market shares increased from 70% to 85%.

[Figure 3 here]

Of all firms that produced below the exit threshold, 22 firms continued to exist after 2003. Among these firms, 12 were not state-owned, so the STMA could not force them to close down. A further 7 were dropped during the data cleaning procedure due to anomalies such as negative intermediate input expenditures. Merely 3 SOEs below the exit threshold survived after 2003, for unknown reasons.

Appendix table A2 shows the ‘first stage’ regressions of how input market Hirschman-Herfindahl indices changed differently between firms with and without competitors below the exit threshold in 2002. In markets which were subject to the consolidation policy, the HHI indices for both materials and labor increased by a third compared to control markets when using province-level market definitions. At the prefecture-level, the concentration indices increased by a fifth, and at the county level by a one seventh.²⁶

Factor revenue shares

Panel (a) of figure 4 plots the evolution of total labor and intermediate input expenditure over total revenue in the industry (all deflated). The aggregate labor share of revenue fluctuated at around 3%, while the aggregate intermediate input share of revenue fell from 40% to 25% between 2000 and 2007. Changes in relative input expenditure can be due to various reasons other than market power, such as technical change. I will come back to these reasons in the next section.

[Figure 4 here]

²⁶The higher increase in concentration at the province-level masks the fact that in many counties, the number of firms dropped to zero, which means that these markets are omitted from the analysis. Accounting for these markets, the consolidation was actually more pronounced the more narrowly markets are defined.

Treatment and control groups

To what extent did consolidation contribute to the fall in the intermediate input share of revenue? Let the number of firms i producing less than 100,000 cases Q per year t in market i be denoted N_{it} :

$$N_{it} = \sum_{f \in i} (\mathbb{I}[Q_{ft} < 100,000])$$

Firms producing less than 100,000 cases were forced to exit in 2003. I construct a consolidation treatment variable C_f which is a dummy indicating whether firm f is located in a county in which there was at least one firm producing below the exit threshold in 2002, at the onset of the size-dependent consolidation.

$$C_f = \mathbb{I}[N_{i,2002} > 0]$$

Panel (a) of table A1 contains some more information about the treatment and control group sizes. Before the policy was implemented in 2003, half of the firms produced less than 100,000 cases, and together earned 8.1% of total industry revenue. The treatment group consists of 15% of the firms producing above the threshold that had at least one competitor below the threshold in the same county. At the prefecture level, 39% of firms were in the treatment group.²⁷ The revenue and output shares of the treatment group were similar to the share of the number of firms.

I compare cigarette prices, leaf prices and production quantities between the treatment and control groups before 2003 in panel (b) of table A1. When defining input markets at a more disaggregated level, the pre-treatment differences in terms of prices between treatment and control groups are larger. Firms in treated counties already charged lower cigarette and leaf prices before the policy started than firms in the control group. At the province-level, this was not the case. This may have to do with announcement effects, which I explore in appendix C.1.

Difference-in-differences model

In order to assess the effects of consolidation on input and product prices, I estimate the difference-in-differences model in equation (1). I compare firms with and without competitors below the exit threshold before and after 2003. The outcome of interest y_{ft} is product prices p_{ft} , leaf prices w_{ft}^M , and wages w_{ft}^L ; all in logs. I also use the log factor revenue shares as the outcome variable. The labor revenue share is defined as $\alpha_{ft}^L \equiv \frac{W_{ft}^L L_{ft}}{P_{ft} Q_{ft}}$ and the intermediate input revenue share as $\alpha_{ft}^M \equiv \frac{W_{ft}^M M_{ft}}{P_{ft} Q_{ft}}$, with labor and intermediate input units L , M , cigarette prices P and quantities Q . The consolidation dummy C_f itself is not included as it is part of the firm fixed effect θ_f^y . The coefficient of interest that

²⁷When defining input markets at the province-level, 86% of firms are in the treatment group.

quantifies the consolidation effects is θ_2^y .

$$y_{ft} = \theta_0^y + \theta_1^y \mathbb{I}[t \geq 2003] + \theta_2^y C_f \mathbb{I}[t \geq 2003] + \theta_3^y t + \theta_f^y + v_{ft}^y \quad (1)$$

$$\text{with } y_{ft} \in \left\{ p_{ft}, w_{ft}^M, w_{ft}^L, \frac{W_{ft}^M M_{ft}}{P_{ft} Q_{ft}}, \frac{W_{ft}^L L_{ft}}{P_{ft} Q_{ft}} \right\}$$

Pre-trends

The usual assumptions for a difference-in-differences model apply. First, the error term v_{ft} has to be conditionally independent from the consolidation dummy $C_f \mathbb{I}[t \geq 2003]$. The error term v^y contains all time series variation in outcome y_{ft} that is not captured by the other control variables. Second, the trends in the outcome variable need to be parallel for both treatment and control groups in the absence of the treatment.

The main question underlying this exclusion restriction is why there were firms operating below the threshold in some markets, but not in others. Figure A1 shows that most firms were distributed close to the exit threshold of 100,000 cases before 2003. Panel (b) of table A1 shows that firms in the treatment and control groups were comparable in terms of leaf and cigarette prices and size before 2003.

I verify the parallel trends assumption by testing whether the treatment and control groups had parallel trends in the outcome variables of interest before 2003. I estimate equation (2) on the time period 1999-2002. The coefficient of interest is the interaction effect between the treatment variable and time, η_2 : if this coefficient is close to zero and insignificant, parallel pre-trends in the outcomes of interest cannot be rejected.

$$y_{ft} = \eta_1^y C_f + \eta_2^y C_f * t + \eta_3^y t + v_{ft}^y \text{ if } t < 2003 \quad (2)$$

The estimates of equation (2) are in panel (d) of table 2 and confirm that it cannot be rejected that the pre-trends for both wages, leaf prices and cigarette prices were parallel before 2003.

Results

I start with visual evidence. Panel (b) of figure 4 compares the evolution of average relative input expenditure between the treatment and control group. The average ratio of intermediate input expenditure over the wage bill fell from 11 to 8 for firms in the treatment group between 2002 and 2006. For firms in the control group, it increased from 11 to 12. Taking the weighted averages by labor usage or the median, in panels (c)-(d), yields very similar patterns.

The drop of relative intermediate input expenditure can be due to two reasons. One explanation could be factor-biased technological change. For tobacco leaf, this is unlikely: the required amount of tobacco leaf per cigarette varies little across firms.²⁸ A factor-biased shock that changed the amount of labor needed per cigarette is possible. In order to generate the patterns in figure 4, however, the required amount of labor per cigarette would have to increase over the sample period. Mechanization of production has the opposite effect. Secondly, these patterns could be due to a change in relative input prices. This could be either due to increased monopsony power on tobacco leaf markets, decreased monopsony power on labor markets, or both.

I examine this further by estimating equation (1) using cigarette prices, leaf prices and wages per employee as the outcome variable. I estimate the same model using different input market definitions at the province, prefecture and county level in panels (a), (b) and (c) of table 2.²⁹ I find that in all three specifications, labor wages fell by 14.7%, 12.0% and 5.4% respectively,³⁰ but this drop was not statistically significant. Leaf prices did, in contrast, fall by 37% when defining leaf markets at the province or prefecture level, and by 53% when using county-level leaf markets. Cigarette prices did, finally, decrease by 23-32%, depending on the leaf market definition used. The narrower markets are defined, the larger the effects. Consolidation hence led both to lower leaf prices and lower factory-gate cigarette prices. The relative drop in leaf prices was significantly larger than the drop in cigarette prices at the prefecture and county level. The effect of a firm exiting in the same county is larger compared to a firm exiting at the other side of the same province, as manufacturers are likely to compete more for leaf with nearby manufacturers.

[Table 2 here]

Variation in prices due to consolidation, which is arguably exogenous in this setting, is not sufficient to draw conclusions about the underlying mechanism. Falling leaf and factory-gate prices could be due to increased markdowns and increased markups, but changes in productive efficiency would also lead to different equilibrium input and product prices.³¹ I therefore combine the natural experiment with a model of production and input market competition.

²⁸More evidence on this is in appendix C.

²⁹In panel (d), I test the parallel pre-trend assumption using equation (2). The time trends in all three outcome variables were not significantly different before the policy was implemented, so parallel pre-trends cannot be rejected.

³⁰ $\exp(-0.160) - 1 = -0.147$

³¹Changes in productivity alone could not explain the combination of falling input prices and falling product prices together, though.

3 A model of production, markups and markdowns

3.1 Production and costs

Production

Cigarette manufacturers f produce Q_{ft} cases of cigarettes using a quantity of tobacco leaf M_{ft} , labor L_{ft} and fixed assets K_{ft} .³² I allow for substitution between labor and capital, but not between tobacco leaf and labor or capital. Tobacco leaf may be substitutable to a limited extent with capital, for instance due to waste reducing technologies. The crucial substitution pattern when inferring markups and markdowns is, however, the substitutability between leaf and labor.³³ I assume it is impossible.³⁴ Let the production function be given by equation (3):

$$Q_{ft} = \min \left\{ \beta_{ft}^M M_{ft}, \Omega_{ft} H(L_{ft}, K_{ft}) \right\} \exp(\varepsilon_{ft}) \quad (3)$$

The amount of tobacco leaf needed to produce a case of cigarettes is given by β_{ft}^M . In the baseline model, I impose that firms do not differ in terms of leaf content, $\beta_{ft}^M = \beta^M$.³⁵ Firms do differ in terms of their productivity level Ω_{ft} . In line with most of the prior literature, this productivity term is assumed to be a Hicks-neutral scalar. Firms use a common production technology $H(\cdot)$ which parametrizes the substitution pattern between labor and capital. I assume $H(\cdot)$ is twice differentiable in both labor and capital. Measurement error in output is denoted ε_{ft} . Equation (3) nests production functions in which all inputs are substitutable: the input requirement β^M would then be zero by definition, and intermediate inputs added as a substitutable input.

Even if tobacco leaf is not substitutable in the cigarette production function, total intermediate inputs could be substitutable with total labor if it were possible to vertically integrate cigarette factories with farms. This is not the case in the Chinese tobacco industry (Peng, 1996; FAO, 2003; Wang, 2013). If firms have large buying power over intermediate input suppliers, they also have less incentives to integrate vertically.

³²Other intermediate inputs, such as paper and filters, are also part of M . I abstract from these in the model as they together represent less than a third of intermediate input costs and are as non-substitutable as leaves.

³³The reason for this is that the first order cost minimization conditions are derived for leaf and labor, not for the capital stock which is assumed fixed.

³⁴Dunn and Heien (1985) found no substitution between agricultural inputs in food processing industries. Sumner and Alston (1987) argued tobacco leaf is substitutable with non-farm inputs, which goes against my Leontief model. They used, however, an input demand approach which rules out monopsony power. If leaf prices are endogenous, regressing relative input usage on relative input prices will naturally result in a positive correlation, even if inputs are not substitutable. I will estimate the substitution elasticity between all inputs in section 6.

³⁵I will extend the model to allow for both factor-augmenting productivity and heterogeneous leaf content in section 6.

The capital stock evolves dynamically with a depreciation rate δ^K and investment I_{ft} :

$$K_{ft} = \delta^K K_{ft-1} + I_{ft-1}$$

I assume firms produce a single but differentiated product, cigarettes, at price P_{ft} .³⁶ Cigarette markets are allowed to be imperfectly competitive, so the price of cigarettes is a function of how much a firm and its competitors produce. I deal with quality differences in cigarettes and tobacco leaf by using a price control in the production function, as in De Loecker et al. (2016).³⁷

Input markets

I assume that both tobacco leaf and labor are variable and static inputs, meaning that they can be adjusted without adjustment costs and are proportional to the number of cigarettes produced.³⁸ Tobacco leaf is sold at least at monthly intervals at the purchasing stations, and more cigarettes require more tobacco. While the NBS survey does not separately report production and non-production workers, the US census does. In 1997, over 70% of US cigarette manufacturing employees were production workers, and hence variable (U.S. Census Bureau, 1997a).³⁹ I choose to model labor as a static variable as hiring and firing flexibility is considerably higher in China compared to the global and regional median countries in terms of labor market flexibility (World Economic Forum, 2019).

The prices for labor and tobacco leaf are denoted W_{ft}^L and W_{ft}^M . They depend on exogenous firm characteristics, some of which are observed, denoted \mathbf{Z}_{ft} , and some of which are latent, denoted ζ_{ft} . Examples of these characteristics are whether a firm is state-owned or not (observed) and how close it is located to a highway (latent). Next, input prices potentially also depend on how many inputs the firm uses in a given year. If this is the case, then the supply curve for input $V \in \{M, L\}$ is upward-sloping. Finally, equilibrium input prices also depend on characteristics and input prices of all other firms in the same market i . This set of firms is denoted as \mathcal{F}_{it} .

$$W_{ft}^V = W^V(\mathbf{Z}_{ft}, V_{ft}, \zeta_{ft}, \mathbf{Z}_{-ft}, V_{-ft}, \zeta_{-ft}) \quad \forall -f \in \mathcal{F}_{it}$$

The price elasticity of input supply is defined as $\psi_{ft}^V - 1$ and can differ across firms and over time. If ψ^V equals one, the input supply function is flat, which means that the price of input V is exogenous

³⁶The model can be generalized to a multi-product setting by using De Loecker et al. (2016), but this is not of first-order importance for the tobacco industry as the average firm earns more than 90% from selling cigarettes.

³⁷More details and the exact empirical specifications used follow in section 4.

³⁸I follow the classification of Akerberg, Caves, and Frazer (2015). In section 6.2, I discuss inventories and dynamic leaf demand.

³⁹These workers accounted for 65% of the wage bill.

to the firm.

$$\psi_{ft}^V \equiv \frac{\partial W_{ft}^V}{\partial V_{ft}} \frac{V_{ft}}{W_{ft}^V} + 1 \geq 1 \quad \text{for} \quad V_{ft} \in \{M_{ft}, L_{ft}\} \quad (4)$$

I allow for input market power on the market for tobacco leaf, so ψ^M can be above one. I assume, however, that firms are price takers on labor and capital markets: $\psi^L = 1$. Manufacturing workers are much more mobile than farmers in China, and can also easily switch to other manufacturing industries, as their skills are not specific to cigarette manufacturing. Tobacco farmers, in contrast, can only supply cigarette manufacturers or switch crops, which is very costly. I will provide more empirical evidence for this assumption in section 6. The method could easily be generalized to allow for endogenous wages W_{ft}^L .⁴⁰

As I estimate the input supply model using data with one-year intervals, a short-run supply elasticity is inferred. I discuss the difference between short- and long-run supply elasticities in section 6.2.

Decision-making

I assume manufacturing firms choose the tobacco leaf price W_{ft}^M each year in order to minimize per-period variable costs, taking the other input prices as given. In theory, the Chinese government centrally regulated leaf prices and procurement up to 2015, after which leaf markets were liberalized (Lan, 2015). In reality, however, manufacturing firms had considerable pricing power on leaf markets from the 1980s onwards. Peng (1996) provides anecdotal evidence for this: leaf prices were, for instance, set for 15 ‘quality grades’.⁴¹ As these grades were mostly subjective, manufacturing firms could set prices by choosing the corresponding grade. Peng (1996) mentions frequent conflicts between peasants and manufacturers over grading, bitter feelings among the farmers and, in some cases, peasants being forced to sell tobacco at a price below their cost of production.

The assumption that firms minimize costs can be questioned. It is often suggested that state-owned enterprises (SOEs) have non-profit objectives such as generating local employment (Lu & Yu, 2015). In the tobacco industry, Peng (1996) notes that cigarette manufacturers have “the purpose of making profits” and “often bargain with each other for better deals”.⁴²

Assumption 1. — Firms choose input prices W_{ft}^M annually to minimize a short-term cost function, taking the other input prices as given.

⁴⁰This would require to impose a similar structure on how firms compete on labor markets to the model of intermediate input competition of this paper.

⁴¹I cope with this input quality differentiation in the production model by adding a price control, similar to De Loecker et al. (2016).

⁴²In appendix B.5, I still extend the model to allow for objective functions other than cost minimization. Different objective functions will change the inferred markup levels. As the vast majority of tobacco manufacturers are SOEs anyway, it is unlikely that the observed changes in markups and markdowns will be driven by differences in firm objectives.

The associated Lagrangian of the cost minimization problem is given by equation (5), with marginal costs λ_{ft} :

$$W_{ft}^{M*} = \arg \min \mathcal{L}_{ft} = W_{ft}^M M(W_{ft}^M) + W_{ft}^L L_{ft} + \lambda_{ft} \left(Q_{ft} - Q_t(M(W_{ft}^M), L_{ft}, K_{ft}, \Omega_{ft}, \beta_{ft}^M) \right) \quad (5)$$

As labor and tobacco leaf cannot be substituted, there is just one first order condition, rather than one for each variable input.⁴³ When firms choose the leaf price, and hence the quantity of tobacco leaf used, they automatically also choose how much labor to use.

3.2 Markups

General case: endogenous input prices and non-substitutable inputs

The markup is defined as the ratio of cigarette prices P_{ft} over marginal costs λ_{ft} :

$$\mu_{ft} \equiv \frac{P_{ft}}{\lambda_{ft}}$$

Solving the first order condition in equation (5) for marginal costs yields the markup expression in equation (6a), which is derived in appendix B.1.

$$\mu_{ft} = \left(\frac{\alpha_{ft}^L}{\beta_{ft}^L} \psi_{ft}^L + \alpha_{ft}^M \psi_{ft}^M \right)^{-1} \quad (6a)$$

with $\alpha_{ft}^V \equiv \frac{V_{ft} W_{ft}^V}{P_{ft} \frac{Q_{ft}}{\hat{\varepsilon}_{ft}}}$ for $V \in \{L, M\}$

The output elasticity of labor β_{ft}^L is retrieved when estimating the production function. The revenue shares α_{ft}^V on the right-hand side of equation (6a) are in principle observed, but I follow De Loecker and Warzynski (2012) and adjust for measurement error ε . This means that I calculate revenue shares using quantities in which the estimated error $\hat{\varepsilon}$ is netted out. As said before, labor wages are exogenous in this paper: $(\psi_{ft}^L - 1) = 0$. Even then, the right-hand side of equation (6a) still contains the input price elasticity of tobacco leaf, ψ_{ft}^M , which is latent. The intuition for the fact that markups depend on the input supply elasticity is that the slope of the input supply curve is part of marginal costs: if the firm increases output by one unit, costs increase by more the steeper the input supply function is, as input prices endogenously increase.

Markups μ_{ft} and the input price elasticity $(\psi_{ft}^M - 1)$ are hence not separately identified without

⁴³This also applied to the beer brewing production function in De Loecker and Scott (2016).

adding more structure to the model. Even if the firm would have more variable inputs with exogenous prices, this does not lead to separate identification of ψ and μ . As none of these inputs would be substitutable with tobacco leaf, the input demand conditions for all these inputs would incorporate the endogenous price effect in the same way.

Special case (i): Exogenous input prices and substitutable inputs

Suppose all inputs have exogenous prices and are mutually substitutable. In that case, the non-substitutable input requirement is by definition zero, $\beta_{ft}^M = 0$, and all input price elasticities are all zero: $(\psi^V - 1) = 0, \forall V$. The markup expression then simplifies to the formula from De Loecker and Warzynski (2012):

$$\mu_{ft} = \frac{\beta_{ft}^L}{\alpha_{ft}^L} \quad (6b)$$

Special case (ii): Endogenous input prices and substitutable inputs

Next, consider a setting in which all inputs are substitutable, but in which input prices are endogenous. This implies upward-sloping input supply functions. The markup is now expressed as the output elasticity of a variable input divided by its revenue share *and* divided by the input price elasticity of supply plus one. This corresponds to the expression from Morlacco (2017).

$$\mu_{ft} = \frac{\beta_{ft}^L}{\alpha_{ft}^L \psi_{ft}^L} \quad (6c)$$

If there is just one variable substitutable input L , markups and markdowns are not separately identified. In case there are multiple variable substitutable inputs of which at least one input has an exogenous input price, markups and markdowns can be separately identified. Suppose, for instance, that there are two variable inputs L and L^c , and that the price of L^c is given from the firm's perspective. Equation (6c) then becomes a system of 2 equations. Because the markdown L^c is assumed to be one, there are two unknowns: the markup μ and the markdown under the price of input L .

$$\begin{cases} \mu_{ft} = \frac{\beta_{ft}^L}{\alpha_{ft}^L \psi_{ft}^L} \\ \mu_{ft} = \frac{\beta_{ft}^{L^c}}{\alpha_{ft}^{L^c}} \end{cases}$$

The markdown can now be found by dividing both equations. This is the markdown expression from Morlacco (2017):

$$\psi_{ft}^L = \frac{\alpha_{ft}^{L^c} \beta_{ft}^L}{\alpha_{ft}^L \beta_{ft}^{L^c}}$$

Special case (iii): Perfect input markets and a non-substitutable input

A final special case holds when all input prices are exogenous, but when there is one input that cannot be substituted for any other input. In this case, $\beta_{ft}^M > 0$, but all input supply elasticities ($\psi_{ft}^V - 1$) are zero. The markup is given by equation (6c), which corresponds to De Loecker and Scott (2016). It is identified even if there is only one substitutable input.

$$\mu_{ft} = \left(\frac{\alpha_{ft}^L}{\beta_{ft}^L} + \alpha_{ft}^M \right)^{-1} \quad (6d)$$

Note that cases (ii) and (iii) can be blended: if the substitutable input prices are endogenous, but non-substitutable input prices are not, the markup is identified as long as there is at least one substitutable input with an exogenous price.

Possible approaches to achieve identification

Back to the general case in equation (6a). In order to separately identify the markup and the intermediate input supply elasticity, additional structure needs to be imposed on how firms compete either on product or input markets. One could, for instance, identify markups μ_{ft} by imposing a model of how firms compete on product markets and back out the input supply elasticity without further assumptions. In this paper, I take the opposite approach and identify the input price elasticity of leaf supply by imposing a model of how firms compete on leaf markets. The reasons why I prefer this second approach are entirely specific to the context of Chinese cigarette manufacturing. Leaf markets are easier to define as they are isolated by law, while cigarette markets are not. Moreover, it is harder to model cigarette demand than leaf supply. Demand for cigarettes is likely to be dynamic due to addiction. Moreover, the wholesaler's conduct and retail prices are unobserved. Even if there is a literature that addresses each of these complications,⁴⁴ modeling leaf supply is easier as leaf is sold on well-delineated static spot markets.

⁴⁴Dynamic demand is modeled in general by Gowrisankaran and Rysman (2012) and specifically for cigarettes consumption by Baltagi and Levin (1986). There is also a literature on conduct identification (Ciliberto & Williams, 2014; Miller & Weinberg, 2017; Michel & Weiergraeber, 2018). Delipalla and O'Donnell (2001) estimate conduct of European cigarette manufacturers.

3.3 Markdowns

Markdown interpretation of the input supply elasticity

The input price elasticity of supply, $(\psi_{ft}^M - 1)$, is intimately related to the concept of the ‘markdown’, the wedge between an input’s marginal revenue product and its price. To see this, the cost minimization problem in equation (5) must be rewritten as a profit maximization problem. Using the derivation in appendix B.2, the following expression is obtained:

$$\psi_{ft}^M W_{ft}^M = \underbrace{\frac{\partial(P_{ft}Q_{ft})}{\partial M_{ft}}}_{\text{MRP leaf}} - \underbrace{\beta_{ft}^M W_{ft}^L \frac{\partial L_{ft}}{\partial Q_{ft}}}_{\text{Marginal labor cost}} \quad (7)$$

The right-hand side of this equation is the marginal revenue product of tobacco leaves (‘MRP leaf’) minus the marginal labor cost of increasing the number of tobacco leaves by an additional unit. This entire term can hence be thought of as the marginal profit gain from an additional unit of leaf, net of leaf costs. The term ψ^M can now be interpreted as a ‘markdown’: it is the ratio of the marginal benefit of leaves divided by the leaf price farmers receive. If the markdown is large, farmers get a small fraction of their contribution to manufacturing profits.

Input competition model

Let there be I_t isolated leaf markets i . Farmers j can sell tobacco leaves to manufacturing firms $f \in \mathcal{F}_{it}$, with $f = 0$ indicating the outside option of not selling to any firm. I assume each firm operates in exactly one market and that farmers sell their entire production to a single firm, which makes sense as there were 20 million household-level farms producing tobacco leaf in 1997 (FAO, 2003), selling to merely 350 firms. I follow S. T. Berry (1994) and the ensuing literature and let the utility function of an farmer j depend on the leaf price, manufacturer characteristics \mathbf{Z}_{ft} and ζ_{ft} and an i.i.d. type-I distributed manufacturer-farmer utility shock ν_{jft} . I allow for the fact that farmers may not accept the highest bid, as it is possible that non-price characteristics ζ influence farm utility as well. The location of the manufacturing plant or who owns the factory could, for instance, influence farmer utility conditional on the leaf price offered.

$$U_{jft} = \gamma^W W_{ft}^M + \gamma^Z \mathbf{Z}_{ft} + \zeta_{ft} + \nu_{jft}$$

I normalize the utility of the outside option to zero, as usually: $U_{j0t} = 0$. I explicitly rule out variation in supplier preferences for input prices or manufacturer characteristics, γ^W and γ^Z . In appendix C.4,

I allow for more flexible substitution patterns by using a nested logit model.⁴⁵ In this section, I also estimate alternative input model specifications, such as a logs-on-logs model.

I assume that the farmer-manufacturer utility shock ν_{jft} is i.i.d. across firms, farmers and time, and follow the usual logit assumption:

Assumption 2. — The farmer-manufacturer utility shock ν_{jft} follows an extreme-value type-I distribution.

As stated before, firms choose tobacco leaf prices simultaneously in order to minimize per-period costs. I denote the tobacco leaf market share of firm f in year t as $S_{ft} = \frac{M_{fjt}}{\sum_{r \in \mathcal{F}_{it}} M_{fjt}}$. The distributional assumption on ν_{jft} allows the intermediate input share S_{ft} to be written as:

$$S_{ft} = \frac{\exp(\gamma^W W_{ft}^M + \gamma^Z \mathbf{Z}_{ft} + \zeta_{ft})}{\sum_{r \in \mathcal{F}_{it}} \exp(\gamma^W W_{rt}^M + \gamma^Z \mathbf{Z}_{rt} + \zeta_{rt})}$$

Dividing this share by the market share of the outside option S_{0t} , whose utility is normalized to zero, as well as taking logs, leads to equation 8, which will be estimated in the next section.

$$s_{ft} - s_{0t} = \gamma^W W_{ft}^M + \gamma^Z \mathbf{Z}_{ft} + \zeta_{ft} \quad (8)$$

The markdown (ψ) can, finally, be expressed as a function of observable input prices and input market shares, and of the estimated valuation parameter γ^W :

$$\psi_{ft}^M \equiv \left(\frac{\partial S_{ft}}{\partial W_{ft}^M} \frac{W_{ft}^M}{S_{ft}} \right)^{-1} + 1 = \left(\gamma^W W_{ft}^M (1 - S_{ft}) \right)^{-1} + 1 \quad (9)$$

4 Empirical analysis

The empirical analysis consists of four main steps. I start by estimating the production function. Secondly, I estimate the input supply function in order to retrieve the markdown. Third, I combine the output elasticity estimates from the first step with the markdown estimates from the second step to calculate markups. Fourth, I examine how the 2003 consolidation policy affected both markdowns, markups and productivity. Finally, I revisit some of the key modeling assumptions in some extensions. The estimated output elasticities are used both when estimating the input supply function and when estimating markups. I therefore bootstrap the entire estimation procedure using 50 bootstrap iterations.

⁴⁵When applying the methods used to labor markets, heterogeneous supplier preferences and all kinds of other frictions seem especially important.

4.1 Production function identification

Identification of the production function serves three purposes. First, the output elasticity of labor is required to identify the markup. Next, it allows quantifying returns to scale. Finally, the productivity residuals will be used as leaf demand shifters in order to identify the input price elasticity of leaf supply.

If input prices would be exogenous and the production function a constant returns to scale Cobb-Douglas production function, the cost shares of each input would simply equal their output elasticity (Syverson, 2004).⁴⁶ In the context of this paper, the Cobb-Douglas function in labor and capital only contains inputs with exogenous input prices. I do not want to impose constant returns to scale, however, as a key motivation of this paper is to allow for increasing returns to scale and efficiency gains from consolidation.⁴⁷

I therefore estimate the production function. Tobacco leaf can be ignored due to the Leontief assumption: only the $H(\cdot)$ function needs to be estimated.⁴⁸ Denoting logs of variables in lowercases, equation (10a) needs to be estimated.

$$q_{ft} = h(l_{ft}, k_{ft}) + \omega_{ft} \quad (10a)$$

A dummy $C_{ft} \in \{0, 1\}$ indicates whether a firm is subject to the consolidation policy, which will be discussed later in more detail. I allow productivity to endogenously change due to consolidation, as in De Loecker (2013); Braguinsky et al. (2015):

$$\omega_t = g(\omega_{ft-1}, C_{ft}) + \xi_{ft}$$

Product and input differentiation

The log production function in equation (10a) has both input and output quantities on both sides of the equation. Cigarettes are, however, differentiated products with various quality levels. As pointed out in De Loecker et al. (2016), this can result in bias as firms which produce high-quality, high-price cigarettes use more labor per unit of cigarette than low-quality producers. Moreover, I observe the number of employees \tilde{l} rather than the total amount of hours worked l , and the skills of workers,

⁴⁶If, however, monopsony power differs across firms, this no longer holds.

⁴⁷In appendix C.3, I impose constant returns to scale and use the cost shares approach to find the output elasticities of labor and capital, and use these to estimate markups and markdowns.

⁴⁸In general, it could be optimal for firms to diverge from the Leontief ‘first order condition’ of intermediate inputs equaling the $H(\cdot)$ function in labor and capital, as argued in Ghandi, Navarro, and Rivers (2018) The assumption that intermediate inputs enter the production function linearly solves this problem, however.

which could differ across firms. The unit wage W^L , which is the wage bill divided by the number of employees, captures variation in both hours worked and worker skills. I follow De Loecker et al. (2016) by adding a function $a(\cdot)$ in both product cigarette prices and unit wages to the production function. The production function in logs, equation (10a), is hence replaced by equation (10b):

$$q_{ft} = h(\tilde{l}_{ft}, k_{ft}) + a(w_{ft}^L, p_{ft}) + \omega_{ft} \quad (10b)$$

Latent markups and markdowns

The usual approach to identify the production function in the presence of a latent productivity scalar has been to invert an input demand function and express the productivity residual ω as a function of observables only (Olley & Pakes, 1996; Levinsohn & Petrin, 2003; Akerberg et al., 2015). Following the production function, equation (3), leaf demand depends on productivity, the leaf content per cigarette and the optimal amount of labor and capital used:

$$M_{ft} = \frac{\Omega_{ft}}{\beta^M} H(L_{ft}, K_{ft})$$

If intermediate input quantities are observed, identification of the production function would be straightforward. As shown in appendix A of Akerberg et al. (2015), the first stage inversion can be achieved by regressing quantities on labor, intermediate inputs and capital. The only reason why intermediate input usage differs across firms, conditional on labor and capital used, is variation in productivity Ω_{ft} .

Intermediate input quantities are, however, latent. It is hence necessary to invert the labor demand equation for productivity. When solving the first order condition in equation (5), firms choose a leaf price that minimizes variable costs. An optimal output level Q_{ft}^* corresponds to this leaf price. Firms choose the mix between labor and capital that minimizes costs

As derived in appendix B.3, labor demand depends on cigarette prices, other inputs and their prices, ownership consolidation, and firm characteristics \mathbf{Z} such as its export status or ownership structure, which are all observed. It also depends, however, on markups μ , markdowns ψ^M and productivity ω_{ft} , which are all latent. If two of these three unobservables are latent, which is very likely, the input demand inversion breaks down. If firms use few inputs, we cannot be certain whether this is due to low productivity, high markups or high markdowns.

$$l_{ft} = l(k_{ft}, W_{ft}^M, W_{ft}^L, \mathbf{Z}_{ft}, \omega_{ft}, \mu_{ft}, \psi_{ft}^M)$$

Solution 1: Input prices and market shares in input demand

The first strategy to identify the production function is to stick to the inversion approach, but to insert drivers of markdown and markup variation across firms in the input demand function. This approach was used in De Loecker et al. (2016). In most models of supply and demand, markup and markdown variation across firms depends on prices and market shares. In the logit model outlined in section 3.3, for instance, markdown variation depends on the tobacco leaf price and leaf market share. If the same holds on the cigarette market, then including leaf and cigarette prices and upstream and downstream market shares in the input demand function should solve the problem. Defining market shares downstream can be difficult and was one of the reasons that the demand estimation approach was not followed, but we can insert multiple market shares at different geographical levels (county, prefecture, province) in the demand function.

In appendix D, I discuss the validity of this approach in more detail using simulated data. I find that the production function can be consistently identified using this approach if prices are observed and market definitions correctly defined, under the condition that there is no unobserved heterogeneity in supplier and consumer valuations of prices and firm characteristics. As soon as farmers differ, for instance, in their valuation of firm characteristics, such as whether it is state-owned or not, the production function estimates can be seriously biased. The intuition for this is that in this case, markdowns differ across firms even after conditioning on input market shares and input prices, as firms take into account these heterogeneous valuations when choosing inputs. Because of this, latent markdowns enter as a persistent unobservable into the input demand function, which is no longer invertible in latent productivity.

I follow the two-stage procedure of Akerberg et al. (2015). In the first stage, I invert the leaf demand function for the latent productivity term. This results in equation (11). The vector of demand shifters z_{ft} includes all observables which may affect input demand: product and ownership types, export dummies and export shares of revenue in the vector of input demand shifters z_{ft} . I use a third-order polynomial in labor and capital in $\Phi(\cdot)$, and include the leaf price and input market share to proxy for markdown variation, as discussed in the previous section.

$$q_{ft} = \Phi(\tilde{l}_{ft}, k_{ft}, w_{ft}^L, w_{ft}^M, z_{ft}, p_{ft}, s_{ft}) + \varepsilon_{ft} \quad (11)$$

In the baseline model, I use a Cobb-Douglas function in labor and capital for $H(\cdot)$. In appendix C.2, I estimate a Translog production function instead, which results in very similar estimates.

$$h(L_{ft}, K_{ft}) = \beta^L l_{ft} + \beta^K k_{ft}$$

In line with the timing assumptions imposed for capital and labor, the moment conditions to identify

the output elasticities are given by:

$$\mathbb{E}\left\{\xi_{ft}(\beta^L, \beta^K) \begin{pmatrix} l_{ft-1} \\ k_{ft} \end{pmatrix}\right\} = 0$$

Solution 2: BLP instruments

An alternative identification strategy is to exploit input price variation across firms. This was already explored in an older productivity literature, as in Griliches and Mairesse (1998). The production function literature has stepped away from using input prices as instruments because it is doubtful that they are orthogonal to total factor productivity.⁴⁹ I rely on an instrumental variables approach instead.⁵⁰ When input markets are oligopsonic, however, a whole new set of possible instruments for input prices and quantities become available, similar to the oligopoly demand identification literature (S. Berry et al., 1995).

In a recent paper, S. Berry et al. (2019) discuss the use of instruments to identify input supply in oligopsony. If these instruments can be used to instrument input prices when identifying the input supply curve, then they can also be used to instrument input quantities when identifying the production function. Characteristics of competing cigarette manufacturers in the same leaf market could, for instance, be used as shifters of the leaf price, and hence leaf quantities, provided that these firm characteristics enter farmer utility or shift input demand.

Consider manufacturer characteristics which enter the farmer utility function $(\tilde{Z}_{ft}, \hat{Z}_{ft}) \in \mathbf{Z}_{ft}$, such as the firm's ownership structure. The characteristics of competing firms in the same market is denoted \tilde{X}_{ft} :

$$\tilde{X}_{ft} = \frac{1}{|\mathcal{F}_{it}| - 1} \left(\sum_{r \in \mathcal{F}_{it}} (\tilde{Z}_{rt}) - \tilde{Z}_{ft} \right)$$

If there are as many suitable characteristics as coefficients for the substitutable inputs (in this paper, 2), the moment conditions are:

$$\mathbb{E}\left\{\omega_{jt}(\beta_l, \beta_k) \begin{pmatrix} \tilde{z}_{jt} \\ \hat{z}_{jt} \end{pmatrix}\right\} = 0$$

In order for the exclusion restriction to be satisfied, the selected characteristics z of competing firms need to be orthogonal to the firm's productivity level, conditional on all observables. There are multiple threats to this identification strategy. First, if there are productivity spillovers across firms or

⁴⁹Moreover, there is not much input price variation within narrow industry-market pairs in countries with centralized wage bargaining.

⁵⁰A more detailed discussion on the use of instruments in production function identification is in Akerberg, Benkard, Berry, and Pakes (2007).

X-inefficiencies, this would violate the exclusion restriction. Secondly, firm characteristics that are strategically chosen as part of the input market game played by the oligopsonists are not to be used as instruments.

I select two sets of variables as the instrument Z . First, I use import tariffs in other manufacturing industries in the same county, sales-weighted across firms in the county, as an instrument. Even if firms cannot set labor wages themselves, equilibrium wages will be affected by input demand in other manufacturing industries. The disadvantage of this instrument is that it is county- and year-specific, not firm-specific. Three out of four cigarette manufacturers are, however, the only producer in their county. Secondly, I add the average export share of revenue of competitors and export participation of competitors in the same prefecture as instruments. As competitors enter the international cigarettes market, their cigarettes demand increases, and hence also leaf prices and optimal input quantities. If there are no ‘productivity spillovers’ from exporting to non-exporting firms, export shocks at other firms would be excluded from the own firm’s productivity level. As exports make up less than 1% of total revenue, this is unlikely to be a strong instrument.

In appendix D, I detail further the identification of production functions when input prices are endogenous, and use simulated data to compare the performance of both the IV and control function approaches.

Estimates

The estimated output elasticities are in panel (a) of table 3. The left columns contains the estimates which are inferred when using ACF: the output elasticities of labor and capital are 0.320 and 0.761 respectively. The standard errors are large, as is usual when estimating ACF on a small sample.⁵¹ The scale parameter is 1.052, which implies modest scale returns. The scale parameter is not significantly different from one, but this may be a result of a lack of power due to the small sample.

The right columns report the estimates when using export participation and export shares of competing firms and the weighted tariff rate in other manufacturing industries in the same leaf market as instruments for labor and capital. The firm’s export status and export share are controlled for in the production function. The estimated output elasticities of labor and capital are now 0.404 and 0.699. The scale parameter of 1.104 suggests modestly increasing returns to scale, but has a relatively large standard error. The sample using IV estimation is larger than when using ACF, as lagged variables are not needed.⁵² Throughout the rest of the paper, I will use the IV estimates.⁵³

⁵¹This was, for instance, also the case in the industry-by-industry estimates from De Loecker et al. (2016)

⁵²In panel (b) of table A11, I re-estimate the IV model on the ACF sample size. Both output elasticities using the IV model are now much closer to the ACF estimates, at 0.365 and 0.804.

⁵³I conduct different robustness checks by using the ACF estimates, a translog production function and a substitutable leaf

[Table 3 here]

4.2 Markdown and markup estimation

Input supply estimation

Next, I estimate the input supply function, equation (8). I control for three observable firm characteristics Z_{ft} which are likely to affect leaf supply. First, I control for cigarette prices, as they are a proxy for high quality. Secondly, I control for manufacturer ownership dummies. As the Chinese tobacco industry is highly regulated and politically influenced, the utility from selling to a manufacturer may differ depending on whether this manufacturer is state-owned, collectively owned or a private firm. Finally, I include prefectural dummies in order to control for the manufacturer's location.

It is unlikely that controlling for these observable supply shifters suffices to identify the leaf supply function. If manufacturers have unobservable characteristics ζ_{ft} which enter farmer utility, then this will affect the equilibrium leaf price.⁵⁴ This results in the same kind of simultaneity bias as when estimating product demand. In order to separately identify input demand and supply, an instrument for input prices is hence needed. This has to be a shifter of input demand which does not enter input supply, i.e. does not enter farmer utility.

The productivity residual ω_{ft} , which was estimated in the previous section, is such an input demand shifter. In order for the exclusion restriction to hold, productivity of the cigarette manufacturer should be excluded from farmer utility, conditional on the leaf price. In other words, conditional on how much they are paid, farmers do not care how efficient the manufacturing firms they are selling to are.

In which cases would this exclusion restriction be violated? First, it could be that manufacturers prefer to buy repeatedly from the same suppliers. The literature on vertical relationships in developing countries has emphasized the importance of relational contracts and repeated interaction (Macchiavello, 2018). As higher productivity manufacturers are less likely to exit in the future, farmers may prefer to sell to these manufacturers even when these do not offer higher leaf prices. In the Chinese tobacco industry, such repeated interaction seems not to be a concern of first-order importance, as leaf markets are highly regulated. In settings where incomplete contracts matter more and where institutions are less strong, other identification strategies may be necessary.

production function in appendix C.

⁵⁴Examples of such characteristics are whether the factory is easily accessible by road, how long it has been producing, political connections of its managers, etc.

Secondly, the exclusion restriction would be violated if there are important switching costs for buyers or sellers. In this case, high productivity buyers would be preferred because they are more likely to buy again in the future. This again does not seem to be crucial in the tobacco industry: leaf markets is auctioned at monthly intervals at purchasing points, so switching suppliers seems very easy Peng (1996); Wang (2013).⁵⁵

A third threat to identification of the input supply curve concerns product differentiation. If firms that are highly productive, in physical terms, choose higher quality, and hence more expensive, tobacco leaf, this would violate the exclusion restriction. I also control for cigarette prices in the input supply function, however, which should pick up differences in cigarette quality. Moreover, using the brand-level data on product characteristics shows, however, that the physical productivity of cigarette manufacturers does not correlate significantly with any product characteristic or quality indicator, as shown in panel c of Appendix table A19.

The consolidation dummies C_f , which were defined earlier in section 2.3 are candidate instruments. Operating in a market subject to the consolidation affected the input market shares and buying power of manufacturers, and hence their equilibrium leaf demand. If the consolidation did not enter farmer utility, conditionally on the leaf price, then these consolidation dummies can be used as instruments for leaf supply.

I use these consolidation dummies to test for overidentifying restrictions in appendix table A5. The Sargan-Hansen χ^2 is 0.988. If the consolidation dummies are excluded instruments, it cannot be rejected that manufacturer TFP is a valid instrument for leaf prices as well.

Once the input supply elasticity is estimated, markdowns ψ^M are inferred by using equation (9). Finally, I use both the output elasticity of labor and the markdown to calculate markups, using equation (6a).

Measurement

As is usually the case in production and cost datasets, I do not observe leaf prices W_{ft}^M , but rather leaf expenditure $M_{ft}W_{ft}^M$. Because of the Leontief production function, dividing leaf expenditure by the output quantity results in the leaf price divided by the leaf concentration β_{ft}^M .

$$\frac{M_{ft}W_{ft}^M}{Q_{ft}} = \frac{W_{ft}^M}{\beta_{M_{ft}}}$$

⁵⁵When applying the same model on manufacturing labor markets, more caution is needed. There are many reasons why employees would prefer to work for highly productive firms, even if these offer lower wages, such as career dynamics or better working conditions.

In the baseline model, I assume all cigarette producers have the same tobacco content per cigarette, β_M . Additional data reveal very little variation in leaf contents per cigarette across firms. Moreover, as long as firms did not adjust the tobacco content in response to consolidation, the estimated treatment effects should not change.

Market definitions

Leaf markets need to be defined in order to estimate the leaf supply function. There is usually at least one purchasing station per county. As became clear from the map in figure 2, however, many counties ended up without a tobacco manufacturing firm, even though tobacco farming continued. The market definition hence has to be wider than the county-level at least for a subset of firms. The market definition also has to be less than or equal to the province-level, as leaf trade across provinces within China is restricted. In the baseline approach, I define input markets at the province-level, because this yields the most conservative estimates for the treatment effects. Firms never compete on leaf markets outside their own province due to the transport restrictions. I do, however, re-estimate the main regressions with prefecture- and county-level leaf market definitions as robustness checks. Input supply becomes more inelastic the more local input markets are defined, implying higher leaf price markdowns.

Input supply estimates

The estimates of the leaf supply function, equation (8) are in panel (b) of table 3. The OLS estimates are in the first column, and indicate a downward-sloping supply curve. As already mentioned, however, this is likely due to simultaneity bias. The IV estimates are in the second column of panel (b), and imply an upward-sloping supply curve. The average leaf price per pack is RMB 1.12. The coefficient implies that an increase in this leaf price per pack of RMB 0.10, a relative increase of 9%, leads to an increase in the leaf market share of 44%. Productivity residuals are a strong instrument, as the first stage F-statistic is 79.14. Converting these estimates to input supply elasticities using equation (9) delivers an average inverse supply elasticity of 0.54.⁵⁶ A survey of estimated tobacco leaf supply elasticities in Askari and Cummings (1977) reports similar elasticities: 0.71 for India, 0.51 for Bangladesh and 0.60 for Nigeria.

Throughout the paper, I have assumed that manufacturing labor wages are exogenous. Although this assumption was motivated in the industry background section, it can be tested empirically. I estimate the elasticity of the labor supply function in the same way I did for tobacco leaf. I use the same specification and instruments as for tobacco leaf, with the exception that I add exports of

⁵⁶The average markdown is 2.869. The average input supply elasticity is hence $1/(2.869 - 1) = 0.535$

competitors as a second instrument, because the first stage is weaker than for tobacco leaf prices. The results are in panel (c) of table 3. The left columns report the OLS estimates and the right columns report the IV estimates. The estimated elasticity of labor supply is very close to zero and statistically insignificant both with and without instruments. This supports the assumption made that manufacturing labor wages are exogenous in this industry.

Markdown and markup estimates

I report selected moments for both markdowns and markups in panel (a) of table 4. The first two columns correspond to the baseline model in which tobacco leaf and other inputs cannot be substituted. The average markdown is 2.87, which means that farmers receive 35% of their marginal contribution to firm profits, according to (equation 7). Markups are on average 1.06, which implies that cigarette prices surpass marginal costs by 6.2%. This combination of large manufacturing margins on leaf markets and small margins on the wholesale market is consistent with the fact that they buy from many small farmers, but sell to a monopsonic wholesaler. The standard errors on the markdown and markup estimates are due to uncertainty about both the estimated input supply elasticity and output elasticities of inputs. The average markdown lies significantly above one, which is the level that implies exogenous input prices, while the average markup is not significantly different from one.

[Table 4 here]

The entire distribution of markups and markdowns is plotted in figure 5. Around half of manufacturing firms sell to the CNTTC at prices below their marginal costs, which does not mean they are loss-making, as they still have positive markdowns. The markdown distribution lies above one, meaning that all manufacturing firms have at least some input market power on the leaf market. There is a long tail of firms with markdowns above 3, meaning that farmers receive less than one third of their marginal revenue product.

[Figure 5 here]

Comparison to the substitutable input model

How important was it to model leaf as being non-substitutable? In panel (b) of table 4, I compare the markdown and markup estimates using the more conventional model with substitutable tobacco leaf,

which is also used in Morlacco (2017). Markups are now estimated to be on average 5.63, but with a very large standard error. Markdowns are, to the contrary, estimated to be 0.77, implying that farmers are paid *more* than their marginal profit contribution. Given the industry background, these markup and markdown estimates are unrealistic.

The markup expression when inputs are substitutable, μ_{ft}^{SUB} , is an overestimate of real markups if there is a non-substitutable input. This can be seen by comparing markup expressions (6b) and (6a):

$$\mu_{ft}^{SUB} = \frac{\beta_{ft}^L}{\alpha_{ft}^L} > \left(\frac{\alpha_{ft}^L}{\beta_{ft}^L} + \alpha_{ft}^M \psi_{ft}^M \right)^{-1} \quad \text{if } \alpha_{ft}^M > 0$$

The intuition for this overestimation is that in the non-substitutable model, marginal costs incorporate both the cost of intermediate inputs *and* labor: both have to be increased simultaneously in order to increase output (De Loecker & Scott, 2016).

The markdown in the substitutable inputs model, $\psi_{ft}^{M,SUB}$ is obtained by dividing the markup over the input with exogenous prices, labor, over the markup using the input with endogenous prices, tobacco leaf, as explained in Morlacco (2017):

$$\psi_{ft}^{M,SUB} = \frac{\beta_{ft}^M \alpha_{ft}^L}{\beta_{ft}^L \alpha_{ft}^M}$$

This expression is wrong in any case in the non-substitutable inputs model, as the first order conditions for tobacco leaf and labor are not linearly independent. Still, the substitutable production function will generally underestimate the markdown as the estimated output elasticity of leaf β^M will be below unity, while it is one in reality due to non-substitutability of tobacco leaf. The substitutable inputs model compares the revenue share of materials to its output elasticity, and will hence underestimate the difference between both.

Comparison to prior monopsony studies

How large are the markdown estimates implied by this paper? Most prior studies of monopsony power estimate the input price elasticity. I convert these estimates into the ratio of input prices over marginal revenue products using equation (9), which is on average 0.35 in this paper. Appendix table A14 compares this ratio to a selection of prior studies about monopsony power. A lower input price to marginal revenue share implies more monopsony power. The ratios range from 0.83 for tenure-track professors in US academia (Goolsbee & Syverson, 2019)⁵⁷ to 0.70 for female grocery clerks in Southern USA (Oaxaca & Ransom, 2010).

⁵⁷Monopsony power was estimated to be higher for associate and full professors.

The estimated degree of monopsony power on Chinese tobacco leaf markets is hence much larger compared to these prior studies. One reason relates to the outside option of the input suppliers. Tobacco farmers sell to a narrow set of cigarette firms, which makes their earnings sensitive to changes in downstream market structure. As discussed in the industry background, tobacco farmers face large crop switching costs, for instance due to crop-specific capital and skill requirements. Supply of inputs that are less industry-specific is more likely to be elastic. Skill requirements for grocery employees seem, for instance, not specific to grocery stores. Depending on their fields of research, it is likely that college professors can be employed outside of academia.

A second key difference lies in the degree of labor mobility costs. Artuc, Lederman, and Porto (2015) estimate labor mobility costs to be higher in China compared to the USA and Germany. Within China, mobility costs are higher in rural areas because of the Hukou permit system and land-rights uncertainty. The more costly it is for farmers to migrate in order to switch industries, the more monopsony power this gives to manufacturing firms buying their products.

More research is needed to confirm the importance of monopsony power on labor markets in other industries and other countries. The approach used in this paper offers a framework to conduct similar research in these other settings. The bulk of research on monopsony power still focuses on the USA. In many countries throughout the developing world, labor mobility costs are higher than in China (Artuc et al., 2015), which makes the existence and importance of monopsony power more likely.

Markdown and markup drivers

Which firm and county characteristics explain variation in markups and markdowns? In panel (c) of table 4, I report some correlations. Both markups and markdowns are increasing in relation to firm size, although the relationship is stronger for markdowns. State-owned enterprises have higher markdowns, but not higher markups. Counties with higher unemployment rates have higher markdown levels: worse outside options for farmers imply higher monopsony power for cigarette producers. Counties with high migration rates, defined as the share of the population that was born in another county, have much lower markdowns. Higher migration rates imply a more mobile labor force. Firms cannot decrease leaf prices too much in these counties, as workers would migrate more rapidly in response. Finally, corporate tax rates, defined as the share of revenue which is taxed by the government, are higher for high markdown firms, but not for high markup firms. This suggests that governments extract part of the rents generated by monopsony power, besides taking their share of firm profits in SOEs.

4.3 The effects of consolidation on markups, markdowns and productivity

With the markup, markdown and productivity estimates at hand, I now examine how these were affected by the ownership consolidation. I use the difference-in-differences model shown earlier, equation (1). Rather than prices, I now use markups, markdowns and productivity as the outcome variables y .

The main identifying assumptions were already outlined in section 2.3. First, the time series variation in the outcomes, conditional on all control variables, should be independent from whether firms had competitors below the exit threshold in 2002 or not. Second, the pre-trends in all outcome variables should be parallel for both treatment and control groups.

Consolidation treatment effects

The treatment effect estimates from equation (1) are in panel (a) of table 5. I use province-level market definitions.⁵⁸ Markdowns increased by 37% for firms affected by consolidation compared to firms in the control group.⁵⁹ Markups did, in contrast, fall by 23%. Productivity is estimated to have increased by 8% more for firms in consolidated markets, but this difference is not statistically significant.

The rise in markdowns and fall in markups is consistent with the fact that manufacturers face many small suppliers, but a single monopolistic buyer. The manufacturer surplus which was generated by market power gains on the leaf market seem to have been partly extracted by the government-controlled wholesaler, which exerted a downward pressure on factory-gate cigarette prices. In the next section, I will examine in more detail who benefited and lost the most from the consolidation policy.

[Table 5 here]

Treatment effects in the substitutable leaf model

Earlier in this section, I showed that markup and markdown levels were not realistic when estimating a production function with substitutable tobacco leaves. In panel (b) of table 5, I compare the

⁵⁸In appendix table A21, I re-estimate the treatment effects in equation (1) when defining input markets at the prefecture and county level. As markets are defined more narrowly, treatment effects on markups and markdowns increase, while the TFP gains remain of similar magnitudes. Firms located in the same county are more likely to be direct competitors for tobacco leaf compared to firms located at the opposite side of the same province.

⁵⁹ $= \exp(0.314) - 1$

consolidation treatment estimates when using a Cobb-Douglas model in labor, leaves and capital to the Leontief model of the previous paragraph. The markup and markdown effects are estimated to be 25% and -14% compared to the Leontief model, but have the same sign. Both models interpret a declining revenue share of intermediate inputs and increasing revenue share of labor as evidence for increased markdowns and decreased markups.

The TFP treatment effects are, in contrast, estimated to be 35% when using the substitutable inputs model, compared to 8% in the Leontief model. The reason for this difference is that input prices and input quantities are not separately observed. De Loecker and Goldberg (2014) discussed how unobserved input quantities led to biased production function coefficients when inputs differ in terms of quality. The source of bias in this paper is, in contrast, monopsony power. A drop in intermediate input prices will be interpreted as rising productivity when not accounting for endogenous input prices:

$$q_{ft} = \beta^L l_{ft} + \beta^K k_{ft} + \beta^M (m_{ft} + w_{ft}^M) + \omega_{ft}$$

In the Leontief model, intermediate inputs do not enter the estimated production function, and hence unobserved leaf prices do not enter the productivity residual. Prior work on SOE privatization and consolidation policies found that they led to large increases in profitability (Gupta, 2005; Brown, Earle, & Telegdy, 2006; Hsieh & Song, 2015; Chen et al., 2018). These profitability gains could be due to both increased monopsony power or TFP growth.

Pre-trends

I again test whether markups, markdowns and productivity had parallel trends between the treatment and control groups before the treatment started by estimating equation (2). The estimated interaction effects of the time trends and treatment categories from equation (2) are in panel (c) of table 5. Parallel pre-trends before 2003 cannot be rejected.

Heterogeneous effects

I interact the consolidation treatment effects on markdowns in equation (1) with firm and county characteristics, in order to see where markdowns increased the most, controlling for province dummies. In panel (a) of table 6, I find that the markdown increase was larger in counties in which there were more firms under the exit threshold of 100,000 cigarette cases in 2002. The change in market structure was larger where there were more firms producing under the exit threshold. I also find that the markdown increase was largest for the smallest firms. As the exiting firms were mainly small, this exit was a relatively larger shock for smaller firms.

In panel (b), I interact the treatment effects with demographic characteristics of each county. In the left columns, I find that the markdown increase was larger in counties with higher unemployment rates in 2000. Farmers had less attractive outside options in high unemployment areas, which allowed manufacturers to increase the leaf price markdown. In column 2, I interact the treatment effect variable with the migrant share of the county population in 2000. Markdowns increased more in counties with relatively more immigrants. Both unemployment and migration are, however, partially decisions made by workers. These correlations therefore do not necessarily imply causality.

[Table 6 here]

5 Distributional consequences and scale economies

5.1 Rural-urban income inequality

What were the distributional consequences of the consolidation policy? I focus on income inequality between manufacturing workers and tobacco farmers, which are based mainly in urban and rural areas respectively. Rural-urban income inequality has risen sharply in China over the past two decades (Yang, 1999; Benjamin et al., 2005; Ravallion & Chen, 2009). This has also been the case in the tobacco industry: while the average manufacturing wage grew by 14.5% per year between 1999 and 2006, average tobacco leaf prices fell by 5% per year on average. Accounting profits increased by no less than 24% per year for cigarette manufacturers. These patterns show that two margins of inequality increased: first, the income gap between manufacturing workers and farmers rose sharply, and second, the gap between firm profits and employee wages rose as well.

Tobacco leaf prices are not the only determinant of income for farmers, as production quantities and farm productivity are as well. Aggregate producer statistics from the Food and Agriculture Organization (FAO) show, however, that farm sizes remained constant and yields per acre grew by merely 1.8% per year during this time period (FAO, 2019), not enough to compensate falling leaf prices.

Back-of-the-envelope calculation

Equation (7) quantified the relationship between the leaf price and the markdown. The marginal gain from an additional unit of leaf equals the marginal revenue product of leaf $MRPM$ minus the marginal labor cost that is needed for this, MLC . The leaf price equals the ratio of this gain over the

markdown:

$$W_{ft}^M = \frac{MRPM_{ft} - MLC_{ft}}{\psi_{ft}^M}$$

I now conduct a back-of-the-envelope calculation of the extent to which the consolidation policy could have contributed to income inequality between tobacco farmers and manufacturing workers. I let markdowns of firms in the treatment group evolve in the same way as they did for firms in the control group. I assume both the marginal revenue product of leaf and marginal labor costs to have remained unchanged. Using notation from equation (1), the alternative leaf price \tilde{W}^M becomes:

$$\tilde{W}_{ft}^M = \begin{cases} W_{ft}^M \exp(\theta_2^{\psi^M}) & \text{if } t \geq 2003 \text{ \& } C_f = 1 \\ W_{ft}^M & \text{otherwise} \end{cases} \quad (12)$$

Figure 6 compares the evolution of leaf prices to manufacturing wages between 1999 and 2006, with both series being normalized at 0 in 2003. Before 2003, manufacturing wages (blue line) already outgrew leaf prices (red solid line). Between 2003 and 2006, manufacturing wages increased by 60%, while leaf prices fell by 10%. The dashed red line shows that without enforcing the exit thresholds, leaf prices would have *grown* by 20% over this time period. The consolidation hence explains 40% of the increase in income inequality between cigarette manufacturing workers and tobacco farmers.

[Figure 6 here]

Caveats

The analysis above is by no means a counterfactual simulation. First, it ignores entry and exit. Higher entry and lower exit of farmers could have led to different equilibrium leaf prices. Second, the analysis is of a partial equilibrium nature. As tobacco represents a large share of economic activity in some provinces, changes to leaf prices would also have affected equilibrium cigarette prices, manufacturing wages and prices and wages in other sectors.

5.2 Margins along the value chain

I now compare the effects of the consolidation on the surplus of all actors throughout the value chain. In equation (13), I decompose the cigarette retail price P^{ret} , which is the total value generated in the industry, into the share that farmers, manufacturers, and both wholesaler and retailers receive.⁶⁰

⁶⁰As I do not observe retail prices, I cannot distinguish wholesale from retail surplus

Farmer income is assumed to be entirely generated from tobacco leaf, and farm input prices are normalized to zero. The variable profit margin of manufacturing firms and their employees receive is the difference between the factory-gate cigarette price and the leaf price. The wholesaler/retailer receives, finally, the difference between the retail price and the factory-gate cigarette price.

$$\underbrace{\frac{W^M}{P^{ret}}}_{\text{Farmer's share}} + \underbrace{\frac{P - W^M}{P^{ret}}}_{\text{Manufacturer's share}} + \underbrace{\frac{P^{ret} - P}{P^{ret}}}_{\text{Wholesaler \& retailer's share}} = 1 \quad (13)$$

As I do not observe firm-level retail prices, I use aggregate cigarette retail price data from Zheng, Wang, Hua, and Marquez (2016) to calculate average wholesaler profit margins. Retail prices could, however, have dropped in consolidated markets, which I cannot pick up using the aggregate data.

Retail price decomposition

In panel (a) of table 7, I decompose which share of the cigarette price is allocated to farmers, manufacturers and wholesalers/retailers,⁶¹ using equation (13). The leaf price is on average 29% of the retail cigarette price. The difference between the factory-gate price and the leaf price is another 35% of the retail price. The remaining 36% of value gets generated at the wholesaling and retailing level.

How were gross profit margins of these different agents affected by consolidation? In panel (b), I estimate how these margins, calculated using equation (13), changed due to the consolidation treatment.⁶² Farmer gross profit margins fell by 39%, due to the drop in leaf prices, while manufacturing profit margins increased by 10%. The profit gain from falling leaf prices is partially undone by the drop in factory-gate cigarette prices. The wholesaler/retailer was, finally, the main winner from consolidation: wholesaling gross profit margins increased by 31%. The consolidation policy was ordered and carried through by the wholesaler's administrative counterpart, the STMA. It therefore seems likely that consolidation served as a tool to increase profits of the wholesaling monopoly, rather than as a means to increase manufacturing efficiency.

As firm-level retail prices are unobserved, I cannot distinguish wholesaler, retailer and consumer surplus. As the government-regulated cigarettes monopoly remained unchanged before and after 2003, it seems likely that the wholesaler internalized the profits from lower factory-gate prices. It cannot be ruled out, however, that this price drop was passed through to retailers and/or consumers.⁶³

⁶¹I do not observe wholesale prices, so cannot distinguish margins of wholesalers from retailers

⁶²I still use province-level market definitions.

⁶³Quantifying this pass-through would be crucial when examining the public health consequences of the consolidation

[Table 7 here]

5.3 Returns to scale vs. scale economies

The official objective for enforcing the exit thresholds was to generate scale economies. In the discussion so far, I have mainly focused on returns to scale in production. Estimating the production function revealed that there were modestly increasing returns to scale, and that physical productivity slightly increased due to the consolidation, although these increases were statistically not very significant.

Forcing small manufacturers to exit may, however, have resulted in scale economies other than increasing returns to scale. Duplicated fixed costs may have been eliminated, for instance. I use the observed profit levels Π_{ft} in order to examine these potential scale economies. Profits are defined as follows, with fixed costs FC_{ft} being all costs other than intermediate input or labor expenditure. Although I call these costs ‘fixed’, some variable costs, such as transport costs, may be part of this term as well.

$$\Pi_{ft} \equiv P_{ft}Q_{ft} - M_{ft}W_{ft}^M - L_{ft}W_{ft}^L - FC_{ft}$$

Consolidation and scale economies

I examine whether total fixed costs per market evolved differently between markets with and without exit threshold enforcement. I re-estimate the difference-in-differences model, equation (1) at the level of input markets i . I use the log total fixed costs per market i in year t , FC_{it} , as the dependent variable. I use the indicator for the market having at least one firm below the exit threshold in 2002, C_i as the independent variable and interact it with the post-treatment period:

$$\log(FC_{it}) = \theta_0^F + \theta_1^F \mathbb{I}[t \geq 2003] + \theta_2^F C_i \mathbb{I}[t \geq 2003] + \theta_3^F C_i + v_{it}^F$$

I estimate this model at the province, prefecture and county level in table 8. In column 1, I use log fixed costs as inferred from the profit figures as the dependent variable. In column 2, I use the log capital stock as the dependent variable. The sign of the treatment effects differs depending on the market definition, and the treatment effects are never significant. There is hence no strong evidence for scale economies generated by the consolidation policy.

policy. If retail cigarette prices fell, cigarette consumption could have increased in response.

6 Alternative explanations

I now revisit three key modeling assumptions. First, I test whether tobacco leaf and labor are really non-substitutable when producing cigarettes. Second, I revisit the assumption that the productivity shifter was Hicks-neutral in labor and capital. Third, I use data on brand-level product characteristics to revisit the assumption that the tobacco content per cigarette was constant across firms.

6.1 Substitutable tobacco leaf

The non-substitutability between labor and tobacco leaf has been a maintained assumption throughout this paper. It can be tested empirically. Suppose the cigarette production function is no longer given by the Leontief production function from (3), but by the following CES production function:

$$Q_{ft} = \left(\left(\beta^M M_{ft}^{\frac{\sigma^M-1}{\sigma^M}} + \beta^L L_{ft}^{\frac{\sigma^M-1}{\sigma^M}} \right)^{\frac{\sigma^M}{\sigma^M-1}} \right)^{\beta^{ML}} K_{ft}^{\beta^K} \Omega_{ft}$$

The substitution elasticity σ^M measures the extent to which labor and tobacco can be substituted. I still assume substitutability between variable inputs and capital.

Solving the first order conditions from equation (5) results in equation (14). Firms use relatively more labor compared to tobacco leaf if wages are lower, if the output elasticity of labor is relatively higher, or if firms have more monopsony power over tobacco leaf.

$$l_{ft} - m_{ft} = \sigma^M (\ln(W_{ft}^M) - \ln(W_{ft}^L)) - \sigma^M (\ln(\beta^M) - \ln(\beta^L)) + \sigma^M \ln(1 + \psi_{ft}^M) \quad (14)$$

Estimating equation (14) is subject to the same simultaneity bias as when estimating the input supply function. The extent of monopsony power of a firm affects its optimal input demand. Moreover, intermediate input prices are not observed separately. I use the same BLP instruments for wages which were introduced in the production function estimation section to estimate the elasticity of substitution between labor and materials, σ^M .

The results are in column 1 of table A17. The elasticity of substitution between intermediate inputs and labor is estimated to be 0.21, but is not significantly different from zero, in line with the Leontief model used throughout the paper. I can also re-estimate equation (14) with capital instead of tobacco leaf on the left-hand side, which allows for flexible substitution between labor and capital, but not between labor and intermediate inputs. This also requires capital to be viewed as a variable input, which I do by means of a first approximation. The elasticity of substitution estimate between labor and capital is in column 2, and is estimated to be 0.89 with a standard error of 0.22. The unitary

substitution between labor and capital which was imposed in the baseline Cobb-Douglas model hence cannot be rejected.

6.2 Static vs. dynamic demand and supply

Dynamic leaf demand

I model both leaf demand and supply decisions as being static. In reality, both are likely to have a dynamic component as well. Manufacturers hold leaf inventories, which I abstract from in the current model. If they dynamically optimize their input sourcing depending on the leaf price, this will change the implied markup estimates.

Inventories can, however, only affect short-run markup fluctuations. I show in appendix table A10 that the negative markup effect of the consolidation becomes smaller when limiting the time frame of the treatment effect from three to one year.

Dynamic leaf supply

The supply elasticities I estimated were short-run elasticities, as variation in prices and quantities using one-year intervals was used to estimate the input supply function. It is likely that leaf supply by farmers becomes more elastic when increasing the time horizon at which input supply is estimated (Hamilton, 1994). There are six months between sowing and harvesting, and it takes another two to eight weeks to cure the leaves.⁶⁴ In the short run, sowing and curing costs are sunk, and farmers will not exit as long as their variable profits are positive.

If tobacco leaf supply would be inelastic only in the short run, but elastic in the longer run, then markdowns in consolidated markups should rise immediately after the exit threshold enforcement, but fall again later, as farmers start exiting. The evidence in appendix table A10 does, however, not align with this. The drop in markdowns in consolidated markets increased during the three years following the reform, rather than falling again. I conclude from this that leaf supply is not only inelastic in the short run, but also over a longer period of at least three years.

6.3 Labor-augmenting productivity

The productivity shifter ω was assumed to be Hicks-neutral throughout the paper. What if there was factor-augmenting technical change? Equation (14) shows that factor-augmenting productivity and

⁶⁴Source: <https://www.pmi.com/glossary-section/glossary/tobacco-curing>

markdowns are not separately identified when using the production approach. Both factor-augmenting productivity and monopsony power are inferred by comparing relative input usage in their respective literatures. This potentially has implications outside of this paper. In a paper about migration in China using the same dataset as this paper, Imbert, Seror, Zhang, and Zylberberg (2018) find that increased immigration led firms to produce at a higher labor intensity. I find, however, that monopsony power was lower in counties with high migration rates, as workers are more mobile there. In a model which takes input prices as given, falling monopsony power would be picked up as lower capital intensity.

In the context of this paper, this is very unlikely to drive the stylized facts concerning input revenue shares. As the industry-wide relative cost share of labor increased relatively to leaf, this would mean that cigarette production became much less capital-intensive over time, in sharp contrast to the general trend in Chinese manufacturing. Both the levels and changes in the estimated markups could, however, change when allowing for factor-biased technical change. I redefine the $H(\cdot)$ function in production function (3) to allow for labor-specific productivity Ω^L and allow for flexible substitution between labor and capital, measured by parameter σ^K . This is similar to the production function used in Doraszelski and Jaumandreu (2017).

$$H(L_{ft}, K_{ft}) = \left(\beta^K K_{ft}^{\frac{\sigma^K-1}{\sigma^K}} + \beta^L (L_{ft} \Omega_f^L)^{\frac{\sigma^K-1}{\sigma^K}} \right)^{\frac{\sigma^K}{\sigma^K-1}} \exp(\omega_{ft})$$

Using the same derivation as for equation (14) and treating capital as a variable input, labor-augmenting productivity can be estimated as the residual of the following equation, in which W^K is the interest rate. In contrast to equation (14) in the previous section, markdowns do not enter relative input demand for capital and labor as the markdown level affects both input decisions to the same extent. The relative input demand equation becomes:

$$k_{ft} - l_{ft} = \sigma^K (\ln(W_{ft}^L) - \ln(W_{ft}^K)) + \sigma^K (\ln(\beta^K) - \ln(\beta^L)) + (1 - \sigma^K)(\omega_{ft}^L)$$

The estimates of this equation were already shown in the second column of table A17. The markup formula, (6a) can be rewritten as follows, using the same derivation as before:

$$\mu_{ft} = \left(\frac{\alpha_{ft}^L}{\beta_{ft}^L \Omega_{ft}^L} \psi_{ft}^L + \alpha_{ft}^M \psi_{ft}^M \right)^{-1}$$

I re-estimate the difference-in-differences model, equation (1), using labor-augmenting productivity as the left-hand side variable. The results are in appendix table A18. Labor-augmenting productivity increased by 26%, 15% and 4% due to the consolidation when using province-, prefecture and county-level market, respectively.

In appendix table A13, I re-estimate the consolidation treatment effects on markdowns, markups

and total factor productivity, but add the log capital to labor ratio as a control variable. Controlling for this variation in capital intensity across firms and over time does not change the treatment effect estimates by much.

6.4 Heterogeneous product characteristics

Finally, I revisit the assumption that the tobacco leaf content per cigarette, β_{ft}^M , was constant across firms. I use the product characteristics data on a subset of firms to show that in reality, there was limited variation in tobacco concentration and in other cigarette characteristics, such as ventilation rates and paper quality. The entire distribution of leaf contents per cigarette lies between 63% and 75%, so the decline in the leaf share of revenue cannot be due to changes in the leaf content alone. Moreover, as long as product characteristics were similar between the control and treatment groups, they do not affect the difference-in-difference estimates. Table A19 shows this is indeed the case: firms in the treatment and control groups did not differ significantly in terms of any product characteristic, except that they were less likely to be ‘ventilated’ cigarettes.

The second table in table A19 shows, however, that markdowns did not correlate with any product characteristic. Hence, the observed markdown increase for the treatment group cannot be due merely to changes in cigarette design. Markups do correlate with product characteristics: cigarettes with higher filter densities and more tobacco content are sold at higher markups, as these are more likely to be of higher quality. The physical cost to produce cigarettes with different product characteristics is very similar as both TFP and the leaf price are not significantly different across cigarette product types.

6.5 Further robustness checks

In appendix C, I carry out multiple other robustness checks. I narrow down the treatment time window from 2004-2006 to 2004 only. I also compare exporting behavior across treatment and control groups, use different functional forms for the production function and include non-wage benefits as a variable labor cost. None of these robustness checks alter the main findings to a meaningful extent.

7 Conclusion

In this paper, I examine the effects of ownership consolidation on buying power, market power and productive efficiency. I find that a large-scale ownership consolidation in the Chinese tobacco industry allowed cigarette manufacturers to increase the tobacco leaf price markdown by 37%. This

monopsony power was both due to large crop switching costs, rural migration impediments, and a large increase in downstream market concentration. For urban manufacturing workers, many of these frictions played a smaller role, and I find no evidence for monopsony power in manufacturing labor markets. In response to the rise in monopsony power on leaf markets, monopsonic wholesalers reacted by increasing their own margins in the wholesaling market. Consolidation hence resulted not only in lower leaf prices, but also in lower factory-gate cigarette prices. I find that markups of cigarette manufacturers fell by 24%.

The main motivation to consolidate the tobacco industry was to increase scale economies. I find weak evidence for increasing returns to scale and for modest gains in total factor productivity. Taking endogenous input prices into account is crucial to interpret these productivity gains, and wrongly imposing exogenous input prices lead to the erroneous conclusion that productive efficiency increased by 30%. I find no evidence for other types of scale economies, such as reduced fixed costs.

The consolidation policy had important distributional consequences. The main beneficiary of consolidation was the government-controlled wholesaler, which monopolizes cigarette markets. Farmer income was hit by the surge in monopsony power, and I find that forcing small manufacturers to exit explained 42% of the rise in income inequality between tobacco farmers and manufacturing employees between 2003 and 2006. As these groups mainly live in rural and urban areas respectively, the consolidation policy contributed to the rise in urban-rural income inequality in the tobacco industry.

On the methodological level, I discuss the role of input substitution patterns for identification of markups, markdowns and productive efficiency using the production and cost approach. I show that low input substitutability limits the identification of markups and markdowns using the production and cost approach. I propose a solution to this challenge by complementing the production and cost model with an input supply model.

Large-scale consolidation waves are a frequently used policy tool in transitioning countries, and there is increased M&A activity in developed countries, such as the U.S. In order to fully understand the effects of consolidation, it is crucial to take into account their effects on both productive efficiency, scale economies, and competition on input and product markets.

References

- Akerberg, D., Benkard, L., Berry, S., & Pakes, A. (2007). Econometric tools for analyzing market outcomes. In *Handbook of econometrics* (Vol. 6, p. 4171-4276).
- Akerberg, D., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6), 2411-2451.
- Artuc, E., Lederman, D., & Porto, G. (2015). A mapping of labor mobility costs in the developing world. *Journal of International Economics*, 95(1), 28–41.
- Ashenfelter, O. C., Farber, H., & Ransom, M. R. (2010). Labor market monopsony. *Journal of Labor Economics*, 28(2), 203-210.
- Askari, H., & Cummings, J. T. (1977). Estimating agricultural supply response with the Nerlove model: a survey. *International economic review*, 257–292.
- Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2017). Concentrating on the fall of the labor share. *American Economic Review: Papers and Proceedings*, 107(5), 180–85.
- Baltagi, B. H., & Levin, D. (1986). Estimating dynamic demand for cigarettes using panel data: the effects of bootlegging, taxation and advertising reconsidered. *The Review of Economics and Statistics*, 148–155.
- Benjamin, D., Brandt, L., & Giles, J. (2005). The evolution of income inequality in rural china. *Economic Development and Cultural Change*, 53(4), 769-824.
- Berry, S., Azar, J., & Marinescu, I. (2019). *Estimating labor market power*.
- Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 64(4), 841-890.
- Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.
- Blonigen, B., & Pierce, J. (2016). *Evidence for the effects of mergers on market power and efficiency* (Working paper No. 22750). National Bureau of Economic Research.
- Braguinsky, S., Ohyama, A., Okazaki, T., & Syverson, C. (2015). Acquisitions, productivity and profitability: Evidence from the Japanese cotton spinning industry. *The American Economic Review*, 105(7), 2086-2119.
- Brandt, L., Van Biesebroeck, J., Wang, L., & Zhang, Y. (2017). WTO accession and performance of Chinese manufacturing firms. *American Economic Review*, 107(9), 2784-2820.
- Brandt, L., Van Biesebroeck, J., & Zhang, Y. (2012). Creative accounting or creative destruction? firm-level productivity growth in Chinese manufacturing. *Journal of Development Economics*, 97(2), 339-351.
- Brooks, W. J., Kaboski, J. P., Yao, A. L., & Qian, W. (2019). *Exploitation of labor? classical monopsony power and labor's share* (Working Paper No. 25660). NBER.

- Brown, J. D., Earle, J. S., & Telegdy, A. (2006). The productivity effects of privatization: Longitudinal estimates from Hungary, Romania, Russia, and Ukraine. *Journal of Political Economy*, 114(1), 61–99.
- Card, D., & Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4), 772-793.
- Chatterjee, S. (2019). *Market power and spatial competition in rural india* (Working Paper No. 2019/04). Cambridge INET.
- Chen, Y., Igami, M., Sawada, M., & Xiao, M. (2018). Privatization and productivity in china. *Available at SSRN 2695933*.
- Ciliberto, F., & Williams, J. W. (2014). Does multimarket contact facilitate tacit collusion? inference on conduct parameters in the airline industry. *The RAND Journal of Economics*, 45(4), 764–791.
- Covarrubias, M., Gutiérrez, G., & Philippon, T. (2019). *From Good to Bad Concentration? US Industries over the past 30 years* (Tech. Rep.). National Bureau of Economic Research.
- Delipalla, S., & O’Donnell, O. (2001). Estimating tax incidence, market power and market conduct: The European cigarette industry. *International Journal of Industrial Organization*, 19(6), 885–908.
- De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics*, 5(3), 1-21.
- De Loecker, J., & Eeckhout, J. (2018). *Global market power* (Working Paper No. 24768). NBER.
- De Loecker, J., Eeckhout, J., & Unger, G. (2018). *The rise of market power and the macroeconomic implications* (Working Paper). <https://sites.google.com/site/deloeckerjan/research/RMPDLEU.pdf>.
- De Loecker, J., Goldberg, P., Khandelwal, A., & Pavcnik, N. (2016). Prices, markups and trade reform. *Econometrica*, 84(2), 445-510.
- De Loecker, J., & Goldberg, P. K. (2014). Firm performance in a global market. *Annu. Rev. Econ.*, 6(1), 201–227.
- De Loecker, J., & Scott, P. T. (2016). *Estimating market power: evidence from the us brewing industry* (Working Paper No. 22957). National Bureau of Economic Research.
- De Loecker, J., & Warzynski, F. (2012). Markups and firm-level export status. *The American Economic Review*, 102(6), 2437-2471.
- Doraszelski, U., & Jaumandreu, J. (2017). Measuring the bias of technological change. *Journal of Political Economy*, 126(3), 1027-1084.
- Dunn, J. E., & Heien, D. (1985). The demand for farm output. *Western Journal of Agricultural Economics*, 10(1836-2016-150642), 13–22.
- Fang, J., Lee, K., & Sejpal, N. (2017). The China National Tobacco Corporation: From domestic to global dragon? *Global Public Health*, 12(3), 315-334.

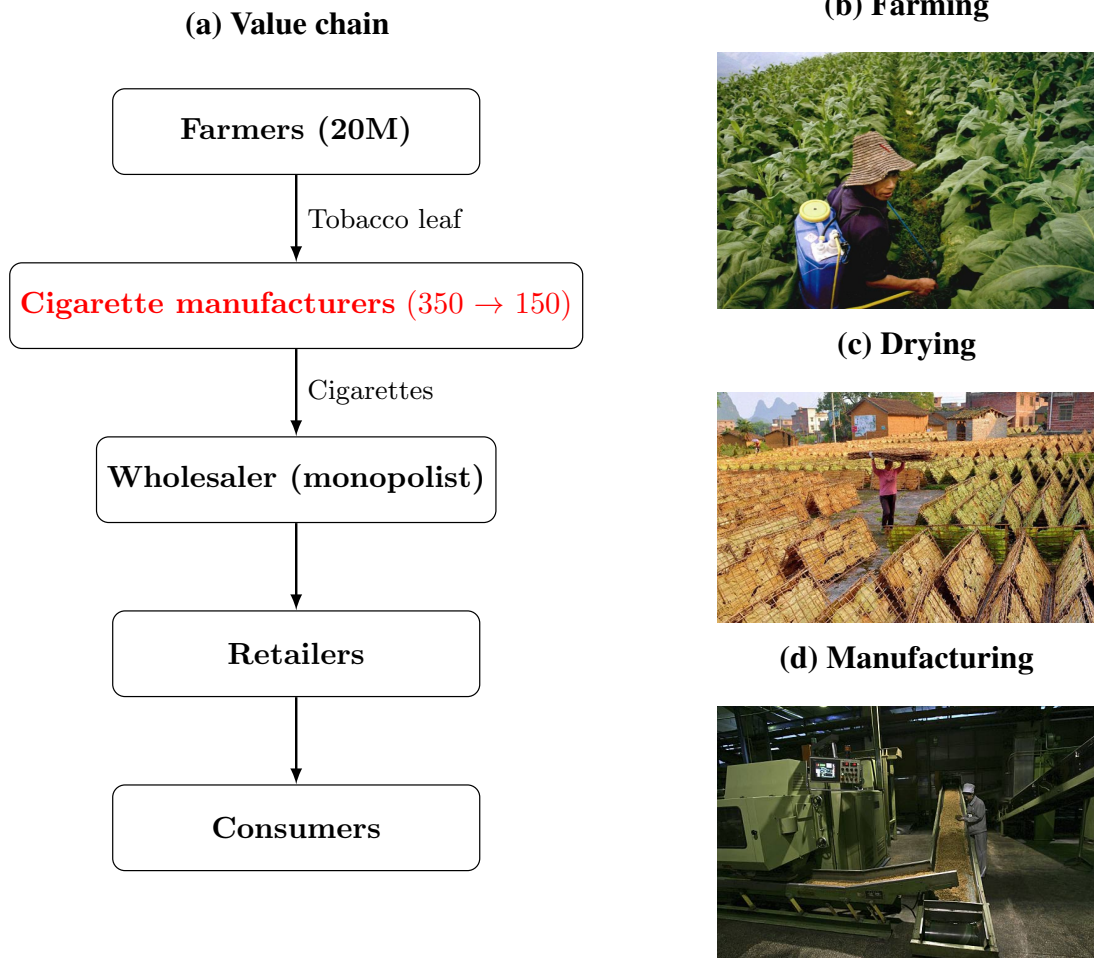
- FAO. (2003). Issues in the global tobacco economy: Selected case studies. In *Fao commodity studies* (Vol. 1810-0783). Food and Agriculture Organization of the United Nations.
- FAO. (2019). *Faostat statistical database*.
- Ghandi, A., Navarro, S., & Rivers, D. (2018). *On the identification of gross output production functions*.
- Goodchild, M., & Zheng, R. (2018). Early assessment of China's 2015 tobacco tax increase. *Bulletin of the World Health Organization*, 96(7), 506-512.
- Goolsbee, A., & Syverson, C. (2019). *Monopsony power in higher education: a tale of two tracks* (Tech. Rep.). National Bureau of Economic Research.
- Gowrisankaran, G., & Rysman, M. (2012). Dynamics of consumer demand for new durable goods. *Journal of Political Economy*, 120(6), 1173–1219.
- Grieco, P., Pinkse, J., & Slade, M. (2017). Brewed in North America: Mergers, marginal costs and efficiency. *International Journal of Industrial Organization*, 59(2018), 24-64.
- Griliches, Z., & Mairesse, J. (1998). Production functions: The search for identification. In S. Strom (Ed.), *Econometrics and economic theory in the twentieth century* (pp. 2413-2415, 2417, 2421). Cambridge University Press.
- Gupta, N. (2005). Partial privatization and firm performance. *The Journal of Finance*, 60(2), 987–1015.
- Hall, R. E. (1986). Market structure and macroeconomic fluctuations. *Brookings Papers on Economic Activity*, 2(1986), 285-322.
- Hamilton, J. L. (1994). Joint oligopsony-oligopoly in the US leaf tobacco market, 1924–39. *Review of Industrial Organization*, 9(1), 25–39.
- Harris, J. E. (1998). *The price of cigarettes and the profits of cigarette manufacturers with and without federal intervention, 1997-2006* (Report). American Cancer Society.
- Hirsch, B., Schank, T., & Schnabel, C. (2010). Differences in labor supply to monopsonistic firms and the gender pay gap: An empirical analysis using linked employer-employee data from Germany. *Journal of Labor Economics*, 28(2), 291–330.
- Hsieh, C.-T., & Song, Z. (2015). *Grasp the large, let go of the small: The transformation of the state sector in China* (Working Paper No. 21006). NBER.
- Hu, T., Mao, Z., Ong, M., Tong, E., Tao, M., Jiang, H., . . . Yurekli, A. (2006). China at the crossroads: the economics of tobacco and health. *Tobacco Control*, 15(1), 37-41.
- Imbert, C., Seror, M., Zhang, Y., & Zylberberg, Y. (2018). Migrants and firms: Evidence from China.
- Lan, L. (2015, 05). *Market will now determine tobacco prices*. Chinadaily.com.cn.
- Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2), 317-341.
- Li, V., Wang, Q., Xia, N., Tang, S., & Wang, C. (2012). Tobacco crop substitution: pilot effort in China. *American Journal of Public Health*, 102(9), 1660-1663.

- Lu, Y., & Yu, L. (2015). Trade liberalization and markup dispersion: evidence from China's WTO accession. *American Economic Journal: Applied Economics*, 7(4), 221-253.
- Macchiavello, R. (2018). A mutually beneficial relationship: relational contracts in developing countries. *A Research Agenda for New Institutional Economics*, 53.
- Manning, A. (2003). *Monopsony in motion: Imperfect competition in labor markets*. Princeton University Press.
- Matsudaira, J. D. (2014). Monopsony in the low-wage labor market? Evidence from minimum nurse staffing regulations. *Review of Economics and Statistics*.
- Michel, C., & Weiergraeber, S. (2018). *Estimating industry conduct in differentiated products markets*.
- Miller, N. H., & Weinberg, M. C. (2017). Understanding the price effects of the millercoors joint venture. *Econometrica*, 85(6), 1763–1791.
- Minale, L. (2018). Agricultural productivity shocks, labour reallocation and rural–urban migration in China. *Journal of Economic Geography*, 18(4), 795–821.
- Morlacco, M. (2017). *Market power in input markets: theory and evidence from French manufacturing* (Tech. Rep.).
- Naidu, S., Nyarko, Y., & Wang, S.-Y. (2016). Monopsony power in migrant labor markets. *Journal of Political Economy*, 124(6), 1735-1792.
- Nargis, N., Zheng, R., Xu, S. S., Fong, G. T., Feng, G., Jiang, Y., ... Hu, X. (2019). Cigarette affordability in China, 2006–2015: findings from International Tobacco Control China surveys. *International journal of environmental research and public health*, 16(7), 1205.
- Naughton, B. (2007). *The Chinese Economy: Transitions and Growth*. Cambridge, MA: MIT Press.
- Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2), 307-342.
- Oaxaca, R. L., & Ransom, M. R. (2010). New market power models and sex differences in pay. *Journal of Labor Economics*, 28(2), 267–289.
- O'Connor, R., Li, Q., Stephens, W. E., Hammond, D., Elton-Marshall, T., Cummings, K. M., ... Fong, G. T. (2010). Cigarettes sold in China: Design, emissions and metals. *Tobacco Control*, 19(2), 47-53.
- Olley, S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6), 1263-1297.
- Peng, Y. (1996). The politics of tobacco: relations between farmers and local governments in China's southwest. *The China Journal*, 36, 67-82.
- Ransom, M. R., & Sims, D. P. (2010). Estimating the firm's labor supply curve in a "new monopsony" framework: schoolteachers in Missouri. *Journal of Labor Economics*, 28(2), 331–355.
- Ravallion, M., & Chen, S. (2009). China's (uneven) progress against poverty. In (p. 65-111). Routledge.

- Rossi-Hansberg, E., Sarte, P.-D., & Trachter, N. (2018). *Diverging trends in national and local concentration* (Tech. Rep.). National Bureau of Economic Research.
- Scott, P. T. (2013). *Dynamic discrete choice estimation of agricultural land use*.
- Song, F., Zhao, J., & Swinton, S. M. (2011). Switching to perennial energy crops under uncertainty and costly reversibility. *American Journal of Agricultural Economics*, 93(3), 768–783.
- State Administration of Taxation. (2015). *Annual report* (Tech. Rep.). State Administration of Taxation.
- State Council of the People's Republic of China. (1997). *Regulation for the implementation of the law on tobacco monopoly of the People's Republic of China*.
- STMA. (2002). *Implementation opinions of the state tobacco monopoly administration on the organizational structure adjustment of cigarette industry enterprises in the tobacco industry*.
- Sumner, D. A., & Alston, J. M. (1987). Substitutability for farm commodities: the demand for us tobacco in cigarette manufacturing. *American Journal of Agricultural Economics*, 69(2), 258–265.
- Syverson, C. (2004). Market structure and productivity: A concrete example. *Journal of Political Economy*, 112(6), 1181–1222.
- Syverson, C. (2019). Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives*, 33(3), 23–43.
- Talhout, R., Richter, P. A., Stepanov, I., Watson, C. V., & Watson, C. H. (2018). Cigarette design features: effects on emission levels, user perception, and behavior. *Tobacco regulatory science*, 4(1), 592–604.
- Tortarolo, D., & Zarate, R. (2018). *Measuring imperfect competition in product and labor markets* (Mimeo). UC Berkeley.
- United Nations. (2019). *Un comtrade database*. (<http://comtrade.un.org/>)
- U.S. Census Bureau. (1997a). *1997 economic census of manufacturing*.
- U.S. Census Bureau. (1997b). *Cigarette manufacturing* (1997 Economic Census - Industry Series). US Census Bureau.
- Van Biesebroeck, J. (2007). The robustness of productivity estimates. *Journal of Industrial Economics*, 55(3), 529-569.
- Wang, J. (2013). *State-Market Interaction in China's Reform Era: Local State Competition and Global Market-building in the Tobacco Industry*. Routledge.
- Wollmann, T. (2019). *Getting away with merger: The case of dialysis clinics in the United States*. (Working Paper)
- World Economic Forum. (2019). *World Economic Forum Global Competitiveness Index*.
- Yang, D. T. (1999). Urban-biased policies and rising income inequality in China. *American Economic Review: Papers and Proceedings*, 89(2), 306-310.

Zheng, R., Wang, Y., Hua, X., & Marquez, P. V. (2016). *Cigarette affordability in China, 2001-2016*.
World Bank.

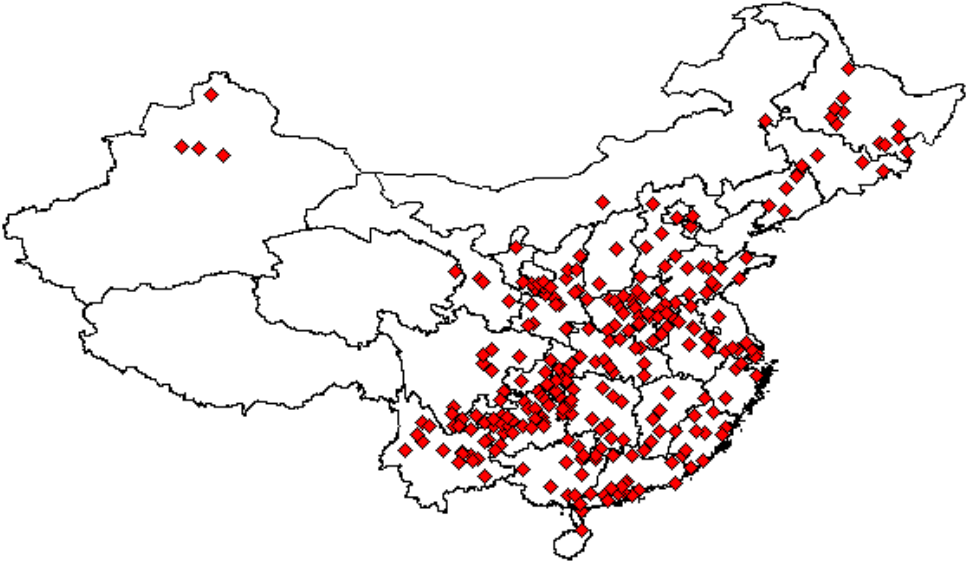
Figure 1: Tobacco industry structure



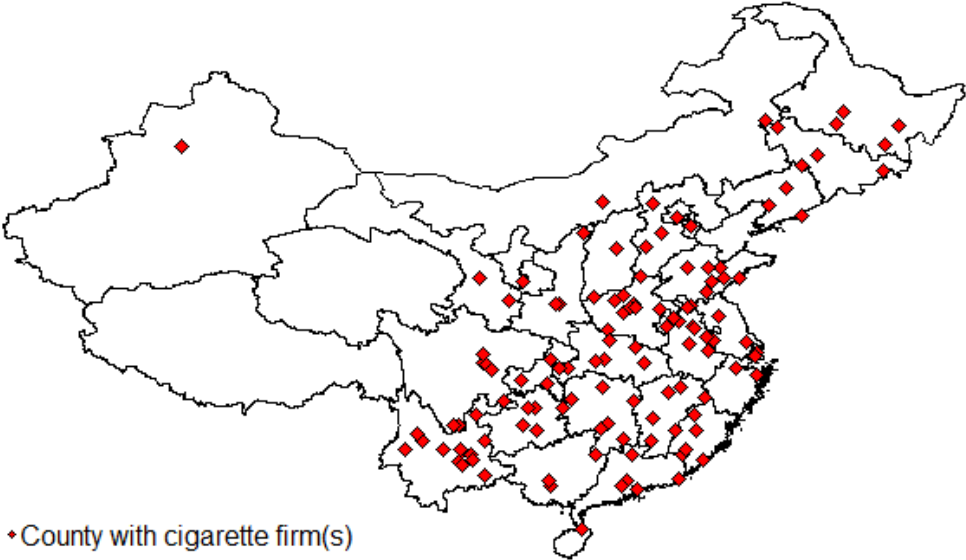
Notes: Figure (a) gives a schematic overview of the consecutive actors in the cigarette value chain in China. “CNTTC” stands for *Chinese National Tobacco Trade Company*, and is the wholesaling arm of the CNTC/STMA. This is a government-controlled monopolist. Figures (b)-(d) show images of the production process steps. Copyrights go to (a) ©Hu(2018), (b) ©news.cn, and (c) ©Getty Images

Figure 2: Cigarette manufacturing locations

(a) 1999



(b) 2006

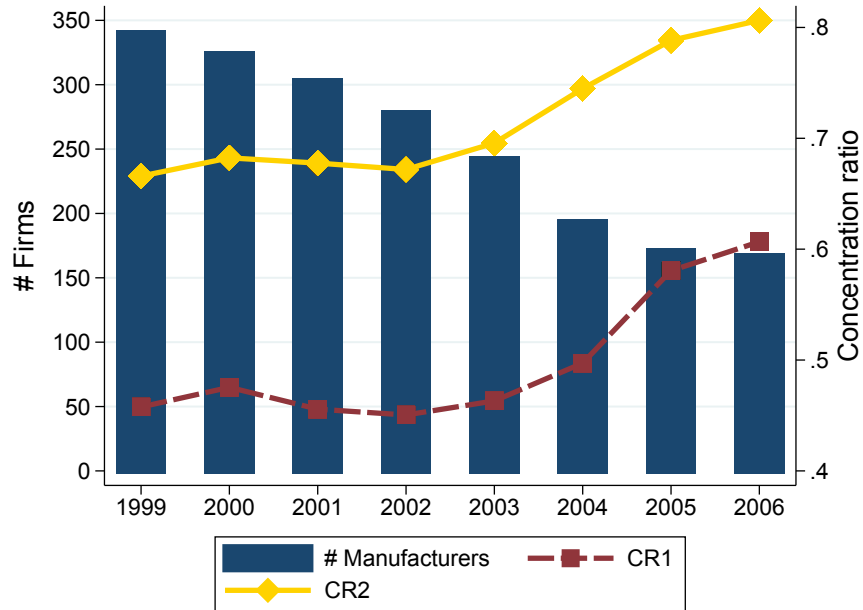


◆ County with cigarette firm(s)

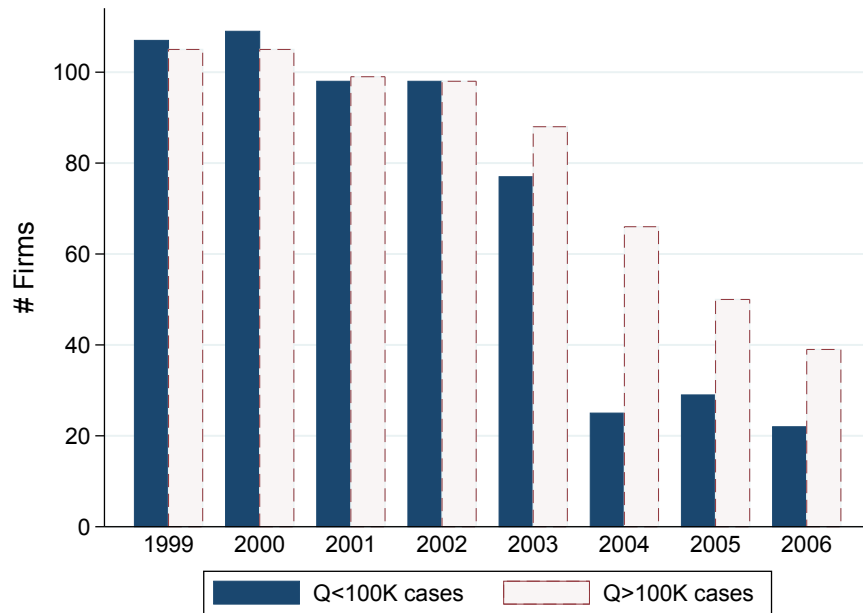
Notes: These maps depict all counties with at least one cigarette manufacturing firm in 1999 and 2006. In counties with at least one cigarette manufacturer, there were on average 1.24 firms.

Figure 3: Market structure

(a) All firms

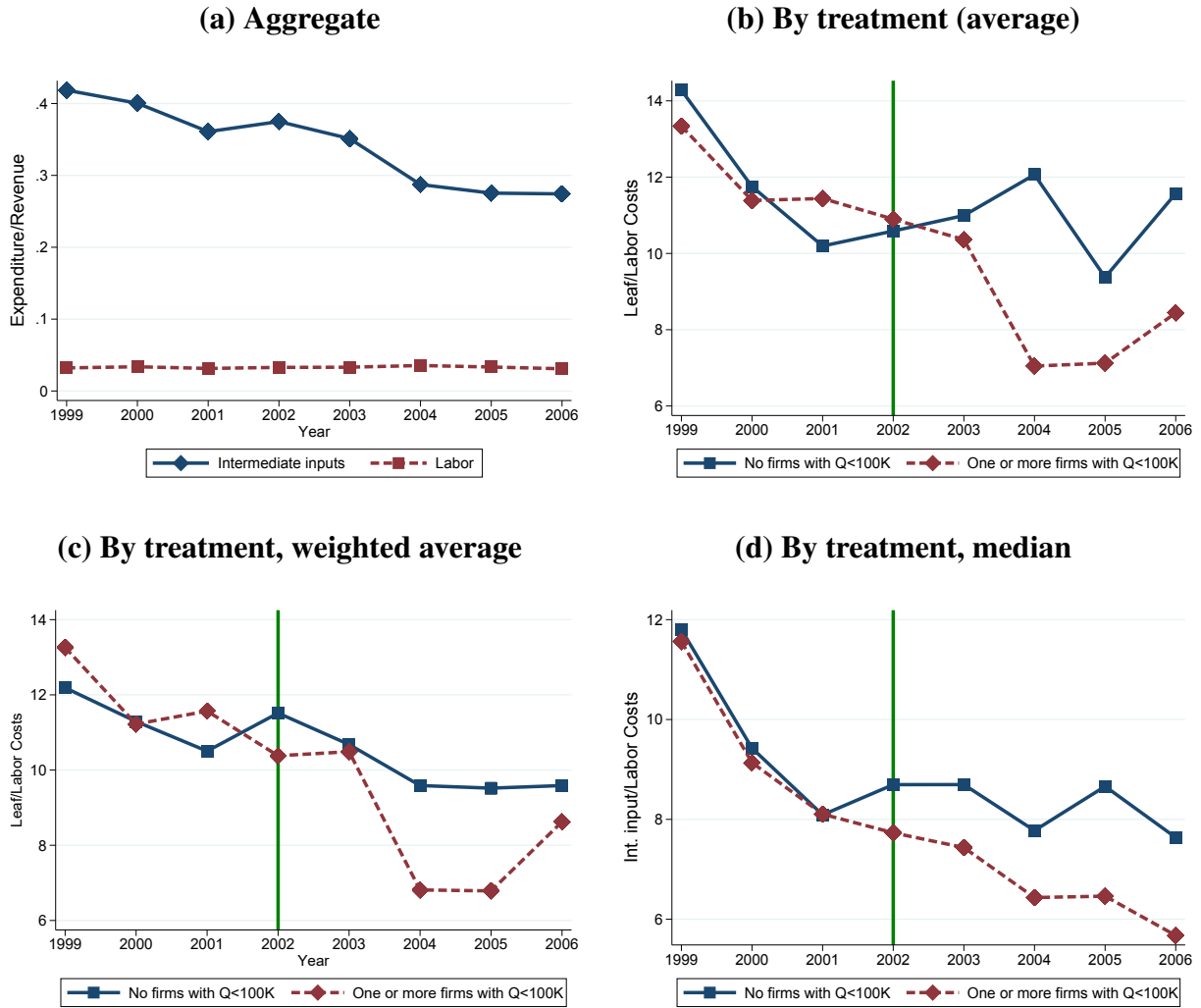


(b) Firms below vs. above size threshold



Notes: Panel (a) shows the evolution of the total number of cigarette manufacturers in China (left axis) and the combined provincial market shares of the two biggest firms per province (right axis). Panel (b) breaks this evolution down into firms below and above the exit threshold of 100,000 cases per year. This graph excludes firms for which quantities are unknown, which is why the total number of firms in panel (b) is lower compared to panel (a).

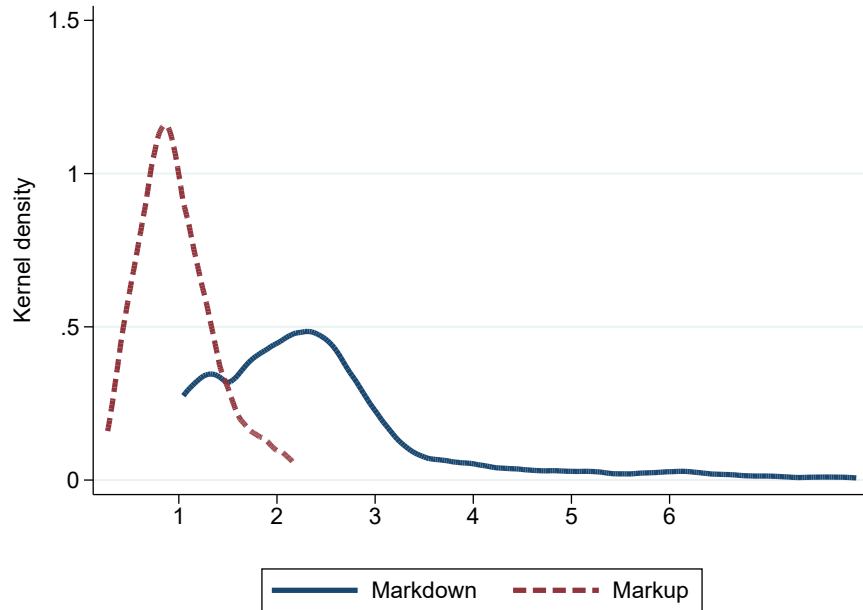
Figure 4: Factor revenue shares



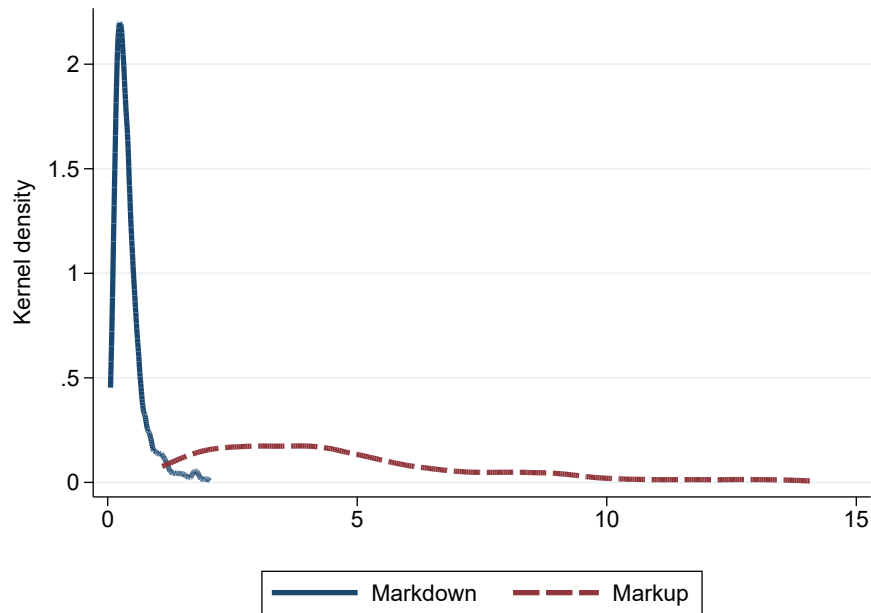
Notes: Panel (a) plots the evolution of the total wage bill and total intermediate input expenditure over industry revenue. Panels (b) to (d) plot the ratio of labor expenditure over intermediate input expenditure over time for two groups of firms. The treatment group includes firms with at least one competitors below the size threshold in 2002, while the control group are firms without any of these competitors in 2002. In panel (b), I calculate the unweighted average for the control and treatment groups. In panel (c), I weight by labor expenditure. In panel (d), I calculate the median. Market definitions are at the prefecture level.

Figure 5: Markups and markdowns

(a) Leontief model

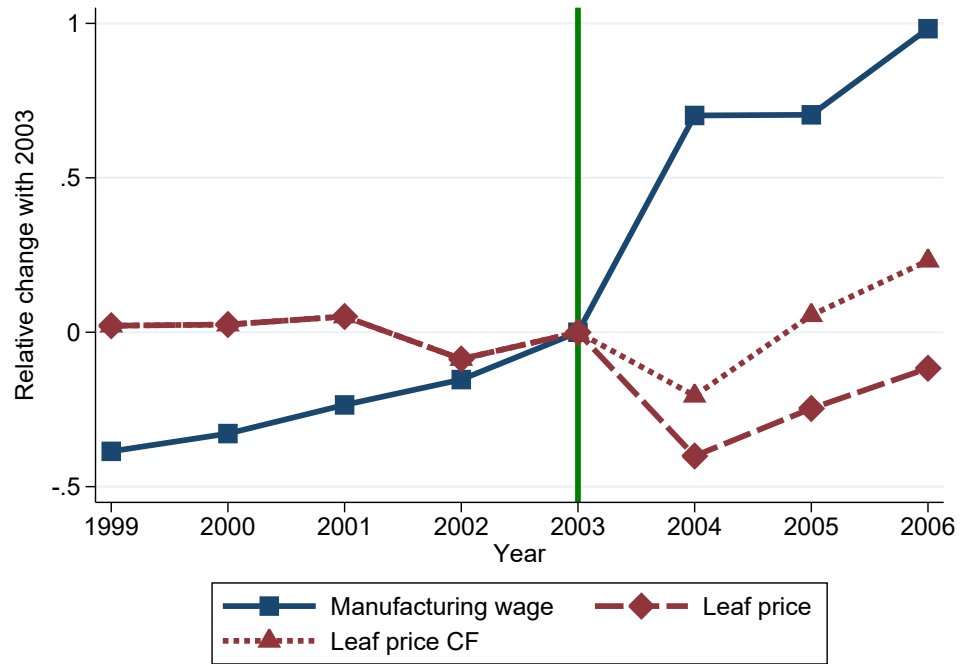


(b) Cobb-Douglas model



Notes: Panel (a) plots markdowns and markups using the baseline model in which tobacco leaf is not substitutable. Panel (b) does the same using the Cobb-Douglas model with substitutable tobacco leaf. All distributions are censored at 5th and 95th percentiles.

Figure 6: No consolidation counterfactual



Notes: The solid lines plot change in average manufacturing wages and leaf prices compared to 2003 (normalization: 2003=0). The dashed line plots the counterfactual leaf price evolution in the counterfactual scenario in which the exit thresholds were not enforced.

Table 1: Selected summary statistics

Variable	Mean	Std. Dev.	N
Revenue (million \$)	104.6	195.23	1109
Quantity (million cases)	0.34	0.42	1109
Price per case (\$)	1623.09	14118.72	1109
Profit (million \$)	11.73	42.52	1109
Wage bill (million \$)	3.46	6.06	1109
Employees (thousands)	1.19	1.06	1109
Material expenditure (million \$)	35.59	51.16	1109
Capital stock (million \$)	47.71	71.14	1109
Export dummy	0.22	0.42	1109
Export share of revenue	0.01	0.05	1109
County population (millions)	71.67	49.52	792
Leaf content per cigarette (mg)	681.47	31.07	185
Filter density (mg/ml)	112.8	3.68	185

Notes: A case contains 50,000 cigarette sticks. Reported prices are factory-gate prices. All monetary variables are denoted in 2006 US dollars.

Table 2: Reduced-form evidence: consolidation and prices

<i>Panel (a): Province-level markets</i>						
	log(Wage)		log(Leaf price)		log(Cigarette price)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	-0.160	(0.086)	-0.467	(0.116)	-0.391	(0.091)
R-squared	0.800		0.862		0.863	
Observations	1,091		1,091		1,091	
<i>Panel (b): Prefecture-level markets</i>						
	log(Wage)		log(Leaf price)		log(Cigarette price)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	-0.128	(0.082)	-0.466	(0.120)	-0.257	(0.100)
R-squared	0.800		0.865		0.863	
Observations	1,091		1,091		1,091	
<i>Panel (c): County-level markets</i>						
	log(Wage)		log(Leaf price)		log(Cigarette price)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	-0.056	(0.117)	-0.747	(0.170)	-0.349	(0.123)
R-squared	0.800		0.870		0.864	
Observations	1,091		1,091		1,091	
<i>Panel (d): Pre-trends</i>						
	log(Wage)		log(Leaf price)		log(Cigarette price)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * Year	-0.031	(0.090)	-0.189	(0.156)	-0.145	(0.137)
R-squared	0.031		0.021		0.020	
Observations	764		764		764	

Notes: I estimate how labor wages, leaf prices and factory-gate cigarette prices change differently between firms with and without any competitors below the exit threshold in 2002. Panel (a) defines the input market at the province-level, panel (b) at the prefecture-level and panel (c) at the county-level. Controls include firm fixed effects, export dummy, ownership type dummies and a linear time trend. Panel (d) estimates the pre-trends for all three outcomes in the period 1999-2003, with treatment effects being defined at the province-level.

Table 3: Structural model estimates

<i>Panel (a): Output elasticities</i>				
	ACF		IV	
	Estimate:	SE:	Estimate:	SE:
Labor	0.320	(0.226)	0.404	(0.096)
Capital	0.761	(0.140)	0.699	(0.107)
Returns to scale	1.081	(0.191)	1.104	(0.079)
R-squared	0.989		0.887	
1st stage F-statistic	–		46.79	
Observations	823		1,108	
<i>Panel (b): Leaf supply semi-elasticity</i>				
	OLS		IV	
	Estimate:	SE:	Estimate:	SE:
Leaf price	-0.172	(0.097)	3.671	(0.862)
R-squared	0.863		0.119	
1st stage F-statistic	–		79.14	
Observations	1,091		1,091	
<i>Panel (c): Labor supply semi-elasticity</i>				
	OLS		IV	
	Estimate:	SE:	Estimate:	SE:
Labor	-0.0001	(0.001)	-0.029	(0.068)
R-squared	0.904		0.793	
1st stage F-statistic	–		21.89	
Observations	1,091		1,091	

Notes: Panel (a) reports the output elasticities which are obtained from estimating the production function. Using ACF reduces the sample size as the lagged variables need to be observed. Panels (b) and (c) report the estimated price semi-elasticities of input supply, both for tobacco leaf and labor. The left-hand side variable is the log province-level market share minus the log outside option market share. The endogenous right-hand side variable is the leaf price for one pack of cigarettes in 1000 RMB. Manufacturing TFP from the Leontief model is used as instrument. I control for prefecture dummies, ownership dummies, and cigarette prices. Unit wages are also controlled for in the leaf supply estimation model. Bootstrapped standard errors are in the parentheses.

Table 4: Markups and markdowns

<i>Panel (a): Leontief model</i>				
	Markdown		Markup	
	Estimate	S.E.	Estimate	S.E.
Mean	2.869	(0.433)	1.062	(0.092)
Median	2.235	(0.286)	0.923	(0.086)
Observations	1,091		1,091	
<i>Panel (b): Cobb-Douglas model</i>				
	Markdown		Markup	
	Estimate	S.E.	Estimate	S.E.
Mean	0.768	(1.510)	5.629	(8.127)
Median	0.371	(0.621)	3.440	(5.946)
Observations	1,091		1,091	
<i>Panel (c): Correlations</i>				
	log(Markdown)		log(Markup)	
	Estimate	S.E.	Estimate	S.E.
log(Output)	0.113	(0.068)	0.064	(0.036)
1(SOE)	1.266	(0.221)	-0.064	(0.119)
log(Unemployment rate)	0.233	(0.106)	-0.112	(0.057)
log(Migration rate)	-0.304	(0.116)	0.129	(0.053)
log(Tax rate)	0.285	(0.054)	-0.055	(0.045)
R-squared	0.577		0.862	
Observations	751		751	

Notes: Panel (a) shows key moments for the markdown and markup distribution using the baseline Leontief model in which tobacco leaf cannot be substituted for labor or capital. Panel (b) does the same for the model that is Cobb-Douglas in leaf, labor and capital. Panel (c) shows correlations of the markup and markdown distribution from the Leontief model with selected firm and county characteristics. Province-dummies are controlled for. Bootstrapped standard errors are in parentheses.

Table 5: Consolidation treatment effects

<i>Panel (a): Leontief model</i>						
	log(Markdown)		log(Markup)		log(Productivity)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	0.314	(0.065)	-0.265	(0.065)	0.075	(0.082)
1(year \geq 2003)	-0.320	(0.071)	0.334	(0.075)	-0.159	(0.091)
R-squared	0.811		0.764		0.867	
Observations	1,091		1,091		1,091	
<i>Panel (b): Cobb-Douglas model</i>						
	log(Markdown)		log(Markup)		log(Productivity)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	0.223	(0.092)	-0.147	(0.068)	0.298	(0.084)
1(year \geq 2003)	-0.083	(0.100)	0.088	(0.078)	-0.339	(0.092)
R-squared	0.731		0.758		0.871	
Observations	1,091		1,091		1,091	
<i>Panel (c): Pre-trend</i>						
	log(Markdown)		log(Markup)		log(Productivity)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment*Year	0.089	(0.084)	-0.077	(0.093)	-0.054	(0.042)
Treatment	-178.452	(167.755)	154.931	(185.117)	107.538	(84.231)
Year	-0.046	(0.082)	0.095	(0.090)	0.042	(0.030)
R-squared	0.316		0.095		0.404	
Observations	756		756		756	

Notes: Panel (a) estimates the differences-in-differences model using the Leontief specification and province-level input market definitions. Controls include firm fixed effects, a linear time trend, ownership type dummies and export dummies. Panel (b) estimates the same regression, but uses the Cobb-Douglas model in labor, capital and tobacco leaf instead. I use the same sample as in panel (a), even though the Cobb-Douglas markups and markdowns are available for a larger sample as they do not require observable quantities. Panel (c) estimates the pre-trends on the period 1999-2002, and uses the Leontief model estimates. Bootstrapped standard errors are in parentheses.

Table 6: Heterogeneous treatment effects

<i>Panel (a): Firm characteristics</i>	log(Markdown)		log(Markdown)	
	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	-0.808	(0.756)	4.925	(1.134)
Treat*1(year \geq 2003)*#firms under 100K	1.379	(0.607)		
Treat*1(year \geq 2003)*firm output			-0.364	(0.096)
R-squared	0.338		0.470	
Observations	779		771	
<i>Panel (b): County characteristics</i>	log(Markdown)		log(Markdown)	
	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	2.326	(0.813)	3.948	(1.017)
Treat*1(year \geq 2003)*log(unemployment rate)	0.457	(0.215)		
Treat*1(year \geq 2003)*log(migration rate)			0.832	(0.268)
R-squared	0.336		0.327	
Observations	779		771	

Notes: Panel (a) interacts the consolidation treatment effects with the number of competitors producing below the exit threshold before 2003 and with the firm's output level. Panel (b) interacts the treatment effects with the county-level unemployment and migration rates. The migration rate is defined as the share of the county population who are born in another province.

Table 7: Consolidation and margins along the chain

<i>Panel (a): Share of industry surplus</i>	Farming		Manufacturing		Wholesale+Retail	
	0.286		0.352		0.362	
<i>Panel (b): Log(GPM*) of:</i>	Farmers		Manufacturers		Wholesalers	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	-0.498	(0.101)	0.091	(0.055)	0.273	(0.089)
Observations	1,091		1,091		697	
R-squared	0.846		0.684		0.873	

Notes: *Gross Profit Margin. Panel (a) shows the division of industry surplus between farmers, manufacturers and wholesalers. Province-level market definitions are used. Panel (b) estimates how the gross profit margins of farmers, manufacturers and wholesalers changed due to the consolidation. Linearly fitted retail prices are used in 1999 and 2000. The three effects do not need to add up to zero as consumer surplus changed as well. Province-level market definitions are used.

Table 8: Scale economies

<i>Panel (a): Province level</i>				
	log(Fixed costs)		log(Capital stock)	
	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	-0.310	(0.356)	-0.051	(0.309)
R-squared	0.295		0.236	
Observations	213		213	
<i>Panel (b): Prefecture level</i>				
	log(Fixed costs)		log(Capital stock)	
	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	0.269	(0.241)	0.043	(0.192)
R-squared	0.154		0.079	
Observations	743		763	
<i>Panel (c): County level</i>				
	log(Fixed costs)		log(Capital stock)	
	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	-0.211	(0.257)	0.151	(0.210)
R-squared	0.266		0.253	
Observations	933		993	

Notes: I estimate how total fixed costs and the capital stock at the input market level were affected by enforcing the exit thresholds. I define input markets at the province, prefecture and county level in panels (a), (b) and (c). The number of observations is lower for fixed costs than for the capital stock as total inferred fixed costs are negative for some markets, which are omitted due to taking logs.

Appendices

A Data appendix

A.1 Production and cost data

As the dataset was already cleaned in Brandt et al. (2012), who checked consistency of firm identifiers, for instance, not much more cleaning is required. I do remove outliers in cigarette and leaf prices by censoring at the 1st and 99th percentiles. I also delete observations with negative intermediate input expenditure. As a result, I retain 2,025 observations, covering 470 firms. When defining input markets and treatment effects and when estimating markdowns, I include all firms in this industry, not just cigarette producers, in order to cover all tobacco leaf buyers. In the remainder of the analysis, I restrict the analysis to cigarette manufacturers only, which account for 95% of industry revenue, and also only focus on firms for which quantities are observed. This reduces the dataset to 1,091 observations.

A.2 Quantity data

Quantities are given at the product-firm-month level during these years by the NBS, and as the firm identifiers are the same as in the ASIF dataset, both can be merged. I only keep product codes that are measured in numbers and aggregate to the yearly level. As firms usually produce just one product, cigarettes, firm-level prices can be inferred simply by dividing firm revenue by the total number of units produced per year. Quantities are observed for just about half of the firms, which reduces the sample size to 1260. Units are defined as cigarette cases, each of which contain 50,000 cigarettes (Fang et al., 2017). From 2004 onwards, the case definition seems to change. Fortunately, I observe both current and lagged quantities in each month, and by comparing both I am able to scale the post-2004 quantities in order to make them consistent with the pre-2004 observations. As the treatment variables are defined based on quantity units in 2002, this does not change cross-sectional variation in monetary variables. More details about this dataset are in Lu and Yu (2015).

I deflate all monetary variables using the relevant deflators but as I study just one industry this only affects the time-series, and not the cross-sectional, variation in the observed variables.

A.3 County data

I retrieve county-level data from the 2000 population census through the *Harvard Dataverse*. The population census contains many variables, of which I use the total county population, the unemployed population, and the number of immigrants per county.

A.4 Product characteristics

I obtain brand-level cigarette characteristics from O'Connor et al. (2010) for a subset of firms in 2009, such as the leaf content per cigarette and other characteristics which affect the smoking experience (Talhout, Richter, Stepanov, Watson, & Watson, 2018). This dataset is observed only for 13% of the observations, but covers 29% of observations by revenue. Again, I only use this data in an extension. I link the brands in O'Connor et al. (2010) to the firms in the dataset. As I do not observe a decomposition of firm sales into brands, I have to aggregate from the brand to the firm-level, and therefore I simply calculate average product characteristics across brands.

A.5 Sample size

After cleaning the data, such as dropping observations with negative input expenditures, I retain 2,025 observations in total, 1,091 of which contain observed quantities. The main regressions which require quantity data will hence have a sample size of 1,091.

B Derivations

B.1 Markups with endogenous input prices and non-substitutable inputs

I start by deriving the markup expression in equation (6a). I start with the cost minimization problem in equation (5). Firms choose input prices in order to minimize per-period costs. The shadow price λ_{ft} is the marginal cost of increasing the input price by one unit. The markup μ is, as always, defined as the ratio of the product price over this marginal cost:

$$\mu_{ft} \equiv \frac{P_{ft}}{\lambda_{ft}}$$

Solving the first order cost minimization condition gives the following expression for λ :

$$\lambda_{ft} = W_{ft}^L \frac{\partial L_{ft}}{\partial Q_{ft}} + \frac{W_{ft}^M M_{ft}}{Q_{ft}} \psi_{ft}^M$$

Substituting the revenue shares $\alpha_{ft}^V \equiv \frac{V_{ft} W_{ft}^V}{P_{ft} Q_{ft}}$ for $V \in \{L, M\}$ and $\beta_{ft}^L \equiv \frac{\partial Q_{ft}}{\partial L_{ft}} \frac{L_{ft}}{Q_{ft}}$ gives:

$$\lambda_{ft} = \frac{\alpha_{ft}^L P_{ft}}{\beta_{ft}^L} + \alpha_{ft}^M P_{ft} \psi_{ft}^M$$

Finally, dividing prices by marginal costs yields equation (6a).

B.2 Markdown interpretation of input supply elasticity

Next, I derive the markdown formula in equation (7). In contrast to the previous paragraph, I now assume that firms choose the leaf price to maximize profits rather than minimize costs (this will, of course, give the same result). The first order condition gives:

$$\frac{\partial P_{ft} Q_{ft}}{\partial M_{ft}} \frac{\partial M_{ft}}{\partial W_{ft}^M} - M_{ft} - W_{ft}^M \frac{\partial M_{ft}}{\partial W_{ft}^M} - W_{ft}^L \frac{\partial L_{ft}}{\partial Q_{ft}} \frac{\partial Q_{ft}}{\partial M_{ft}} \frac{\partial M_{ft}}{\partial W_{ft}^M} = 0$$

Dividing by $\frac{\partial M_{ft}}{\partial W_{ft}^M}$ and substituting the expression for ψ^M leads to equation (7).

B.3 Demand for substitutable inputs with endogenous prices

Firms choose the optimal leaf price W_{ft}^{M*} that minimizes their costs, as shown in equation (5). This leaf price corresponds to an optimal output level Q_{ft}^* , which is a function of the the markdown ψ^M , all input prices, Hicks-neutral productivity, and the leaf requirement β^M .

Conditional on this optimal output level, firms choose the mix between labor and capital. Denoting the marginal cost of labor and capital as λ_{ft}^H , and the interest rate as W_{ft}^K , the choices for labor and capital are given by the following cost minimization problem:

$$\min_{L_{ft}, K_{ft}} W_{ft}^L L_{ft} + W_{ft}^K K_{ft} - \lambda_{ft}^H \left(H(L_{ft}, K_{ft}) - Q_{ft}^* \right)$$

Assuming a Cobb-Douglas function $H(L, K)$, solving the first order conditions results in labor demand $l(k_{ft}, w_{ft}^L, w_{ft}^M, \beta^L, \beta^K, W_{ft}^K, \omega_{ft}, \mu_{ft}, \psi_{ft}^M)$.

The optimal amount of labor used depends on the leaf price markdown. The reason for this is that

if leaf prices react more to the leaf quantity used, firms choose a lower equilibrium output level, and hence also use less labor.

B.4 Input market equilibrium in oligopsony

The conditions for the leaf market to be in equilibrium are needed to solve equilibrium leaf prices in each iteration of the Monte Carlo simulation. Let product prices P be normalized to one and let firms f simultaneously choose input prices W_{ft}^M every period by solving the same static cost minimization problem as before:

$$\begin{aligned} W_{ft}^M &= \arg \min \left(W_{ft}^L L_{ft} + W_{ft}^M M_{ft} - Q(\cdot) \right) \\ \Leftrightarrow W_{ft}^L \frac{\partial L_{ft}}{\partial M_{ft}} \frac{\partial M_{ft}}{\partial W_{ft}^M} + W_{ft}^M \frac{\partial M_{ft}}{\partial W_{ft}^M} + M_{ft} - \frac{\partial Q_{ft}}{\partial L_{ft}} \frac{\partial L_{ft}}{\partial M_{ft}} \frac{\partial M_{ft}}{\partial W_{ft}^M} &= 0 \\ \Leftrightarrow M_{ft} + \frac{\partial M_{ft}}{\partial W_{ft}^M} \left(W_{ft}^M + \beta_{ft}^M \frac{W_{ft}^L}{\frac{\partial Q_{ft}}{\partial L_{ft}}} - \beta_{ft}^M \right) &= 0 \end{aligned}$$

Dividing by market size \bar{M} and defining $S_{ft} \equiv \frac{M_{ft}}{\bar{M}}$, this gives the following equilibrium condition, which is very similar to those in oligopoly demand models:

$$S_{ft} + \frac{\partial S_{ft}}{\partial W_{ft}^M} \left(W_{ft}^M + \beta_{ft}^M \frac{W_{ft}^L}{\frac{\partial Q_{ft}}{\partial L_{ft}}} - \beta_{ft}^M \right) = 0 \quad (15)$$

B.5 Different objective functions

Assumption (1) stated that firms decide on their input prices to minimize per-period variable costs. SOEs could, however, have different objectives, such as achieving ‘social stability’ through high and countercyclical employment, as argued by Li et al. (2012). There are two ways in which such size objectives can enter the firm’s minimization problem.

Output size objective

Suppose first that firms value being large and are willing to sacrifice some profits to achieve this. Such an objective enters the shadow price λ_{ft} . Let this shadow price in equation (5) be denoted $\hat{\lambda}_{ft} = \frac{\lambda_{ft}}{\tau_{ft}}$. Firms with a larger preference for producing a lot have a larger parameter τ_{ft} . The markup μ_{ft} is now given by:

$$\mu_{ft} = \tau_{ft} \frac{P_{ft}}{\lambda_{ft}}$$

If firms value being large rather than profitable, the true markup will hence be larger than the estimated markup. The reason for this is that the cost minimization model infers large input usage as an indication of low markups, while in reality, this is due to a preference towards large size.

The same holds for the markdown ψ_{ft}^M . If firms value a large size, they will set higher input prices. Through the markdown expression, the model interprets this as evidence for low monopsony power. In reality, though, this merely reflects size preferences.

$$\psi_{ft}^M = \gamma^W W_{ft}^M (1 - S_{ft})^{-1} + 1$$

Input size objective

Now suppose that firms specifically want to employ a lot of manufacturing workers, but do not have such preferences for farming employment (or the other way around). In this case, the true input price \hat{W}_{ft}^L is different from the measured input price. If firms value employing many workers, the implicit wage is lower than the observed wage, so $\tau_{ft}^L < 1$:

$$\hat{W}_{ft}^L = W_{ft}^L \tau_{ft}^L$$

As firms do not choose labor and tobacco leaf separately, this has the same effects on markup and markdown estimates as a different shadow price λ . The derivation in appendix B shows that marginal costs are linear in both input prices.

Remedies

Throughout the text, I always allow for different objective functions across ownership types, because I systematically add ownership dummies to all regressions. These never change the results much, as there is not much variation in ownership: most firms are SOEs anyway. I do not allow for different objective functions within ownership classes, as all SOEs are, for instance, assumed to have the same objective function.

C Robustness checks

C.1 Difference-in-differences robustness checks

Continuous treatment measures

Throughout the main text, a dummy variable was used to indicate the treatment groups, namely the presence of firms below the exit threshold in a market before 2003. I re-estimate the treatment model using different treatment measures. First, I use the share of firms in a market producing below the threshold in 2002 as a treatment measure. Second, I weight these firms by employment, rather than taking unweighted averages. Besides the treatment indicator definition, I keep all model specifications fixed. The results are in panel (a) of table A8. The interpretation of the estimates is as follows: in provinces in which the share of firms below the threshold was 10 percentage points higher, markdowns were 8.3% higher. The results are consistent with the baseline regression: increasing markdowns in consolidated markets, decreased markups and no strong evidence for productivity gains.

Firm and year fixed effects

In the baseline model, firm fixed effects were included as controls, while year fixed effects were not. A linear time trend was used instead. In panel (b) of table A8, I use two different specifications. First, I no longer control for firm dummies. The coefficients are similar to the model with firm fixed effects, except that the drop in markups is less pronounced. Next, I include both firm and year fixed effects. This does not change the estimates much.

Different moments

The baseline difference-in-differences model was estimated using regular unweighted OLS. In the first part of panel (c) of table A8, I use a quantile regression instead. I report the treatment effects on the median of each outcome variable. The estimates are now smaller compared to when averages are used, and are less dependent on the market definitions used. The broad trend of rising markdowns and falling markups is still present in all specifications, though. Finally, I weight observations by their employment size. The estimated coefficients are very similar to the baseline regression.

Self-selection above exit threshold

The exit thresholds of 100,000 cases per year were based on production from 2002. If firms knew these exit thresholds beforehand, there should be ‘bunching’ of firms just above this exit threshold. Figure A1 plots the distribution of the number of cases between 2000 and 2004. Up to 2003, most firms are around the exit threshold of 100,000 cases per year, and there is no bunching just above this threshold.

Dropping firms under the 100K threshold

In table A11, I re-estimate the difference-in-differences regression when dropping the firms under the production threshold which survived after 2002. The results are similar to the main specifications.

Narrowing the treatment time window

As was shown in figure 3, the number of firms producing less than 100,000 cases a year fell relatively faster mainly between 2002 and 2004. After 2004, there was also significant exit among firms producing more than 100,000 cases a year.

I narrow down the timeframe to the periods 1999-2005 and 1999-2004, which means that the treatment effects are estimated over just two and one years respectively, in table A10. The effects become a bit smaller when narrowing the time frame, but are still in the same order of magnitude as the original results, and significant.

Announcement effects

The industry consolidation and specific size thresholds were announced by the STMA on May 2, 2002. It may be, however, that firms were anticipating the reform and decreased their leaf prices already before the policy was implemented. In table A9, I test for such announcement effects. I re-estimate the difference-in-differences model from equation (1) with markdowns, markups and productivity as the outcome variables. In the main specification, the start of the policy was defined to be 2003. I also re-estimate the difference-in-differences model using 2001 and 2002 as treatment starting years. I use province-, prefecture- and county-level market definitions in panels (a), (b) and (c).

At both the province- and county-level, markdowns decreased significantly already in 2002, which is consistent with announcement effects as the policy was announced in that year. Markups seem to have decreased, however, already before 2002. These markups reflect pricing behavior by the government-owned wholesaler. As the STMA, which ordered the consolidation, and the CNTC,

which operates the wholesaling monopoly, are heavily intertwined, it is possible that they adjusted their prices in expectation to the policy change. The modest rise in productivity after 2003 in the treatment group was, finally, not present before the treatment started.

C.2 Translog production function

Throughout the main text, I used a Cobb-Douglas specification for the labor-capital term $H(\cdot)$ in the production function. As the elasticity of substitution estimate between labor and capital was not significantly different from one, this seems to be the correct production function. Nevertheless, I also use a translog specification for $H(\cdot)$ as a robustness check. The corresponding functional form of $h(\cdot)$ in logarithms is given by:

$$h(L_{ft}, K_{ft}) = \beta^L l_{ft} + \beta^K k_{ft} + \beta^{LK} l_{ft} k_{ft} + \beta^{2L} l_{ft}^2 + 2\beta^{2K} k_{ft}^2$$

The moment conditions to estimate this translog production are given by:

$$\mathbb{E} \left\{ \xi_{ft}(\beta^L, \beta^K, \beta^{LK}, \beta^{L^2}, \beta^{K^2}) \begin{pmatrix} l_{ft-1} \\ k_{ft} \\ l_{ft-1} k_{ft} \\ l_{ft-1}^2 \\ k_{ft}^2 \end{pmatrix} \right\} = 0$$

Table A3 compares the estimates for the main coefficients of interest and the markdown and markup averages between the Cobb-Douglas and Translog model. The estimates are very similar and never significantly differ from each other.

C.3 Using cost shares as output elasticities

Suppose there are constant returns to scale, exogenous input prices for labor and capital, and a Cobb-Douglas function for $H(\cdot)$ in the production function, equation (3). The output elasticities of labor and capital are then equal to their cost shares. Denoting capital investment as I_{ft} , this gives:

$$\begin{cases} \beta_{ft}^L = \frac{W_{ft}^L L_{ft}}{W_{ft}^L L_{ft} + I_{ft}} \\ \beta_{ft}^K = \frac{I_{ft}}{W_{ft}^L L_{ft} + I_{ft}} \end{cases}$$

The output elasticities, markup and markdown estimates and consolidation treatment effects using this approach are in table A4. They are very similar to the baseline estimates.

C.4 Alternative labor supply specifications

Log-logs and inverse supply

The baseline supply model was a log on levels labor supply model, in equation (8). Alternatively, I estimate the following log-on-logs model:

$$s_{ft} - s_{0t} = \hat{\gamma}^W \log(W_{ft}^M) + \hat{\gamma}^Z \mathbf{Z}_{ft} + \hat{\zeta}_{ft}$$

I can also estimate an inverse supply model, as in Goolsbee and Syverson (2019):

$$\log(W_{ft}^M) = \tilde{\gamma}^W \log(s_{ft} - s_{0t}) + \tilde{\gamma}^Z \mathbf{Z}_{ft} + \tilde{\zeta}_{ft}$$

Nested logit

It is possible to allow for more flexible substitution patterns in supplier choices by using a nested logit model. Each farmer j now chooses a manufacturer f within nest g in market i . The set of manufacturers in nest g in market i at time t is denoted \mathcal{F}_{it}^g . The error structure in the utility function now differs, with utility being parametrized as:

$$U_{jft} = \underbrace{\gamma^W W_{ft}^M + \gamma^Z \mathbf{Z}_{ft} + \zeta_{jt}}_{\delta_{ft}} + (1 - \sigma) \nu_{jft}$$

The preference shocks ν_{jft} still follow a type-I extreme value distribution. Following S. T. Berry (1994), the input market share is given by:

$$S_{ft} = \frac{\exp(\frac{\delta_{ft}}{1-\sigma})}{D_{gt}^\sigma [\sum_g D_{gt}^{1-\sigma}]}$$

with $D_{gt} \equiv \sum_{f \in \mathcal{F}_{it}^g} \exp(\frac{\delta_{ft}}{1-\sigma})$ The markdown is now expressed as:

$$\psi_{ft}^M \equiv \left(\frac{\partial S_{ft}}{\partial W_{ft}^M} \frac{W_{ft}^M}{S_{ft}} \right)^{-1} + 1 = \left(\gamma^W W_{ft}^M \left(\frac{1}{1-\sigma} - \frac{\sigma}{1-\sigma} S_{fgt} - S_{ft} \right) \right)^{-1} + 1$$

I use two types of nests. In the first specification, I define three nests, according to the manufacturer's ownership types. I divide manufacturers into (i) SOEs, (ii) private firms and collectives, and (iii) foreign firms. Secondly, I define prefectural markets as nests of province-level markets.

Results

The estimates for these alternative labor supply specifications are in table A6. In panel (a), I report the log-on-logs estimates. These are consistent with the baseline specification: the leaf supply curve is upward-sloping, the labor supply curve flat. Panel (b) reports the inverse log-on-log specification. If leaf demand increases by 10%, leaf prices increase by 1.5%. For labor, the coefficient is negative but insignificant. Panels (c) and (d) contain the estimates for the nested logit models. The nesting elasticity between different prefectural sub-markets is 0.621. When defining the nests as ownership types, the elasticity is 0.545. Both elasticity estimates are significantly below one and above zero, meaning that the nests are neither perfect complements nor substitutes.

The average and median markup and markdown estimates from both nested logit models are in table A7. Markdowns are on average higher in the nested logit models: 2.85 with the ownership nests and 3.05 with the prefectural nests, compared to 2.63 in the standard logit model. Markups are - mechanically - lower, at 1.05 and 0.96. The treatment effects of the consolidation are very similar to those in the baseline model. Finally, I compare the markup and markdown distributions from both nested logit models with the baseline logit model in figure A2. The markdown distribution becomes flatter when allowing for more flexible substitution patterns, which is logical: more variation in markdowns is allowed for, both across firms and over time.

C.5 Export participation

The Chinese economy globalized rapidly throughout the 1990s and 2000s, and the accession to the WTO in 2003 affected total factor productivity and markup growth (Brandt, Van Biesebroeck, Wang, & Zhang, 2017; Li et al., 2012). As was already argued, the tobacco industry remained largely domestic: less than 1% of total industry revenue came from exports. For the 16% of firms who do export, exports represent merely 6% of their revenues on average. In table A20, I test whether exporting behavior changed through the consolidation, and conclude that it did not.

D Monte Carlo simulations

D.1 Motivation

In this section, I simulate a production and cost dataset with oligopsonistic input markets. This serves two objectives. First, I compare consistency and efficiency of existing production function estimation approaches with the IV approach proposed in this paper, under different models of input market

competition. Second, I use this data to show that the input supply curve can be consistently estimated using productivity shocks of buyers as instruments. The estimation strategy in the paper consists of three broad steps:

- (i) Estimate the production function, equation (10a) \rightarrow retrieve $(\hat{\omega}_{ft}, \hat{\beta}_{ft})$
- (ii) Use productivity residuals ω_{ft} to estimate markdowns by using re the input supply function and markdowns \rightarrow retrieve $(\hat{\psi}_{ft}^M)$
- (iii) Retrieve markups using $\hat{\beta}_{ft}$ from (i) and $\hat{\psi}_{ft}^M$ from (ii).

D.2 Data generating process

I create a panel of 200 local intermediate input markets with exactly three firms in each market. All firms are observed during 3 periods and there is no entry or exit. As a result, the dataset contains 1800 observations. I impose a standard logit model for intermediate input supply, as in the model section. I simulate the dataset using 50 iterations.

Simulation procedure

In each iteration loop, the simulation procedure is as follows:

1. Draw $\omega_{f1}, \xi_{f1}, \zeta_{f1}, K_{f1}$
2. Find equilibrium wages W_{f1}^M using conditions in equation (15)
3. Calculate optimal investment I_{f1}
4. Use transition equations to get $\omega_{f2}, x_{f2}, K_{f2}$
5. Repeat steps (2)-(4) until final period

I start by drawing initial productivity ω , productivity shocks ξ , firm characteristics ζ (which enter supplier utility) and capital K from their respective distributions, all in the first year. Next, I search for equilibrium input prices and intermediate input quantities for all firms in order to have an equilibrium in all input markets. Based on these equilibrium input prices, firms decide on investment, and optimal investment is calculated. In a fourth step, the capital stock is updated for the next year, using the capital accumulation equation and productivity is updated in the next period using the equation of motion for productivity. In the next period, the draws and equilibrium calculation are repeated. Due to the upward-sloping intermediate supply curve, analytical expressions for labor usage no longer

exist (in contrast with, for instance, (Van Biesebroeck, 2007)). Equilibrium wages and employment levels are jointly obtained by solving the equilibrium conditions in equation (15) for each market in each year.

Parametrization

The parametrization of all coefficients and distributions are in table A15. I let the output elasticity of labor and capital be 0.4 and 0.6. I use a lognormal distribution for TFP, with a serial correlation of 0.7 and a cross-sectional standard deviation of 0.5. Input market sizes are distributed uniformly between 0.5 and 1.5, meaning that the largest markets are three times larger than the smallest ones. Working conditions are drawn from a uniform distribution between 0 and 3. Valuation for wages and working conditions is 2.5 and 0.5 respectively, and in the second DGP there is a random coefficient on wages which varies 15% compared to the average wage valuation. Investment costs are normalized to one and the annual depreciation rate is 10%.

[Table A15 here]

D.3 Results

The estimated output elasticities using the simulated dataset are in panel (a) of table A16. The OLS estimates are, as usual, biased due to simultaneity. When input prices and market shares do not feature in the first stage regression of ACF, the ACF estimates are as biased as OLS, which confirms the non-identification result in the theoretical model. When properly controlling for both market shares and input prices, in column 3, ACF delivers consistent estimates.

[Table A16 here]

Secondly, I use the productivity residuals as instrument for input prices in the supply estimation. In the simulated model, the wage valuation coefficient is 2.5. Panel (c) in table A16 compares the OLS and IV estimates of equation (8). The results are very similar to the real estimates in the paper: The OLS estimate is negative, while the IV estimate is positive and close to the truth, at 2.25. The true value of 2.50 lies within the 95% confidence interval of the estimates. The first stage F-statistic is around 100, meaning that TFP is a strong instrument for endogenous input prices.

Random coefficients

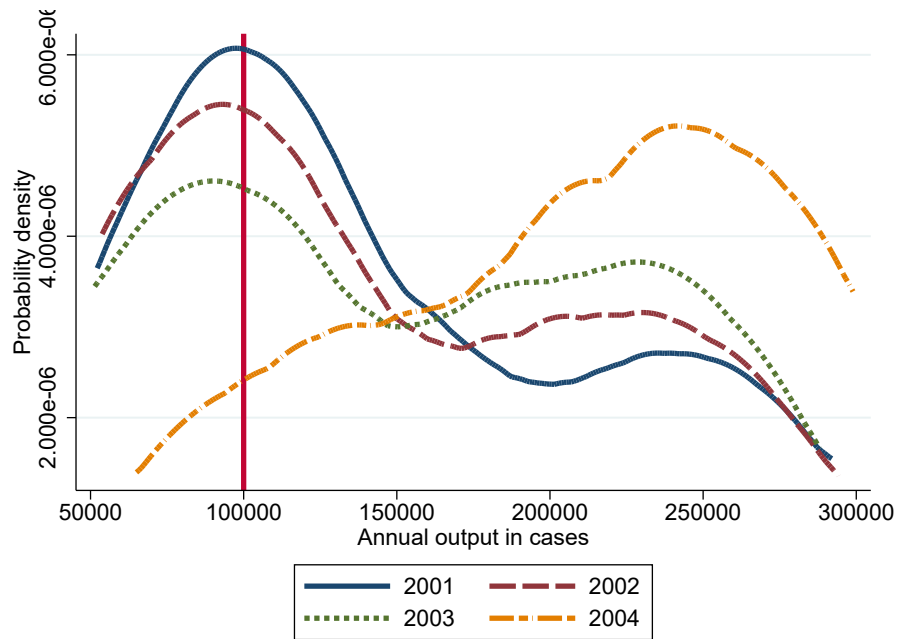
In the model, it was assumed that input supply follows a standard logit model. This implies that input suppliers all value input prices in the same way. It could be, however, that there is unobserved heterogeneity in supplier preferences for input prices and other firm characteristics. In the case of labor, for instance, it seems likely that some workers prefer higher wages while others value working conditions or job security more. While estimating such a random coefficients input supply function is possible using mixed logit models, such as S. Berry et al. (1995), it poses problems for identifying the production function. Let the input supply elasticities be distributed across suppliers, and hence across firms with mean $\bar{\gamma}^W$ and σ^γ :

$$\gamma_{ft}^W \sim \Gamma(\bar{\gamma}^W, \sigma^\gamma)$$

In this case, markdown variation across firms is not captured by observed input prices and market shares alone. Including these in the input demand function hence does not suffice in solving the problem of serially correlated unobservables which prevent inverting unobserved scalar productivity. Panel (b) in table A16, in the appendix, shows the simulated results for both production function estimators when input supply follows a random coefficients logit model. The estimates when using Akerberg et al. (2015) are almost as biased as the OLS results, even when including market shares and wages in the first stage regression.

Solving this problem is beyond the scope of this paper: for agricultural spot markets, it seems reasonable to assume that farmers all value input prices in the same way. For labor markets, this seems more problematic. The solutions to this challenge are reserved for further work.

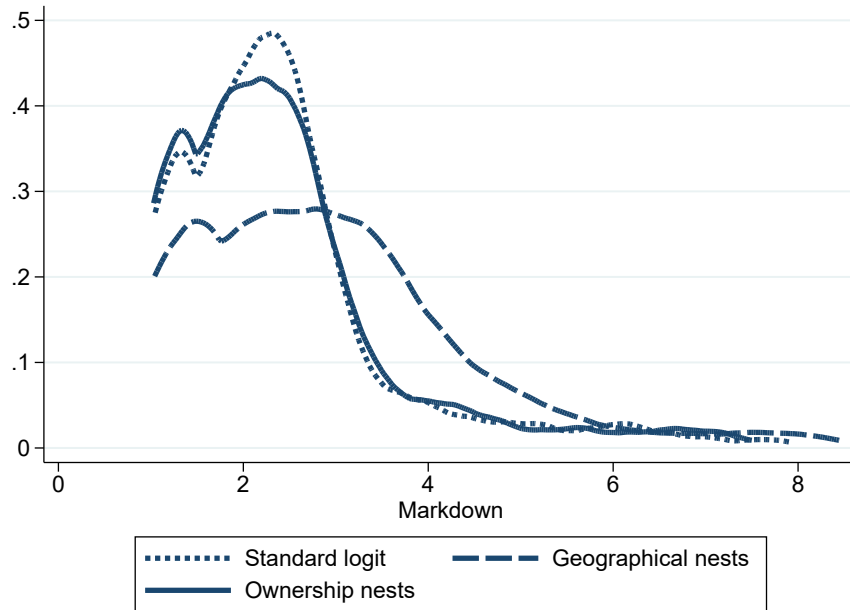
Figure A1: Quantity distribution and bunching



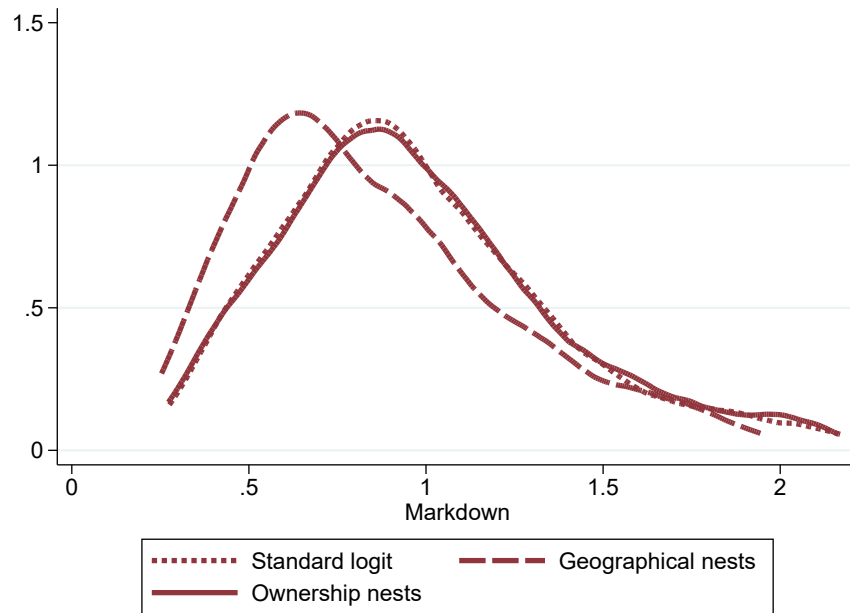
Notes: This graph plots the distribution of the number of cigarette cases in 2001, 2002 and 2003. There is no evidence for 'bunching' just above the exit threshold of 100,000 cases per year.

Figure A2: Markups and markdowns: nested logit

(a) Markdowns



(b) Markups



Notes: Panel (a) plots markdowns using the baseline logit model and both nested logit models, with nests being defined as ownership types and as prefectural nests. Panel (b) does the same for markups. All distributions are censored at 5th and 95th percentiles.

Table A1: Treatment and control groups before policy

Panel (a) Treatment group size	% Firms		% Revenue		% Output	
Firms producing less than 100K cases	50.4		8.1		4.3	
Treatment (province)	86.1		89.2		87.7	
Treatment (prefecture)	38.8		38.3		40.4	
Treatment (county)	14.9		11.7		15.9	
Panel (b) Observable characteristics	Cigarette price*		Leaf price**		Quantity***	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Control group (province)	1,384	(613)	546	(177)	568	(433)
Treatment group (province)	1,412	(688)	602	(249)	491	(279)
Control group (prefecture)	1,448	(733)	626	(251)	575	(232)
Treatment group (prefecture)	1,346	(5 67)	559	(120)	626	(251)
Control group (county)	1,447	(711)	705	(243)	601	(255)
Treatment group (county)	1,186	(428)	535	(152)	559	(120)

Notes: *Price per case of 50,000 cigarettes in RMB, **Leaf price per case, ***Annual output in thousands of cases. All comparisons are on the period 1999-2002.

Table A2: Treatment effects and HHIs

	log(Materials HHI)		log(Labor HHI)	
	Estimate:	SE:	Estimate:	SE:
<i>Panel (a): Province level</i>				
Treatment*1(year \geq 2003)	0.269	(0.081)	0.293	(0.086)
R-squared	0.332		0.400	
<i>Panel (b): Prefecture level</i>				
Treatment*1(year \geq 2003)	0.181	(0.044)	0.186	(0.044)
R-squared	0.172		0.156	
<i>Panel (c): County level</i>				
Treatment*1(year \geq 2003)	0.122	(0.031)	0.094	(0.037)
R-squared	0.094		0.047	

Notes: In this table, I regress Hirschman-Herfindahl (HHI) indices at different geographical levels on the consolidation treatment variable interacted with the post-treatment dummy. The panel is defined at the market-year level. I control for market fixed effects and a linear time trend.

Table A3: Translog model in labor and capital

<i>Panel (a): Output elasticities</i>		Labor		Capital			
		Estimate:	SE:	Estimate:	SE:		
		0.353	(0.226)	0.690	(0.140)		
<i>Panel (b): Markdowns and markup</i>		Markdown		Markups			
		Estimate:	SE:	Estimate:	SE:		
Mean		2.428	(1.28)	1.143	(0.230)		
<i>Panel (c): Consolidation effects</i>		log(Markdown)		log(Markup)		log(TFP)	
		Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
Treatment *1(year \geq 2003)		0.291	(0.061)	-0.247	(0.061)	0.078	(0.083)
R-squared		0.799		0.751		0.866	
Observations		1,091		1,091		1,091	

Notes: Panel (a) reports the output elasticities which are obtained from estimating the production function. Panels (b) and (c) report the estimated price semi-elasticities of input supply, both for tobacco leaf and labor. The left-hand side variable is the log province-level market share minus the log outside option market share. The right-hand side variable is the input price in levels. Manufacturing TFP from the Leontief model is used as instrument. Bootstrapped standard errors are in the parentheses.

Table A4: Using cost shares as output elasticities

<i>Panel (a): Output elasticities</i>		Labor		Capital			
		Estimate:	SE:	Estimate:	SE:		
		0.427	()	0.573	()		
<i>Panel (b): Markdowns and markup</i>		Markdown		Markups			
		Estimate:	SE:	Estimate:	SE:		
Mean		2.196	()	1.25	()		
<i>Panel (c): Consolidation effects</i>		log(Markdown)		log(Markup)		log(TFP)	
		Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
Treatment *1(year \geq 2003)		0.268	(0.057)	-0.222	(0.073)	0.081	(0.081)
R-squared		0.802		0.774		0.883	
Observations		1,091		1,091		1,091	

Notes: This table reports the output elasticities, markdowns, markups and consolidation effects when assuming constant returns to scale and when using the cost shares of labor and capital as output elasticities. Panel (a) reports the output elasticities, panel (b) the markdown and markup levels, and panel (c) the estimated consolidation treatment effects at the province level.

Table A5: Overidentifying restrictions

	log(Leaf market share)		log(Leaf market share)	
	Estimate:	SE:	Estimate:	SE:
	3.668	(0.862)	3.659	(1.793)
Instruments:				
Log(TFP)		Yes		Yes
Consolidation dummies		No		Yes
Sargan-Hansen test:				
χ^2			4.332	
p-value			0.115	
R-squared	0.053		0.147	
Observations	1,091		1,091	

Notes: I estimate the semi-elasticity of leaf supply using either only manufacturing TFP as instrument for prices, or both the consolidation dummies and TFP as instruments. I report the Sargan-Hansen test statistic for overidentifying restrictions.

Table A6: Alternative input supply models

<i>Panel (a): Log-on-logs</i>				
	log(Leaf market share)		log(Labor market share)	
	Estimate:	SE:	Estimate:	SE:
log(Input price)	1.657	(0.207)	0.032	(0.187)
R-squared	0.723		0.838	
1st stage F-statistic	176		64	
Observations	1,091		1,091	
<i>Panel (b): Prefectural nests</i>				
	log(Leaf market share)		log(Labor market share)	
	Estimate:	SE:	Estimate:	SE:
Input price	1.836	(0.198)	0.025	(0.024)
Nesting elasticity	0.621	(0.060)	0.849	(0.201)
R-squared	0.780		0.865	
Observations	1,091		1,091	
<i>Panel (c): Ownership type nests nests</i>				
	log(Leaf market share)		log(Labor market share)	
	Estimate:	SE:	Estimate:	SE:
Input price	3.320	(1.590)	-0.013	(0.117)
Nesting elasticity	0.545	(0.032)	1.253	(0.201)
R-squared	0.382		0.688	
Observations	1,091		1,091	

Notes: In panel (a), I re-estimate the input supply models, but with log input prices on the right-hand side, rather than input prices in levels. In panel (b), I define provinces as markets and prefectures as sub-markets among which farmers can choose to be active in. In panel (c), I estimate the nested logit models with nests being defined as three ownership types: SOEs, private firms and collectives, and foreign firms.

Table A7: Markups and markdowns in nested logit models

<i>Panel (a): Prefectural nests</i>				
	Markdown		Markup	
	Estimate:	SE:	Estimate:	SE:
Mean	3.049	(0.103)	0.960	(0.668)
	log(Markdown)		log(Markup)	
	Estimate:	SE:	Estimate:	SE:
Treatment * 1(year \geq 2003)	0.342	(0.060)	-0.285	(0.055)
<i>Panel (b): Ownership nests</i>				
	Markdown		Markup	
	Estimate:	SE:	Estimate:	SE:
Mean	2.852	(0.164)	1.052	(0.429)
	log(Markdown)		log(Markup)	
	Estimate:	SE:	Estimate:	SE:
Treatment * 1(year \geq 2003)	0.370	(0.059)	-0.309	(0.057)

Notes: Panel (a) reports average markups, markdowns and treatment effects when using the geographically nested logit leaf supply model. Panel (b) does the same with the ownership nested logit model.

Table A8: Alternative treatment measures

<i>Panel (a): Other treatment measures</i>	log(Markdown)		log(Markup)		log(TFP)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
i. Share firms under 100K						
Province level	0.795	(0.144)	-0.349	(0.154)	0.353	(0.203)
Prefecture level	0.764	(0.140)	- 0.411	(0.129)	0.154	(0.141)
County level	1.120	(0.184)	-0.579	(0.145)	0.069	(0.124)
ii. Employment share under 100K						
Province level	0.270	(0.261)	-0.353	(0.272)	0.482	(0.291)
Prefecture level	0.719	(0.215)	-0.465	(0.163)	-0.156	(0.205)
County level	1.035	(0.192)	-0.559	(0.154)	0.002	(0.142)
<i>Panel (b): Different fixed effects</i>	log(Markdown)		log(Markup)		log(TFP)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
ii. No firm fixed effects						
Province level	0.403	(0.141)	-0.267	(0.147)	0.059	(0.102)
Prefecture level	0.261	(0.114)	-0.038	(0.106)	0.123	(0.112)
County level	0.729	(0.205)	-0.332	(0.167)	0.210	(0.167)
iv. Adding year fixed effects						
Province level	0.270	(0.059)	-0.213	(0.056)	-0.020	(0.068)
Prefecture level	0.379	(0.070)	-0.199	(0.066)	0.056	(0.067)
County level	0.802	(0.144)	-0.380	(0.113)	0.037	(0.102)
<i>Panel (c): Different moments</i>	log(Markdown)		log(Markup)		log(TFP)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
v. Median						
Province level	0.149	(0.022)	-0.140	(0.051)	-0.033	(0.052)
Prefecture level	0.333	(0.081)	-0.114	(0.065)	0.120	(0.063)
County level	0.369	(0.078)	-0.283	(0.142)	0.109	(0.099)
vi. Employment-weighted average						
Province level	0.296	(0.064)	-0.240	(0.068)	-0.057	(0.106)
Prefecture level	0.226	(0.086)	-0.073	(0.083)	0.111	(0.075)
County level	0.635	(0.131)	-0.406	(0.121)	0.169	(0.101)

Notes:

Table A9: Announcement effects

<i>Panel (a): Province level</i>	log(Markdown)		log(Markup)		log(TFP)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
Treatment * 1(year \geq 2001)	0.011	(0.083)	-0.161	(0.064)	-0.081	(0.065)
Treatment * 1(year \geq 2002)	0.164	(0.046)	-0.008	(0.055)	-0.065	(0.089)
Treatment * 1(year \geq 2003)	0.144	(0.057)	-0.128	(0.060)	0.067	(0.106)
<i>Panel (b): Prefecture level</i>	log(Markdown)		log(Markup)		log(TFP)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
Treatment * 1(year \geq 2001)	0.143	(0.081)	-0.214	(0.075)	-0.096	(0.057)
Treatment * 1(year \geq 2002)	0.123	(0.075)	0.031	(0.070)	-0.058	(0.075)
Treatment * 1(year \geq 2003)	0.224	(0.082)	-0.124	(0.083)	0.141	(0.089)
<i>Panel (c): County level</i>	log(Markdown)		log(Markup)		log(TFP)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
Treatment * 1(year \geq 2001)	0.057	(0.085)	-0.129	(0.082)	-0.125	(0.060)
Treatment * 1(year \geq 2002)	0.268	(0.112)	-0.150	(0.100)	-0.029	(0.097)
Treatment * 1(year \geq 2003)	0.529	(0.168)	-0.185	(0.137)	0.111	(0.120)

Notes: I re-estimate the difference-in-differences model from equation (1). In contrast with the main specification, I define the start of the treatment effect to take place in 2000, 2001 and 2002. Province-, prefecture- and county-level market definitions are used in panels (a), (b) and (c). The specific size thresholds were announced in 2002.

Table A10: Narrowing the time frame

<i>Panel (a): 1999-2005</i>	log(Markdown)		log(Markup)		log(TFP)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
Treatment * 1(year \geq 2003)	0.271	(0.071)	-0.239	(0.068)	0.094	(0.093)
1(year \geq 2003)	-0.274	(0.074)	0.252	(0.073)	-0.166	(0.100)
Observations	984		984		984	
R-squared	0.806		0.787		0.869	

<i>Panel (b): 1999-2004</i>	log(Markdown)		log(Markup)		log(TFP)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
Treatment * 1(year \geq 2003)	0.199	(0.084)	-0.160	(0.081)	0.063	(0.116)
1(year \geq 2003)	-0.214	(0.083)	0.186	(0.081)	-0.150	(0.115)
Observations	924		924		924	
R-squared	0.817		0.796		0.883	

Notes: In panels (a) and (b), I re-estimate the consolidation treatment effects using shorter time frames for the post-treatment period. I reduce this treatment period from three to two and one years, respectively.

Table A11: Sample size robustness

<i>Panel (a): Dropping firms $Q < 100K$</i>	log(Markdown)		log(Markup)		log(TFP)	
Treatment * 1(year \geq 2003)	0.223	(0.065)	-0.243	(0.065)	-0.006	(0.081)
1(year \geq 2003)	-0.165	(0.073)	0.170	(0.073)	-0.051	(0.093)
Observations	626		626		626	
R-squared	0.767		0.817		0.668	
<i>Panel (b): Using ACF sample size</i>	Labor coefficient		Capital coefficient		Scale economies	
Treatment * 1(year \geq 2003)	0.365	(0.155)	0.804	(0.148)	1.168	()
Observations			825			
R-squared			0.903			

Notes: In panel (a), I drop the surviving firms under the exit threshold and re-estimate the treatment effects model on this sample. In panel (b), I re-estimate the production function using the IV approach but reduce the sample to the ACF estimation sample.

Table A12: Exit and mergers

<i>Panel (a) Province-level:</i>	log(Markdown)		log(Markup)		log(TFP)	
Exit treatment * 1(year \geq 2003)	0.320	(0.113)	-0.261	(0.103)	0.034	(0.128)
Merger treatment* 1(year \geq 2003)	-0.043	(0.117)	0.017	(0.113)	0.081	(0.141)
Observations	1,091		1,091		1,091	
R-squared	0.801		0.756		0.894	
<i>Panel (b) Prefecture-level:</i>	log(Markdown)		log(Markup)		log(TFP)	
Exit treatment * 1(year \geq 2003)	0.381	(0.103)	-0.197	(0.087)	0.133	(0.094)
Merger treatment* 1(year \geq 2003)	0.044	(0.107)	-0.077	(0.090)	-0.029	(0.094)
Observations	1,091		1,091		1,091	
R-squared	0.880		0.844		0.895	
<i>Panel (c) County-level:</i>	log(Markdown)		log(Markup)		log(TFP)	
Exit treatment * 1(year \geq 2003)	0.721	(0.144)	-0.353	(0.120)	0.100	(0.108)
Merger treatment* 1(year \geq 2003)	0.192	(0.104)	-0.151	(0.094)	-0.010	(0.079)
Observations	1,091		1,091		1,091	
R-squared	0.879		0.868		0.894	

Notes: In this table, I re-estimate the consolidation treatment model with two types of treatment effects. The ‘exit treatment’ is the same consolidation dummy as used throughout the main text. It indicates the presence of at least one firm below the exit threshold of 100,000 cases produced in 2002. The ‘merger treatment’ indicates the presence of at least one firm below the merger threshold of 300,000 cases per year. Panels (a)-(c) re-estimate the treatment effects for all three outcomes of interest at the province, prefecture and county-level respectively.

Table A13: Consolidation treatment effects, controlling for the capital-labor ratio

<i>Panel (a): Province level</i>						
	log(Markdown)		log(Markup)		log(Productivity)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	0.299	(0.063)	-0.256	(0.064)	0.076	(0.083)
log(Capital/labor)	0.007	(0.028)	0.044	(0.050)	-0.199	(0.059)
R-squared	0.808		0.762		0.870	
Observations	1,091		1,091		1,091	
<i>Panel (b): Prefecture level</i>						
	log(Markdown)		log(Markup)		log(Productivity)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	0.407	(0.071)	-0.246	(0.069)	0.101	(0.073)
log(Capital/labor)	0.081	(0.058)	-0.029	(0.056)	-0.197	(0.059)
R-squared	0.880		0.847		0.870	
Observations	1,091		1,091		1,091	
<i>Panel (c): County level</i>						
	log(Markdown)		log(Markup)		log(Productivity)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
log(Capital/labor)	0.002	(0.063)	0.036	(0.060)	-0.198	(0.059)
Treatment * 1(year \geq 2003)	0.788	(0.134)	-0.412	(0.111)	0.082	(0.102)
R-squared	0.880		0.870		0.870	
Observations	1,091		1,091		1,091	

Notes: In this table, I re-estimate the difference-in-differences model using various input market definitions, while controlling for capital intensity differences across firms.

Table A14: Comparison to prior monopsony studies

Paper	Country	Industry	Input	Input price / MRP
Goolsbee and Syverson (2019)	USA	Universities	Professors	0.83
Ransom and Sims (2010)	USA	Schools	Teachers	0.82 (M) - 0.75 (F)
Hirsch, Schank, and Schnabel (2010)	Germany	All	Employees	0.78
Oaxaca and Ransom (2010)	USA	Grocery stores	Clerks	0.74 (M) - 0.70 (F)
This paper	China	Tobacco	Farmers	0.38

Note: I report the average price of an input over its marginal revenue product, which I calculate based on the reported labor supply elasticity using equation (9). ‘M’ stands for male and ‘F’ for female workers.

Table A15: Parametrization of Monte Carlo simulations

Parameter	Description	Value
β_l	Output elasticity of labor	0.4
β_k	Output elasticity of capital	0.6
ω_{jt}	TFP	$\sim \exp(N(0, 0.5))$
ζ_{jt}	TFP shock	$\sim \exp(N(0, 0.5))$
ρ	TFP serial correlation	0.7
σ^a	TFP cross-sectional standard deviation	0.5
\bar{M}_m	Input market size	[0.5,1.5]
ξ_{jt}	Working conditions	$\sim \mathcal{U}[0, 3]$
α^ξ	Working condition valuation	0.5
α^w	Wage valuation	2.5
γ^w	Input price valuation by supplier (DGP 1)	0
$\bar{\gamma}^w$	Input-price valuation by supplier (DGP 2)	$\sim \mathcal{U}[-0.375, 0.375]$
\bar{K}_0	Initial capital stock	$\sim \exp(N(0, 0.25))$
κ	Investment cost coefficient	1
δ	Depreciation	0.1

Table A16: Monte Carlo simulations

<i>Panel (a): Output elasticities</i> – standard logit supply model	OLS		ACF		ACF bis*	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
Labor (truth = 0.4)	0.544	(0.028)	0.582	(0.070)	0.401	(0.113)
Capital (truth = 0.6)	0.555	(0.025)	0.543	(0.030)	0.600	(0.043)
<i>Panel (b): Output elasticities</i> – mixed logit supply model	OLS		ACF bis*			
	Estimate:	SE:	Estimate:	SE:		
Labor (truth = 0.4)			0.518	(0.029)	0.489	(0.127)
Capital (truth = 0.6)			0.562	(0.024)	0.564	(0.041)
<i>Panel (c): Input supply elasticity</i>	OLS		IV			
	Estimate:	SE:	Estimate:	SE:		
Labor (truth = 2.5)			-0.159	(0.083)	2.249	(0.641)
First-stage F-stat			–	–	97.949	(76.044)

Notes: *ACF with input market shares and input prices in first stage. Instrument: productivity residuals. Controls: market dummies. Simulated using 50 iterations.

Table A17: Elasticity of substitution

	log(Materials/Labor)		Log(Capital/Labor)	
	Estimate:	SE:	Estimate:	SE:
Elasticity of substitution	0.146	(0.281)	0.909	(0.222)
First-stage F-stat	64.58		64.58	
Observations	1,091		1,091	
R-squared	0.283		0.478	

Notes: Controls: ownership type, product type, year dummies, export dummy, export revenue share. IVs: avg. export share of revenue of competitors, export participation of competitors.

Table A18: Ownership consolidation and labor-augmenting productivity

Panel (a): Province-level markets		log(LAP*)	
	Estimate:	SE:	
Treatment * 1(year \geq 2003)	0.228	(0.089)	
Observations	1,091		
R-squared	0.662		
Panel (b): Prefecture-level markets		log(LAP*)	
	Estimate:	SE:	
Treatment * 1(year \geq 2003)	0.141	(0.089)	
Observations	1,091		
R-squared	0.661		
Panel (c): County-level markets		log(LAP*)	
	Estimate:	SE:	
Treatment * 1(year \geq 2003)	0.042	(0.120)	
Observations	1,091		
R-squared	0.661		

Notes: *Labor-augmenting productivity. Dependent variables in logs. Markup re-calculated with labor-augmenting productivity taken into account.

Table A19: Product characteristics

<i>Panel (a): Cigarette content</i>						
	log(Leaf weight)		log(Filter density)		log(Rod density)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
Treatment	0.000107	(0.00685)	0.00571	(0.0101)	0.0142	(0.0181)
Observations	353		353		353	
R-squared	0.804		0.702		0.550	
<i>Panel (b): Cigarette design</i>						
			log(Paper permeability)		log(Ventilation)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
Treatment			0.0331	(0.0732)	-0.891	(0.491)
Observations			353		353	
R-squared			0.586		0.848	
<i>Panel (c): Correlations</i>						
	log(Markup)		log(Markdown)		log(TFP)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
Ventilation	0.0149	(0.0281)	0.0261	(0.0303)	-0.00883	(0.0573)
Rod density	4.685	(6.984)	-0.253	(7.524)	1.224	(14.24)
Filter density	22.46	(6.346)	-8.990	(6.836)	-9.010	(12.93)
Leaf weight	11.66	(3.651)	-5.984	(3.933)	-9.399	(7.441)
Paper permeability	1.981	(1.355)	-1.391	(1.459)	-0.786	(2.761)
Observations	137		137		137	
R-squared	0.687		0.700		0.488	

Notes: Panel (a) compares the cigarette contents between the treatment and control groups. Panel (b) does the same for two cigarette design features. Panel (c) reports the correlations between markups, markdowns, productivity and cigarette characteristics. Province dummies are controlled for.

Table A20: Consolidation and exporting

	Export dummy		Log(Exports/Revenue)	
	Estimate:	SE:	Estimate:	SE:
Treatment * 1(year \geq 2003)	-0.048	(0.043)	-0.011	(0.009)
Treatment	-0.030	(0.042)	-0.006	(0.010)
1(year \geq 2003)	0.076	(0.025)	0.008	(0.003)
Observations	2,638		2,486	
R-squared	0.040		0.011	
1(year \geq 2003)	0.076	(0.025)	0.008	(0.003)

Note: Controls include a time trend, product type, ownership type and firm fixed effects.

Table A21: Consolidation effects, alternative market definitions

<i>Panel (a): Province level</i>	log(Markdown)		log(Markup)		log(Productivity)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	0.299	(0.063)	-0.255	(0.064)	0.074	(0.082)
R-squared	0.801		0.788		0.862	
Observations	1,091		1,091		1,091	
<i>Panel (b): Prefecture level</i>	log(Markdown)		log(Markup)		log(Productivity)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	0.405	(0.071)	-0.245	(0.069)	0.107	(0.074)
R-squared	0.879		0.847		0.867	
Observations	1,091		1,091		1,091	
<i>Panel (c): County level</i>	log(Markdown)		log(Markup)		log(Productivity)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Treatment * 1(year \geq 2003)	0.788	(0.134)	-0.413	(0.111)	0.084	(0.103)
R-squared	0.718		0.763		0.862	
Observations	1,091		1,091		1,091	

Notes: Estimates of treatment effect in difference-in-differences regression reported. Same specification as before. Dependent variable in logs. Treatment and control groups calculated by defining input markets at different geographical levels.

Table A22: Market power and ownership

<i>Panel (a): Buying power</i>				
	log(Markdown)		log(Markdown)	
	Estimate:	SE:	Estimate:	SE:
Log(% Equity state-owned)	0.135	0.0240	0.0168	(0.0618)
Log(% Equity owned by legal person)	0.246	(0.079)	0.101	(0.077)
Observations	1,091		1,091	
R-squared	0.061		0.877	
Firm FE	No		Yes	
<i>Panel (b): Selling power</i>				
	log(Markdown)		log(Markdown)	
	Estimate:	SE:	Estimate:	SE:
Log(% Equity state-owned)	0.310	(0.0240)	0.0135	(0.0618)
Log(% Equity owned by legal person)	0.246	(0.0413)	0.101	(0.125)
Observations	1,091		1,091	
R-squared	0.044		0.743	
Firm FE	No		Yes	

Notes: Omitted ownership categories are 'private person', 'foreign' and 'collective'.