Market structure, oligopsony power and productivity

Michael Rubens*

July 26, 2020

Abstract

I examine how changes in market structure affect input market power, product market power, and total factor productivity. Using a structural model of production and input market competition, I study the effects of a large-scale consolidation policy in the Chinese cigarette manufacturing industry, which forced small-scale factories to close down. Although this policy was aimed at increasing productivity, I find that its main effect was to increase oligopsony power on tobacco leaf markets, as leaf price markdowns increased by 28%. This had important aggregate distributional consequences: the markdown rise explains 20% of the increase in income inequality between manufacturing workers and tobacco farmers between 2003 and 2006.

Keywords: Market Power, Monopsony, Market Structure, 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 advice. I also thank Frank Verboven, Chad Syverson, Otto Toivanen, Dan Ackerberg, Myrto Kalouptsidi, and Benjamin Vatter for helpful suggestions, and participants at various conferences and workshops. Financial support from the Research Foundation Flanders (Project 11A4618N), Fulbright Belgium and the Belgian American Educational Foundation is gratefully acknowledged.

1 Introduction

There is an ongoing debate about the aggregate evolution of market concentration and market power, both in the U.S.A. and globally (Rossi-Hansberg et al., 2018; Autor et al., 2017; Covarrubias et al., 2019; De Loecker et al., 2020). Increasing market concentration can lead to changes in aggregate productive efficiency, due to scale economies and returns to scale, but can also affect both oligopoly power of firms on their *product* markets, and oligopsony power on their *input* markets.[1] Existing empirical work on market structure, market power and productivity tends to focus on a subset of these three effects, while assuming away the others. In order to fully understand the aggregate consequences of changes in market structure, however, it is important to study its effects on all these channels together.

This paper fills this gap by empirically examining the effects of changes in market structure on both market power, buyer power, and productive efficiency. For this purpose, I construct a structural model to separately identify *markups*, i.e. the wedge between marginal costs and product prices, from *markdowns*, the wedge between marginal factor products and input prices, and total factor productivity. The typical approach in the so-called 'cost-side approach' has been to combine production and cost data to model only the input demand side, and compare marginal factor products to prices (Hall, 1986; De Loecker and Warzynski, 2012). Although this approach usually assumes exogenous input prices, extensions to allow for endogenous input prices due to buyer power were made by De Loecker et al. (2016); Morlacco (2017). I show, however, that this approach fails to identify markups from markdowns as soon as a subset of inputs are non-substitutable.[2] This is the case for many intermediate inputs across various industries, such as beer brewing (hop), coffee roasting (beans) and consumer electronics (rare earth metals), among many others.[3] In order to solve this identification challenge, a model of the input supply side needs to be combined with the production model, which yields input demand conditions. I therefore impose

[1]These relationships are theoretically ambiguous, cfr. Syverson (2019).

[2]In 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.

[3]Even 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.

a discrete choice model of input suppliers choosing producers in a similar vein to demand-side approaches in empirical IO such as Berry (1994). I show that identification of both the production function and of the input supply model leads to separately identified markups, markdowns and total factor productivity.

Next, I examine how markups, markdowns and productivity are affected by changes in market structure. For this purpose, I study the Chinese cigarette manufacturing industry, which provides an 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.[4] That being said, the identification approach for markups, markdowns and productivity is broadly applicable outside this specific industry, and studying the effects of changes in market structure on markups and markdowns could be achieved even without observing exogenous changes in market structure by imposing more structural assumptions, similarly to the merger counterfactual literature in empirical IO.[5]

In addition, the fact that concentration increases dramatically along the value chain makes buyer power likely in the cigarette industry: around 20 million farmers sell leaves to manufacturers, with the number of manufacturers decreasing from 350 to 150 during the consolidation. These manufacturers in turn sell cigarettes domestically to a monopsonistic government-controlled wholesaler.[6] We can thus expect buyer power to be present along the chain.[7] The key driver of monopsony power on leaf markets is the legal obligation of tobacco farmers to sell their entire output locally, and large switching costs towards other crops and occupations. Both of these features characterize rural labor markets across the developing world, which means that the evidence for buyer power found is this paper are likely to apply to other industries and countries as well.[8]

---

[4]Besides this feature, this industry is also interesting merely due to its size: annual industry revenue exceeds \$7 billion, and 40% of the world's cigarettes are made in China.

[5]Examples of such an approach are Nevo (2001) and Miller and Weinberg (2017).

[6]The 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.

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

[8]Localized agricultural markets due to internal trade regulations are, for instance, a driver of monopsony power on Indian agricultural markets (Chatterjee, 2019), and switching costs are key in the agricultural economics literature

The analysis is structured in three steps. First, I providing reduced-form evidence for the effects of the consolidation on both input and product 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 leaf prices fell sharply for the treatment group, factory-gate cigarette prices fell to a lower extent, and wages fell just slightly. Although these effects are arguably a consequence of the consolidation, they do not suffice 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.

In a second step, I therefore estimate a structural model to recover cigarette price markups, leaf price markdowns, and total factor productivity of manufacturing plants. 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 31% of their marginal revenue product, which is a much larger wedge compared to most prior work focusing on high-income countries. The combination of low markups and high markdowns shows that manufacturers mainly had market power on their input markets, which is consistent with the fact that they bought from many small farmers, but sold to a single large buyer.[9]

Finally, I combine all these estimates to estimate how the consolidation of cigarette manufacturing firms affected market power downstream and upstream, and manufacturing productivity. I find that the change in market structure mainly led to increased oligopsony power on leaf markets: leaf price markdowns increased on average by 28% more in the treatment group relative to the control group. The markdown increase was the largest in leaf markets that were highly frictional. I find some evidence for a drop in cigarette price markups, which is consistent with a monopsonistic wholesaler used its own buying power to push down factory-gate prices. Finally, I find no strong evidence for the policy to have spurred productivity growth, in contrast to the official policy objective. I find, however, that total output fell in the consolidated leaf markets, which is consistent

(Song et al., 2011).

[9] High 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).

with the classical monopsony model.

I use the estimates of the structural model to quantify the extent to which the consolidation policy contributed to rural-urban income inequality. By increasing markdowns on tobacco leaf markets, but not on manufacturing labor markets, income inequality between rural farmers and urban manufacturing workers increased. I find that the markdown increase due to the consolidation explains 20% of the increase in income inequality between farmers and manufacturing workers.[10]

The two key contributions of this paper are to examine how changes in market structure affect both product price markups, input price markdowns and productivity, and to provide an empirical framework that allows separate identification of these three dependent variables when not all production inputs are substitutable. In addition to the literature on market power and ownership consolidation mentioned earlier, this paper relates to three related strands of literature. First, there is a series of recent and contemporaneous papers that use discrete choice models of input supply with differentiated firms to identify labor market power, such as Card et al. (2018); Berry et al. (2019).[11] These papers do, however, not combine an input supply model with a production-cost model, and hence focus only on competition on input markets.[12] Moreover, I use the estimated input demand shocks from the production model to help identify the parameters of the input supply model, and discuss the assumptions that are necessary for this approach.

Secondly, I contribute to the literature on the efficiency gains from consolidation (Braguinsky et al., 2015; Blonigen and Pierce, 2016; Grieco et al., 2017) and from State-Owned Enterprise (SOE) reform and privatization (Hsieh and Song, 2015; Chen et al., 2018). In contrast to these papers, I allow for input prices to be endogenous choices of the firm, and for buyer power to change in response to changes in market structure. This changes the interpretation of the productivity residual: assuming exogenous leaf prices leads to the conclusion that average TFP increased by 20% due to the consolidation. In reality, leaf prices fell as monopsony power increased on leaf markets. A part of the large TFP gains from consolidation and SOE privatization found in this

[10]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. Such transfer schemes may not have been necessary in the absence of a consolidation.

[11]Other recent work on monopsony power, but with a different modelling strategy and research question, includes Naidu et al. (2016); Goolsbee and Syverson (2019); Berger et al. (2019); Jarosch et al. (2019).

[12]Tortarolo and Zarate (2018) does combine a production model with an input supply model, but with a different identification strategy, assuming substitutable inputs, and with a different research question.

literature could therefore be due to increased buyer power, rather than increased efficiency.[13] This has important policy implications: large-scale consolidation policies, such as the one studied in this paper, are increasingly common both in China,[14] and other countries such as Indonesia. If these reforms lead to rising monopsony power, could even lead to decreasing productivity growth through reduced allocative efficiency.

Finally, 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; Ravallion and Chen, 2009). I show that consolidation of SOEs can be an important driver of income inequality by increasing buyer power of these SOEs on agricultural markets.[15]

This paper is structured as follows. I discuss the industry background, data, and reduced-form evidence on the consolidation policy in section 2. Next, I present the model and discuss identification of markups, markdowns and productivity in section 3. The main results are discussed in section 4. I end with quantifying the aggregate consequences of the consolidation policy for income redistribution and productivity growth at the industry level in section 5.

2 Key facts on the Chinese tobacco industry

2.1 Industry setting

Farming

The value chain of the production of cigarettes in China is visualized in panel (a) of figure 1. At the start of the panel, in 1998, 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). After being harvested and dried, tobacco leaf needs to be 'cured'.[16]

[13] Hsieh and Song (2015) finds, for instance, 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.

[14] China consolidated many of its SOEs into industrial giants in various important industries such as energy, transport utilities, telecommunication and defense industries. These policies are known as *"Grasping the large and letting the small go"* (Naughton, 2007).

[15] Although the relationship between market power and income inequality has been studied before, e.g. in De Loecker et al. (2020), their focus has been mainly on product market power rather than on monopsony power.

[16] Various alternative processes are possible, such air curing, fire curing and flue curing.

Farmers sell tobacco leaf to local 'purchasing stations', which are operated by the cigarette manufacturers. Tobacco leaves are sorted into quality 'grades', each of which sells at a different price. Chinese tobacco farms became less profitable during the time period studied: they dropped from being the median cash crop in terms of farm profitability in 1997 to the last place in 2004. (FAO, 2003; 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).

Manufacturing

Cigarette manufacturers turn tobacco leaf and other intermediate inputs, such as paper and filters, into cigarettes using labor and capital.[17] Intermediate inputs make up for 90% of variable input expenditure, and tobacco leaf accounts for two thirds of intermediate input expenditure.[18] A picture of the consecutive steps in the cigarette production process are in panels (b)-(d) of figure 1. A map of tobacco manufacturing locations in 1999 and 2006 is in panels (b) and (c).

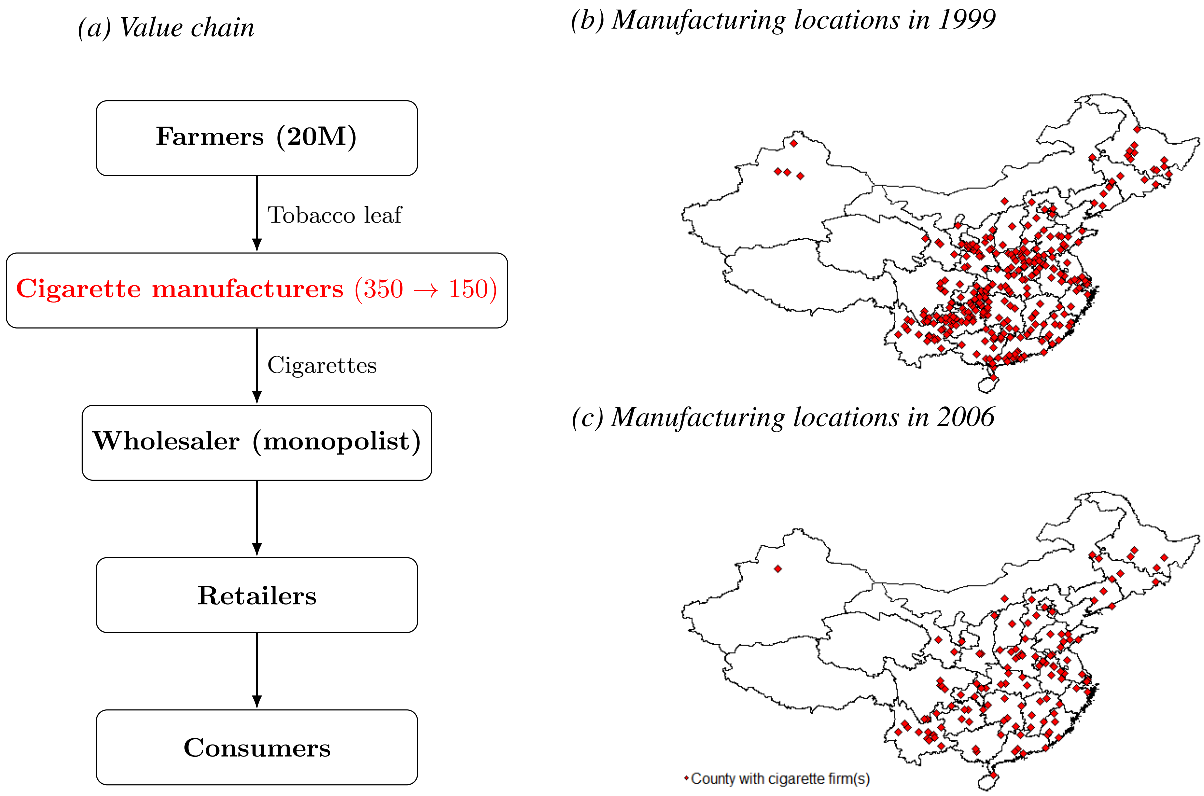
Wholesaling

Manufacturers sell their cigarettes to wholesalers which are controlled by the State Tobacco Monopoly Administration (STMA) through its commercial counterpart, the *Chinese National Tobacco Trade*

[17]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.

[18]The 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, 1997)

Figure 1: Tobacco industry structure



Notes: Panel (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. Panels (b)-(c) map the 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.

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.[20] The fiscal importance of the tobacco industry may be an important reason for this protection: in 1997,

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

[20]Using UN Comtrade data, accessed at http://comtrade.un.org/

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 (State Administration of Taxation, 2015).

Market definitions

Farmers are obliged to sell their leaf output at purchasing stations in their own county. Tobacco leaf cannot be transported across county borders without the approval of the provincial board of the industry regulator, the *State Tobacco Monopoly Administration* (STMA). Leaf markets are therefore in theory restricted to the county-level (State Council of the People's Republic of China, 1997). In practice, there is some tobacco trade across counties as not all tobacco-growing counties contain a cigarette factory, and as cigarette manufacturers frequently locate purchasing stations near county boundaries to attract nearby farmers from other counties (Peng, 1996). I therefore define leaf markets at the prefectural level, which is the administrative level above the county. Another motivation for using prefectural leaf market definitions is that leaf prices significantly fall with the number of firms in a prefecture, while this relationship does not exist at the province or county levels, as shown in appendix E.3. In appendix D.1, I re-estimate the model using both narrower and broader leaf market definitions as a robustness check. There were on average 1.9 cigarette manufacturers per prefecture throughout the sample, and 193 prefectures with at least one cigarette factory. The average Hirschman-Herfindahl index was 0.795, so leaf markets were highly concentrated.

In contrast to tobacco leaf, cigarette markets are not isolated, although not fully integrated either because of transportation costs and provincial home bias. I do not take a stance on cigarette market definitions, as this is not necessary to estimate the model.

2.2 Data

I combine multiple datasets. First, I use establishment-level production and cost data on the manufacturers between 1999 and 2006 from the *Annual Survey of Industrial Firms*, which is conducted by the National Bureau for Statistics (NBS). I retain all manufacturers of "Tobacco and Manufactured Tobacco Substitutes".[21] This results in 478 establishments and 2139 observations. The above-scale survey includes non-SOEs with sales exceeding 5 million RMB and all SOEs irrespective of their size.[22] The 'establishments' in the NBS data usually refer to subsidiaries, rather than to independent firms: almost all factories observed in the data are formally subsidiaries of the *Chinese National Tobacco Corporation* (CNTC). However, cigarette factories of the corporation "operate as separate enterprises responsible for their own losses and profits" (Peng, 1996). They are autonomous in how they operate and set input prices and hence compete against each other (Wang, 2013). 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.[23] Third, I obtain county-level population statistics from the 2000 census. Finally, I use brand-level cigarette characteristics for some robustness checks. More details about all datasets and cleaning thereof is in appendix A. Summary statistics on some variables of interest are in appendix table A1.

2.3 Reduced-form evidence: consolidation and prices

The consolidation policy

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).[24] The official motivation for this policy was to "enable China's cigarette industry to achieve scale and efficiency" (STMA, 2002). The left graph in figure 2 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,[25] while firms with

[21] These correspond to CIC codes 1610, 1620 and 1690.

[22] I refer to Brandt et al. (2012) for a comprehensive discussion of this dataset.

[23] Some 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.

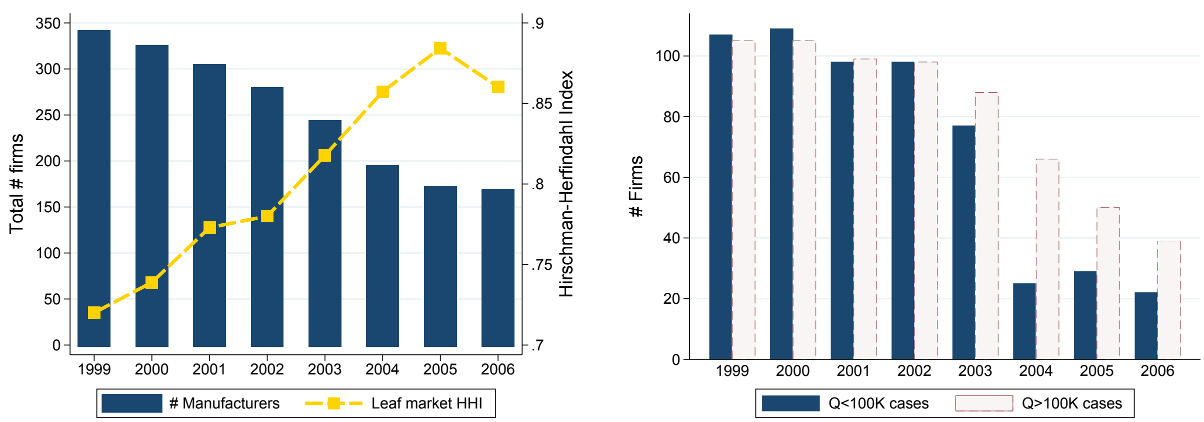
[24] I test for announcement effects in appendix D.1

[25] One case contains 50,000 sticks of cigarettes (Fang et al., 2017)

an annual production below 300,000 cases were encouraged to merge with larger firms.[26] The graph on the right of figure 2 compares the number of firms which produce less and more than 100,000 cases per year.[27] The number of firms under the exit threshold fell sharply between 2002 and 2004, from 98 to 25, compared to 98 to 66 above the threshold. 22 firms continued to exist after 2003, despite producing below the exit threshold.[28] The firms producing less than 100,000 and 300,000 cases represented a third and one half of all firms respectively in 2002, generating 6% and 11% of industry revenue. As figure 2 shows, average leaf market HHIs increased from 0.72 to 0.86 between 1999 and 2006.

Figure 2: 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 leaf market HHIs at the prefectural level levels (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).

Factor revenue shares

Panel (a) of figure 3 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

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

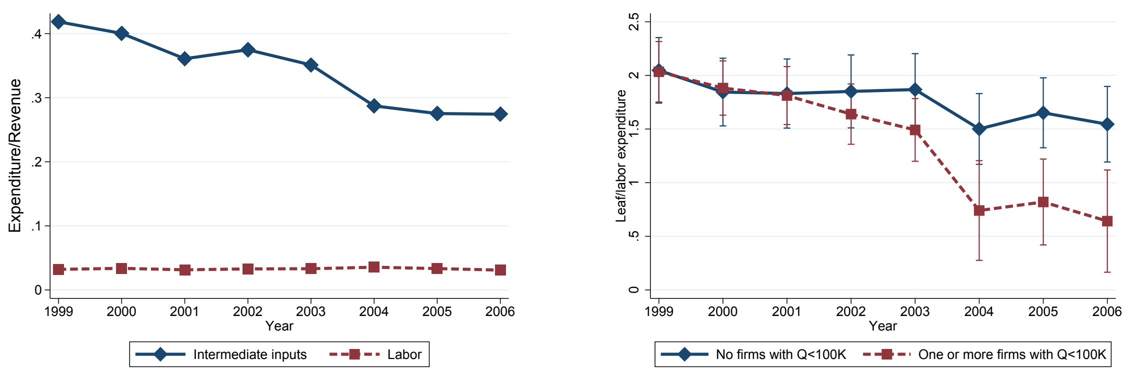
[27] As quantities are observed for only a subset of firms, the annual number of firms reported is lower compared to the previous graph.

[28] Among these, 12 were privately owned, and could hence not be forced to exit. Another 7 firms were dropped during the data cleaning procedure due to anomalies such as negative intermediate input expenditures. That leaves 3 'non-compliers', which did not exit for unknown reasons.

3%, while the aggregate intermediate input share of revenue fell from 40% to 25% between 2000 and 2007. The variable cost share of tobacco leaf hence dropped sharply. One explanation for this could be a fall of the amount of tobacco leaf needed to produce a cigarette relatively to labor. This is very unlikely, however: the required amount of tobacco leaf per cigarette varies little across firms.[29] The amount of labor needed per cigarette could have changed due to mechanization, but in order to generate the patterns in figure 3, the required amount of labor per cigarette would have had to *increase* over the sample period, while mechanization has the opposite effect. A second explanation for panel (a) is falling leaf prices relatively to labor. This can be due to rising oligopsony power on leaf markets relatively to labor markets, or to many other reasons, such as general equilibrium price changes, as China's manufacturing sector grew quickly during this period.

Figure 3: Factor revenue shares

(a) Aggregate *(b) By treatment (average)*



Notes: Panel (a) plots the evolution of the total wage bill and total intermediate input expenditure over industry revenue. Panels (b) compares the average ratio of labor expenditure over intermediate input expenditure over time for the consolidation treatment and control groups. Confidence intervals are plotted at the 95%.

Treatment and control groups

To what extent did increased concentration contribute to the fall in the intermediate input share of revenue? Let the number of firms producing less than 100,000 cigarette cases in market i and year

[29]More evidence on this is in appendix C.6.

t be denoted N_{it}, with firms denoted by f and the number of cases produced as Q_{ft}.

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

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, when the reform started.

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

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. Appendix table A6 contains more information about the characteristics of the treatment and control groups.

Changes in market structure and prices

In order to assess the effects of changes in downstream market structure on both 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 subsequently the log factory-gate cigarette price p_{ft}, log wages per employee w^L_{ft}, and the log revenue share of intermediate inputs, $\log(\frac{\text{material expenditure}}{\text{revenue}})$. Intermediate input prices are not observed, so I use the intermediate input revenue share as a first indicator of leaf prices. The consolidation dummy C_f itself is not included as it is subsumed into the firm dummy θ^y_f. The coefficient of interest that quantifies the consolidation effects is θ^y_2.

$$y_{ft} = \theta^y_0 + \theta^y_1 \mathbb{I}[t \geq 2003] + \theta^y_2 C_f \mathbb{I}[t \geq 2003] + \theta^y_3 t + \theta^y_f + v^y_{ft} \quad (1)$$

$$\text{with } y_{ft} \in \{p_{ft}, w^L_{ft}, \log(\frac{\text{material expenditure}}{\text{revenue}})\}$$

I start with visual evidence. Panel (b) of figure 3 compares the evolution of average relative input expenditure $\log(\frac{\text{material expenditure}}{\text{labor expenditure}})$ between the treatment and control group. This is hence the ratio of the solid blue line over the dotted red line in panel (a), broken into treatment and control groups. 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. Next, I estimate equation (1) in panel (a) of table 1.[30] Labor wages increased by 13%,[31] but this increase was not statistically significant. Cigarette prices dropped by 7.9%, but again not significantly. The intermediate input share of revenue fell, however, by 16% on average, in line with the visual evidence shown above.

Table 1: Reduced-form evidence: consolidation and prices

(a): Consolidation treatment effects	log(Wage)		log(Leaf rev. share)		log(Price)	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Treatment * 1(year≥2003)	0.124	(0.0872)	-0.175	(0.0710)	-0.082	(0.095)
R-squared	0.350		0.127		0.391	
Observations	1,086		1,086		1,086	
(b): Pre-trends						
Treatment * year	-0.0172	(0.0936)	-0.0675	(0.0885)	-0.146	(0.140)
R-squared	0.0281		0.0199		0.0129	
Observations	752		752		752	

Notes: Left-hand variables are log wages per employee, log intermediate input expenditure per revenue and log cigarette prices at the firm-year level. Panel (a) reports the average treatment effects from equation (1), with the following controls: export dummy, ownership type, product dummies, linear time trend. Panel (b) estimates the pre-trends for all three outcomes on the period 1999-2003.

This evidence tells us that the change in market structure seemed to have mainly affected prices on leaf markets, rather than on labor markets. This reduced-form evidence is, however, not sufficient to draw conclusions about the underlying mechanism. Falling leaf prices could be due to increased markdowns, but changes in productive efficiency would also lead to different equilibrium input and product prices. Moreover, in order to know how markups changed, observing price variation is not sufficient, marginal costs need to be recovered as well. I therefore construct a more

[30] I use prefecture-level market definitions. In panel (d), I test the parallel pre-trend assumption using equation (18). The time trends in all three outcome variables were not significantly different before the policy was implemented, so parallel pre-trends cannot be rejected.

[31] $= \exp(0.124) - 1$

structural model of competition and production in the next section.

3 A model of production, markups and markdowns

3.1 Input demand

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.[33] Let the production function be given by equation (2):

$$Q_{ft} = \min\left\{\beta^M_{ft} M_{ft}, \Omega_{ft} H(L_{ft}, K_{ft}, \boldsymbol{\beta})\right\} \exp(\varepsilon_{ft}) \tag{2}$$

The amount of tobacco leaf needed to produce a case of cigarettes is assumed to be a scalar β^M_{ft}. Manufacturers differ in terms of their productivity level Ω_{ft}. In the baseline specification, this productivity term is assumed to be a scalar, but this can be generalized.[34] Firms use a common production technology $H(.)$ with parametrization $\boldsymbol{\beta}$ for the substitution pattern between labor and capital. I assume $H(.)$ is twice differentiable in both labor and capital. Measurement error in output is denoted ε_{ft}. Equation (2) 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.

Cigarette prices

I assume manufacturers produce a single product, cigarettes, at price P_{ft}.[35] Cigarettes are vertically differentiated across firms, with an unobserved, firm-level quality index ζ_{ft}. Cigarette quality is

[32] Other 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.

[33] One reason why leaf could be substitutable for labor or capital would be waste-reducing technologies. I estimate the elasticity of input substitution in appendix C.1, however, and find no evidence for leaf to be substitutable with either labor or capital.

[34] I extend the model to allow for factor-augmenting productivity in appendix C.3.

[35] The 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.

assumed to be exogenous. In section 4.1, I discuss more in detail how endogenous quality choices could affect identification of the model.

Assumption 1. — Cigarette quality ζ_{ft} is exogenous from the point of view of an individual manufacturer f.

I do not impose a specific model of competition or demand on either the wholesale or retail markets. The cigarette demand function is given by equation (3), of which the parameters $\boldsymbol{\tau}_{ft}$ are heterogeneous across firms and time periods. If manufacturers have pricing power on the wholesale market, then cigarette prices are endogenous to how many cigarettes they sell.

$$P_{ft} = P(Q_{ft}, \zeta_{ft}; \boldsymbol{\tau}_{ft}) \tag{3}$$

Input sourcing

As the amount of leaf per cigarette is assumed constant, leaf M is a variable input. I assume manufacturers can flexibly adjust their leaf stock without adjustment costs or inventories, which makes leaf a static input.[36] The flexibility assumption is made because leaf is sold at a monthly frequency without the use of forward contracts.[37] Labor is assumed to be a variable input as well: cigarette manufacturing plants rely mainly on production workers; which are likely to be variable in output.[38] I do allow for adjustment costs in labor, however, which makes it a dynamic input. This is in line with the prior literature on Chinese SOE-dominated industries, such as Chen et al. (2018): there are important hiring and firing costs on Chinese labor markets, as in most countries, and especially in state-owned enterprises where employees enjoy more job security compared to private firms. The capital stock is, finally, assumed to be fixed and evolves dynamically with a depreciation rate ρ^K and investment I_{ft}: $K_{ft} = \rho^K K_{ft-1} + I_{ft-1}$.

[36] I follow the input classification of Ackerberg et al. (2015).

[37] In section C.9, I discuss departures from these assumptions, such as inventories and dynamic leaf demand.

[38] The NBS surveys does not distinguish production from non-production workers, but 70% of US cigarette manufacturing employees and 65% of the wage bill were production workers, and hence variable, in 1997 (U.S. Census Bureau, 1997)

Input prices

The prices of leaf, labor and capital are denoted W_{ft}^M, W_{ft}^L and W_{ft}^K. The extent of buyer power of a manufacturer f over an input $V \in \{L, M, K\}$ is parametrized by the input supply elasticity ψ_{ft}^V.

$$\psi_{ft}^V \equiv \frac{\partial W_{ft}^V}{\partial V_{ft}} \frac{V_{ft}}{W_{ft}^V} + 1 \geq 1$$

If the price of input V is exogenous to a manufacturer, then increasing the input quantity purchased does not lead to a change in the price of this input, which implies that $\psi_{ft}^V = 1$. If a manufacturer has buyer power over input V, the input price W^V increases when more inputs are purchased, meaning that $\psi_{ft}^V > 1$. In the baseline model, the prices of labor and capital are assumed to be exogenous with respect to individual manufacturers because labor wages did not adjust much in response to the consolidation, and because manufacturing labor and capital markets do not share the leaf markets' institutional feature of being geographically isolated due to transportation restrictions. The model can, however, easily be extended to allow for buyer power over either labor or capital.[39]

Tobacco leaf is a differentiated product as well. In order to produce high-quality cigarettes, high-quality leaf is needed, which is more expensive. I assume that leaf quality is a strictly increasing function of cigarette quality ζ_{ft}, which means that ζ_{ft} indicates both output and input quality. The leaf price therefore depends on the quality index ζ_{ft}. The leaf price also depends on the leaf quantity if there is buyer power, and on manufacturer characteristics that affect the payoff of the leaf farmers. These manufacturer characteristics are denoted $\mathbf{X}$ if observed, and $\boldsymbol{\xi}_{ft}$ if latent. The leaf supply function is given by equation (4) and is parametrized by firm-year specific coefficients $\boldsymbol{\gamma}_{ft}$:

$$M_{ft} = M(W_{ft}^M, \zeta_{ft}, \mathbf{X}_{ft}, \boldsymbol{\xi}_{ft}; \boldsymbol{\gamma}_{ft}) \tag{4}$$

[39] I do this in appendix C.5, but find no evidence for such buyer power over manufacturing workers.

Manufacturer decisions

Variable profits are given by $\Pi_{ft} = P_{ft}Q_{ft} - W^M_{ft}M_{ft} - W^L_{ft}L_{ft}$. Using the production and product demand functions (2)-(3), variable profits can be rewritten as:

$$\Pi_{ft} = P(M_{ft}, \zeta_{ft}, \boldsymbol{\tau}_{ft})\beta^M M_{ft} - W^M_{ft}M_{ft} - W^L_{ft}H^{-1}(K_{ft}, M_{ft}, \Omega_{ft})$$

I assume manufacturing firms choose the tobacco leaf price W^M_{ft} each year in order to maximize per-period variable profits. Prior production-cost approaches to markup identification assume cost minimization conditional on output levels, rather than profit maximization. As intermediate inputs are non-substitutable, however, choosing their price can only be done by also changing output, which hence cannot be conditioned upon.[40]

***Assumption 2.* —** Firms simultaneously choose input prices W^M_{ft} annually to maximize per-period variable profits Π_{ft}.

As leaf quantities are a function of leaf prices, equation (4) can be substituted into the variable profit function. The profit maximization problem is therefore given by equation (5). As labor and tobacco leaf cannot be substituted, there is just one first order condition, rather than one for each variable input.[41] When firms choose the leaf price, and hence the quantity of tobacco leaf used, they automatically also choose how much labor to use.

$$\max_{W^M_{ft}} \Big(P(W^M_{ft}, \psi^M_{ft}, \zeta_{ft}, \boldsymbol{\tau}_{ft})\beta^M M(W^M_{ft}, \zeta_{ft}, \mathbf{X}_{ft}, \boldsymbol{\xi}_{ft}; \boldsymbol{\gamma}_{ft}) - W^M_{ft} M(W^M_{ft}, \zeta_{ft}, \mathbf{X}_{ft}, \boldsymbol{\xi}_{ft}; \boldsymbol{\gamma}_{ft}) \\ - W^L_{ft}H^{-1}(K_{ft}, W^M_{ft}, \zeta_{ft}, \mathbf{X}_{ft}, \boldsymbol{\xi}_{ft}; \boldsymbol{\gamma}_{ft}, \Omega_{ft})\Big) \quad (5)$$

Both the assumption that firms maximize profits and that they choose input prices can be questioned. It is often suggested that state-owned enterprises (SOEs) have non-profit objectives such as generating local employment (Lu and Yu, 2015). In the tobacco industry, however, Peng (1996) notes that cigarette manufacturers have "the purpose of making profits" and "often bargain with each other for better deals".[42] Next, leaf prices were in theory regulated by the government. In

[40] This also implies that increasing monopsony power translated into decreasing output, which I verify in appendix 5.2.

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

[42] In appendix C.7, I still extend the model to allow for objective functions other than cost minimization. Different

reality, however, manufacturing firms had considerable pricing power on leaf markets from the 1980s onwards: Peng (1996) mentions, for instance, frequent conflicts between peasants and manufacturers over leaf pricing, with farmers being, in some cases, forced to sell tobacco at prices below their cost of production when showing up at a purchasing point at moments of oversupply on the market.[43] I discuss leaf pricing strategies under price regulations in the Chinese setting in more detail in appendix E.5. Finally, purchasing stations are formally an additional intermediary between manufacturers and farmers, but as discussed above they are not independent from the manufacturers, and therefore not separately modelled from the manufacturers.

3.2 Markups and markdowns

Definitions

The markup ratio μ is the ratio of factory-gate cigarette prices P_{ft} over marginal costs MC_{ft}, which are defined as $MC_{ft} \equiv \dfrac{\partial(W^M_{ft} M_{ft} + W^L_{ft} L_{ft})}{\partial Q_{ft}}$:

$$\mu_{ft} \equiv \frac{P_{ft}}{MC_{ft}} \tag{6}$$

It can be shown that the inverse supply elasticity ψ^V is equal the ratio of the marginal revenue product of input V over its price, as in equation (7).[44] This supply elasticity therefore has the interpretation of a *markdown ratio*: the larger ψ^V, the wider the gap between the input price and marginal product, which indicates higher buyer power.

$$\psi^V_{ft} = \frac{\frac{\partial(P_{ft} Q_{ft})}{\partial V_{ft}}}{W^V_{ft}} \tag{7}$$

Finally, the markdown *wedge* δ^V is defined as the relative wedge between the marginal revenue product of input V and the price of that input. It is a function of the input price elasticity of supply

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.

[43] This can, for instance, be due to coordination failures.

[44] This is derived in appendix E.1.

ψ^V.[45]

$$\delta^V_{ft} \equiv \frac{\frac{\partial(P_{ft}Q_{ft})}{\partial V_{ft}} - W^V_{ft}}{\frac{\partial(P_{ft}Q_{ft})}{\partial V_{ft}}} = \frac{\psi^V_{ft} - 1}{\psi^V_{ft}}$$

From this point onwards, I will shorten *markdown ratio* ψ^V to *markdown*, and report and use this statistic, rather than the markdown *wedge* δ^V, for two reasons. First, ψ^V has a similar interpretation as the markup ratio μ, being that a value of one corresponds to the exogenous price case. Related to this, the magnitude and distribution of the markup ratio μ has the same support as $\psi^V \in [0, \infty]$, while the markdown wedge δ^V has support on $[-1, 1]$. Second, the product of ψ^V and μ has the interpretation of the variable profit margin. This also implies that firms can operate at a positive variable profit even if the markup μ is below one: there is a wedge both between the product price and marginal costs, and between marginal costs and input prices.

General case: endogenous input prices and non-substitutable inputs

Solving the first order conditions from equation (5) for marginal costs yields the markup expression in equation (8a), which is derived in appendix E.1.

$$\mu_{ft} = \left(\frac{\alpha^L_{ft}}{\beta^L_{ft}}\psi^L_{ft} + \alpha^M_{ft}\psi^M_{ft}\right)^{-1} \tag{8a}$$

with $\alpha^V_{ft} \equiv \dfrac{V_{ft}W^V_{ft}}{P_{ft}Q_{ft}}$ for $V \in \{L, M\}$

The markup expression in (8a) looks different compared to the typical markup expression from De Loecker and Warzynski (2012), in which it is the ratio of an output elasticity over a revenue share, for two reasons. First, equation (8a) has an additive structure, because of the complementarity of labor and intermediate inputs. Each variable input cannot be changed without changing the other input as well. Second, the markup expression contains the input price elasticities ψ^L and ψ^M. 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. As said before, labor wages are assumed to be exogenous in the context of Chinese tobacco, meaning that

[45]Cfr. appendix E.1.

$(\psi_{ft}^L - 1) = 0$, but this can easily be relaxed.

The revenue shares α_{ft}^V on the right-hand side of equation (8a) are observed. The output elasticity of labor β_{ft}^L is latent, but can be retrieved by estimating the production function. This does, however, not suffice to identify the markup: the markdown ratio ψ_{ft}^M is still latent. Markups μ_{ft} and markdowns are hence not separately identified when only identifying the production function. 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 markdowns are equal to one: $\psi^V = 0$, $\forall V$. The markup expression then simplifies to the formula from De Loecker and Warzynski (2012):

$$\mu_{ft} = \frac{\beta_{ft}^L}{\alpha_{ft}^L} \tag{8b}$$

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 markdown. This corresponds to the expression from Morlacco (2017). If the price of at least one variable input is exogenous, then the markdown of all other inputs can be found by dividing the markup of these other inputs by the markup derived using the input with the exogenous price.

$$\mu_{ft} = \frac{\beta_{ft}^L}{\alpha_{ft}^L \psi_{ft}^L} \tag{8c}$$

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.[46] In this case, $\beta_{ft}^M > 0$, but all markdowns (ψ_{ft}^V) are one. The markup is given by equation (8c), which corresponds to De Loecker and Scott (2016). It is identified even if there is only one substitutable input.

$$\mu_{ft} = \left(\frac{\alpha_{ft}^L}{\beta_{ft}^L} + \alpha_{ft}^M\right)^{-1} \tag{8d}$$

Possible approaches to achieve identification

Back to the general case in equation (8a). There are two options to separately identify the markup from the markdown. A first is to identify markups μ by imposing a 'full' model of how firms compete on their product market and back out the markdown without modeling how they compete on their input markets. A second is to identify markdown using a model of input market competition. In the context of this paper, I take this second approach because leaf markets are easier to model and define than cigarette markets. Demand for cigarettes is inherently dynamic due to addiction, and cigarette markets are geographically not delineated, while both these issues do not apply for tobacco leaf.

3.3 Input supply

In this section, I specify a 'full' model of how manufacturers compete on their leaf markets, in order to identify the input supply elasticity ψ_{ft}^M. I rely on a discrete choice model in the tradition of Berry (1994).

Farmer utility and decisions

Farmers j sell tobacco leaves on an isolated market i 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

[46] 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.

market and that farmers sell their entire production to a single firm, which makes sense as there were 20 million household-level farms producing but merely 350 firms in 1997 (FAO, 2003). The utility of a farmer j depends on the leaf price, manufacturer characteristics $\mathbf{X}_{ft}$ and $\boldsymbol{\xi}_{ft}$, cigarette quality ζ_{ft}, and an i.i.d. type-I distributed manufacturer-farmer utility term ν_{jft}. Examples of manufacturer characteristics that enter farmer utility could be state ownership, which is observed, or the distance between the factory and a major highway, which is latent.[47] The utility derived from the outside option is normalized to zero. The cigarette quality scalar ζ_{ft} enters farmer utility as higher quality leaves are costlier to grow.

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

I assume farmers periodically choose which manufacturer to sell to by maximizing their static utility. They may choose not to sell to the manufacturer offering the highest price due to the non-price characteristics in the utility function. In the baseline model, I assume all farmers have the same preferences over input prices and manufacturer characteristics, γ^W and γ^X. This is reasonable because there is not much of a relationship between the farmers and the manufacturers other than transacting money; farmers are therefore mainly likely to care about the price they receive for their leaf, and about the latent cost of transporting leaf to the market.[48] Farmer choices are assumed to be static, with the model being specified using one-year intervals. The elasticities that are recovered are, hence, short-run elasticities.[49] I assume that the farmer-manufacturer utility shock ν_{jft} is i.i.d. across firms, farmers and time, and impose the usual logit assumption:

***Assumption 3.* —** The farmer-manufacturer utility shock ν_{jft} follows an extreme-value type-I distribution.

[47] An example of the farmer-manufacturer specific utility shock ν_{jft} could be accidental encounters between farmers and manufacturing employees that facilitate trading relationships.

[48] In appendix C.4, I allow for more flexible substitution patterns by using a nested logit model. When applying the methods used to labor markets, heterogeneous supplier preferences and all kinds of other frictions seem especially important. In that appendix, I also estimate alternative input model specifications, such as a logs-on-logs model.

[49] I discuss the difference between short- and long-run supply elasticities in section C.9.

Competition on leaf markets

As was stated in assumption 2, firms simultaneously choose tobacco leaf prices each period in order to minimize their variable costs. The leaf market share of firm f in year t is denoted as $S_{ft} = \frac{M_{fjt}}{\sum_{r \in \mathcal{F}_{it}} M_{frt}}$. Assuming a pure strategy interior equilibrium exists, and making use of the distributional assumption about ν_{jft}, the first order condition for every firm can be rewritten as:

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

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 9, which will be estimated in the next section.

$$s_{ft} - s_{0t} = \gamma^W W_{ft}^M + \boldsymbol{\gamma}^X \mathbf{X}_{ft} + \boldsymbol{\xi}_{ft} + \zeta_{ft} \tag{9}$$

The leaf price markdown ψ^M, being the inverse input supply elasticity, can be expressed as a function of observable input prices and input market shares, and of the estimated valuation parameter γ^W:

$$\psi_{ft}^M \equiv \Big(\frac{\partial S_{ft}}{\partial W_{ft}^M} \frac{W_{ft}^M}{S_{ft}}\Big)^{-1} + 1 = \Big(\gamma^W W_{ft}^M (1 - S_{ft})\Big)^{-1} + 1 \tag{10}$$

4 Empirical analysis

In this section, I start by discussing identification and estimation of markups, markdowns and productivity. Next, I examine how all three changed in response to the consolidation policy using the difference-in-differences model from section 2.

4.1 Production function: empirics

Identification

The logarithmic version of the production function, equation (2), is given by equation (11). As tobacco leaf is assumed to be non-substitutable and a linear function of the number of cigarettes, it can be omitted from the production function.[50] The production coefficients $\boldsymbol{\beta}$ need to be identified.

$$q_{ft} = h(l_{ft}, k_{ft}, \boldsymbol{\beta}) + \omega_{ft} + \varepsilon_{ft} \tag{11}$$

There are various possible identification strategies for the production function, each with its merits and shortfalls. I combine assumptions on the timing input choices of firms from (Ackerberg et al., 2015; Olley and Pakes, 1996) with the dynamic panel data estimator of Blundell and Bond (2000) to identify the production function. I refer to appendix B for a comprehensive discussion on the choice of the identification strategy, and for a comparison of the estimates when using 'proxy-variable' approaches. In short, the main benefit is that I do not need to impose additional structure on the distribution of latent markups and markdowns to identify the production function, which would be unappealing as this would partly answer the research question using hard-to-test functional form assumptions.

The main drawback of the dynamic panel estimator is that a restrictive linear auto-regressive process on the productivity evolution needs to be imposed.[51] As the productivity effect of the consolidation is one of the research questions, it is crucial to model the productivity evolution as endogenous to the consolidation. I therefore rewrite productivity as a linear function of the consolidation dummy C_{ft}, which indicates that the manufacturer is located in a market subject to the consolidation policy.[52] The productivity residual $\tilde{\omega}_{ft}$ is productivity net of any consolidation effects: $\omega_{ft} = \beta_c C_{ft} + \tilde{\omega}_{ft}$. I impose an AR(1) process on the evolution of $\tilde{\omega}_{ft}$, with serial correlation ρ and unexpected productivity shock υ_{ft}.

$$\tilde{\omega}_{ft} = \rho\tilde{\omega}_{ft-1} + \upsilon_{ft} \tag{12}$$

Identification is achieved by imposing timing assumptions on the choices of the inputs. As labor and capital were assumed to be dynamic inputs with adjustment costs, I assume that both labor and

[50] In general, it could be optimal for firms to diverge from the Leontief 'first order condition' of intermediate inputs equaling the $H(.)$ function in labor and capital (Ghandi et al., 2018) The assumption that intermediate inputs enter the production function linearly solves this problem, as explained in Ackerberg et al. (2015).

[51] I re-estimate the model using Ackerberg et al. (2015) in appendix B, which allows for more flexible productivity transitions, which yields very similar results.

[52] This specification is related to, but different from, the productivity model in Braguinsky et al. (2015), where mergers enter the transition equation for productivity, but not the production function directly.

capital are chosen by the manufacturer one period prior to observing the unexpected productivity shock υ_{ft}. I assume manufacturers cannot choose whether to be subject to the consolidation, so the consolidation dummy C_{ft} is exogenous. The corresponding moment condition result from taking ρ differences are given by equation (13), and are based on Blundell and Bond (2000) and its simplified version in Ackerberg et al. (2015).

$$\mathbb{E}\Big[\upsilon_{ft} + \varepsilon_{ft} - \rho\varepsilon_{ft-1} | (l_{ft}, k_{ft}, C_{ft})\Big] = 0 \tag{13}$$

Product differentiation

Estimating the production function of differentiated products can be biased due to the fact that high-quality inputs, of which prices are unobserved, are needed to produce high-quality, high-price products (De Loecker et al., 2016). Although cigarettes are definitely differentiated products with unobserved quality differences ζ_{ft}, the main determinant of cigarette quality is leaf quality, and not labor or capital characteristics. As leaf does not enter the production function, the 'input price bias' of De Loecker et al. (2016) should not be a first-order issue in this paper. I still extend the model to allow for unobserved quality differences in labor and capital in appendix E.5.

Estimation

In the baseline model, I use a Cobb-Douglas function in labor and capital, meaning that $\boldsymbol{\beta} = (\beta_l, \beta_k)$ and $h(l_{ft}, k_{ft}) = \beta^L l_{ft} + \beta^K k_{ft}$.[53] Rewriting equation (13) using the Cobb-Douglas form yields the following moment conditions:

$$\mathbb{E}\Big[(q_{ft} - \rho q_{ft-1}) - \beta_0(1-\rho) - \beta_k(k_{ft} - \rho k_{ft-1}) - \beta_l(l_{ft} - \rho l_{ft-1}) - \beta_c(C_{ft} - \rho C_{ft-1}) | (l_{ft}, k_{ft}, C_{ft}, q_{ft-1})\Big] = 0$$

One of the drawbacks of the dynamic panel approach compared to the proxy variable approaches is that is harder to account for endogenous exit. Exit in the industry was, however, mainly based on enforcement of the consolidation, which is assumed to be exogenous with respect to individual manufacturers. Moreover, I do not find evidence for a significant correlation between productivity levels and exit probabilities, as discussed in appendix E.2.

[53] In appendix C.2, I estimate a translog production function instead.

Results

The estimated output elasticities are in panel (a) of table 2. The first two columns reports the OLS estimates of the output elasticities of labor and capital, which are 0.450 and 0.831, but these are subject to the usual simultaneity bias in production function estimation. The dynamic panel estimates are in the right columns, and are 0.426 and 0.572 for labor and capital respectively. The scale parameter is estimated at 0.998, which indicates constant returns to scale, although it is imprecisely estimated.

Table 2: Structural model estimates

(a): Production function	OLS		Blundell-Bond (2000)	
	Est.	S.E.	Est.	S.E.
Output elasticity of labor, β_l	0.678	(0.0585)	0.350	(0.143)
Output elasticity of capital, β_k	0.744	(0.0455)	0.495	(0.167)
Returns to scale, $(\beta_l + \beta_k)$	1.422	(0.0306)	0.845	(0.160)
R-squared	0.768		0.891	
Observations	819		819	

(b): Leaf supply function	OLS		IV	
	Est.	S.E.	Est.	S.E.
Leaf supply price semi-elasticity γ^W	-0.195	(0.0246)	2.227	(0.890)
1st stage F-statistic			20.20	
R-squared	0.694		0.176	
Observations	1,086		1,086	

Notes: Panel (a) reports the estimated output elasticities using OLS and the dynamic panel estimator. Both are estimated on the sub-sample of observations for which lagged variables are observed. Panel (b) reports the leaf supply semi-elasticities. The right-hand side variable in equation (9) is the leaf price for one pack of cigarettes in 1000 RMB. Standard errors are bootstrapped.

4.2 Input supply function: empirics

Identification

Next, I turn to the identification of the input supply function, equation (9). Leaf prices W^M and quantities M_{ft} are not observed separately in the data, as in most production-cost datasets. I impose, however, that manufacturers do not differ in terms of leaf content, $\beta^M_{ft} = \beta^M$.[54] This allows recovering the leaf price up to a constant by dividing leaf expenditure by the number of cigarettes produced:[55] $W^M_{ft} = \frac{W^M_{ft} M_{ft}}{Q_{ft}} \beta^M \exp(\varepsilon_{it})$.

As the manufacturers know that the latent manufacturer characteristics ξ_{ft} affect the utility of the suppliers, they take this into account when setting their leaf prices. In order to separately identify input demand and supply, an input demand shifter can be used as an instrument for input prices. I rely on manufacturing productivity ω_{ft}, which was estimated in the previous section, as an instrumental variable.[56] As productivity enters the input demand function, it is by definition relevant. The exclusion restriction is that the productivity term does not enter the supplier utility function, that is, that it is independent from the supply function residual $\xi_{ft} + \zeta_{ft}$, which includes both latent manufacturer 'attractiveness' ξ_{ft} and cigarette/leaf quality ζ_{ft}.[57]

$$\mathbb{E}[\xi_{ft} + \zeta_{ft} | \omega_{ft}, \mathbf{X}_{ft}] = 0$$

In other words, the exclusion restriction means three things. First, farmers do not care how efficient the manufacturing firms they are selling to are, conditional on how much they are paid and on observable manufacturer characteristics. Productivity differences between manufacturers can have many reasons, such as differences in managerial ability. As the farmers do not work at the manufacturers, but only interact with these firms through monetary transactions at the leaf markets, it seems reasonable that the farmers do not care about how productive their buyers are, conditional on the price they receive for their leaf. Second, product quality is conditionally independent from total factor productivity. This is a reasonable assumption for this industry, as cigarette quality is determined mainly by leaf quality, which does not enter the production function. Third, the observable manufacturer characteristics that enter the farmer supply function $\mathbf{X}_{ft}$, such as in which market it is located, are exogenous.

[54] Additional brand-level data reveal very little variation in leaf contents per cigarette across manufacturers. I discuss the consequences of leaf content heterogeneity in appendix C.6.

[55] Measurement error ε is also part of the inferred leaf price, but is assumed to be i.i.d. across firms and time.

[56] To be clear, I use the productivity term ω, not the residual after netting out the consolidation effects $\tilde{\omega}$.

[57] Productivity is by definition uncorrelated with the farmer-utility specific utility term ν_{jft}, which was already assumed to be i.i.d. across manufacturers and over time.

Threats to input supply identification

There are, nevertheless, multiple threats to the validity of the exclusion restriction, but I argue that they are not a first-order concern in the context of the Chinese tobacco industry. First, it could be that suppliers prefer to sell repeatedly to the same buyers. This is the case in many industries that are characterized by incomplete contracts or weak contract enforceability.[58] Another driver of interlocked relationships would be important search or switching costs on the seller's side.[59] In all these cases, sellers would prefer to sell to more productive buyers as these are less likely to exit in the future, even if they offer a lower price. In the tobacco industry setting, such repeated interaction seems not to be a concern of first-order importance, as leaf markets the form of frequent auctions rather than long-term contracts.[60] Moreover, as was already mentioned, exit is mainly driven by government policies, rather than by productivity differences in this industry.

A second threat to identification of the input supply curve concerns product differentiation. If manufacturers that are highly productive, in physical terms, choose higher quality, and hence more expensive, tobacco leaf, this would violate the exclusion restriction. This is the reason for assuming exogenous quality levels in assumption 1. I do, however, control for cigarette prices in the input supply function, which should pick up variation in cigarette quality. Moreover, using the brand-level data on product characteristics shows that the physical productivity of cigarette manufacturers does not correlate significantly with any product characteristic or quality indicator.[61] Finally, the variation in leaf cost shares could be due to endogenous quality choices, which were abstracted away from in the model. If this were true, then there would be both a leaf price markdown and a quality markdown. In order to reconcile falling leaf cost shares, quality would have had to drop sharply over the time period studied, while all consumer surveys report that Chinese cigarette quality improved over time (Hu, 2008).

[58]The literature on vertical relationships in developing countries has emphasized the importance of relational contracts and repeated interaction (Macchiavello, 2018).

[59]When 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.

[60]Especially in the Chinese setting, where leaf markets are highly regulated and geographically clustered in narrow locations (Peng, 1996; Wang, 2013). Even in the U.S., tobacco leaf contract have only been used since the early 2000s Dimitry (2003).

[61]This evidence is shown in panel c of appendix table A5.

Estimation

I estimate equation (9) using the manufacturing productivity residual Ω as an instruments for the leaf price W^M. I include three manufacturer characteristics in the vector $\mathbf{X}_{ft}$, which are likely to affect leaf supply. First, I control for cigarette prices, as they are a proxy for quality. Second, I control for manufacturer ownership types, as the Chinese tobacco industry is under political influence; farmers may therefore derive a different utility from selling to manufacturers that are state-owned rather than private. Finally, I include prefectural dummies to control for the geographical differences.

The outside option needs to be defined, meaning how many tobacco farmers could have been farming tobacco, but chose not to. As there is barely any crop switching towards or from tobacco leaf (Li et al., 2012), I model the outside option of tobacco farming as being employed in non-agricultural occupations. I therefore set the share of the population choosing the outside option as the share of the population that works in non-agricultural sectors, using the population census data. This share is 36.7% on average, as the tobacco farms are predominantly located in rural provinces.

Results

The estimates of the leaf supply function from equation (9) are in panel (b) of table 2. The coefficient on the leaf price, γ^W, is 2.23 and significantly above zero, so the leaf supply curve is upward-sloping. When not instrumenting, the coefficient estimate is negative, which shows that the manufacturer characteristics endogeneity problem matters. This difference mirrors the results in Berry et al. (2019). I discuss the magnitude of the supply estimate by transforming it into an input supply elasticity and markdown in the next section, where I also compare the input supply estimates to the related literature.

4.3 Markups, markdowns and consolidation

Markup and markdown distributions

Markups can be computed using equation (8a), as the output elasticity of labor β^L and leaf supply elasticity ψ^M are estimated, and the revenue shares α^L and α^M observed. I do not follow

De Loecker and Warzynski (2012) by netting out measurement error ε from the input revenue shares, as the dynamic panel estimator does not allow separate identification of productivity and measurement error in output. Leaf price markdowns can be calculated from the input supply elasticity, leaf prices and market shares, using equation (10).

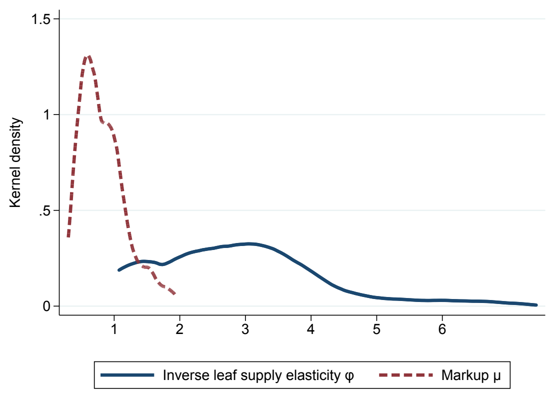
The markup and markdown distributions are in panel (a) of figure 4, with some selected moments in panel (b). The distributions are winsorized at the 1th and 99th percentiles on the graph, but the moments are based on the full distributions. The average markdown ratio is 3.318, which corresponds to a markdown wedge of 69.9%: the farmer selling to the average manufacturer receives 31.1% of his marginal revenue product. The standard error on the markdown estimate is 0.836, which implies that the markdown ratio lies between 1.68 and 4.96 with 90% confidence. This corresponds to a markdown wedge confidence interval between 40% and 80%.

Most of the related literature focuses on U.S. labor markets, and find much smaller markdowns. Both Berry et al. (2019) and Goolsbee and Syverson (2019) find a wedge δ^M of around 0.17 for U.S. online job board vacancies and tenure-track professors respectively, which is four times smaller. Ransom and Sims (2010) find a wedge of around 30% for Southern U.S. grocery clerks. Naidu et al. (2016), which studies immigrant construction workers in the United Arab Emirates, finds an average wedge of 50% for new recruits, which is already much closer to the one in this paper. The reason for these differences most likely relates to the level of frictions on local labor markets. As was discussed earlier, rural labor markets are highly frictional in China due to immigration restrictions and crop switching costs. The worse the outside employment options of farmers, the higher markdowns should be. I present evidence for this by regressing markdowns on local market characteristics in appendix E.6.

The markup ratio is on average 0.906, with a standard error of 0.233. Cigarette prices are hence on average roughly equal to manufacturing marginal costs. The markup ratio μ lies below one for more than half of the observations, which implies that these manufacturers sell to the wholesaler at prices below their marginal costs. As was explained in section 3.2, this does not imply negative variable profit margins: the variable profit margin is equal to the product of the markup μ and the markdown ψ^M. This product lies above one for 90% of the observations. The markdown distribution lies at the right of the markup distribution, which means that there is a larger gap between the leaf price and marginal costs, then between marginal costs and cigarette prices. In other words, the main profit source of manufacturers comes from pushing leaf prices down, rather than cigarette prices up. The markdown ratio is significantly higher than the markup ratio, which implies that manufacturers have more market power upstream than downstream. This is consistent with the fact that manufacturers buy from many small farmers on local markets, but sell to a monopsonistic wholesaler on an integrated national market, which presumably use their own buyer power to push cigarette prices towards manufacturer marginal costs.

Figure 4: Markups and markdowns

(a) Distribution



(b) Moments

	Markdown		Markup	
	Est.	S.E.	Est.	S.E.
Mean	3.318	(0.836)	0.906	(0.233)
Median	2.821	(0.656)	0.756	(0.175)
Obs.	1,086		1,086	

Notes: Bootstrapped standard errors are in parentheses.

The effects of consolidation on markups, markdowns and productivity

With the markup, markdown and productivity estimates at hand, I now feed these back into the difference-in-differences model in order to know how the consolidation policy affected market power, buyer power and productive efficiency, which are now the outcome variables y in equation (1). The identifying assumptions for this model are still the same as outlined in section 2.3. The consolidation treatment effect estimates are in panel (a) of table 3. I refer to appendix C.8 for the pre-trends. Markdowns increased by 27% for manufacturers affected by consolidation compared to manufacturers in the control group, and this increase is highly significant. The exit of the smaller manufacturers therefore mainly resulted in an increase in buyer power of the manufacturers. This was to be expected, as leaf markets were already imperfectly competitive with high leaf markdowns prior to the reform. Markdowns did not increase by the same amount for all manufacturers and markets. In appendix E.7, I show that markdowns rose more in areas where workers had fewer outside options due to lower educational attainment and higher unemployment rates.

Table 3: Consolidation treatment effects

	log(Markdown)		log(Markup)		log(Productivity)	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Treatment * 1(year≥2003)	0.239	(0.0588)	-0.0676	(0.0593)	0.0955	(0.0797)
Within R-squared	0.138		0.187		0.112	
Observations	1,086		1,086		1,086	

Notes: Controls include manufacturer fixed effects, a linear time trend, ownership type, product and exporting dummies.

Markups did, in contrast, fall by 7%; although this drop was not significant. As I show in appendix table A7, however, the markup fall was significant at the province-level, which is closer to the cigarette market scope than prefectures are. At first sight, such a fall in markups seems less intuitive than the rise of markdowns: an increase in product market concentration usually results in rising, not falling markups. One has to keep in mind, however, that the wholesaler is a monopsonist of its own, and can use its buyer power to push down factory-gate cigarette prices. As the manufacturers increased their profit margins due to their increased monopsony power over farmers, it is logical that the wholesaler strategically reacted to this by using its own monopsony power over manufacturers to push down factory-gate cigarette prices. In appendix E.4, I specify and estimate a bargaining model between the manufacturers and wholesaler that rationalizes why an increase in downstream market concentration led to lower, rather than higher, cigarette price markups.

Total factor productivity is estimated to have increased by 10%, but this increase is again not significant. The production function estimates already rejected increasing returns to scale, but the fact that the productivity residual did not change also shows that the consolidation did not lead to important scale economies. This does not mean that scale economies are absent in this industry: they could have been netted out by diseconomies of scale, such as increased transporation costs because of increased distances between factories and farms.[62] The average treatment effects are, finally, not informative about the *aggregate* effects of the consolidation policy, which I address in the next section.

[62]It is also possible that transportation costs were entirely paid by the farmers, but they would be priced into the leaf price, which could not explain falling leaf prices in consolidated leaf markets.

5 Aggregate consequences

5.1 Distributional consequences

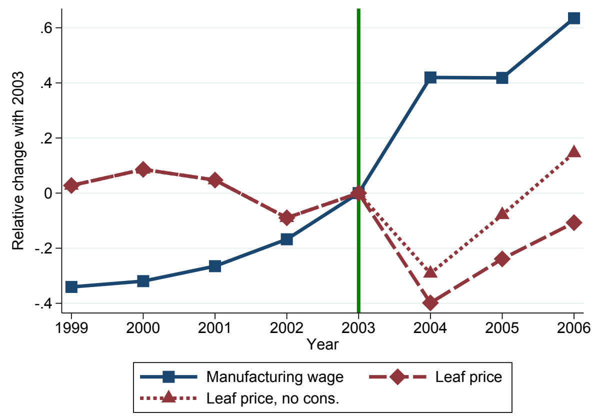
The analysis so far focused on the *average* effects of the consolidation within surviving manufacturers over time. In this section, I turn to the consequences of the policy at a more aggregate level. I start with the effects on the distribution of income between factors. By increasing buyer power on leaf markets, but not manufacturing labor markets, the consolidation of cigarette manufacturing contributed to income inequality between these two inputs. As farmers and factory workers are live mainly in rural and urban areas, respectively, this inequality margin relates to the urban-rural income gap, which has risen sharply in China over the past two decades (Yang, 1999; Ravallion and Chen, 2009). The tobacco industry was no exception to this evolution: while the average manufacturing wage grew by 14.5% per year between 1999 and 2006, average tobacco leaf prices fell by 5% per year. Accounting profits increased by no less than 24% per year for cigarette manufacturers.

I 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 manufacturers in the treatment group evolve in the same way as they did for manufacturers 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:

$$\tilde{W}_{ft}^M = \begin{cases} W_{ft}^M \exp(\theta_2^{\psi^M}) & \text{if} \quad t \geq 2003 \quad \& \quad C_f = 1 \\ W_{ft}^M & \text{otherwise} \end{cases} \tag{14}$$

Figure 5 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 5: Consolidation and income inequality



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.

The analysis above is by no means a counterfactual simulation, and comes with a nmuber of caveats. 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. Besides tobacco leaf prices, farm productivity and agricultural input costs matter as well for farm profits. 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), which was not enough to compensate falling leaf prices.[63]

[63] This fall in farmer profits is consistent with the micro-level evidence discussed earlier.

5.2 Productivity and output growth

Aggregate productivity growth

Even if the consolidation did not lead to productivity changes within manufacturers over time, the consolidation could affect aggregate productivity levels through input reallocation or through a reducation in fixed costs duplication. I compute prefecture-level aggregate productivity $\bar{\omega}_{it}$ by weighting manufacturer-level productivity by labor usage.

$$\bar{\omega}_{it} \equiv \sum_{f \in \mathcal{F}_{it}} \Big(\frac{\omega_{ft} L_{ft}}{\sum_{f \in \mathcal{F}_{it}} (L_{ft})} \Big)$$

Average prefectural productivity is denoted as $\hat{\omega}_{it}$, with $|\mathcal{F}_{it}|$ being the number of cigarette manufacturing firms in prefecture i and year t. The level of reallocation is the difference between aggregate and average productivity (Olley and Pakes, 1996).

$$\hat{\omega}_{it} \equiv \sum_{f \in \mathcal{F}_{it}} \Big(\frac{\omega_{ft}}{|\mathcal{F}_{it}|} \Big)$$

I estimate how both aggregate and average productivity were affected by the consolidation using equation (1) at the province-level. As shown in panel (a) table 4, aggregate productivity fell by 34%, which is in line with the distortionary effects of oligopsony power, but this drop was not statistically significant. Average productivity increased, in line with the consolidation results presented before, but again not significantly so.

Aggregate output growth

In the classical monopsony model presented throughout the paper, higher monopsony power should lead to lower equilibrium amounts of cigarettes produced. This is a testable implication of the model: I estimate how total cigarette production evolved differently between treatment and control markets in panel (b) of table 4. Total cigarette production fell on average by 24% in the treated provinces compared to the control provinces. Manufacturer-level cigarette production did not increase significantly on average, so the output decrease was solely due to the exit of the smaller manufacturers, whose disappearing output was not entirely reallocated to the incumbents as they

Table 4: Aggregate productivity and production

(a) Productivity	log(Aggregate TFP)		log(Average TFP)	
	Est.	S.E.	Est.	S.E.
Treatment * 1(year≥2003)	-0.331	(0.205)	0.148	(0.134)
Within R-squared	0.0211		0.0739	
Observations	767		767	

(b) Output	log(Total output)		log(Average output)	
	Est.	S.E.	Est.	S.E.
Treatment * 1(year≥2003)	-0.269	(0.0874)	0.0428	(0.193)
Within R-squared	0.0901		0.0261	
Observations	767		767	

Notes: Panel (a) compares the evolution of aggregate and average productivity between treatment and control groups, at the province level. Panel (b) does the same comparison for output.

produced less due to their increased monopsony power. This did not necessarily mean that Chinese consumers consumed less cigarettes, or that product market cigarette prices increased: there may have been a compensating effect through increased cigarette imports and/or illegal cigarette production. Both of these variables are unobserved, though, so these mechanisms cannot be verified.

6 Conclusion

In this paper, I examine the effects of changes in market structure on buyer power, product market power and productive efficiency. I discuss identification of markups, markdowns and productivity using production and cost data when a subset of inputs are non-substitutable. I use a structural model of input demand and supply to study the effects of a large-scale consolidation program in the Chinese tobaco industry that was aimed at spurring productivity growth. I find no strong evidence for such an increase in productivity, but find that the consolidation mainly led to increased oligopsony on rural factor markets. This increase in buyer power contributed to increased urban-rural income inequality to an important extent.

References

Ackerberg, D., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2017). Concentrating on the fall of the labor share. *American Economic Review: Papers and Proceedings 107*(5), 180–85.

Berger, D. W., K. F. Herkenhoff, and S. Mongey (2019). Labor market power. Technical report, National Bureau of Economic Research.

Berry, S., J. Azar, and I. Marinescu (2019). Estimating labor market power.

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

Blonigen, B. and J. Pierce (2016). Evidence for the effects of mergers on market power and efficiency. Working paper 22750, National Bureau of Economic Research.

Blundell, R. and S. Bond (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric reviews 19*(3), 321–340.

Braguinsky, S., A. Ohyama, T. Okazaki, and C. Syverson (2015). Acquisitions, productivity and profitability: Evidence from the Japanese cotton spinning industry. *The American Economic Review 105*(7), 2086–2119.

Brandt, L., J. Van Biesebroeck, and Y. Zhang (2012). Creative accounting or creative destruction? firm-level productivity growth in Chinese manufacturing. *Journal of Development Economics 97*(2), 339–351.

Brown, J. D., J. S. Earle, and A. Telegdy (2006). The productivity effects of privatization: Longitudinal estimates from Hungary, Romania, Russia, and Ukraine. *Journal of Political Economy 114*(1), 61–99.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 36*(S1), S13–S70.

Chatterjee, S. (2019). Market power and spatial competition in rural india. Working Paper 2019/04, Cambridge INET.

Chen, Y., M. Igami, M. Sawada, and M. Xiao (2018). Privatization and productivity in china. *Available at SSRN 2695933*.

Covarrubias, M., G. Gutiérrez, and T. Philippon (2019). From Good to Bad Concentration? US Industries over the past 30 years. Technical report, National Bureau of Economic Research.

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics 5*(3), 1–21.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics 135*(2), 561–644.

De Loecker, J., P. Goldberg, A. Khandelwal, and N. Pavcnik (2016). Prices, markups and trade reform. *Econometrica 84*(2), 445–510.

De Loecker, J. and P. T. Scott (2016). Estimating market power: evidence from the us brewing industry. Working Paper 22957, National Bureau of Economic Research.

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *The American Economic Review 102*(6), 2437–2471.

Dimitry, C. (2003). Contracting in tobacco? contracts revisited. *Tobacco Outlook 254*(01), 1–8.

Doraszelski, U. and J. Jaumandreu (2017). Measuring the bias of technological change. *Journal of Political Economy 126*(3), 1027–1084.

Doraszelski, U. and J. Jaumandreu (2019). Using cost minimization to estimate markups. Mimeo.

Fang, J., K. Lee, and N. Sejpal (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*, Volume 1810-0783. Food and Agriculture Organization of the United Nations.

FAO (2019). Faostat statistical database.

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, firm turnover, and efficiency: selection on productivity or profitability? *American Economic Review 98*(1), 394–425.

Ghandi, A., S. Navarro, and D. Rivers (2018). On the identification of gross output production functions.

Goolsbee, A. and C. Syverson (2019). Monopsony power in higher education: a tale of two tracks. Technical report, National Bureau of Economic Research.

Grieco, P., J. Pinkse, and M. Slade (2017). Brewed in North America: Mergers, marginal costs and efficiency. *International Journal of Industrial Organization 59*(2018), 24–64.

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.

Hsieh, C.-T. and Z. Song (2015). Grasp the large, let go of the small: The transformation of the state sector in China. Working Paper 21006, NBER.

Hu, T., Z. Mao, M. Ong, E. Tong, M. Tao, H. Jiang, K. Hammond, K. Smith, J. de Beyer, and A. Yurekli (2006). China at the crossroads: the economics of tobacco and health. *Tobacco Control 15*(1), 37–41.

Hu, T.-w. (2008, 01). Tobacco control policy analysis in china: Economics and health.

Jarosch, G., J. S. Nimczik, and I. Sorkin (2019). Granular search, market structure, and wages. Technical report, National Bureau of Economic Research.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies 70*(2), 317–341.

Li, V., Q. Wang, N. Xia, S. Tang, and C. Wang (2012). Tobacco crop substitution: pilot effort in China. *American Journal of Public Health 102*(9), 1660–1663.

Lu, Y. and L. Yu (2015). Trade liberalization and markup dispersion: evidence from China's WTO accession. *American Economic Journal: Applied Economics 7*(4), 221–253.

Macchiavello, R. (2018). 6 a mutually beneficial relationship: relational contracts in developing countries. *A Research Agenda for New Institutional Economics*, 53.

Miller, N. H. and M. C. Weinberg (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. Technical report.

Naidu, S., Y. Nyarko, and S.-Y. Wang (2016). Monopsony power in migrant labor markets. *Journal of Political Economy 124*(6), 1735–1792.

Nargis, N., R. Zheng, S. S. Xu, G. T. Fong, G. Feng, Y. Jiang, Y. Wang, and X. Hu (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.

O'Connor, R., Q. Li, W. E. Stephens, D. Hammond, T. Elton-Marshall, K. M. Cummings, G. A. Giovino, and G. T. Fong (2010). Cigarettes sold in China: Design, emissions and metals. *Tobacco Control 19*(2), 47–53.

Olley, S. and A. Pakes (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. and D. P. Sims (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. and S. Chen (2009). *China's (uneven) progress against poverty*, pp. 65–111. Routledge.

Rossi-Hansberg, E., P.-D. Sarte, and N. Trachter (2018). Diverging trends in national and local concentration. Technical report, National Bureau of Economic Research.

Song, F., J. Zhao, and S. M. Swinton (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. Technical report, 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. and J. M. Alston (1987). Substitutability for farm commodities: the demand for us tobacco in cigarette manufacturing. *American Journal of Agricultural Economics 69*(2), 258–265.

Syverson, C. (2019). Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives 33*(3), 23–43.

Talhout, R., P. A. Richter, I. Stepanov, C. V. Watson, and C. H. Watson (2018). Cigarette design features: effects on emission levels, user perception, and behavior. *Tobacco regulatory science 4*(1), 592–604.

Tortarolo, D. and R. Zarate (2018). Measuring imperfect competition in product and labor markets. Mimeo, UC Berkeley.

U.S. Census Bureau (1997). Cigarette manufacturing. 1997 economic census - industry series, US Census Bureau.

Wang, J. (2013). *State-Market Interaction in China's Reform Era: Local State Competition and Global Market-building in the Tobacco Industry*. Routledge Contemporary China Series. Routledge.

Yang, D. T. (1999). Urban-biased policies and rising income inequality in China. *American Economic Review: Papers and Proceedings 89*(2), 306–310.

Online appendix (not for publication)

A Data

A.1 Production and cost data

I use the NBS above-scale industrial survey (ASIF). I refer to Brandt et al. (2012) for a detailed description of the data. I keep all establishments with CIC codes 1610, 1620 and 1690. In theory, only CIC code 1620 includes cigarette manufacturers, but the product descriptions of the other CIC codes included also mostly contained cigarette manufacturers, which is the product that accounts for 95% of total revenue across all three CIC codes anyway.

A.2 Quantity data

Quantities are observed 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 manufacturers 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. Units are defined as cigarette cases, each of which contain 50,000 cigarettes (Fang et al., 2017). From 2004 onwards, the case unit definition changes, however. Fortunately, I observe both current and lagged quantities in each month. By comparing both, I 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 the quantity data are in Lu and Yu (2015).

A.3 County data

I retrieve county-level population data from the 2000 population census through the *Harvard Dataverse*[64]. The population census contains many variables, of which I use the total county population,

[64] https://dataverse.harvard.edu/dataverse/chinacensus

the unemployed population, the number of immigrants per county, and the population by educational attainment.

A.4 Product characteristics

I obtain brand-level cigarette characteristics from O'Connor et al. (2010) for a subset of manufacturing firms in 2009, such as the leaf content per cigarette and other characteristics which affect the smoking experience (Talhout et al., 2018). This dataset is observed only for 13% of the observations, but covers 29% of observations by revenue. I only use this data in an extension. I link the brands in O'Connor et al. (2010) to the manufacturers 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 Data cleaning and number of observations

The raw dataset contains 2,638 observations, with 508 establishments and 10 years (1998-2007). The data were cleaned in accordance with the procedures described in Brandt et al. (2012): I deflate all monetary variables (profits, revenues, intermediate input expenditure, wages, and export revenue) using the industry output and input deflators. As I study a single industry, however, this only affects the time-series variation in these variables, not their cross-sectional variation. I remove outliers in cigarette and leaf prices by winsorizing the 1st and 99th percentiles, and deleted observations with negative intermediate input expenditure. I also restricted the panel to 1999-2006, as quantities were not observed for 1998 and 2007. This cleaning reduces the dataset to 2,025 observations, covering 470 establishments over 8 years. Quantities, which are needed to estimate the production function, are reported for just 1,260 observations and 274 establishments. Together with the data cleaning, this reduces the sample size to 1,086 observations and 247 establishments. This selected sample covers 52% of total revenue in the raw dataset.

A.6 Summary statistics

Some summary statistics are in table A1. The average manufacturing 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.[65]

Table A1: 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.

B Alternative production function identification strategies

B.1 General discussion

Benefits of the dynamic panel estimator

In the main text, I relied on the dynamic panel estimator of Blundell and Bond (2000), with the timing assumptions of (Ackerberg et al., 2015) and (Olley and Pakes, 1996). For the setting of

[65] Harris (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.

the Chinese tobacco industry and for the main research questions of this paper, the dynamic panel estimator has two main benefits over the 'proxy variables' approaches of Olley and Pakes (1996); Levinsohn and Petrin (2003); Ackerberg et al. (2015). First, as it does not rely on inversion of the input demand function, the dynamic panel estimator does not require additional structural assumptions on the distribution of latent markups and markdowns. In the proxy estimator with input demand inversion, markup and markdown variation needs to be controlled for in the input demand function: high input demand can hence be due to either high productivity, low markups, low markdowns, or a combination of all three (Doraszelski and Jaumandreu, 2019). The proxy variable approach can cope with variation of markups and markdowns under certain models of product demand and input supply, if their drivers are included in the input demand conditions, but this requires additional parametric assumptions on the demand and supply functions (De Loecker et al., 2016).[66] Second, although the Ghandi et al. (2018) critique does not apply for the Leontief function used in this paper, materials need to be used as the invertible proxy variable. (Ackerberg et al., 2015) does not allow, however, for serially correlated shocks to material prices. As materials are endogenous in the context of this paper, serial correlation in material prices is very probable: manufacturers with high monopsony power today are likely to still have a lot of monopsony power in the next period. As material prices can be inferred from the Leontief function, including these prices in the input demand conditions can help, but these prices are not exogenous.

Drawbacks of the dynamic panel estimator

The dynamic panel estimator has, however, two main drawbacks compared to Ackerberg et al. (2015). First, it is harder to account for endogenous exit. Second, linearity is required for the transition equation of the productivity term, which is a strong parametric assumption. I re-estimate the model using ACF(2015) and compare its results to the dynamic panel estimator in the next section.

[66] For instance, with logit product demand and input supply functions, it is sufficient to include input and product market shares and prices in the input demand conditions.

ACF without first-stage inversion

Following the production function, equation (2), 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 would be observed, identification of the production function would not even require inversion of the input demand function. As was argued in Ackerberg et al. (2015), one could simply regress output quantities on material quantities, the residual of which would be measurement error ε_{ft}. This error could then be used directly to form the moment conditions. The problem is, however, that intermediate input quantities are latent in this paper. The leaf demand function was already used to back out leaf prices. The residual of the regression quantities on leaf expenditure hence contains both endogenous leaf prices and measurement error, which cannot be separately identified from each other.

B.2 Production function estimation using ACF (2015)

Identification and estimation

I now rely on the identification strategy of Ackerberg et al. (2015). I impose a different equation of motion for productivity than equation (12): rather than having a linear AR(1) model in productivity net of consolidation effects, $\tilde{\omega}$, I now include the consolidation dummies in the equation of motion for productivity ω, as in Braguinsky et al. (2015); De Loecker (2013):

$$\omega_{ft} = \rho g(\omega_{ft-1}, C_{ft}) + \upsilon_{ft}$$

Intermediate inputs are used as the flexible input for the first stage inversion. As derived in appendix E.1, leaf demand depends on cigarette prices, other inputs and their prices, on the consolidation treatment C_{ft} (as this affects productivity), , and input demand shifters $\mathbf{Z}$ 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. I control for leaf and cigarette market shares, which are measured at the prefecture and province-level respectively and denoted by the vector $\mathbf{s}_{ft}$. The first stage regression is given by equation (15), and is used to recover measurement error ε:

$$q_{ft} = \Phi_t(l_{ft}, m_{ft}, k_{ft}, C_{ft}, \mathbf{Z_{ft}}, w^m_{ft}, w^l_{ft}, p_{ft}, \mathbf{s}_{ft}) + \varepsilon_{ft} \tag{15}$$

Productivity can now be recovered as a function of data and parameters to be estimated:

$$\omega_{ft} = \hat{\phi}_{ft} - h(l_{ft}, k_{ft}, \boldsymbol{\beta})$$

The productivity innovation υ_{ft} is given by the difference between productivity and its expected value from the equation of motion.

$$\upsilon_{ft} = \omega_{ft} - \mathbb{E}\Big(\omega_{ft}|\omega_{ft-1}, C_{ft-1}\Big)$$

The moment conditions to identify $\boldsymbol{\beta}$ are, still assuming both labor and capital to be dynamic, pre-determined inputs, given by:

$$\mathbb{E}\big[\upsilon_{ft}(l_{ft}, k_{ft})\big] = 0$$

For estimation, I use the same Cobb-Douglas functional forms as those used in the main text. I use a third-order polynomial in the inputs for the first-stage regression.

Endogenous exit and input price bias

I also re-estimate two extensions of the model. First, I account for endogenous selection into exit by including estimated exit probabilities in the first-stage regression, as in Olley and Pakes (1996). Concretely, I regress an exit dummy $1(\text{exit}_{ft+1})$ on the log capital stock, exporting behavior, and ownership type dummies using a probit model. I include the fitted exit probabilities in $\Phi(.)$.

Second, I take into account the input price bias concern of De Loecker et al. (2016) by adding a price control to the production function. As explained in the main text, input price bias should not be a large problem in the context of the tobacco industry as the main differentiated input is leaf, which does not enter the production function, and as labor wages are observed and controlled for. The only problem can be that different capital goods are necessary to produce higher quality cigarettes. I therefore follow De Loecker et al. (2016) by adding a function $a(.)$ of log cigarette prices to equation (11), assuming that cigarette prices are monotonically increasing in quality.

$$q_{ft} = h(l_{ft}, k_{ft}, \boldsymbol{\beta}) + a(p_{ft}, \boldsymbol{\beta}) + \omega_{ft} + \varepsilon_{ft}$$

I use a linear function for the price term: $a(p_{ft}) = \beta_p p_{ft}$. The moment conditions are now also a bit different. As the price coefficient β_p needs to be idnetified as well, an additional instrument is needed. I assume prices can be changed flexibly, and therefore add lagged cigarette prices as an instrument in the moment conditions:

$$\mathbb{E}\big[\upsilon_{ft}(l_{ft}, k_{ft}, p_{ft-1})\big] = 0$$

Results

The estimated model parameters using ACF are in table A2. The first column contains the basic ACF model, the second column takes into account endogenous exit, and the third column adds the price control to the production function. The first two models yield nearly identical estimates for both the production function in panel (a), for markups and markdowns in panel (b), and for the consolidation treatment effects in panel (c). This shows that endogenous exit does not seem of a first order concern in this industry, which is reassuring for the dynamic panel estimator used in the main text. The output elasticities are somewhat higher than those using the dynamic panel estimator, and markdowns are higher and markups lower, but again not significantly so. Although output elasticities are lower in the dynamic panel estimator, markups are higher in that model: this is due to the fact that markdowns are estimated to be lower in the baseline model, which enter marginal costs. The consolidation treatment effects look very similar across both the ACF specifications and the dynamic panel model.

The model with the price control yields higher output elasticities, especially for capital, but these are very imprecisely estimated. Markdowns are also higher, but again with large standard errors. The consolidation treatment effects are again very similar and not significantly different to both the other ACF model specifications and the dynamic panel estimator.

Table A2: Structural model estimates using ACF(2015)

Model	ACF		ACF		ACF	
Endogenous exit	No		Yes		No	
Price control in P.F.	No		No		Yes	
(a): Production function	Est.	S.E.	Est.	S.E.	Est.	S.E.
Labor	0.442	(0.136)	0.439	(0.134)	0.451	(0.152)
Capital	0.624	(0.0841)	0.621	(0.0789)	0.928	(0.257)
Scale parameter	1.066	(0.143)	1.032	(0.134)	1.379	(0.367)
(b): Market power						
Markup (mean)	0.889	(0.148)	0.893	(0.146)	0.733	(0.153)
Markdown (mean)	3.615	(0.969)	3.584	(0.940)	4.944	(1.620)
(c): Treatment effects						
Markdown change	0.251	(0.0615)	0.250	(0.0591)	0.289	(0.0703)
Markup change	-0.0722	(0.0544)	-0.0715	(0.0537)	-0.102	(0.0654)
TFP change	0.109	(0.0844)	0.109	(0.0821)	0.132	(0.0933)

Notes: The first column uses the estimation procedure of Ackerberg et al. (2015). The second column adds estimated exit probabilities in the first stage of ACF(2015), following Olley and Pakes (1996). The third column adds a price control in the production function, as in De Loecker et al. (2016).

C Revisiting the assumptions

C.1 Intermediate input substitutability

Estimating the elasticity of input substitution

Throughout the paper, it has been assumed that tobacco leaf cannot be substituted with either labor or capital. Tobacco leaf may be substitutable to a limited extent with capital, for instance due to waste reducing technologies Another reason why intermediate inputs could be substitutable with labor, even if leaf and labor are non-substitutable, would be vertical integration between cigarette factories and farms.[67] The elasticity of substitution between tobacco leaf and the other inputs can,

[67] This is, however, not a feature of the Chinese tobacco industry (Peng, 1996; FAO, 2003; Wang, 2013).

however, be estimated. Let the cigarette production no longer take the Leontief form from equation (2), but the following CES production function instead:

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

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 (16a). Manufacturers use relatively more labor compared to tobacco leaf if wages are lower, if the output elasticity of labor is relatively higher, or if manufacturers have more monopsony power over tobacco leaf.

$$l_{ft} - m_{ft} = \sigma^M \left(w_{ft}^M - w_{ft}^L \right) - \sigma^M \left(\ln(\beta^M) - \ln(\beta^L) \right) + \sigma^M \ln(1 + \psi_{ft}^M) \tag{16a}$$

Leaf prices can no longer be recovered from the Leontief production function, and are hence latent. Equation (16a) hence has to be estimated using intermediate input expenditure $W^M M$, using equation (16b).

$$l_{ft} - m_{ft} - w_{ft}^M = -\sigma^M w_{ft}^L + (\sigma^M - 1) w_{ft}^M - \sigma^M \left(\ln(\beta^M) - \ln(\beta^L) \right) + \sigma^M \ln(1 + \psi_{ft}^M) \tag{16b}$$

The only observed variable in the right-hand side of equation (16b) is the log labor wage w^L. Estimating this equation is subject to two types of endogeneity bias. First, the extent of oligopsony power over tobacco leaf affects optimal input demand, as explained earlier. Second, variation in intermediate input prices due to reasons other than oligopsony power now enters the residual of equation 16b, and is by definition correlated with intermediate input expenditure. An instrument for labor usage is hence needed. I rely on the average export share of revenue and average export participation in other industries than tobacco manufacturing in the same county as an instrument for wages. Wages increased by much more in areas affected by increasing exports compared to areas with less trade penetration during this time period. The exclusion restriction is that export participation and behavior in other manufacturing industries did not affect either leaf market monopsony power or the production function coefficients in the cigarette manufacturing industry. This seems fine, as leaf markets are domestic, and as productivity was assumed to be Hicks-neutral anyway.

The substitution elasticity estimates are in panel (d) of table A3. The estimated elasticity of substitution between intermediate inputs and labor is 0.0424, and not significantly different from zero, but significantly different from one. This supports the Leontief model used throughout the paper over the traditional gross output Cobb-Douglas production function in labor, capital and materials. Next, I estimate the elasticity of substitution between labor and capital as well by estimating equation (16a) with capital instead of tobacco leaf on the left-hand side. This requires the additional assumption that both labor and capital are variable. This elasticity is estimated to be 0.927, and is significantly above zero but not significantly different from one. Hence, the Cobb-Douglas substitution between labor and capital which was imposed in the baseline model cannot be rejected. Sumner and Alston (1987) found a different substitution elasticity which was greater than zero, but used 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.

Substitutable leaf model

Suppose a gross output Cobb-Douglas production function in labor, capital and materials would have been used, as in equation (17a). How would this affect the markup and markdown estimates compared to the baseline model in which materials are non-substitutable?

$$q_{ft} = \beta^M m_{ft} + \beta^L l_{ft} + \beta^K k_{ft} + \omega_{ft} + \varepsilon_{ft} \tag{17a}$$

If all inputs are substitutable, the markup is given by equation (8b): $\mu_{ft} = \dfrac{\beta^L}{\alpha^L}$. The leaf price markdown ψ_{ft}^M is equal to the ratio of the markup of the variable of which the price is exogenous over the markup of the input of which the price is endogenous (Morlacco, 2017):

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

It is clear that the markup μ and markdown ψ^M from the Cobb-Douglas model are different from the markup and markdown expressions in the Leontief model. The direction of the bias is not obvious. The Cobb-Douglas markup μ_{ft} is an overestimate of the true Leontief markup μ_{ft} if:

Table A3: Alternative production models

Model:	Substitutable leaf		Translog in L,K		Labor-augm. prod.	
(a) Output elasticities:	Est.	S.E.	Est.	S.E.	Est.	S.E.
β^L	0.169	(0.116)	0.282	(0.309)	0.114	(0.00102)
β^K	0.379	(0.127)	0.535	(0.532)	0.886	(0.00102)
β^M	0.296	(0.0404)				
(b) Markups and markdowns:						
ψ^M (average)	0.431	(1.480)	3.454	(2.797)	4.212	(1.777)
μ (average)	5.154	(3.650)	0.854	(0.726)	0.621	(0.0688)
(c) Consolidation treatment effects:						
θ^{ψ^M}	0.274	(0.0857)	0.245	(0.319)	0.270	(0.0494)
θ^{μ}	-0.0613	(0.0533)	-0.0765	(0.784)	-0.116	(0.0441)
θ^{ω}	0.208	(0.0679)	0.0842	(0.0831)	0.0596	(0.105)
θ^{β^L}					-0.0601	(0.0742)
(d) Input substitutability:						
σ^M	0.0424	(0.290)				
σ^K	0.927	(0.215)				

Notes: The first two columns use a Cobb-Douglas model in leaf, labor and capital. The middle columns use a translog model in labor and capital. The right two columns allow for the labor coefficient to vary flexibly. Panels (a)-(c) report the production and supply model coefficients and markup and markdown moments, while panel (d) shows the estimates consolidation treatment effects.

$$\mu_{ft} \geq \mu_{ft} \Leftrightarrow \beta^L > (\frac{1}{\beta^L} - \frac{\alpha^M_{ft}}{\alpha^L_{ft}}\psi^M_{ft})$$

If the estimated output elasticities of labor are the same for the Leontief and Cobb-Douglas models, then the markup from the Cobb-Douglas model always overestimates the markup from the Leontief model. The reason for this is that the Cobb-Douglas model does not take into account that the marginal cost of labor also depends on the input price elasticity of materials due to the complementarity between labor and materials. Marginal costs are hence underestimated, and markups overestimated. The estimated labor output elasticity will, however, not be the same: the

Cobb-Douglas will most likely yield a lower output elasticity of labor compared to the Leontief model, as materials are controlled for in the production function. This leads to a lower markup estimate. Whether the markup from the Cobb-Douglas model is an over- or underestimate of the actual markup depends on which of both biases dominates.

For the concrete setting of Chinese tobacco, the markup and markdown estimates using the Cobb-Douglas model are given in column 1 of panel (b) in table A3. The markup is estimated to be higher than 5, while the markdown ratio is estimated to be below one. Both estimates have extremely large standard errors, though, because the output elasticities in panel (a) are imprecisely estimated. These markup and markdown estimates are unrealistic for the Chinese setting: they imply that farmers have a lot of bargaining power, and end up being paid well above their marginal product, and that manufacturers have high market power on the cigarette markets, while the monopsonistic wholesaler has no buyer power.

Next, I compare the consolidation treatment estimates when using the production function with substitutable leaves from equation (17a). The estimated treatment effects are in the first column of panel (c) in table A3. The markup and markdown are estimated to have increased by 32% and decreased by 6% in response to consolidation, which are slightly larger effects compared to the Leontief model. The sign of the effects is similar, though: 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, very different. Productivity is estimated to have increased by 23% with the substitutable inputs model, and this increase is signficant, unlike the one in the Leontief model. The reason for this difference is that input prices and input quantities are not separately observed. The empirical production function is therefore not equation (17a), but equation (17b), with material expenditure rather than material quantities on the right-hand side. A drop in latent intermediate input prices due to increased monopsony power will be interpreted as rising productivity in the substitutable leaf model.[68]

$$q_{ft} = \beta^L l_{ft} + \beta^K k_{ft} + \beta^M (m_{ft} + w_{ft}^M) + \omega_{ft} + \varepsilon_{ft} \tag{17b}$$

[68] De Loecker and Scott (2016) 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 rather than input quality variation.

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 et al., 2006; Hsieh and Song, 2015; Chen et al., 2018). These profitability gains could be due to both increased monopsony power or TFP growth.

C.2 Translog production function

Throughout the main text, I used a Cobb-Douglas specification for the labor-capital term $H(.)$ 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 $H(.)$ as a robustness check. The corresponding functional form of h(.) 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 identification strategy follows Ackerberg et al. (2015) and is described in appendix **??**, with the only difference being that the moment conditions are now given by:

$$\mathbb{E}\Big\{\Delta\omega_{ft}(\beta^L, \beta^K, \beta^{LK}, \beta^{L2}, \beta^{K2}) \begin{pmatrix} l_{ft-1} \\ k_{ft} \\ l_{ft-1}k_{ft} \\ l_{ft-1}^2 \\ k_{ft}^2 \end{pmatrix} \Big\} = 0$$

Columns 3-4 of A3 contain the estimated coefficients of interests using the translog model. All results are very similar to the baseline model.

C.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? I redefine the production function to allow for labor-specific productivity β_{ft}^L in equation (17c). This is the Cobb-Douglas version of the production function used in Doraszelski and Jaumandreu (2017).

$$q_{ft} = \tilde{\beta}^L_{ft} l_{ft} + \tilde{\beta}^K_{ft} k_{ft} + \tilde{\omega}_{ft} \tag{17c}$$

Latent variation in the output elasticity of labor leads to biased estimates of the markup: variation in factor revenue shares can be due to variation in markups or output elasticities. In the context of this paper, labor-augmenting productivity difference sare 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 assume that capital is variable and that there are constant returns to scale, meaning that $\beta^L_{ft} + \beta^K_{ft} = 1$. This implies that the output elasticities of labor and capital are equal to their cost shares (not taking into account intermediate input expenditure), as in Foster et al. (2008).[69] Because there is measurement error, I average output elasticities of labor and capital by year t and prefecture i. Denoting the set of manufacturers in a leaf market as $\mathcal{F}_i$ and investment as I, this means that output elasticities are given by: As this paper aims to separately identify the effects of consolidation on markups, markdowns and productivity, however, imposing constant returns to scale would answer part of this question through a functional form assumption, which is not appealing. This is why the cost shares approach was not taken as the baseline identification strategy, together with the fact that capital is not a variable input.

$$\begin{cases} \beta^L_{it} = \frac{1}{|\mathcal{F}_i|} \sum_{f \in \mathcal{F}_r} \left(\frac{W^L_{ft} L_{ft}}{W^L_{ft} L_{ft} + I_{ft}} \right) \\ \beta^K_{it} = \frac{1}{|\mathcal{F}_i|} \sum_{f \in \mathcal{F}_r} \left(\frac{I_{ft}}{W^L_{ft} L_{ft} + I_{ft}} \right) \end{cases}$$

The resulting output elasticities and markups and markdowns are in columns 5 and 6 of table A3. Panel (c) shows that the treatment effects on both markdowns are markups are similar to the baseline model, although the markup drop is now larger and statistically significant. Labor-augmenting productivity, as measured by the flexible labor elasticity β^L_{ft}, did not change significantly in response to the consolidation.

[69] This would not hold if there would be monopsony power over labor or capital, but this was already ruled out.

C.4 Nested logit leaf supply model

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

$$U_{jft} = \underbrace{\gamma^W W^M_{ft} + \gamma^X \mathbf{X}_{ft} + \xi_{jt}}_{\delta_{ft}} + (1-\sigma)\nu_{jft}$$

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

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

with $D_{gt} \equiv \sum_{f \in \mathcal{F}^g_{it}} \exp(\frac{\delta_{ft}}{1-\sigma})$ I define markets as provinces and sub-markets as prefectures. The markdown is now expressed as:

$$\psi^M_{ft} \equiv \Big(\frac{\partial S_{ft}}{\partial W^M_{ft}} \frac{W^M_{ft}}{S_{ft}}\Big)^{-1} + 1 = \left(\gamma^W W^M_{ft}\Big(\frac{1}{1-\sigma} - \frac{\sigma}{1-\sigma} S_{fgt} - S_{ft}\Big)\right)^{-1} + 1$$

The results are in the first column of table A4. The elasticity of substitution between selling inside or outside the own prefecture is 0.510 and significantly different from both 0 and 1. From the farmer's point of view, different prefectures are hence neither perfect substitutes or complements. This most likely stems from transport costs and from the internal leaf trade regulations. The markdowns and markups are respectively lower and higher compared to the baseline model, at 2.2 and 1.2. There is, however, a very large standard error on both market power estimates, as the dataset may be too small and/or the instruments too weak to identify the nested logit model and the production function jointly. The consolidation treatment effects are, however, still much in line with those in the baseline model: markdowns rise significantly by around 20% while markups fall slightly, but not significantly so at the prefecture level.

C.5 Endogenous manufacturing wages

Throughout the text, manufacturing wages were assumed to be exogenous. This assumption can be tested by estimating the slope of the labor supply function. I adapt the leaf supply function (9) to labor, denoting log labor market shares as s^L_{ft} and the idiosyncratic labor supply shocks as ξ^L_{ft}:

$$s^L_{ft} - s^L_{0t} = \gamma^L W^L_{ft} + \boldsymbol{\gamma}^{XL}\mathbf{X}_{ft} + \xi^L_{ft}$$

I use the same specification and instruments as for tobacco leaf. The results are in the third column of table A4. The estimated elasticity of labor supply is very close to zero and statistically insignificant. This supports the assumption that manufacturing labor wages are exogenous in this industry. This is reasonable given the industry setting: manufacturing workers are based in urban areas, and hence more mobile than farmers who have rural *Hukou* permits. Their skills are also much less specific to cigarette manufacturing compared to tobacco farmers, who can only sell their good to this industry, or switch crops, which was already argued to be very costly.

C.6 Heterogeneous intermediate input requirements

Variation in product characteristics

I revisit the assumption that the tobacco leaf content per cigarette, β^M_{ft}, was homogeneous. The product characteristics data shows that variation in tobacco concentration and other cigarette characteristics, such as ventilation rates and paper quality, was very limited across manufacturer. As shown in table A1, the average manufacturer uses 681 mg of tobacco leaf per cigarette of 1000 mg, and the standard deviation of this content is merely 31 mg. The entire distribution of leaf contents lies between 630 and 750 mg. This range is much too small to explain the observed decline in the leaf share of revenue. Moreover, as long as product characteristics were similar between the control and treatment groups, they do not affect the difference-in-difference estimates. Panel (a) of table A5 compares all product characteristics between the treatment and control groups. Both groups did not differ in any of the observable characteristics, and barely any of the variation in product characteristics is explained by the treatment dummies.

Table A4: Alternative input supply models

Model:	Nested logit		Labor monopsony	
(a): Input supply	Est.	S.E.	Est.	S.E.
Leaf price	2.614	(0.742)		
Nesting elasticity	0.510	(0.175)		
Wage			0.0465	(0.224)
(b): Markdowns and markups				
Leaf markdown ψ^M	2.215	(8.563)		
Markup μ	1.170	(0.412)		
(c): Consolidation treatment effects				
Markdown change θ^{ψ^M}	0.185	(0.0790)		
Markup change θ^{μ}	-0.0329	(0.151)		

Notes: In the first column, I estimate a nested logit leaf supply model with prefectural nests. In the second column, I estimate the manufacturing labor supply function using TFP as instrument for wages. In the third column, I regress a log-linear supply model using log(TFP) as instrument for log leaf prices and the same controls as before.

Correlations with markups, markdowns and productivity

The second table in table A5 compares how markups, markdowns and productivity correlate with the product characteristics. Markups do not correlate significantly with any of the characteristics, while markdowns do. Manufacturers with higher leaf contents per cigarette have lower markdowns. The reason for this is that the leaf price $W^M = \frac{M}{Q}$ was measured to be higher for manufacturers with a higher leaf content β_f^M. This higher leaf price is wrongly interpreted as being due to a lower markdown in the model with homogeneous leaf contents. The last column shows the correlation between TFP and product characteristics. Manufacturers with a higher leaf content per cigarette again have lower TFP estimates. Higher leaf contents are likely to be correlated with higher labor and capital usage as well, due to unobserved product quality differences, which explains the lower productivity estimate in the quantity production function.

Table A5: Product characteristics

(a) Comparisons	log(Leaf weight)		log(Filter density)		log(Rod density)	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Treatment	0.00253	(0.0171)	0.0106	(0.0141)	-0.00706	(0.0160)
Observations	289		289		289	
R-squared	0.0008		0.0239		0.0062	

	log(Paper permeability)		log(Ventilation)	
	Estimate:	SE:	Estimate:	SE:
Treatment	0.000275	(0.0501)	-1.979	(1.984)
Observations	289		289	
R-squared	<0.0001		0.0290	

(b) Correlations	log(Markup)		log(Markdown)		log(TFP)	
	Estimate:	SE:	Estimate:	SE:	Estimate:	SE:
log(Leaf weight)	1.746	(1.419)	-2.145	(0.867)	-6.466	(1.328)
Ventilation	0.00307	(0.00551)	0.00784	(0.00410)	0.0321	(0.00633)
log(Rod density)	-0.460	(1.384)	3.006	(0.903)	5.705	(1.645)
log(Filter density)	0.786	(1.335)	0.0844	(0.762)	-0.365	(1.760)
Paper permeability	0.00187	(0.315)	-0.148	(0.270)	-0.967	(0.762)
Observations	168		168		168	
R-squared	0.0311		0.0759		0.133	

Notes: Panel (a) compares the cigarette contents between the treatment and control groups. Panel (b) reports the correlations between markups, markdowns, productivity and cigarette characteristics.

C.7 Non-profit maximizing firms

Throughout the paper, it was assumed that manufacturers maximize per-period profits, as stated in assumption 2. Chinese firms, and especially those that are state-owned, may have different objectives. Achieving ‘social stability’ through high and countercyclical employment is, for instance, frequently mentioned in the literature (Li et al., 2012). As was discussed in the main text, various industry sources confirm that the cigarette manufacturers compete against each other on their input markets, and have incentives to achieve lower costs. Nevertheless, I now discuss two ways in which such size objectives can enter the manufacturer’s profit function and how this affects the estimates throughout the paper.

Output size objective

A first deviation from profit maximization could be that manufacturers value being large, and are willing to sacrifice some profits to achieve this larger output. Such an objective changes marginal costs MC_{ft}: the additional cost of producing more is lower if manufacturers value being large. Let the altered marginal costs be denoted $\hat{M}C_{ft} = \frac{MC_{ft}}{\varsigma_{ft}}$. Manufacturers with a larger preference for producing a lot have a larger parameter ς_{ft}. Consistently with equation (6), the markup μ_{ft} is now given by $\mu_{ft} = \varsigma_{ft}\dfrac{P_{ft}}{MC_{ft}}$. If manufacturers 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 logic holds for the markdown ψ^M_{ft}. If manufacturers 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 could also reflect size preferences.

Input size objective

Now suppose that manufacturers 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 $\hat{W}^L_{ft}$ is different from the measured input price. If manufacturers value employing many workers, the implicit wage is lower than the observed wage, so $\varsigma^L_{ft} < 1$: $\hat{W}^L_{ft} = W^L_{ft}\varsigma^L_{ft}$. As manufacturers do not choose labor and tobacco leaf separately, this has the same effects on markup and markdown estimates as a different marginal cost MC. Marginal costs are linear in both input prices, as shown in appendix E.1.

Interpretation of the consolidation treatment effects

Firm objectives that diverge from static profit maximization lead to biased markup and markdown estimates. To which extent is this problematic for the consolidation treatment effects described in the paper? First of all, even if manufacturers have an objective function that is not entirely consistent with profit maximization, it is only differences between manufacturers that matter. As 98% of the market is under some type of state control, large objective differences are not that likely. Moreover, both ownership type dummies and the equity structure of manufacturers were controlled for in all regressions of interest, without changing the results much. Secondly, even if manufacturers differ in their objectives, this is not a problem if their objective function is stable over time: such differences get absorbed in the manufacturer fixed effects that were included in the difference-in-differences model. Finally, even if manufacturers' objectives change over time, this is fine as long as changes in manufacturers' objectives are uncorrelated to the consolidation treatment. It is unclear why exit of competing manufacturers would change the objective function of the incumbents.

C.8 Difference-in-differences model assumptions

Two assumptions need to be fulfilled for the difference-in-differences model to be valid. First, the error terms υ_{ft}^{y} has to be conditionally independent from the consolidation dummy $C_f\mathbb{I}[t \geq 2003]$ for each outcome variable $y \in \{\mu, \psi^M, \omega\}$. The error term υ^y contains all time series variation in markups, markdowns and productivity that is not captured by the other control variables. The selection into markets with competitors below the exit threshold should therefore be unrelated to the evolution of markups, markdowns and productivity within manufacturers over time. Second, the trends in markups, markdowns and productivity need to be parallel for both treatment and control groups in the absence of the treatment.

Pre-trends and announcement effects

I verify whether the trends in the dependent variables prior to the reform were parallel by estimating equation (18) 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.

$$y_{ft} = \eta_1^y C_f + \eta_2^y C_f * t + \eta_3^y t + \nu_{ft}^y \quad \text{if} \quad t < 2003 \tag{18}$$

The results are in panel (a) of table A6. For the time period 1999-2002, the markup and markdown trends were not parallel for the treatment and control groups. They were parallel until 2001, however. Markups and markdowns started changing already in 2002, while the size thresholds were imposed only starting in 2003. The consolidation policy was already announced by the STMA on May 2, 2002, which may have led to changed pricing behavior by the manufacturers and wholesaler before small manufacturers actually exited. The pre-trends in productivity are parallel on the entire pre-treatment period.

Treatment and control group comparison

Panel (b) of table A6 shows that the control group contains 14% of manufacturers when defining leaf markets at the province-level. The control group becomes larger when defining narrower leaf markets. One half of all manufacturers before 2003 produced below the exit threshold, but they were responsible for only 8% of industry revenue. Panel (c) of table A6 shows that manufacturers in the treatment and control groups were not significantly different in terms of leaf and cigarette prices and size before 2003.

Table A6: Treatment and control groups before policy

(a) Pre-trends	log(Markup)		log(Markdown)		log(TFP)	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Treatment*year (1999-2002)	0.0863	(0.0318)	-0.0363	(0.0259)	-0.0924	(0.0495)
Treatment*year (1999-2001)	0.0288	(0.0369)	-0.0898	(0.480)	-0.110	(0.0676)

(b) Treatment and exiters group size	% Firms	% Revenue	% Output
Treatment group	38.8	38.3	40.4
Firms with $Q < 100K$	50.4	8.1	4.3

(c) Alternative market definitions	% Firms	% Revenue	% Output
Treatment (province)	86.1	89.2	87.7
Treatment (county)	14.9	11.7	15.9

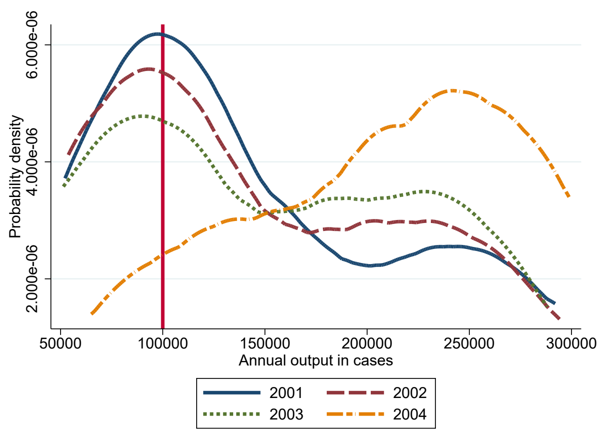
(d) Observable characteristics	log(Cigarette price)		log(Wage)		log(Quantity)	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Treatment group dummy	-0.0758	(0.0641)	0.0641	(0.0704)	0.0355	(0.0915)

Notes: Panel (a) compares the pre-trends between treatment and control groups before 2002 and 2003. Panel (b) reports the treatment group sizes for different market definitions. Panel (c) compares the treatment and control groups on the period 1999-2002.

Self-selection into treatment and control groups

The difference-in-differences assumptions would be violated if manufacturers could self-select into producing below or above 100,000 cases. If this were to be the case, then there should be 'bunching' in the firm-size distribution just above the exit threshold. Figure A1 shows this is not the case. Moreover, most manufacturers were distributed close to the exit threshold of 100,000 cases before 2003, which makes it more realistic that these firms were comparable.

Figure A1: Firm size distribution by year



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.

C.9 Dynamic leaf supply and demand

Dynamic leaf supply

The estimated supply elasticities are 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.[70] 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

[70]Source: https://www.pmi.com/glossary-section/glossary/tobacco-curing

rise immediately after the exit threshold enforcement, but fall again later, as farmers start exiting. In panel (e) of table A7, I limit the time frame to end in 2005 and 2004, in order to compare the short- to long-term effects of the consolidation. The estimates in the first column show that markdowns in consolidated markets rose during the three years following the reform, rather just rising initially and 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.

Dynamic leaf demand

If manufacturers hold leaf inventories, then leaf demand would be dynamic, and not only depend on current leaf prices, but also on expected future leaf prices. This would change both the markup estimates, and the effects of the consolidation on markups. Inventories can, however, only affect short-run markup fluctuations. The estimates in the second column of panel (d) in table A7 show that the markup reaction to the consolidation also increased in size with the time since the consolidation. If the estimated effects would merely be due to changes in inventories as a reaction to an expected price change, then markups should only fall temporarily in reaction to the consolidation. In order to establish the longer-term effects of the consolidation, though, a longer panel would be needed.

D Robustness checks

D.1 Alternative difference-in-difference model specifications

Alternative leaf market definitions

Leaf markets were defined at the prefecture level throughout the paper. In table panel (a) of table A7, I redefine leaf markets at the province and county levels. The consolidation treatment effect for markdowns is estimates to be 45% at the county level, and 35% at the province level: the prefecture-level estimates are thus the most conservative ones, but the pattern of increasing markdowns holds across various market definitions. The estimated markup effect is very similar at the county-level, but much larger at the province level, where markups dropped by 21%. As was explained in the main text, the drop in markups can be rationalized by monopsony power of the wholesaler. The fact cigarette markets extend across multiple cities could be the reason why the markup effect is noticeable only at the more aggregate province-level. The productivity effects of the consolidation are, finally, very similar across the different market definitions.

Continuous treatment measures

Throughout the main text, a dummy variable was used to indicate the treatment groups, namely the presence of manufacturers 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 manufacturers in a market producing below the threshold in 2002 as a treatment measure. Second, I weight these manufacturers by employment, rather than taking unweighted averages. Besides the treatment indicator definition, I keep all model specifications fixed. The results are in panel (b) of table A7. The average share of firms below the treshold was 22%. The interpretation of the estimates is therefore that markdowns were around 7% higher in prefectures with twice as many firms below the exit threshold after 2003, compared to the average prefecture. The results are consistent with the baseline regression: increasing markdowns in consolidated markets, decreasing markups and increasing productivity, although insignificantly for the last two variables. The results in the specification in which the employement share below the exit threshold is taken instead yields very similar results, except for the productivity change coefficient, which becomes very close to zero.

Firm and year fixed effects

In the baseline model, manufacturing firm fixed effects were included as controls, while year fixed effects were not. A linear time trend was used instead. In panel (c) of table A7, I use two different specifications. First, I no longer control for manufacturer dummies. The coefficients are similar to the model with manufacturer fixed effects, except that the drop in markups is less pronounced. Next, I include both manufacturer 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 (d) of table A7, 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

Table A7: Robustness checks

Consolidation treatment effect for:	log(Markdown)		log(Markup)		log(Productivity)	
(a) Market definitions	Est.	S.E.	Est.	S.E.	Est.	S.E.
Provinces	0.298	(0.0678)	-0.231	(0.0664)	0.0991	(0.0784)
Prefectures	0.239	(0.0588)	-0.0676	(0.0593)	0.0955	(0.0767)
Counties	0.373	(0.0789)	-0.0852	(0.0965)	0.0804	(0.112)
(b) Different treatment measures						
Share of manufacturer under 100K	0.488	(0.110)	-0.135	(0.114)	0.147	(0.161)
Employment share under 100K	0.403	(0.107)	-0.118	(0.145)	<0.001	(0.216)
(c) Different fixed effects						
No manufacturer fixed effects	0.212	(0.0560)	0.0158	(0.0528)	0.135	(0.0862)
Adding year fixed effects	0.172	(0.0527)	0.0536	(0.0512)	0.0985	(0.0817)
(d) Different moments						
Median	0.126	(0.0379)	-0.0758	(0.0474)	0.199	(0.0709)
Employment-weighted average	0.164	(0.0505)	-0.0114	(0.0434)	0.135	(0.0751)
(e) Different sample sizes						
Time frame = 1999-2004	0.156	(0.0682)	0.00467	(0.0640)	-0.0043	(0.0877)
Time frame = 1999-2005	0.204	(0.0632)	-0.0374	(0.0572)	0.0670	(0.0805)
(f) Exporting behavior	Export dummy		log(Exp./Rev.)			
	Est.	S.E:	Est.	S.E:		
Treatment * 1(year≥2003)	<0.001	(<0.001)	-0.0112	(0.00629)		

Notes: Panels (a)-(f) report the consolidation treatment effect θ_2 of equation (1) for a variety of robustness checks. The regressor is treatment * 1(year≥2003) in all cases. Regressands are log markups, markdowns and productivity in panels (a)-(e), and an export dummy and the log export share of revenue in panel (f).

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.

D.2 Exporting behavior

The Chinese economy globalized rapidly throughout the 1990s and 2000s, which could be a confounding factor as globalization can affect both productivity, market power and input reallocation. As was already argued, the tobacco industry remained largely domestic: less than 1% of total industry revenue came from exports. For the 16% of manufacturers who do export, exports represent merely 6% of their revenues on average. In panel (f) of table A7, I test whether exporting behavior changed through the consolidation, which it did not.

E Additional results

E.1 Derivations

Markup expression

I derive the markup expression in equation (8a). Marginal costs MC_{ft} can be expressed as follows:

$$MC_{ft} = W^L_{ft}\frac{\partial L_{ft}}{\partial Q_{ft}} + \frac{W^M_{ft}M_{ft}}{Q_{ft}}\psi^M_{ft}$$

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

$$MC_{ft} = \frac{\alpha^L_{ft}P_{ft}}{\beta^L} + \alpha^M_{ft}P_{ft}\psi^M_{ft}$$

Finally, dividing prices by marginal costs yields equation (8a).

Markdown expression

Next, I derive the markdown formula in equation (7). Consider the simplified model in which manufacturers choose input quantities to maximize profits, and in which there is only one input V:

$$\max_{V_{ft}}[P_{ft}(Q_{ft})Q_{ft}(V_{ft}) - W^V_{ft}(V_{ft})V_{ft}]$$

The first order conditions imply that the marginal revenue product is equal to marginal costs:

$$\underbrace{\frac{\partial(P_{ft}(Q_{ft})Q(V_{ft})}{\partial V_{ft}}}_{\text{Marginal revenue product}} = \underbrace{W_{ft}^{V}\psi_{ft}^{V}}_{\text{Marginal cost}}$$

Rearranging terms shows that the inverse input supply elasticity ψ^V is equal to the ratio of the marginal revenue product of V divided by the price of V:

$$\psi_{ft}^{V} = \frac{\frac{\partial(P_{ft}(Q_{ft})Q_{ft}(V_{ft})}{\partial V_{ft}}}{W_{ft}^{V}}$$

Using the markdown wedge definition $\delta_{ft}^{V} \equiv \frac{\frac{\partial(P_{ft}Q_{ft})}{\partial V_{ft}} - W_{ft}^{V}}{\frac{\partial(P_{ft}Q_{ft})}{\partial V_{ft}}}$, it is trivial to show that the markdown wedge δ^V is the following function of the markdown ratio ψ^V:

$$\delta_{ft}^{V} = \frac{\psi_{ft}^{V} - 1}{\psi_{ft}^{V}}$$

All expressions above hold equally when manufacturers choose input prices rather than input quantities.

Applying these general expressions to the tobacco industry setting results in the following leaf markdown wedge:

$$\delta_{ft}^{M} \equiv \frac{\frac{\partial(P_{ft}Q_{ft})}{\partial M_{ft}} - W_{ft}^{M}}{\frac{\partial(P_{ft}Q_{ft})}{\partial M_{ft}}}$$

The marginal revenue product of tobacco leaf is equal to the product of the cigarette price and markup with the leaf content per cigarette β^M:

$$\frac{\partial(P_{ft}Q_{ft})}{\partial M_{ft}} = \mu_{ft}\beta^{M}P_{ft}$$

The leaf markdown wedge is hence equal to:

$$\delta_{ft}^{M} = \frac{\mu_{ft}\beta^{M}P_{ft} - W_{ft}^{M}}{\mu_{ft}\beta^{M}P_{ft}}$$

If the leaf markdown wedge would be zero, meaning that the inverse leaf supply function is flat, then farmers receive their marginal product $\mu_{ft}\beta^{M}P_{ft}$.

Inputs demand with endogenous prices

Manufacturers choose the profit-maximizing leaf price W_{ft}^{M*}, 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, manufacturers choose the mix between labor and capital. Denoting the marginal cost of labor and capital as MC_{ft}^{H}, and the interest rate as W_{ft}^{K}, the choices for labor and capital are given by the following cost minimization problem:

$$\min_{L_{ft},K_{ft}} W_{ft}^{L} L_{ft} + W^{K} K_{ft} - MC_{ft}^{H}\Big(H(L_{ft}, K_{ft}) - Q_{ft}^{*}\Big)$$

Assuming a Cobb-Douglas function $H(L, K)$, solving the first order conditions results in labor demand $l(k_{ft}, w_{ft}^{L}, w_{ft}^{M}, \beta^{L}, \beta^{K}, W^{K}, \omega_{ft}, \mu_{ft}, \psi_{ft}^{M})$. 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, manufacturers choose a lower equilibrium output level, and hence also use less labor.

E.2 Endogenous exit

It is less straightforward to allow for endogenous exit in the dynamic panel estimator compared to Ackerberg et al. (2015). As was argued in the main text, endogenous exit is less of a concern in the Chinese tobacco industry setting, as exit was mostly a consequence of a centrally imposed policy. As an additional check, I regress the productivity residuals on a dummy indicating exit in the next period in using OLS. The results are in table A8: in the right column, manufacturer fixed effects are added. Both the coefficient estimates and the R-squared are near zero for both specifications.

E.3 Market structure and leaf prices

In table A9, I regress log leaf prices $\log(W^{M})$ on dummies for the presence of one, two or three manufacturers at the province, prefecture, and county level. There is no systematic relationship between leaf prices and market structure at the province and county levels. At the prefectural level, however, leaf prices are 60% lower when there is just one manufacturer, and 35% lower when two or three manufacturers operate, compared to prefecture with four or more manufacturers. Besides

Table A8: Exit and productivity

	$1(\text{Exit}_{ft+1})$		$1(\text{Exit}_{ft+1})$	
	Est.	S.E.	Est.	S.E.
log(TFP)	-0.00227	(0.00160)	-0.0145	(0.00763)
R-squared	0.00210		0.0.0119	
Manufacturer fixed effects	No		Yes	
Observations	1,086		1,086	

Notes: Within R-squared reported for fixed effects model.

the institutional details discussed in the main text, this is another motivating fact for the choice to model leaf markets at the prefectural level.

Table A9: Market structure and leaf prices

	log(Leaf price)	
	Est.	S.E.
Province:		
1 firms	0.0509	(0.137)
2 firms	-0.0439	(0.115)
3 firms	0.429	(0.120)
Prefecture:		
1 firms	-0.892	(0.116)
2 firms	-0.431	(0.110)
3 firms	-0.503	(0.134)
County:		
1 firms	-0.0585	(0.240)
2 firms	-0.139	(0.251)
3 firms	-0.260	(0.288)
R-squared	0.0711	
Observations	1,086	

E.4 Bargaining model between manufacturers and wholesalers

The fact that province-level markups fell in response to the consolidation can be easily explained by a bargaining model. Let the factory-gate prices P_{ft} be determined by a Nash bargaining game between the manufacturers and the wholesaler. Let $\lambda \in [0, 1]$ denote the share of the total surplus received by the manufacturer: if λ is zero, the wholesaler gets all the surplus, if it is one, the manufacturer does, and if $\lambda = 0.5$, they get an equal share of the surplus. Denoting the wholesale price of a cigarette produced by manufacturer f as P_{ft}^W, the Nash product is given by equation (19a).

$$\lambda(\underbrace{P_{ft}^W Q_{ft} - P_{ft}Q_{ft}}_{\text{Wholesaler's profit}}) = (1-\lambda)(\underbrace{P_{ft}Q_{ft} - W_{ft}^M M_{ft} - W_{ft}^L L_{ft} - W_{ft}^K K_{ft}}_{\text{Manufacturer's profit}}) \tag{19a}$$

Suppose relative bargaining power of the manufacturers and the wholesaler remained constant over time, which seems likely as wholesale markets were dominated by a single monopsonistic wholesaler before and after the reform. From equation (19a, it is easy to see that a drop in the leaf price W^M due to increased monopsony power of the manufacturer on leaf markets must have resulted in falling factory-gate cigarette prices P, and hence also in manufacturer markups μ. The extent to which factory-gate cigarette prices fell depends on the bargaining weight λ, which can be estimated by rewriting the Nash product equation as equation (19b).

$$P_{ft}Q_{ft} = (1-\lambda)(W^L L_{ft} + W^M M_{ft} + W^K K_{ft}) + \lambda P_{ft}^W Q_{ft} \tag{19b}$$

If latent wholesale prices are a function of factory-gate prices, this leads to endogeneity bias when estimating equation (19b) with OLS. Assuming the consolidation only affected buyer power on the leaf market, and not on the wholesale market, the consolidation treatment effects can be used to instrument for total costs. This is a reasonable assumption: while leaf markets are local and concentrated due to leaf trade restrictions, this does not apply to the cigarette market, which extends across China. The consolidation therefore did not lead to a very concentrated cigarette market: this market evolved from 350 manufacturers selling to a single wholesaler towards 150 manufacturers selling to the same wholesaler. Estimating this equation by 2SLS delivers a bargaining weight of $(1-\lambda) = 0.901$, with a standard deviation of 0.265. This means that the wholesaler received on average 90% of the total surplus, which is consistent with the fact that there is only one wholesaler buying from many manufacturers. The strong bargaining position of the wholesaler explains the fact that cigarette price markups fell almost to the same extent as leaf price markdowns increased.

E.5 Regulated leaf prices and pricing decisions

In theory, leaf prices were regulated in China until 2015, meaning that prices per quality grade were determined by provincial STMA boards (Wang, 2013; Peng, 1996). In practice there were, however, still many ways in which manufacturers could choose leaf prices, as assumed in the model. First of all, grading criteria were subjective, and manufacturers frequently changed leaf prices by gaming the quality grade under which farmers' leaves are attributed (Peng, 1996). More formally, denote $\tilde{\zeta}_{ft}$ the subjective grade communicated by manufacturing firm f to its farmers, while the real quality grade being ζ_{ft}. As leaf prices are a direct function of the subjective grade, $W^M(\tilde{\zeta}_{ft})$, choosing the subjective grade corresponds to choosing the leaf price, holding the true quality grade ζ_{ft} fixed.

Second, the official grade-price schedules were determined by provincial STMA boards, but these were populated by executives of the CNTC cigarette factories: especially at lower geographical levels, the industry regulator and manufacturing firms were de facto the same organization (Wang, 2013). If only one manufacturer survived in a province, it also controlled the provincial STMA board, and could therefore set the leaf price per grade schedule.

E.6 Markdown and markup drivers

Which firm and county characteristics explain variation in markups and markdowns? In panel (a) of table A10, I report some correlations. Markdown ratios correlate positively with state ownership, and with the local unemployment rate. The higher markdown at SOEs may be due to the fact that non-monetary benefits are high at SOEs. High monopsony power leads to unemployment, which is consistent with the reported correlation. The markup correlations are the opposite from the markdown correlations: as markdowns enter marginal costs, and as manufacturers have little market power downstream, it is logical that markups and markdowns are negatively correlated.

E.7 Heterogeneous treatment effects

In panel (b) of table A10, I compare the consolidation treatment effects across different prefectures of the same province. The markdown increase in response to the consolidation is higher in prefectures with more manufacturers under the exit threshold, which is an indication for the size of the consolidation. It is also higher in prefectures in which the share of workers without any formal schooling is higher. These workers have lower outside options, which may explain why the increase in buyer power is higher in areas populated by these workers. As employment, schooling, and migration decisions are all endogenous choices by workers, these correlations are, however, not necessarily causal.

Table A10: Heterogeneity analysis

Panel (a): Correlations	log(Markdown)		log(Markup)	
	Est.	S.E.	Est.	S.E.
1(SOE)	0.307	(0.0815)	-0.289	(0.0948)
Log(Revenue)	-0.0307	(0.0195)	0.184	(0.0372)
Log(Unemployment rate)	0.118	(0.0515)	-0.214	(0.0663)
Log(Sales tax / Revenue)	0.127	(0.0373)	-0.0635	(0.0469)
R-squared	0.498		0.518	
Observations	765		765	
Panel (b): Heterogeneous treatment effects				
Treat*1(year$\geq$ 2003)*#manufacturers under 100K	0.201	(0.051)	-0.0231	(0.0476)
Treat*1(year$\geq$ 2003)*log(migration rate)	0.0494	(0.0712)	$<$0.001	(0.102)
Treat*1(year$\geq$ 2003)*log(unemployment rate)	0.147	(0.0743)	-0.0880	(0.100)
Treat*1(year$\geq$ 2003)*log(no schooling rate)	0.555	(0.193)	-0.227	(0.241)

Notes: Panel (a) regresses markups and markdowns on manufacturer and market characteristics. Controls: year and consolidation treatment dummies, market population size, percentage of urban Hukou permit holders. Panel (b) compares the consolidation treatment effects between prefectures of the same province. Controls include treatment effect and prefecture dummies.