Missing Data and the Effects of Market Deregulation: Evidence from Chinese Coal Power

Tom Eisenberg

January 25, 2023

Abstract

A series of market reforms were introduced in 2002 in the Chinese wholesale coal power sector. The period immediately after is widely recognized as having had inconsistent service and many blackouts in China, and it is generally accepted that many of the reforms were not fully enacted. Yet, researchers consistently find that these reforms resulted in efficiency gains for power plants. Using new physical and matched financial data, as opposed to only financial data, I find no evidence that there were efficiency gains at the plant-level. I also find that in the aggregate there were large productivity declines over this period. A lack of efficiency gains would also imply that the reforms had no effect on reducing pollution. Any measurable gains in either case are mainly due to input and output price fluctuations.

Key Words: Electricity, China, Competition, Productivity, Efficiency, Deregulation

JEL classifications: Q48, Q41, D24, Q43, L33, P28

1 Introduction

China has undergone many privatization and market reform efforts over the past two decades, so it provides an extremely useful context for investigating if deregulation promotes competition, and in turn productive efficiency. Of particular interest is the Chinese electricity sector, specifically coal, which is the largest energy market (and largest source of carbon emissions) on the planet. Chinese coal power is central to debates about climate change and environmental policy, extremely important to the lives and welfare of Chinese people, but also extremely volatile.

Quantifying firm- or plant-level efficiency, and in turn aggregate efficiency, can be hampered by relying on expenditure or revenue data rather than separate prices and outputs/inputs, famously argued in Klette & Griliches (1996). This paper applies this idea to electricity deregulation efforts in China. Using the series of reforms to China's electricity generation sector in 2002, I study, using production data, whether there are measurable efficiency gains at the plant level in response to these policy changes. Such gains are commonly found in the previous literature that largely relies on revenue and expenditure data. I then extend this analysis to aggregate productivity measures, which have yet to be thoroughly studied in this market.

This paper is the first to empirically investigate, via a case study with a novel dataset, whether measured gains from electricity market reforms are sensitive to biases from missing price and quantity data. In the case of China, where firm-side control over prices is ambiguous and subject to policy influences outside of plants themselves [1], this may be particularly problematic. Since prices are determined largely by local and federal authorities, for example, price changes may be systematically correlated with key variables like state ownership status.

The 2002 reforms are especially at risk of this type of bias: China made multiple policy changes in 2002 to reform the electricity generation market, some of which may have had direct effects on prices. While the government made initial steps toward electricity deregulation and lowering administrative costs for power plants in 2002, they simultaneously deregulated the coal input market (which had treated state-owned firms preferentially with a "two track" system), which may have altered prices systematically (Gao & Van Biesebroeck, 2014). Parallel to this were multiple

[1] This also makes it difficult to use productivity methods that rely on explicit optimization equations from firms, like De Loecker & Warzynski (2012).

technology mandates (Ma & Zhao, 2015) and high-level administrative changes (Gao & Van Biesebroeck, 2014). However, if efficiency gains are found empirically using expenditure-based measures and not physical ones, it suggests there was no improvement in technical efficiency, and only that prices fluctuated in response to policy changes. The central contribution of this paper is to demonstrate this fact at both the individual plant and aggregate levels. Using physical and financial data together, I also directly demonstrate mechanisms that alter the measurement of efficiency gains in Chinese electricity.

In the context of China, this paper is the first to examine the effects of electricity reforms on aggregate efficiency measures. The elimination of unnecessary bureaucracies in plant-level decision making could make for more efficiently run plants, but the relative allocation of production across plants may remain unchanged due to government controls.

Many papers study these reforms, such as Du *et al.* (2009), Du *et al.* (2013), Gao & Van Biesebroeck (2014), Ma & Zhao (2015) and Wei *et al.* (2018). Many are limited to expenditure and revenue data, or have either limited samples or limited input data. Thus, this paper also unifies several strands of analysis of Chinese coal power reforms by attempting to combine several data-related or methodological advances.

There are three major methodological components to this paper. The first is a set of partial factor productivity models, where inputs are regressed on outputs, a series of control variables, and a difference-in-differences variable. This model is in the vein of Fabrizio *et al.* (2007) and Gao & Van Biesebroeck (2014). Gao & Van Biesebroeck (2014) argue that state-owned firms were more exposed to the reforms of 2002 and function as a treatment group for a difference-in-differences analysis. I combine this insight with a version of their model that is adapted to the presence of physical data on output and inputs. Estimating equations are derived from a production function and cost minimization assumptions.

These PFP models do not directly estimate a plant's total factor producitivty (TFP) residual in full, as they only seek to capture the components of production that change for state-owned plants in response to the reforms. I thus also apply techniques from Ackerberg *et al.* (2015) (and Gandhi *et al.* (2020) for robustness) and use the same difference-in-differences analysis.

An advantage of the TFP estimation is I can use the estimated residuals to do the third portion of the paper: aggregate productivity analysis for both physical and revenue/expenditure data. I do this via a set of decompositions from Olley & Pakes (1996) and Melitz & Polanec (2015), which allow me to gauge the role of intensive margin productivity increases, entry, exit, and most importantly, reallocation of production among different firms over time.

I find little evidence that the 2002 electricity market reforms in China caused any differential efficiency gain for treated (state-owned) plants when physical measures of inputs and outputs are used. This result applies for both the TFP and PFP models. Together, they imply that any previously estimated welfare gains or reductions in pollution may not have been realized. For the PFP models, I am able to demonstrate directly that the shift in results from the previous literature is due to price movement that could not be omitted from revenue/expenditure data. Since I do find that some outcomes, like prices, are responsive to the DID analysis, the 2002 electricity reforms in China are likely an example of "imperfect regulation" in energy markets, as discussed in Cicala (2022) and Joskow (2008).

In the case of aggregate productivity, I find a stark contrast between expenditure-based and physical measures of total factor productivity. This is a common channel for benefits in electricity restructuring, as documented in Cicala (2015) and Cicala (2022). Papers looking at other industries in China, like Hsieh & Klenow (2009), find a substantial remaining role for reallocation. For this sector specifically, Wei *et al.* (2018) and Kahrl *et al.* (2013) show in the aggregate that there is some room for productivity growth, so it is unlikely that this sector was already operating at maximum possible efficiency.

Broadly, this paper contributes to the debate on the general effectiveness of electricity reforms happening in many different countries. For example, Cicala (2015), Fabrizio *et al.* (2007), Newbery & Pollitt (1997), Joskow (2008), and Cicala (2022) have all found broadly positive, if complex, results from restructuring in western countries. Han *et al.* (Forthcoming) have recently found conflicting effects in the United States, but the balance of the evidence is that restructuring does tend to increase plant-level and aggregate productivity.

The contrasting results between physical and revenue-based measures of productivity also contribute to the literature on productivity measurement. Foster *et al.*

(2008) established that in the US, revenue-based measures may obscure the important and separate roles that demand and technical efficiency play in analyses of aggregate productivity and firm survival. Similarly, Haltiwanger *et al.* (2018) show that misallocation analysis such as Hsieh & Klenow (2009) that rely on revenue and expenditure data are extremely sensitive to model misspecification and require very specific assumptions on supply and demand. Klette & Griliches (1996) and Ornaghi (2006) prove that conventional TFP estimation methods will misspecify production functions if revenue and expenditure data is used.

The paper proceeds as follows: section 2 discusses the institutional and historical details for this analysis. Section 3 presents and summarizes the data used in the paper. Section 4 presents models and estimation. Section 5 features parameter estimates and decomposition results, while section 6 has robustness checks and extensions. Section 7 concludes.

2 Background and Motivation

2.1 History and Reforms

China features many institutional specifics that inform an analysis of electricity market restructuring. There is a mix of state-owned, jointly owned, and purely private power plants in China's wholesale power market (Liu, 2013), though there is no ex ante clear, systematic way that this status affects their production or pricing outcomes.

The key reforms from 2002 considered in this paper are in the *Notice of the State Council on the Issuance of the Reform Plan for the Electricity System*. Several major reforms were considered or enacted in a short timeframe. They involved breaking up a major state-owned enterprise into five smaller companies and separating administrative functions at the federal level for transmission and generation. There was also a contemporaneous deregulation of the input (coal) market, and pilot programs in select provinces for market-based pricing [2]. Not all of the reforms were made permanent, or fully committed to. Assessing the market in 2017, 15 years after the initial reforms, a Resources for the Future report said that Chinese wholesale

[2] See, i.e., Liu (2013) or Ho *et al.* (2017).

electricity "has no spot market" (Ho *et al.*, 2017).

Central to this paper is a difference-in-differences specification that considers state-owned plants to be the "treated" plants. The previous literature on the subject has thoroughly argued why this is a useful framework for casual analysis. Gao & Van Biesebroeck (2014) cite a host of contemporary evidence that the 2002 reforms should boost competition and have particular causal effects on state-owned plants and firms. For example, in 2003 (as part of the process of the 2002 reforms) the State-owned Asset Supervision and Administration Commission was created to better manage and develop performance metrics for state-owned enterprises in particular.

State-owned plants also had more preferential access to lower coal prices under the "two track" pricing system prior to 2002. Essentially, power plants were promised explicit volumes of coal at guaranteed prices, and these terms were often more favorable for SOEs (Gao & Van Biesebroeck, 2014). That is, all plants were input price takers for a large portion of their coal, with SOEs receiving generally lower prices After 2002, a higher proportion of coal began to be sold in actual markets, which could effectively raise prices more for SOEs given their prior preferential access. However, through either equilibrium effects or other contemporaneous reforms, it is ambiguous what the ultimate effect would be. Parallel to this, as documented in Ma (2011), output prices did not shift accordingly in response to input price changes.

Gao & Van Biesebroeck (2014) explain that electricity reforms were done alongside many reforms in the broader manufacturing sector that tightened standards for SOEs and forced them to "focus more narrowly on their core business". Thus, in addition to the direct reforms undertaken on SOEs, the government had also made clear its intentions to increase competition and technical efficiency among SOEs in the broader economy. Ma & Zhao (2015) discuss how parallel to this several technology mandates happened after 2002 that may contaminate a differences-in-differences estimation. These include targeting of small generation units for decommissioning and the promotion of new capacities.

It is clear that whatever reforms took place, qualitatively no "full" deregulation of wholesale coal-powered electricity happened. As explained in Liu *et al.* (2013), some of China's reluctance to go fully into market-based reforms at this time has to do with the mixed results that restructured markets began seeing, especially in the United States. This qualitative pictures contrasts with the now substantial body of work, including Gao & Van Biesebroeck (2014), Ma & Zhao (2015), Wei *et al.* (2018), Du *et al.* (2009), and Du *et al.* (2013), that has empirically found efficiency gains in response to these reforms.

In the aggregate, news sources from the time do not depict a productive, efficiently run sector. Power outages over large areas were commonly reported over this time, and the real focus of the Chinese government was seen as prioritizing capacity expansion to meet rapidly growing demand (Shunkun *et al.*, 2013). There are many examples of news coverage from the time documenting power shortages, for example in December 2003, factory workers were made to work nights in Shanghai to accommodate power shortages [3]. Shunkun *et al.* (2013) document that the shortages were worst among "economically developed provinces in coastal areas, such as Zhejiang, Jiangsu, and Shanghai."

Demand was growing so quickly during this period that planners may not have been focused on efficiency. A planned market like this can clearly lead to x-inefficiencies. If plants do not directly influence prices or quantities, they will not make costly investments in streamlining their production processes. Similarly, plants with existing efficiency advantages are unable increase their production share due to the lack of a spot market. If national priorities are on meeting consumer demand or sustainable development as argued in Liu (2013), then the distribution of production across plants may not sufficiently take into account their individual efficiencies.

Between the ongoing debate on how successful electricity restructuring is overall, and the debate on whether the reforms in China specifically succeeded, it is clear that this event is an important case study to address in light of increasing data availability. This is true for direct policy reasons and for our ability to understand deregulation and competition in general. While this paper is primarily focused on the Chinese experience, a positive relationship between electricity deregulation and technical efficiency has been documented in many other settings, as demonstrated in Cicala (2022), Cicala (2015), Fabrizio *et al.* (2007), and Newbery & Pollitt (1997), among others.

Joskow (2008) provides a thorough review of this literature, and emphasizes

[3] See Watts (2003)

that several concurrent steps are necessary for electricity deregulation to achieve its desired goals. Joskow (2008) argues that successful implementation of several measures, including vertical separation of transmission and generation, and horizontal restructuring of the regeneration market, are key to the success of deregulation. Joskow argues electricity market liberalization has been successful in "the UK, the Nordic countries, Argentina [...], Chile, Texas, portions of Australia and other countries and regions." The 2002 reforms in China certainly aimed to accomplish several of these steps, but it remains doubtful how successful they have been to date.

2.2 Theoretical Background

The theoretical debate on how to measure and relate productivity (or efficiency), privatization, and competition extends well beyond the electricity sector. As documented in Backus (2020), the correlation between competition and productivity is well established across many industries. Backus (2020) also finds that this is usually due to firm-level responses to competitive conditions. This is the crux of the theoretical argument for potential efficiency gains in this context: as competition is introduced to Chinese electricity, individual plants will become more efficient in response, and both individual and aggregate productivity will increase. However, recent theory on productivity measurement suggests that it is extremely sensitive to a researcher's assumptions and methods.

A seminal paper on productivity measurement is Klette & Griliches (1996), which shows that productivity estimation which uses revenue as a proxy for output is contaminated by firm-level prices. More often than not, these will be correlated with objects of interest and are difficult to assume away. These types of biases are further explored in Ornaghi (2006). This tension has been leveraged recently by papers like Foster *et al.* (2008), who explicitly model the fact that price effects, when separated from technical efficiency, should provide valuable information on firm-specific demand levels and shocks.

Despite theoretical advances in productivity measurement, many (if not most) papers in the productivity literature are constrained by data availability and have to rely on revenue and/or input expenditures. Papers like Hsieh & Klenow (2009), which relies on a series of input first-order conditions, or De Loecker & Warzynski

(2012) and Grieco *et al.* (2016), which combine first-order conditions with structural productivity estimation, have attempted to address this issue via additional modeling assumptions. However, Haltiwanger *et al.* (2018) and Bond *et al.* (2020) show that both of these approaches will be insufficient in many cases. Given realistic complications (adjustment costs, relationships between inputs and demand), these methods' measures of both productivity and markups may not contain useful information on either object. The analysis in Ornaghi (2006) shows specifically for modern methods TFP estimation that the use of revenue and expenditure data may lead to biased production function and residual estimates, so I include these in the analysis as well.

3 Data and Descriptive Evidence

The key dataset for this paper is a confidential survey of coal power plants conducted by the Chinese government. It is a subset of the data contained in the Chinese Environmental Survey (CES), which has only recently become available to researchers, but has been used in Ma & Zhao (2015) and Gowrisankaran *et al.* (2021). The dataset is meant to cover 85% of electricity production in China. I observe the data from 1997-1998, 2000, and 2002-2011 [4]. Major variables include a plant's name, power generated, coal used, and nameplate capacity.

Because this is a newly available dataset, it has had little time under public scrutiny. Gowrisankaran *et al.* (2021), who observe additional pollution variables not in my sample but otherwise draw from the same source, have found that the CES contains at least 85% of the amount of sulfur dioxide emissions reported by aggregate sources in China. Thus, this dataset does not appear to understate pollution relative to other Chinese sources, which would be a primary source of concern. I extend the verification of this data to additional sources: the total coal energy production reported in this dataset is very close to 85% of the aggregate figures reported by the International Energy Agency in publicly available years. These numbers are reported in the appendix in Table 22. Given that the IEA is an external source of reporting, this strongly reinforces the dataset's veracity.

Plants are then linked to the second major dataset used in this paper, the more

[4]Thank you to Shanjun Li, Deyu Rao, and many others for preparing this data and allowing me to access it.

commonly used financial census from the National Bureau of Statistics in China. This dataset contains key financial variables for state-owned firms and all firms above 5 million RMB in earnings. Note that employment data comes from the financial census rather than the confidential survey. Thus, prior papers have had more extensive use of employment data as compared to the coal use and production data.

The financial dataset is stored at the firm level, while the physical data is stored at the plant level. For most aspects of the analysis, this is only a minor issue: specifications that use physical data only can just be run at the plant level [5], and specifications that use the financial data are at least as accurate in this with the previous literature that only observed financial variables. To the extent that the plant-based analyses differ in their units of observation than the firm-based ones, this is a positive feature of the data [6]. As Gao & Van Biesebroeck (2014) point out, the assumption that most firms are just one plant is largely accurate anyway, and thus the distinction is minimal. For analyses that use price indices directly, these indices are necessarily at the firm level.

The fullest version of the (physical) dataset contains 21,121 plant-year observations, though I focus on the much smaller subset that is matched to the financial data. The primary reason for this is that the preferred recent method of identifying state-owned enterprises, from Brandt *et al.* (2017) and others, requires financial data. Table 1 shows what proportion of the near-universe of plants included in the physical dataset survive the matching process. Given that 33% of plants are represented but 49% of capacity is represented, my sample skews toward larger plants.

While any selection is undesirable, there are several reasons why this sample is still extremely useful for analysis. First, previous literature such as Gao & Van Biesebroeck (2014) have found that large plants responded **more** to the 2002 reforms in several key areas. Second, as documented in Ma & Zhao (2015), many smaller plants were targeted for shutdown and were not included in long-term deregulation efforts as a result. Third, while it is not directly testable, the financial census data

[5] The only variable of concern in these analyses would be state ownership status, which is derived from the financial dataset. However, this would generally be defined at the firm level anyway, so it would apply to all plants under the same firm equally.

[6] Gao & Van Biesebroeck (2014), for example, specify that their model would work best for plant-based specifications.

Table 1: Matched Sample Characteristics

Source	Materials
% of Production	50%
% of Production from plants over 50 MW Capacity	51%
% of Capacity	49%
% of Plants	33%

Notes: Depicts the percentage of key totals that are captured in the sample where financial and physical variables are matched. Key matching variables include company name, ownership information and province. Percentages are of the totals in the physical dataset.

that previous results rely upon likely selects for larger firms since it has a lower revenue cutoff for inclusion. Fourth, as shown later in the paper, this sample is able to replicate key qualitative results from other papers that use financial data, suggesting this sample is not underpowered to detect changes from the reforms.

From the physical data I can derive a plant's "heat rate", a standard measure of efficiency calculated by dividing coal input by power output. For my analyses, I omit plants that never exceeded 50 MW [7] of capacity during their entire lifespan, though since they are used in some regressions I include them in the full sample. I omit plants with unrealistic heat rate and price indices [8]. These result in the sample sizes seen in the following tables. In Table 2 are a series of summary statistics:

[7] According to Ma & Zhao (2015), all plants of this size are targeted for closure. Thus, many of the plants missed in the matching of the two datasets would also be excluded anyway.

[8] This includes heat rates below .05, above 2, input prices greater than 2 or below .1, and output prices below .08 or above 2. This results in about 6% of the sample being dropped, though results are qualitatively similar when they are included.

Table 2: Means of Major Physical Variables (Matched Sample), 1995-2011

Year	Cap (MW)	Prod (MW)	Heat Rate (tons/MWh)	N
1995	224.0	180.3	.768	226
1997	220.9	141.0	.748	276
1998	239.1	134.6	.694	317
2000	269.0	153.0	.660	295
2002	336.6	206.0	.651	319
2003	305.0	203.1	.657	403
2004	341.5	236.0	.658	406
2005	368.9	242.2	.634	447
2006	456.7	279.5	.610	467
2007	490.0	289.4	.592	516
2008	541.8	306.3	.563	517
2009	589.9	335.1	.552	453
2010	644.8	385.0	.549	415
2011	674.8	418.0	.548	374

Notes: Table depicts summary statistics for years 1995-2011. Physical variables are from confidential power plant survey, financial variables are from a combination of physical dataset and financial variables from annual NBS manufacturing census. One RMB is roughly .15 dollars, so the output price in 1998 of .26 000 RMB/MWh would equal about 40 dollars per MWh, while the 2007 output price would be more like 47 dollars.

Average plant size grows rapidly over time, utilization (production/capacity) stays relatively stable from 1997 on, and there is large net entry over time based on the growth of the number of plants. There is a general trend downward in heat rates which continues after 2002 and 2004. However, the total decline from 1995-2002 is slightly larger than the total decline from 2002-2011, so it is difficult to detect any effect from this presentation of the data. The financial variables, featured in Table 3, are considerably more limited in terms of their period of coverage, but provide valuable insight.

Table 3: Means of Major Financial Variables, 1998-2007

Year	Price (000 RMB/MWh)	(Linear) Marginal Cost (000 RMB/MWh)	Input Price (000 RMB/Ton)	N
1998	.374	.262	.411	294
2000	.370	.246	.392	279
2002	.372	.257	.414	304
2003	.384	.270	.421	382
2004	.341	.271	.433	377
2005	.385	.317	.530	419
2006	.374	.305	.517	441
2007	.385	.337	.597	490

Notes: Table depicts summary statistics for years 1998-2007. Physical variables are from confidential power plant survey, financial variables are from a combination of physical dataset and financial variables from annual NBS manufacturing census. Figures are for sample where revenue and physical data is matched.

Post-restructuring, using either 2002 or 2004 as a baseline year there is a clear upward trend in input prices, but less so in output prices. This translates into higher marginal costs as well. This is consistent with the findings in Liu *et al.* (2013), where output prices remained more tightly regulated than input prices, resulting in potentially large losses for power plants. This dramatic but asymmetric shift in prices may differ by state ownership status and could have a large effect on the efficiency of plant operations.

My primary method of defining state ownership is plants with majority capital ownership from the state (argued for in Brandt *et al.* (2014)). In robustness checks I also define SOE status to be plants currently or formerly owned by the state monopolist firm broken up in 2002 (as in Zhang *et al.* (2001)). Both measures rely primarily on the financial census data, which contains more detailed ownership information. It is straightforward to match almost all of the matched observations to observations other years for the physical data via their name, location information, and size, but difficult to verify their status directly. I assume that these firms were not privatized or socialized in 1995, 1997, 2010, or 2011, which extends their identification as SOEs into years that are missing financial data. Because these are strong assumptions, I verify the results that use these years are not sensitive to them in robustness checks.

Table 4: Summary Statistics By Ownership Category (2002 and prior)

	State-Owned Firms	Control Firms	Difference in Means
Production (Million MWhs)	2.30	1.79	.51
	(.16)	(.08)	(.17)
Capacity (MW)	409.3	346.0	63.32
	(16.2)	(15.8)	(22.6)
Utilization	63%	62%	.5%
	(1.6%)	(1.3%)	(2%)
N_1	619	694	
Input Price	.37	.40	.03
	(.012)	(.018)	(.023)
N_2	371	150	

Notes: Table depicts summary statistics for years 2002. Physical variables are from confidential power plant survey, financial variables are from a combination of physical dataset and financial variables from annual NBS manufacturing census. N_1 and N_2 reflect different samples for different variables. Standard errors in parentheses. Sources: NBS annual data and confidential survey of power plants. Excludes plants under 50 MW capacity.

Given that the paper's central analyses rely on difference-in-differences estimators, trends and baseline levels of key variables before and after 2002/2004 are central objects in the analysis. Table 4 shows that SOEs were on average larger and produced more before 2002, but utilizations and input prices were comparable.

An important trend to consider is heat rates, input to output ratios that are often considered the basic measure of a coal plant's efficiency. Figure 1 (a) presents these estimate in terms of tons of coal vs megawatt-hours of electricity produced [9].

While they are not identical, the SOE and private lines both decrease from the start to the end of the sample and largely move together after 2002. The private trend is a bit more volatile, but with the exception of 2000, there is no obvious large divergence from the SOE trend. Regardless, the physical evidence in figure 1 (a) does not suggest any particularly strong decrease (which would correspond to increased efficiency) for SOEs around the time of restructuring in . A graph that uses a comparable measure for financial variables tells a very different story, however.

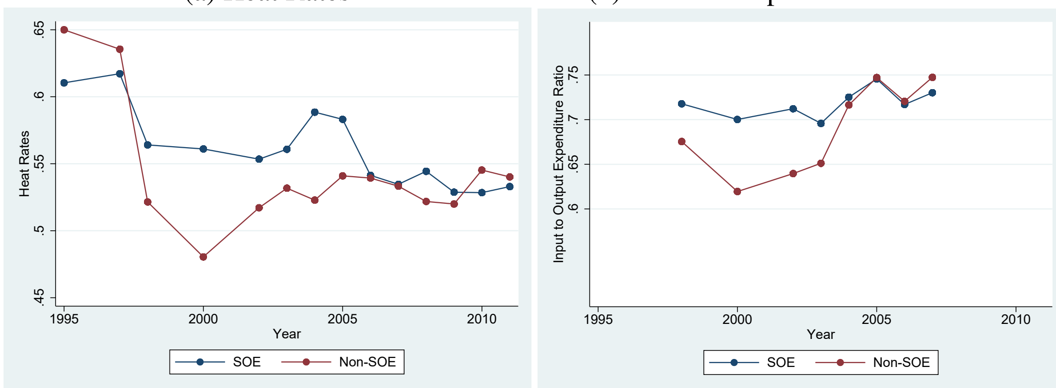
According to Figure 1(b) there is a marked asymmetric shift around 2004 [10].

[9]These are of a similar magnitude to estimates from the US like 1.13 pounds/kWh at https://www.eia.gov/tools/faqs/faq.php?id=667&t=2

[10]Prior to that, the two lines are on similar trends, though it is slightly difficult to tell with the limited pre-trend data in this particular sample since the expenditure data does not extend to 1995 or 1997. However, this is not a major concern since expenditure data features in prior analyses in the literature, and Gao & Van Biesebroeck (2014) show fairly conclusively that there is little evidence of a pre-trend when the full set of fixed effects and controls is accounted for in the DID analysis

Figure 1: Input to Output Ratios, SOE vs Private Plants

(a) Heat Rates (b) Material Expenditure to Revenue



Notes: Heat rates determined by dividing coal use by power produced (lower is better). Graphs only uses the sample of plants matched between physical and financial datasets. Excludes plants below 50 MW maximum capacity. Source: confidential coal power plant survey and financial census data.

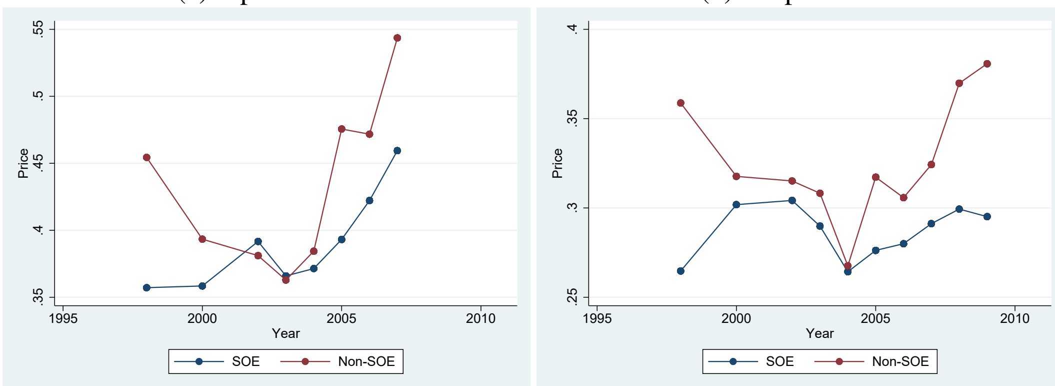
This stands in stark contrast to the prior graph with only physical terms, where SOE and private heat rates were . This provides the first evidence that we will see a divergence between physical and revenue-based assessments of the 2002 reforms. Once the trends for both types of variables are separated, they imply substantially different behavior for private plants and SOEs than earlier data would suggest. The analysis for prices is presented in figures 2.

While the pre-trend input price data is limited, input prices between SOEs and private plants appear to be moving in opposite directions before 2002. The post-2002 increase for private plants is more dramatic, although both types of plants appear to experience a change. Figure 2 (b) shows that output prices demonstrate an even more extreme version of this story: after 2004, prices rise much more sharply for private plants than they do for SOEs. In the case of SOEs, the output prices are even held below pre-2002 levels. Thus, while SOE input prices may also be rising slower than those of private plants, there is no clear indication that they gained from this in their average margins.

How exactly we should expect either input or output prices to differ, and what exactly their bias will be, is discussed in Gao & Van Biesebroeck (2014). The relative change in baseline subsidized state prices vs those resulting from increased

Figure 2: Input and Output Prices, SOE vs Private Plants

(a) Input Prices (b) Output Prices

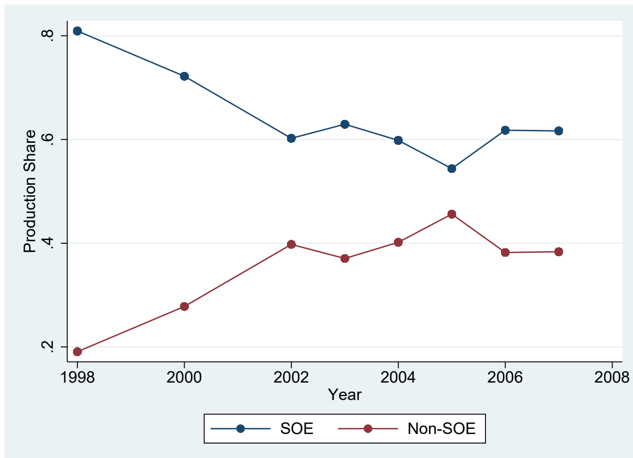


Notes: Excludes plants below 50 MW maximum capacity. Input prices are calculated by dividing input expenditure by coal use. Output prices are calculated by dividing revenue by power generated. SOE status is derived from majority capital ownership status in the financial census data. Points represent unweighted means across power plants. Source: confidential coal power plant survey and financial census data.

market activity are difficult to predict. There is no ex ante obvious pattern these two trends should take. We can now empirically verify that input and output prices both increased for both types of plants, input prices rose faster, and that there was a greater divergence in output prices. These turn out to lead to competing biases in the data, as will be demonstrated with the fuller empirical models.

Thus far, the graphs speak very little to any aggregate change due to efficient reallocation. If this happened along ownership lines, we can get a sense of this with a graph of production shares by ownership status in figure 3, which is also newly available with the data in this paper. 2002 corresponds to a stabilization of previous trends in the data according to this figure, but there is no dramatic movement afterwards. Thus, there is a possibility that efficient reallocation occurred in response to restructuring, but it would depend on which particular plants participated in this shift.

Figure 3: Production Shares by SOE Status



Notes: Shares are total amount of MW produced by all plants that share an ownership status divided by all production by year.

4 Model and Estimating Equations

With the empirical context established, this section outlines the models and estimating equations that will be used to measure efficiency and productivity. Chief concerns are controlling for plant-level heterogeneity, size, year effects, and baseline differences across ownership categories in analyzing the relationships seen in the previous section.

4.1 Model Description

There are two major models in this paper: one using "partial factor productivity" (PFP) models and one using "total factor productivity" (TFP) models. The PFP models are a common way of testing restructuring in electricity markets, and a variant of this analysis has been used in Gao & Van Biesebroeck (2014), in turn derived from Fabrizio *et al.* (2007) and Du *et al.* (2009). Efficiency gains have been found using full TFP estimation in Chinese electricity as well (for example in Ma &

Zhao (2015)). Thus, I use both methods to check the robustness of the main results to different assumptions.

The TFP models attempt to estimate a production function and Solow residuals using structural methods, while the PFP models rely on input first-order conditions derived from cost minimization. In principle, the models should be capturing different effects: the PFP model will capture all relevant changes solely due to SOE status, both through changes in the production function and TFP residuals, for each input at a time. However, it cannot estimate a Solow residual directly. The TFP model should explicitly isolate a plant's residuals and measure change in those distinctly from their production function. While there is overlap in the efficiency changes that the two models capture, they make different enough assumptions about the production process that they would not necessarily produce the same results. Additionally, the TFP estimation procedure allows for the assumption of a Cobb-Douglas production function [11], and for decompositions that measure aggregate productivity patterns.

As mentioned before, the analysis in this paper takes place at the plant (as opposed to firm) level. Plants are assumed to have no control over their assigned prices or quantities. Thus, much like previous papers in the literature, the PFP model relies on cost minimization assumptions. As in Gao & Van Biesebroeck (2014), plants are assumed to produce according to a Leontief functional form:

$$Q = \min_{M,L}\{f_1(M,\beta,\varepsilon_M), f_2(K,L,\alpha,\varepsilon_l)\} \tag{1}$$

$$s.t. Q \geq \bar{Q} \tag{2}$$

This functional form is particularly appropriate for power plants: power generators run according to a mechanical process that burns fuel, and it is nearly impossible to substitute for this with labor or capital in the short run. However, I relax this assumption in robustness checks to make sure it is not driving the results.

[11] Though in the appendix I also present results for translog

4.1.1 Partial Factor Productivity

The PFP models rely on a derived input demand equation for materials, which in this case will be coal. Given the Leontief assumption, the intermediate input demand equation can be approximated the following way, assuming f_1 is monotonically increasing:

$$lnM_{it} = \gamma_i + \gamma_t + \gamma_1 lnQ_{it} + \varepsilon_{it}^M \tag{3}$$

Where M is a plant's material inputs, i indexes plants, t indexes years, and Q is a plant's output. For a revenue-based equation M can represent all expenditures on materials. This provides a basic regression framework to identify the effects of restructuring. Because I have access to physical input data, I am able to estimate this equation directly, with a slight modification to detect the effects of restructuring:

$$lnM_{it} = \gamma_i + \gamma_t + \gamma_1 lnQ_{it} + \gamma_r STATE_{0i} * Restruc_{it} + \varepsilon_{it}^M \tag{4}$$

$STATE$ refers to state ownership status which serves as the treatment variable, while $Restruc$ is a pre- or post-restructuring variable (usually 2002, but also 2004 in robustness checks). Figure **??** suggests that material input to output expenditure ratios increased more for private plants post restructuring. So, taking SOEs as the treatment group means we expect to find a **negative** sign to reflect the fact that their input/output expenditure ratio went up less (ie, they became less inefficient) when we use financial variables. Whether this translates to the physical setting remains to be seen.

f_2 is assumed to be CES in the case of the PFP models:

$$Q = \gamma(\alpha K^\rho + (1-\alpha)\nu L^\rho)^{\frac{\nu}{\rho}} e^{\varepsilon_L} \tag{5}$$

Labor input demand from cost minimization nets the following equation:

$$W = (1-\alpha)\nu\lambda Q^{\frac{\nu-\rho}{\nu}} L^{\rho-1} \tag{6}$$

Here, as in Gao & Van Biesebroeck (2014), λ is the Lagrange multiplier from the constraint of setting Q equal to $\bar{Q}$. Taking logarithms, λ, which will vary by individual and over time, becomes additively separable. Following Fabrizio *et al.*

(2007) and Gao & Van Biesebroeck (2014) I include capital stock, log wages, and fixed effects to attempt to absorb it. Other parameters will be absorbed by the intercept and fixed effects. Incorporating relevant fixed effects and policy changes, this can then be rearranged and rewritten as:

$$lnL_{it} = \alpha_i + \alpha_t + \alpha_1 lnQ_{it} + \alpha_W lnW_{it} + \alpha_K lnK_{it} + + \alpha_r STATE_{0i} * Restruc_{it} + \varepsilon^L_{it} \quad (7)$$

Since changes over time are the variable of interest, there is no need to consistetly estimate ν, α or ρ. W are plant wages and K is a plant's capital stock [12]. Previously in the literature, Q and M had to be replaced by revenue and expenditure variables due to data limitations. Even with extremely detailed controls to proxy for prices, it may not be possible to perfectly purge these models of plant-level heterogeneity in prices.

In terms of estimation, this model is linear in parameters. A major source of potential endogeneity comes from the inclusion of output on the right hand side. While this is not the variable of interest, the results can be sensitive to the choice of instrument. I present specifications using several instruments for logged output in robustness checks. The baseline specification in Table 5 uses the lagged value (with 2nd lag for years with missing data) as an instrument for output. Lagged output is one of the instruments that finds efficiency gains in Gao & Van Biesebroeck (2014), so it allows for a fairly direct comparison.

Figures **??** and 2 indicate that this method will likely lead to different estimates between financial and physical variables. The exact sign and size of the difference is difficult to gauge before running the model, because there are two asymmetric shifts in prices happening: input prices go up more for private plants, but so do output prices. These will have opposing effects: if revenues (conditioned on all relevant controls) get extremely inflated for private plants relative to SOE plants, more so than input expenditures, it will appear like they became relatively more efficient. However, if the input inflation is faster, then they will appear to have become less

[12] The residuals ε^L absorb any productivity shocks, which I do not quantify directly in this particular model. The only way they are addressed is via instrumenting any potentially correlated variables. Technically speaking, the coefficient on quantity α_1 is equal to $(\nu - \rho)/(\nu(1-\rho))$ and α_W equaling $-1/(1-\rho)$.

efficient in response to the reforms.

4.1.2 Total Factor Productivity

The TFP analysis has two stages: a first stage, where production functions are estimated directly and each plant's TFP residual ω is recovered, and a second stage where these residuals are used as dependent variables in a difference-in-differences analysis.

I begin by assuming a Cobb-Douglas production function for the f_2 function:

$$Q = AL^{\beta_l}K^{\beta_k}e^{\omega}e^{\varepsilon} \tag{8}$$

Taking logs, f_2 is represented the following conventional way:

$$q = \beta_0 + \beta_L l_{it} + \beta_K k_{it} + \omega_{it} + \varepsilon_{it} \tag{9}$$

Where lower case letters are logged inputs and outputs, and ω represents a productivity shock that plants observe before they make their input decisions.

The basic conflict in TFP estimation is that if plants can observe ω, their TFP residual, then input choices should be correlated with it, presenting an endogeneity problem. Proxy variable approaches are widely used to estimate TFP within an industry for this reason. Among the most popular and robust of these is Ackerberg *et al.* (2015), which resolves identification issues of earlier entries in the literature. A key assumption in the proxy variable approach is that plants (or in this case, possibly regulators) observe ω when production decisions are made, but not ε_{it}. Ackerberg *et al.* (2015) allow this function to be estimated accounting for the endogeneity of ω via a series of assumptions of input timing, the invertibility of a plant's choice of materials depending on its productivity level, and the dynamic process that ω follows. This allows the researcher estimate a production function and the associated TFP residual in two steps: first isolate idiosyncratic errors ε from TFP ω, then use the dynamic assumptions to form moments based on lags and changes in productivity to estimate the production elasticities. In robustness checks I extend this approach to more recent methods for gross production functions with 3 inputs from Gandhi *et al.* (2020).

The expected sign and magnitude of the bias due to using expenditure data using this particular technique is ambiguous. Ornaghi (2006) documented that relying on industry-level deflators rather than individual prices to convert to physical data tends to lead to downward-biased scale estimates. Given the previously discussed competing asymmetric price shifts in addition to this general phenomenon, the expected effect on TFP residuals in particular is more difficult to characterize ex ante compared to the PFP analysis.

5 Results

5.1 PFP Model Results

Table 5: PFP Regressions Using Physical Measures - Coal

	(1)	(2)	(3)	(4)
Restruc x SOE	.040	.088	.028	.080
	(.025)	(.051)	(.023)	(.051)
SOE	-.026	-.115	-.011	-.091
	(.023)	(.062)	(.019)	(.051)
Log Output	.952	1.49	.951	1.49
	(.022)	(.124)	(.021)	(.159)
Restructuring Year	2002	2002	2004	2004
Power Instrument	Lag Power	Coal Input Price	Lag Power	Coal Input Price
N	3,144	1,878	3,144	1,878
Plants	544	424	544	424

Notes: Dependent variable is logged coal input use. Includes plant fixed effects and year fixed effects. Standard errors are clustered at the plant level. All first-stage F statistics are above the 10% Stock-Yogo weak ID test threshold. Standard errors in parentheses. All regressions include year fixed effects. Plants with maximum capacity below 50 MW are omitted. "Lagged power" uses twice lagged data for missing years.

Table 5 contains parameter results from the baseline material PFP regressions using physical measures (coal, output, capacity as capital stock) where appropriate. I present four specifications: two for each quantity instrument (lagged power and coal input prices), combined with 2002 and 2004 as the choices for the treatment year. Given the new availability of plant-level price indices, I use this opportunity to test them as an instrument, in addition to lagged revenue which has achieved negative point estimates in the previous literature.

For all four specifications, the point estimate of the treatment effect is positive, and in columns (2) and (4) it is even significant at the 10% level. In the input price specifications, the coefficient on log output may be implausibly large. In columns (1) and (3), where the estimate is more reasonable the lower bound of the 95% confidence interval is under -2% in magnitude, which is less than half of the estimates found in the previous literature, as seen in table 7.

To summarize: the baseline specifications present very little evidence that restructuring caused state-owned plants to become disproportionately more efficient than private ones in terms of coal use. In fact, the available evidence is mildly suggestive toward increased **inefficiency**, though not precisely. In later robustness checks, I show that this finding is robust to many other instruments previously used in the literature, like 6-digit area code level production, 6-digit area code level employment, province-level revenue/output.

The first column of table 7 and table 5 together provide instructive comparisons for these results. While my sample differs substantially from that in Gao & Van Biesebroeck (2014) and Du *et al.* (2009), both the 2002 and 2004 restructuring year specifications find negative point estimates that are significant at the 5% level when I use financial data. Due to sample selection concerns, the most useful comparison is with the "large firms" specification from Gao & Van Biesebroeck (2014), which is of a similar magnitude and sign. My sample is actually **more** favorable toward a negative result using financial data than the previous literature, so the dramatic switch to significantly positive coefficients in some specifications suggests that the change is not just due to sampling noise. Note that my preferred choice of instrument, lagged revenue, also results in negative coefficients in previous papers, although as seen in robustness checks, any instrument commonly used in the previous literature achieves the same result.

Table 6: PFP Regressions Using Physical Measures - Employment

	(1)	(2)	(3)	(4)
Restruc x SOE	.077	.008	.038	-.05
	(.043)	(.043)	(.041)	(.040)
SOE	.007	.065	.044	.090
	(.052)	(.051)	(.048)	(.046)
Log Output	.636	-.281	.629	-.280
	(.275)	(.100)	(.276)	(.098)
Log Wage	.060	.185	.061	.184
	(.073)	(.074)	(.074)	(.074)
Log Capacity	-.342	.340	-.340	.339
	(.195)	(.094)	(.196)	(093)
Restructuring Year	2002	2002	2004	2004
Power Instrument	Lag Employment	Input Price	Lag Employment	Input Price
N	1,578	1,906	1,587	1,838
Plants	448	496	448	418

Notes: Dependent variable is logged physical production. Includes plant fixed effects and year fixed effects. Standard errors are clustered at the plant level. All first stage F statistics are above the 10% Stock-Yogo weak ID test threshold. Standard errors in parentheses. All regressions include year fixed effects. Plants with maximum capacity below 50 MW are omitted. "Lagged employment" uses twice lagged data for missing years.

Table 6 presents results for the employment specification. The previous literature is more mixed on the effects of restructuring on employment, but prior papers such as Du *et al.* (2009) find significant efficiency gains using two periods of data. At least one specification for the physical results, column (1) of Table 6, finds a significant **decrease** in efficiency for SOEs, but no other regression finds a precise result. While the employment results represent less of a nullification of previous findings, they do very little to show that SOEs improved in response to restructuring.

With the full physical and financial data, it is possible to do further diagnostics on what causes the discrepancy between the material regression results in Tables 5 and 7. In Table 8, I run four regressions to fully decompose the difference across the two versions of a baseline regression with 2002 as the chosen restructuring year and lagged output as the instrument. Columns (1) and (4) represent the original financial and physical regressions, respectively, while column (2) only uses physical data for output and column (3) only uses physical data for material use.

Table 8 clearly shows that only accounting for input price biases in materials does not change the qualitative results from using only expenditure/revenue data, but correcting for output price biases does. In fact, while the standard error is larger in column (2) than column (4), the highest positive point estimate comes from using

Table 7: Key PFP Results Using Financial Measures

Source	Materials	Employment
Table 5 (1) replaced with financial variables	-.078	.044
	(.0397)	(.037)
Table 5 (3) replaced with financial variables	-.073	.005
	(.026)	(.035)
Gao & Van Biesebroeck (2014) Large Firms 2004	-.060	-.016
	(.026)	(.027)
Du *et al.* (2009)	.0041	-.2939
	(.0055)	(.103)
Gao & Van Biesebroeck (2014) using Lagged Revenue 2004	-.039	-.067
	(.029)	(.028)

Notes: Du *et al.* (2009) is based off of financial data from 1995 and 2004 only. All estimates are from regressions of logged inputs on a difference-in-difference variable and other controls.

log physical output and log input expenditure together. Together, these results help explain the mechanisms behind the results in the previous literature and the reversal with new data: First, there were asymmetric input **and** output price shifts between SOEs and private plants in response to restructuring around 2002, which lead to biases in different directions. Second, the different sensitivities of input and output prices in Chinese electricity, as documented in Ma (2011), meant that the net effect was to overestimate the pro-competitive effects of restructuring.

Taken together, the two mechanisms suggest that it is valid in principle for the literature to have chosen ownership status as a treatment to evaluate China's market restructuring. The data supports the idea that outcome variables like prices did respond to SOE status, which makes it ex ante plausible that incentive structures may have also changes along these lines. The reforms that were successfully undertaken were only partial, however. With the best available physical data, there is little evidence that these reforms introduced meaningful competition into Chinese electricity. Instead, they mostly shifted around two sets of prices that likely remained under government control, and did not shift plant-level incentives for efficiency. Price shifts in the aggregate were documented in Ma (2011) and others, but it had not been established using plant-level evidence that the shift varied across ownership lines.

These findings bridge the discrepancy between the quantitative findings in the

Table 8: Regressions Measuring Which Price Causes More Bias

VARIABLES	(1) Input Expenditure	(2) Input Expenditure	(3) Materials	(4) Materials
Restruc x SOE	-.078	.052	-.10	.040
	(.040)	(.052)	(.041)	(.025)
Log Revenue	.871		.947	
	(.085)		(.069)	
Log Output		.762		.952
		(.094)		(.022)
Restructuring Year	2002	2002	2002	2002
N	1,391	1,618	1,984	3,144
Plants	411	456	489	544

Notes: Dependent variables are logged. Output is instrumented using lags. Includes plant and year fixed effects. Standard errors are clustered at the plant level. Standard errors in parentheses. Plants with maximum capacity below 50MW are omitted.

literature so far and the ongoing view that China has yet to significantly deregulate wholesale electricity markets established in Xu & Chen (2006), Liu (2013), and many others. While price changes are important, Joskow (2008) and many others explain that such reforms will not have their full effect unless a full process of restructuring is undertaken.

Notable robustness checks in Table 13, discussed further in the next section, include a variety of alternative instruments and sample specifications, such as including small plants, or using other instruments from Gao & Van Biesebroeck (2014) such as area-level production and employment.

5.1.1 Implications

If we take the weight of the PFP evidence to say that there is no effect on plant-level efficiency, this implies that any direct welfare effects due to physical efficiency changes are limited. However, it is worth noting that these PFP techniques using financial data have found significant implications in the past: Gao & Van Biesebroeck (2014) estimate that coal consumption would have been 27.4 million tons higher without restructuring, with an associated loss of 78 million tons of carbon dioxide.

For labor, they estimate a decrease in necessary workers of 22,100, which the PFP analyses using fully physical data also do not support.

5.2 TFP Results

5.2.1 Plant-level Regressions

Other papers in the literature have found efficiency gains using TFP residual estimation. The closest analysis to this in the literature is Ma & Zhao (2015), who also have physical data. This paper makes several different methodological choices: namely the use of the ACF method for estimate, the inclusion of labor as an input, and the possible inclusion of fixed effects in the residual analysis. Other notable entries in this literature include Zhao & Ma (2013) and Du *et al.* (2013).

Proxy methods like Ackerberg *et al.* (2015) have not been applied to Chinese electricity. Van Biesebroeck (2007) indicates that ACF may be the most appropriate option for this type of analysis compared to other estimation methods such as Stochastic Factor Analysis (SFA) and Data Envelope Analysis (DEA). The paper argues that DEA is "never the ideal method for estimating productivity growth", while SFA is most appropriate "when one has good reason to believe that productivity differences are constant over time", which there is no reason to suspect in this case.

Table 9 features two sets of results using logged TFP residuals (ω) from the Cobb-Douglas production function as dependent variables. Borrowing from Ma & Zhao (2015), I include a plant's entry status, SOE status, log age, and exposure to technology mandates as controls (represented by "Decommission" and "Entry"). All of these controls are also endogenized in the plant's productivity process in the initial ACF estimation, as recommended by De Loecker & Syverson (2021). "Decommission" refers to plants below 100 MW capacity being targeted for closure, and "entry" refers to plants that just entered in the current year[13]. Note that the PFP results are robust to the inclusion of these controls in Table 13 later in the paper.

[13] Because I do not have generator-level data, I can not test for the effects of the three technology mandates directly as in Ma & Zhao (2015). However, the mandates are indirectly captured by the two control variables. "decommission" captures the targeting of small plants and "entry" captures (partially) whether a plant has recently installed capacity, which was encouraged by the other two mandates.

Table 9: ACF Second Stage Results (Cobb-Douglas)

	Financial	Physical	Financial	Physical
Restruc x SOE	-.072	-.024	-.033	-.004
	(.081)	(.039)	(.061)	(.039)
SOE	-.116	-.019	-.000	.035
	(.087)	(.037)	(.103)	(.084)
Entry	-.399	-.341	-.345	-.345
	(.063)	(.043)	(.062)	(.045)
Decommission	-.483	-.044	-.040	-.142
	(.101)	(.040)	(.056)	(.052)
Log Age	.080	.004	.091	.003
	(.028)	(.008)	(.038)	(.027)
Log Capacity/Fixed Assets	-.035	-.018	-.406	-.026
	(.036)	(.008)	(.080)	(.027)
Restructuring Year	2002	2002	2002	2002
Plant FEs	No	No	Yes	Yes
N	1,866	2,522	1,866	2,522

Notes: Dependent variables are productivity measures from prior ACF estimation using a Cobb-Douglas specification. Standard errors in parentheses. Standard errors are clustered at the plant level. All regressions include year fixed effects. Excludes plants with maximum capacity below 50,000 MW.

In this model, a negative sign for the difference-in-differences coefficient corresponds to less efficiency. In all specifications, the treatment effect is never significantly positive, regardless of which dataset is used or whether fixed effects are included [14]. As explained earlier, the expected sign of the bias due to the inclusion of financial data is much more ambiguous in this case given the nonlinear estimation method. While the bias would appear to run slightly in the opposite direction of the PFP results, no specification has a positive point estimate. Thus, the TFP estimation method contributes to the lack of evidence that restructuring introduced competition and thus efficiency to the market. After incorporating proxy variable methods and labor as an input the differences-in-differences specification is unable to detect any particular efficiency gains for SOEs regardless of fixed effects or which type of data is used. I present results using a translog function form in the appendix, which similarly only generates negative results (including a specification that is significant at the 5% level).

The two technology mandates, to the degree that they can be measured, are consequential. "Entry" is significantly negative in every specification, while "de-

[14] As per De Loecker & Syverson (2021), fixed effects generally should not be included in ACF residual regressions, but the findings are qualitatively similar whether they are included or not.

Table 10: Key Previous TFP Results

Source	Estimate	Note
Ma & Zhao (2015)	.0305***	Data envelope analysis on full physical dataset
Ma & Zhao (2015)	-.0936***	Stochastic frontier analysis on full physical dataset
Zhao & Ma (2013)	.0474**	Data envelope analysis on a panel of 34 plants
Du *et al.* (2013) two inputs	-.19***	Stochastic frontier analysis, financial data
Du *et al.* (2013) three inputs	-.18*	Stochastic frontier analysis, financial data

Notes: "Full physical dataset" refers to the physical dataset from this paper. "Financial data" means only revenues and expenditure data were available. Negative estimates reflect efficiency improvements for stochastic frontier analysis, positive the same for DEA.

commission" is more sensitive to the choice of dependent variables and fixed effects. Thus, there is some evidence that the concurrent technology mandates had their desired effects, as decommissioned plants produced less, all else equal. The effects of entry policy are more difficult to determine, since entering plants would naturally produce less in annual data.

Table 10 surveys the previous TFP results from the literature. For all estimates that used data envelope analysis, the interaction on SOE and restructuring is significantly positive, and for those that use stochastic factor analysis it is negative. All reflect supposed efficiency improvements. As previously argued, the methodological choices of proxy variable methods and using labor as an input are supported in the productivity literature, so the results of this paper can be taken as a potential nullification of prior results, especially alongside the PFP analysis from earlier.

5.2.2 Decompositions

The plant-level regressions describe whether individual plants became more productive over time. While seemingly they did not, it is possible that restructuring lead to improved aggregate production. This question is not addressed in the previous literature on Chinese coal power. To my knowledge, only aggregate patterns in Chinese coal mining have been formally decomposed by Zhou *et al.* (2019).

Market restructuring could plausibly reallocate production from or to state-owned plants. The former follows similar logic for why they should become more productive on the intensive margin: before restructuring they did not face competitive pressure from private firms and just produced at whatever level was assigned to them. If production becomes more freely allocated, then private plants may produce at higher levels if state policies were unfairly favoring state-owned plants. As to why state-owned plants may gain in their production share if competition and efficiency are improved: they tend have lower marginal costs (about 7% lower on average, significant at the 1% level) through the entire sample, so a net gain by SOEs could result in aggregate improvement.

Given a distribution of productivity ω_{it} and shares of production s_{it}, one can generate aggregate measures of productivity, its changes over time, and the relative contributions of different groups to the aggregate figures. The Olley & Pakes (1996) decomposition quantifies the role of both individual plant productivity and the allocation of production in determining aggregate productivity. Here I apply both their decomposition and the methodology developed in Melitz & Polanec (2015) that extends it to incorporate the contributions of entering and exiting firms. The latter works off of pairs of production periods: entrants E are firms present only in the first period, exiters X are only in the second, and survivor S s are present in both.

A firm's production share $s_{it} = \frac{q_{it}}{Q_t}$ is the relative amount of total production it accounts for. Let $s_{Gt} = \sum_{i \in G} s_{it}$. The central terms in the decomposition will be $\Phi_{Gt} = \sum_{i \in G}(s_{it}/s_{Gt})\omega_{it}$. These are group-level measures of aggregate productivity. Define overall aggregate productivity to be $\sum_i (s_{it})\omega_{it}$ and select two years (1 and 2) for comparison. With groups S for survivors, X for exiters, and E for entrants:

$$\phi_1 = s_{S1}\Phi_{S1} + s_{X1}\Phi_{X1} \tag{10}$$

and

$$\phi_2 = s_{S2}\Phi_{S2} + s_{E2}\Phi_{E2} \tag{11}$$

That is, aggregate productivity in the first period is composed of contributions from firms who are either in both periods or the first only. Aggregate productivity in the second period can only be from firms present in either both periods or only the second.

The Olley & Pakes (1996) decomposition uses an equivalent figure of Φ for aggregate productivity. This is split into two simple terms: the covariance between a

Table 11: Log Productivity Percent Changes Relative to 1998 (OP Decomposition)

Year	Total		Mean		Covariance	
	TFPR	TFPQ	TFPR	TFPQ	TFPR	TFPQ
2000	-1.2	1.1	0.2	-2.5	-1.4	3.7
2002	2.8	-5.2	5.1	-12.9	-1.4	7.8
2003	4.8	-14.5	4.9	-18.7	-0.1	4.3
2004	8.3	-19.1	7.6	-18.5	0.7	-0.5
2005	10.0	-18.6	11.5	-19.9	-1.5	1.3
2006	11.9	-16.6	13.0	-18.6	-1.1	2.0
2007	13.6	-14.9	14.2	-18.6	-0.7	3.7

Notes: Aggregate productivity is determined by the weighted sum of individual productivities. Shares are weighted by output. Terms are based on the ω residual from the logged version of a Cobb-Douglas production function. "TFPR" refers to productivity from revenue-based estimation, while "TFPQ" is from physical measures.

firm's production share and its productivity, and the average unweighted productivity across all firms. The former measures the role of allocation, while the latter measures the role of average productivity regardless of production. The key distinction between the OP and MP decompositions is the treatment of entry and exit: in OP, the productivity of every firm is considered together, regardless of its status. As explained in Backus (2020), hypotheses about productivity and competition do not necessarily include or exclude entering or exiting firms. Table 11 the results of the OP decomposition, with the terms converted into changes since 1998:

Because these are all changes relative to 1998, a negative should be interpreted to mean that there a decline in term as of the date in question. The two types of productivity do not have directly comparable units, so we can only look at their trajectories. In Table 11, we can see that physical productivity predicts a massive overall decline from 1998-2007, while revenue-based measures predict growth. Both trends are driven, however, by the "Mean" term, which does not capture reallocation. The OP decomposition results are not consistent with any substantial efficient reallocation of production for either TFPR or TFPQ.

It is important to note that the drastic decline overall in TFPQ is happening despite some observed effects from input price deregulation and technology mandates earlier in the paper. The evidence is consistent with the idea that without successful market liberalization, or at least some realigning of the incentives of plants, other reforms

Table 12: Log Productivity Percent Changes Relative to 1998 (MP Decomposition)

Year	Total		Survivors		Entering		Exiting	
	TFPR	TFPQ	TFPR	TFPQ	TFPR	TFPQ	TFPR	TFPQ
2000	-1.2	1.1	1.5	-3.4	-1.6	2.5	-1.1	2.0
2002	2.8	-5.2	3.5	-3.7	0.1	-1.8	-0.8	0.3
2003	4.8	-14.5	6.3	-16.0	-1.4	0.5	-0.1	1.0
2004	8.3	-19.1	10.1	-20.5	-2.7	2.6	0.8	-1.0
2005	10.0	-18.6	12.3	-20.4	-2.9	2.8	0.6	-1.0
2006	11.9	-16.6	13.1	-17.2	-2.3	1.3	1.0	-0.7
2007	13.6	-14.9	13.7	-13.3	-1.0	1.0	1.0	-2.6

Notes: Aggregate productivity is determined by the weighted sum of individual productivities. Shares are weighted by output. Terms are based on the ω residual from the logged version of the translog production function. "TFPR" refers to productivity from revenue-based estimation, while "TFPQ" is from physical measures.

are not translating into aggregate efficiency on their own. For any of the reforms to be successful, it is likely the case that they need to be undertaken together, as argued in in Joskow (2008).

While it is difficult to say whether the decline in TFPQ is due to restructuring directly or not, we can further decompose some of the sources. Table 12 features results from running the MP decomposition on every year in the sample with 1998 as the first year to highlight the role of entry and exit.

The dramatic TFPQ decline is largely accounted for by surviving plants. While there are likely many explanations, a response to the effects of restructuring would make sense: price incentives, net entry, and the administrative structure changed significantly around 2002. Power plants may take years to adjust to the new regulatory environment and uncertainty. This result also aligns with qualitative reports of the time: 2003-2006 is notably seen as a time of widespread power outages across China due to supply shortages, for example in Shunkun *et al.* (2013). If power plants are operating at near full capacity to meet demand, their productivity may suffer as their marginal cost curves rise near capacity constraints. While these supply shortages may not have been caused by restructuring itself, the TFPR trends on their own provide very little corroborating evidence of the widespread power shortages. This may be because the government enacted relevant price changes to offset these effects.

5.2.3 Implications

According to Gao & Van Biesebroeck (2014), aggregate statistics suggest that coal intensity for the electricity sector as a whole declined from 2002 to 2006, while the share of coal plants in total generation increased. This would suggest that in the aggregate, there had to be some form of productivity gain. The results thus far suggest that it is not due to any gains among surviving plants, and among all plants no gains are driven on the intensive margin be SOEs. One drawback of the proxy variable method to TFP estimation is it is extremely difficult to calculate counterfactual paths of productivity and input use, so it is difficult to say exactly how much coal or labor could have been saved in the aggregate under alternative policies.

This leaves only a few plausible channels for the suggested productivity growth seen in other sources. First is technological innovation or increased scale from entering plants during the period. The MP decomposition supports this to some extent: entering plants contribute to productivity gains in all years after 2002. The decomposition does not calculate the cumulative gain from entering plants across all years since it only compares 2 years at a time, but it is possible that by 2006 this contribution was large. After 2002, entering plants are significantly larger in their capacity (by about 50 MW) than the average incumbent plant. The other possibility is that there are large productivity gains from the elimination of plants under 50 MW, which are not included in the decomposition. This hypothesis is supported by the findings in Ma & Zhao (2015). Thus, the measured productivity losses among surviving plants may possible to reconcile with other aggregate statistics when considering entry and exit policy at the time.

6 Extensions and Robustness Checks

6.1 PFP Results

The top of Table 13 provides a series of alternative coefficient and standard error values for the restructuring term for 2002. Checks include OLS, using small plants, including the full set of price heterogeneity controls from Gao & Van Biesebroeck (2014), and using only a series of firms matched in the covariate space via the Mahalanobis metric. I also exclude years for which SOE values were imputed, and explore alternative state ownership definitions and measurement. In the case of the materials regressions, all but two specifications rule out even a 4% effect at the 95% level, while a strong majority rule out even 3%, both of which are well under the main findings in the literature. Employment is significantly noisier, but there are also now several specifications that are significantly **positive** at the 10% level.

I also investigate alternative instruments, given that many have been tested in the literature. No instrument offers a significantly negative, even at the 10% level, coefficient on either input. For any specification that passes both weak instrument and overidentification tests, the coefficient is positive. In particular, the 6 digit area code employment and production variables have been used in Gao & Van Biesebroeck (2014) and found significantly negative results. Here, while these instruments do not strongly pass a weak instrument test, the materials coefficient is actually significantly positive at the 10% level. Combined with the evidence from the baseline specifications, the robustness checks further confirm that there is very little evidence that restructuring produced relative efficiency gains. Results are broadly similar for 2004 (available upon request).

Table 13: Robustness Checks for $STATE_0 * POST2002$

Dependent variable:	(1) ln(COAL)	(2) ln(EMPLOYMENT)
OLS	.009	.030
	(.017)	(.039)
OLS no plant FEs	.003	.022
	(.022)	(.067)
Including firms under 50 MW capacity	.028	.054
	(.020)	(.046)
Including technology mandate controls	.035	.073
	(.028)	(.042)
Including controls for price heterogeneity	.021	.092
	(.021)	(.048)
Excluding imputed SOE years	.049	.078
	(.026)	(.043)
Mahalanobis matching	.043	.127
	(.030)	(.065)
SOE defined by "big 5" status	.011	−.042
	(.025)	(.043)
SOE Status does not vary	.025	.054
	(.025)	(.052)
Alternative Instruments		
Log 6 digit area code employment and production	.111[#]	.026[#]
	(.062)	(.049)
Log input prices and lagged power	.063[!]	.052[!]
	(.034)	(.038)
Log input prices and twice lagged power	.075[!]	.053[!]
	(.044)	(.039)
Estimated in first differences with twice lagged revenue as instrument	-.176	.007
	(3.43)	(.057)
Market production	.054[!]	.127
	(.030)	(.196)
Market revenue	.055	.058
	(.040)	(.052)
Lagged and twice lagged power	.020[!#]	.070[#]
	(.023)	(.044)
Lagged power and log 6 digit area code total production	.038[!#]	.078[#]
	(.028)	(.043)

Notes: Dependent variables are logged. Power is instrumented with lags unless otherwise specified. Includes plant fixed effects. Standard errors are clustered at the plant level. ! Cragg-Donald F statistic passes Stock-Yogo test (true for all specifications in upper half of table). # overidentification test succeeds at 5%. "Mahalanobis matching" refers to only using SOE and private plants matched on the set of independent variables in the regression. "Big 5" refers to the 5 major state-owned companies that got broken up in 2002. Standard errors in parentheses.

6.2 Gross Production Functions for TFP

All of the estimated functions thus far give very small elasticities for labor. The Leontief assumption may overstate the true lack of substitutability of labor compared to capital and coal use in the power plant's process. A gross output production function that estimates the elasticity with respect to materials may serve as a sensitivity test for the Leontief assumption.

Gandhi *et al.* (2020) have argued proxy variable methods that use materials as their proxy may not be able to identify an elasticity for materials without further assumptions, and the authors develop a new estimator that relies on profit-maximization assumptions by firms. Their estimator can be summarized briefly in three steps: First, given a functional form for production, take a plant's first order condition with regard to materials to estimate the materials elasticity as a function of their input expenditure share. Second, with this information in hand, set up a partial differential equation that relates the integral over materials of the first order condition [15] to the production function and a constant of integration that relies on the other inputs (capital and labor). Third, use Markov assumptions of productivity as in Ackerberg *et al.* (2015) to form moments based on current and lagged inputs that can identify the remaining production function parameters. The GNR results in Table 14 provide analogous TFP residuals to ACF in table 21: a positive number means a productivity increase, a negative number means a productivity decrease. Estimated elasticities are available in the appendix.

[15] In the case of Cobb-Douglas, the partial derivative that gets integrated is just a constant, which will result in an estimate of m_{it} multiplied by the estimated materials elasticity for every plant.

Table 14: GNR Second Stage Results

Restruc x SOE	-.033	-.029
	(.021)	(.028)
Entry	-.034	-.013
	(.021)	(.021)
Decommission	.037	.002
	(.027)	(.025)
Log Age	.017	.049
	(.009)	(.025)
Log Capacity	.006	-.075
	(.006)	(.023)
Plant FEs	No	Yes
Restructuring Year	2002	2002
N	2,864	2,864

Notes: Dependent variables are productivity measures from prior GNR estimation using physical output. Standard errors in parentheses. Standard errors are clustered at the plant level. Includes year fixed effects.

Neither specification finds a significant improvement for treated plants, and the point estimate for output is over 3 times more negative than that for revenue. Thus, the results from a Cobb-Douglas gross output production function are consistent with the rest of the paper. The results are also robust to using stochastic factor analysis using Greene's "true fixed effects" method (Greene, 2005) [16]. Thus, neither the exact assumptions of the ACF estimator, or the Leontief assumption, appear to be driving any particular set of results in the paper.

7 Conclusion

I have demonstrated how the series of reforms undertaken in China's electricity sector in 2002 did not incentivize power plant-level or aggregate efficiency through competition. In turn, I find no simultaneous welfare increases or pollution decreases as in the previous literature. This was done using two methods. The first was the difference-in-differences method pioneered by Fabrizio *et al.* (2007), modified for China by Gao & Van Biesebroeck (2014), and augmented with new data on prices

[16] Results available upon request.

and physical inputs and outputs. For these partial factor productivity methods, the results apply to both materials and labor inputs, though the results for these differed in the previous literature.

Newly measurable patterns in input and output prices show that for these outcomes, there were observable shifts across ownership categories, and not in any measure of technical efficiency. This provides a nuanced interpretation of the PFP results in the previous literature: while measurement error may have lead to the overstating of any pro-competitive effects of the market restructuring, state-owned plants were a reasonable choice of treatment group that did receive special exposure to the 2002 policies. However, the substantial welfare gains found in the previous literature from shifted coal use and employment do not appear to have been realized.

The second major set of findings involved full TFP residual estimation using a series of proxy variable methods. While the exact direction of the bias from missing prices is much more difficult to predict using this method, specifications that use physical data fail to find any positive results regarding efficiency gains. In addition, I find that concurrent technology mandates had a significant effect on plant-level productivity measures, but their measurement does not affect the results regarding the 2002 reforms.

Finally, productivity decompositions following from the TFP estimation also established that there was no aggregate effect via the reallocation of production to more efficient plants. In fact, revenue-based measures find dramatically more productivity growth than output-based ones. Any aggregate productivity losses may not necessarily be due to restructuring, but the findings indicate there are not efficiency gains to be found in the data at the plant or industry level.

Both the plant-level and aggregate evidence are consistent with the qualitative findings on China at this time. The prevailing writing today among think tanks and government sources today is that 2002 featured only a partial implementation of China's intended reforms, and that very little competition was introduced. It is clear that in the present, wholesale electricity prices are not completely market-driven in China. The aggregate findings also line up with retrospective writing about the period that my sample covers: 2003-2006 was a period of widespread power shortages and blackouts in China, which exactly lines up to the most severe aggregate productivity losses that I measure.

As the country's focus shifts more toward climate change and emission reduction, it is possible that introducing competition into the coal power sector may become less important than using less coal entirely. However, as China is still extremely dependent on coal, efficient use of inputs both within power plants and in the aggregate are important policies to consider moving forward. The role of transmission infrastructure, like Ryan (2014) writes about in India, would also be an exciting direction for future research as it is generally acknowledged as a significant bottleneck and focus of the current Chinese government.

8 Acknowledgments

This work is derived from research related to my second dissertation chapter. I am thankful to my dissertation committee Panle Jia Barwick, Julieta Caunedo, Jean-Francois Houde, and Shanjun Li. I am additionally thankful to Francisco Garrido, Todd Gerarden, Miriam Larson-Koester, Debi Prasad Mohapatra, Mar Reguant, Deyu Rao, Ivan Rudik, Flavio Stanchi, and participants at the 2020 EAERE and 2021 MEA conferences. I am also grateful to the Cornell East Asia Program for financial support from the Hu Shih Fellowship in Chinese Studies.

References

Ackerberg, Daniel A, Caves, Kevin, & Frazer, Garth. 2015. Identification Properties of Recent Production Function Estimators. *Econometrica*, **83**(6), 2411–2451.

Backus, Matt. 2020. Why Is Productivity Correlated with Competition? *Econometrica*, **88**(6), 2415–2444.

Bond, Steve, Hashemia, Arshia, Kaplan, Greg, & Zoch, Piotr. 2020. Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data.

Brandt, Loren, Van Biesebroeck, Johannes, & Zhang, Yifan. 2014. Challenges of working with the Chinese NBS firm-level data. *China Economic Review*, **30**, 339–352.

Brandt, Loren, Van Biesebroeck, Johannes, Wang, Luhang, & Zhang, Yifan. 2017. WTO Accession and Performance of Chinese Manufacturing Firms. *American Economic Review*, **107**(9), 2784–2820.

Cicala, Steve. 2015. When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation. *American Economic Review*, **105**(1), 411–444.

Cicala, Steve. 2022. Imperfect Markets vs Imperfect Regulation in US Electricity Generation. *American Economic Review*, **112**(2), 409–441.

De Loecker, Jan, & Syverson, Chad. 2021. An Industrial Organization Perspective on Productivity. *NBER Working Paper 29229*.

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

Du, Limin, Mao, Jie, & Shi, Jinchuan. 2009. Assessing the impact of regulatory reforms on China's electricity generation industry. *Energy Policy*, **37**(2), 712–720.

Du, Limin, He, Yanan, & Yan, Jianye. 2013. The effects of electricity reforms on productivity and efficiency of China's fossil-fired power plants: An empirical Analysis. *NBER Working Paper 24199*, **40**, 804–812.

Fabrizio, Kira R, Rose, Nancy L, & Wolfram, Catherine D. 2007. Do Markets Reduce Costs ? Assessing the Impact of on US Electric Generation Regulatory Restructuring Efficiency. *American Economic Review*, **97**(4), 1250–1277.

Foster, Lucia, Haltiwanger, John, & Syverson, Chad. 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, **98**(1), 394–425.

Gandhi, Ami, Navarro, Salvador, & Rivers, David A. 2020. On the Identification of Gross Output Production Functions. *Journal of Political Economy*, **128**(5), 2973–3016.

Gao, Hang, & Van Biesebroeck, Johannes. 2014. Effects of deregulation and vertical unbundling on the performance of China's electricity generation sector. *Journal of Industrial Economics*, **62**(1), 41–76.

Gowrisankaran, Gautam, Greenstone, Michael, Horacsu, Ali, Liu, Mengi, Shen, Caixia, & Zhang, Bing. 2021. Discharge Fees, Pollution Mitigation, and Productivity: Evidence from Chinese Power Plants.

Greene, William. 2005. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, **126**, 269–303.

Grieco, Paul, Li, Shengyu, & Zhang, Hongsong. 2016. Production Function Estimation with Unobserved Input Price Dispersion. *International Economic Review*, **57**(2), 665–690.

Haltiwanger, John, Kulick, Robert, & Syverson, Chad. 2018. Misallocation Measures: The Distortion That Ate the Residual. *NBER Working Paper 24199*.

Han, Jin Soo, Houde, Jean-Francois, van Benthem, Arthur A, & Abito, Jose Miguel. Forthcoming. Robust Estimates of the Effect of Power Plant Divestitures on Fuel Procurement, A Comment on Cicala (2015). *American Economic Review*.

Ho, Mun S., Wang, Zhongmin, & Yu, Zichao. 2017. China's Power Generation Dispatch. *Resources for the Future*.

Hsieh, Chang-Tai, & Klenow, Peter. 2009. of Economics. *Quarterly Journal of Economics2*, **CXII**(November), 1–55.

Joskow, Paul. 2008. Lessons Learned from Electricity Market Liberalization. *The Energy Journal*, **Special volume**, 9–42.

Kahrl, Fredrich, Williams, James, & Hu, Junfeng. 2013. The Political Economy of Electricity Dispatch Reform in China. *Energy Policy*, **53**, 361–369.

Klette, T.J., & Griliches, Zvi. 1996. The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous. *Journal of Applied Econometrics*, **11**, 343–361.

Liu, Ming-Hua, Margaritis, Dimitris, & Zhang, Yang. 2013. Market-driven coal prices and state-administed electricity prices in China. *Energy Economics*, **40**, 167–175.

Liu, Zhenya. 2013. *Electric Power and Energy in China*. Wiley.

Ma, Chunbo, & Zhao, Xiaoli. 2015. China's Electricity Market Restructuring and Technology Mandates: Plant-level evidence for changing operational efficiency. *Energy Economics*, **47**, 227–237.

Ma, Jinlong. 2011. On-Grid Electricity Tariffs in China: Development, Reform and Prospects. *Energy Policy*, **3**, 2633–2645.

Melitz, Marc J., & Polanec, Saso. 2015. Dynamic Olley-Pakes productivity decomposition with entry and exit. *RAND Journal of Economics*, **46**(2), 362–375.

Newbery, David M., & Pollitt, Michael G. 1997. The Restructuring and Privatisation of Britain's Cegb – Was It Worth It? *Journal of Industrial Economics*, **45**(3), 269.

Olley, Steven, & Pakes, Ariel. 1996. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, **64**(6), 1263–1297.

Ornaghi, Carmine. 2006. Assessing the Effects of Measurement Errors on the Estimation of Production Functions. *Journal of Applied Econometrics*, **21**, 879–891.

Rambachan, Ashesh, & Roth, Jonathan. 2022. A More Credible Approach to Parallel Trends. *Mimeo*.

Ryan, Nicholas. 2014. The Competitive Effect of Transmission Infrastructure in the Indian Electricity Market. *Working Paper*.

Shunkun, Yu, Lisha, Zhou, & Chen, Li. 2013. China Wrestles With Power Shortages. May.

Van Biesebroeck, Johannes. 2007. Robustness of Productivity Estimates. *Journal of Industrial Economics*, **55**(3), 529–569.

Watts, Jonathan. 2003. Shanghai on nights to avoid blackout. December.

Wei, Yi-Min, Chen, Hao, Chyong, Chi Kong, Kang, Jia-Ning, Liao, Hua, & Tang, Bao-Jun. 2018. Economic dispatch savings in the coal-fired power sector: an empirical study of China. *Energy Economics*, **74**, 330–342.

Xu, Shaofeng, & Chen, Wenying. 2006. The Reform of Electricity Power Sector in the PR of China. *Energy Policy*, **34**, 2455–2465.

Zhang, A., Zhang, Y., & Zhao, R. 2001. Impact of Ownership and Competition on the Productivity of Chinese Enterprises. *Journal of Comparative Economics*, 327–346.

Zhao, Xiaoli, & Ma, Chunbo. 2013. Deregulation, vertical unbundling and the performance of China's large coal-fired power plants. *Energy Economics*, **40**, 474–483.

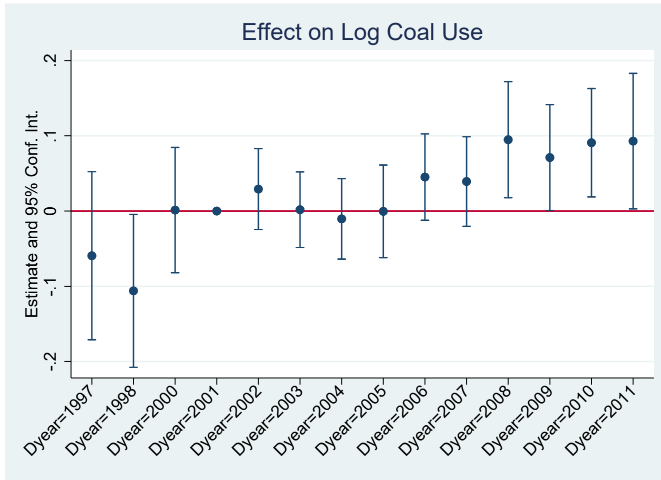
Zhou, Lin, Li, Jianglong, Dan, Yangqing, Xie, Chunping, Long, Houyin, & Lin, Hongxun. 2019. Entering and Exiting: Productivity Evolution of Energy Supply in China. *Sustainability*, **11**.

9 Appendix

9.1 Pre-Trend Analysis

Below are the annual treatment effect estimates from an event-study model of the partial factor productivity regression in column (1) of Table 5:

Figure 4: Coefficients by Year For Materials PFP Regression



Notes: Presents annual coefficient estimates for treatment effects from the material use partial factor productivity regression. Log power is instrumented using lags. Includes plant and year fixed effects.

While there is not a stark visual pre-trend before 2002, there is at least one significantly negative point estimate in 1998 and a steep decline between 1997 and 1998. On average the pre-2002 points seemingly do not violate the parallel trends assumption, but to check the results' robustness I also run the "honest DID" approach from Rambachan & Roth (2022).

The "honest DID" approach works the following way: estimate the worst violation of parallel trends in the pre-treatment part of the sample between two consecutive years, called δ. Next, fix a constant $\bar{M}$ and bound the treatment effect above and below assuming the post-treatment observations violate parallel trends (in either direction) by no more than $\bar{M}\delta$. In this case I restrict the estimate to the 2002 effect.

Table 15: Rambachan & Roth (2022) Bounds - Coal Use

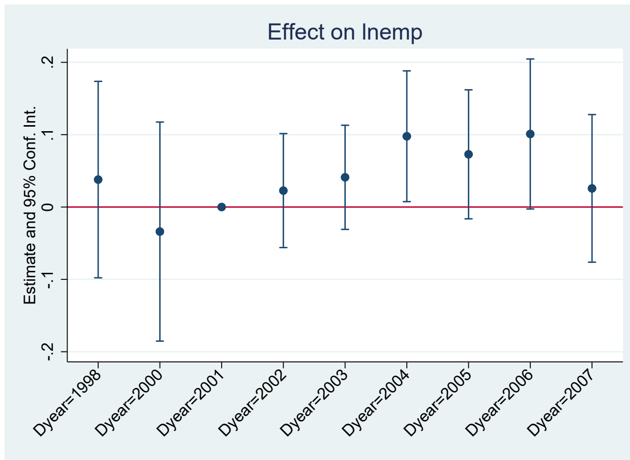
$\bar{M}$	Lower	Upper
0	-.025	.083
.25	-.048	.107
.5	-.082	.143
.75	-.121	.183
1	-.163	.225
1.25	-.206	.268
1.5	-.249	.311

Notes: Bounds represent results from Rambachan & Roth (2022) method used on the coal use PFP regressions for the treatment effect in 2002. $\bar{M}$ represents how much SOE and private plants can violate trends post-treatment as a function of the worst-case scenario estimated trend pre-treatment. For example, $\bar{M} =$ represents the estimate with no parallel trends, and $\bar{M} = 1$ represents bounds where the most extreme change in slope between two points prior to treatment is assumed to be the trend difference (positive or negative).

These bounds are unfortunately imprecise about the robustness of a null result since similar amounts will be added to both ends of the interval as $\bar{M}$ increases. However, it can be useful to examine intervals which include estimates from the previous literature, which start at values of $\bar{M}$ around .25. These intervals also admit as a possibility that there were double digit increases in SOE inefficiency in 2002. Thus, they confirm the assessment from earlier in the paper that the empirical evidence is at best uncertain for the pro-efficiency effects of restructuring.

Below I repeat the two exercises for the employment regressions:

Figure 5: Coefficients by Year For Employment PFP Regression



Notes: Presents annual coefficient estimates for treatment effects from the material use partial factor productivity regression. Log power is instrumented using lags. Includes plant and year fixed effects.

Table 16: Rambachan & Roth (2022) Bounds - Employment

$\bar{M}$	Lower	Upper
0	-.056	.101
.25	-.078	.115
.5	-.113	.145
.75	-.158	.187
1	-.206	.234
1.25	-.254	.280
1.5	-.304	.330

Notes: Bounds represent results from Rambachan & Roth (2022) method used on the employment PFP regressions for the treatment effect in 2002. $\bar{M}$ represents how much SOE and private plants can violate trends post-treatment as a function of the worst-case scenario estimated trend pre-treatment. For example, $\bar{M} =$ represents the estimate with no parallel trends, and $\bar{M} = 1$ represents bounds where the most extreme change in slope between two points prior to treatment is assumed to be the trend difference (positive or negative).

A very similar picture emerges as in the materials regressions, though with somewhat more uncertainty and a shorter timeframe.

9.2 Additional Production Function Estimates

9.2.1 Cobb-Douglas

The elasticity results from the Cobb-Douglas TFP estimation are below:

Table 17: ACF Production Function Parameters - Cobb-Douglas

VARIABLES	(1) Physical	(3) Financial
k	.975	.552
	(.003)	(.045)
l	.052	.247
	(.004)	(.047)
N	2522	1866

Notes: Estimates are derived from a full translog production function using methods from Ackerberg *et al.* (2015). Log age, decommission status, entry, SOE status, and SOE x restructuring are endogenized in the TFP process. "Physical" refers to using physical input, output, and capacity variables, "Financial" refers to using revenues, input expenditures and capital stock.

9.2.2 Translog

Table 18 shows the second-stage residual regression results using a translog specification. The point estimates are far more negative using translog than Cobb-Douglas, and the financial specification is slightly more negative than the physical one in this case too. As with Cobb-Douglas, the precise degree and direction of bias due to the purging of both input and output prices is extremely difficult to predict, in addition to movement from the methodological and sample changes compared to the previous literature.

Table 18: ACF Second Stage Results (Translog)

	Financial	Physical	Financial	Physical
Restruc x SOE	-.147	-.057	-.085	-.031
	(.069)	(.046)	(.055)	(.040)
SOE	-.185	.049	-.085	.064
	(.077)	(.099)	(.055)	(.087)
Entry	-.278	-.266	-.331	-.330
	(.063)	(.045)	(.062)	(.046)
Decommission	-.292	-.053	-.093	-.140
	(.076)	(.050)	(.058)	(.054)
Log Age	-.048	-.112	.065	-.016
	(.024)	(.012)	(.039)	(.028)
Log Capacity/Fixed Assets	.169	.028	-.018	.055
	(.021)	(.013)	(.024)	(.029)
Restructuring Year	2002	2002	2002	2002
Plant FEs	No	No	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
N	1,866	2,522	1,856	2,522

Notes: Dependent variables are productivity measures from prior ACF estimation using a translog specification. Standard errors in parentheses. All regressions include controls for log age and l

Table 19 features the elasticities from the Ackerberg *et al.* (2015) estimation using a translog production function. SOE status, entry, and log age are included as inputs. Much like in the Cobb-Douglas specification in the body of the paper, controls are endogenous to the productivity process in estimation.

Table 19: ACF Production Function Elasticities - Translog

VARIABLES	(1) Physical	(3) Financal
Log Capacity	.87	
Log Employment	.35	.679
Log Fixed Assets		.315
F for higher order terms	467.3	28,000
N	2522	1866

Notes: Estimates are derived from a full translog production function using methods from Ackerberg *et al.* (2015). Elasticities are reported at median input values. Log age, decommission status, entry, SOE status, and SOE x restructuring are endogenized in the TFP process. "Physical" refers to using physical input, output, and capacity variables, "Financial" refers to using revenues, input expenditures and capital stock.

Table 20 shows the actual parameters from the translog specification to give a sense of how elasticities may vary over the distribution of labor and capital. The value added specification tends to attribute higher returns to scale at higher levels of capital and labor, while the effect is more ambiguous using physical measures.

Table 20: ACF Production Function Parameters - Translog

VARIABLES	(1) Physical	(3) Financial
k	1.06	-.683
	(.002)	(.001)
l	-.082	.024
	(.003)	(.001)
kl	.010	-.047
	(.004)	(.0001)
k^2	-.01	.051
	(.002)	(.002)
l^2	.025	.101
	(.002)	(.001)
N	2522	1866

Notes: Estimates are derived from a full translog production function using methods from Ackerberg *et al.* (2015). Log age, decommission status, entry, SOE status, and SOE x restructuring are endogenized in the TFP process. "Physical" refers to using physical input, output, and capacity variables, "Financial" refers to using revenues, input expenditures and capital stock.

9.2.3 Gross Production Functions

Table 21 features the production function estimates using the Gandhi *et al.* (2020) method. Results are extremely close to constant returns to scale for the physical production specification, although the coefficient on labor is extremely low. As argued earlier in the paper, a gross production function approach that relies on first-order conditions may be less appropriate than a value added one with broader assumptions about input optimization, especially in contrast to the PFP models. However, the qualitative results based on productivity residuals change very little across methods.

Table 21: GNR Production Function Results (Cobb-Douglas)

	(1)
Log Capital	.389
	(.043)
Log Labor	-.04
	(.059)
Log Materials	.674
	(.003)
N	1,874

Notes: Assumes restructuring happens in 2002. Output equation uses physical measures. Standard errors in parentheses. Results come from matched financial-physical dataset.

9.3 Aggregate Comparisons

Table 22 compares estimated total production from coal power plants to aggregate data collected by the IEA for available years. The reported 85% of production figure is within 3% in the available years. This, paired with the aggregate pollution analysis on data from the same source in Gowrisankaran *et al.* (2021), shows that the physical input and output dataset closely matches all available public sources. Thus, while the dataset has not been independently verified on the individual observation level, the best available estimates suggest that it is trustworthy.

Table 22: Sample vs IEA Production Estimates (TWh)

Year	Sample (N)	IEA (MW)	%
2001	915 (2000 data)	1115	82
2004	1490	1713	86
2005	1691	1972	86
2008	2352	2733	86
2009	628.1	2913	86
2010	672.0	3273	88

Notes: Table shows comparisons between all production in original dataset (before sample exclusions) and figures for total coal electricity generation in yearly reports from the International Environmental Agency. IEA reports accessed variously through the University of Delaware library's access to the OECD archives, the wayback machine at the internet archive, or http://observ.nucleaire.free.fr