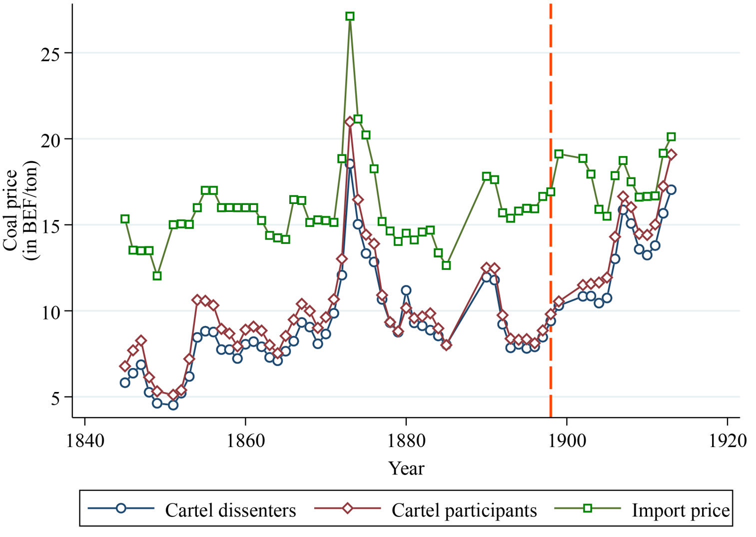

FIG. 2.—Prices and the Liège coal cartel, 1845–1913. The dashed vertical line represents the start of the coal cartel, the Syndicat de Charbonnages Liégeois.

The effect of this cartel can be clearly seen by comparing the Liège coal price to the import price of coal in Belgium.[11] We plot this import price in figure 2. Up to 1897, the Liège coal price was well below the import price of coal at the Belgian border. Following the cartel introduction in 1897, the Liège coal price increased up to the level of imported coal. A cartel would not price above this import price, as this would induce coal buyers to substitute toward imported coal. The cartel also seems to have had implications for the cost share of labor: as was shown in figure 1*B*, the cost share of labor dropped after 1897, indicating that the cartel could have had labor market implications as well. We will examine this hypothesis in the empirical model.

III. Empirical Model

In this section, we set up an empirical model of labor supply and demand to identify collusive conduct by Belgian coal firms. Our approach consists in comparing wage markdown estimates from a production model, which does not impose conduct assumptions on the labor market, with wage markdown bounds that are derived from a labor supply model, both in the absence of collusion and under fully collusive employer behavior.

[11] This import price is computed as total value of imported coal at the border divided by imported quantity of coal; hence, it includes transport costs from foreign mines to the border.

A. *Production Function*

We start with a model of coal production. Output Q_{ft} indicates the tonnage of coal extracted during a given year t by firm f. In this analysis, we assume coal to be a homogeneous product, as there is generally limited variation in coal quality. To do so, we sum the output of coal across the different coal categories of caloric content that the historical sources differentiate.[12] Mines often extracted a combination of these coal types, which are a function of the geological characteristics of the mine's location. We argue that this is innocuous because the caloric content of coal does not affect mining productivity.[13]

Firms use two variable inputs: L_{ft}, which captures the amount of effective labor throughout the year, and the amount of intermediate inputs purchased, M_{ft}. The capital stock consists of steam engines used for water pumping, coal hauling, and ventilation. The value of total capital used at each mine is denoted K_{ft}. Logarithms of variables are denoted in lowercase. As our baseline specification, we assume a Cobb-Douglas production function (1) with output elasticities β^{l}, β^{m}, and β^{k}, and log total factor productivity ω_{ft}:

$$q_{ft} = \beta^{\mathrm{l}} l_{ft} + \beta^{\mathrm{m}} m_{ft} + \beta^{\mathrm{k}} k_{ft} + \omega_{ft}. \tag{1}$$

We specify a Cobb-Douglas production function in labor, capital, and materials, rather than specifying a production function with nonsubstitutable material inputs, because material inputs were to some extent substitutable.[14] For materials and labor, this can be illustrated with the example of mine tunnel excavation, an important activity in nineteenth-century coal production. Firms can vary their materials-to-labor usage ratio by digging tunnels using explosives to open up new areas for coal extraction, or by relying more heavily on labor.[15] We assume that the output elasticities $\boldsymbol{\beta}$ are constant over time, which we relax in appendix section C.1.4. In appendix section C.1.3, we extend the production model to a more flexible

[12] Four coal types are distinguished in the dataset based on volatile content percentiles: 13%–18% (houille maigre sans flamme; i.e., anthracite coal), 18%–26% (houille sèche courte flamme), 26%–32% (houille maigre longue flamme), and >32% (houille grasse longue flamme). The first type was mainly used by households for heating purposes, the second for powering steam engines, and the latter two types for railroad locomotives.

[13] To assert this assumption, we regress the estimated total factor productivity (TFP) residual on the share of high-quality coal and obtain an R^2 below 10^{-5}. We extend the model to allow for differentiated output in app. sec. C.1.5. Appendix sec. C.1.10 provides an extension to multiproduct firms.

[14] Nonsubstitutable intermediate inputs would imply a production function such as $Q_{ft} = \min\{L_{ft}^{\beta^{\mathrm{l}}} K_{ft}^{\beta^{\mathrm{k}}} \Omega_{ft}; M_{ft}\beta^{\mathrm{m}}\}$, as in Ackerberg, Caves, and Frazer (2015). We refer to Rubens (2023b) for factor price markdown identification in settings with nonsubstitutable inputs.

[15] Although firms can substitute between labor, capital, and materials under Cobb-Douglas, we note that the Cobb-Douglas functional form implies that these inputs are Q-complements (Stern 2011).

functional form by estimating a translog production function. In our baseline model, we do not impose any assumptions on the returns to scale (RTS) in the production process, but we also present an extension in which we impose an RTS parameter in the main text. In appendix section C.1.2, we present further details on the latter approach.

The Cobb-Douglas specification rules out factor-biased technological change. We see this as an innocuous assumption because, as was explained in section II.B, capital investment in Liège mines was mainly limited to mining locomotives and lifts, ventilation fans, and water pumps. Ventilation fans and water pumps are safety investments, which can be seen as a sunk cost to operate the mine, but which do not affect labor productivity specifically. Rubens (2023a) did not find evidence for labor augmenting effects of mining locomotives. The main factor-biased technology in mining was the mechanized cutting machine, which was not skill-biased (Rubens 2024). However, such machines were barely adopted in the mines in our dataset due to too narrow coal veins, as mentioned earlier. We defend this assumption further in our setting using detailed technology data in section V.B.

We assume that the TFP transition is given by the first-order Markov process in equation (2), with an unexpected productivity shock v_{ft} and serial correlation ρ. The main benefits of this Markov process relate to the identification of the production function, as will be explained later. Of course, there are also costs to this approach: we rule out richer productivity processes that arise due to cost dynamics (we test this assumption in app. sec. C.1.11):

$$\omega_{ft} = \rho\omega_{ft-1} + v_{ft}. \tag{2}$$

We assume that both labor and intermediate inputs are variable and static inputs, meaning that they are not subject to adjustment frictions and affect only current profits. Capital is, in contrast, assumed to be a dynamic and fixed input: we assume capital investment is chosen 1 period in advance and affects both current and future profits, as capital does not depreciate immediately. We test these timing assumptions in section V.A by looking at the impulse-response functions of the different inputs after the coal demand shock of 1871.

B. *Labor Supply*

1. Labor and Intermediate Input Supply

Firms face a labor supply function with an inverse firm-level elasticity

$$\psi^1_{ft} \equiv \frac{\partial W^1_{ft}}{\partial L_{ft}} \frac{L_{ft}}{W^1_{ft}}.$$

If firms are wage takers on the labor market, this implies that $\psi^{\mathrm{l}}_{ft} = 0$, whereas labor market power implies that $\psi^{\mathrm{l}}_{ft} > 0$. We assume that firms are price takers on their intermediate input markets, meaning that

$$\psi^{\mathrm{m}}_{ft} \equiv \frac{\partial W^{\mathrm{m}}_{ft}}{\partial M_{ft}} \frac{M_{ft}}{W^{\mathrm{m}}_{ft}} = 0.$$

The Belgian coal industry was well integrated in the manufacturing sector and had to compete with other industrial sectors for material inputs such as tools, explosives, and black powder, so it seems reasonable to assume that these input markets were indeed competitive. We defend this assumption further and estimate an alternative model that allows for endogenous intermediate input prices in appendix section C.1.6.

2. Labor Supply Function

For our labor supply model, we rely on a static homogeneous firms model. The main reason to model firms as not being differentiated is that there is very limited wage variation across firms within towns: municipality-year fixed effects explain 93% of miner wage variation. Adding firm fixed effects increases the R^2 only to 94%. If firms would be differentiated in terms of nonwage amenities, this should translate into within-market wage differences. We present more evidence on the standard deviation and explanatory power of firm fixed effects for wages in appendix section C.2.1.[16] We provide a more formal test of employer differentiation in appendix section C.2.2. However, we emphasize that the assumption of homogeneous employers does not reduce the broader applicability of our approach to identify collusion. In appendix section C.2.3, we illustrate this by estimating a model with differentiated employers instead. Similarly, other sources of imperfect labor market competition, such as search costs, could be incorporated into the labor supply model, possibly introducing dynamics.

We assume a log-linear labor supply curve, equation (3), with inverse market-level elasticity Ψ^{l}. In the main specification, we assume that this elasticity is homogeneous across markets and time.[17] Wages W^{l}_{it} are the same for all firms within a labor market i in each year t. Market-level employment is denoted L_{it}, and a market-specific residual ν_{it} reflects variation in the relative attractiveness of different labor markets, for instance due to variation in outside options available to workers. The upward slope of the market-level labor supply curve can have different sources. For instance, even if local labor markets were nonfrictional, heterogeneity in

[16] Although a model of monopsonistic competition with amenities, such as a constant elasticity of substitution model, could result in homogeneous markdowns even with differentiation, this would still lead to wage heterogeneity due to differences in marginal labor products across firms.

[17] We relax this assumption in app. sec. C.2.4.

reservation wages across workers due to outside option differences would lead to an upward-sloping market-level labor supply curve:

$$W_{it}^{\mathrm{l}} = L_{it}^{\Psi^{\mathrm{l}}} \nu_{it}. \tag{3}$$

3. Markdowns and Markups

We define the ratio of the marginal revenue product of labor over the wage as μ_{ft}^{l}, and refer to this ratio as a markdown,

$$\mu_{ft}^{\mathrm{l}} \equiv \frac{\mathrm{MRPL}_{ft}}{W_{ft}^{\mathrm{l}}},$$

where the marginal revenue product of labor MRPL_{ft} is defined in the usual way, $\mathrm{MRPL}_{ft} \equiv \partial(P_{ft} Q_{ft})/\partial L_{ft}$.

Alternatively, the wage markdown is often expressed as a percentage markdown of wages below the marginal revenue product δ_{ft}^{l}, which is a simple function of μ_{ft}^{l}:

$$\delta_{ft}^{\mathrm{l}} \equiv \frac{\mathrm{MRPL}_{ft} - W_{ft}^{\mathrm{l}}}{\mathrm{MRPL}_{ft}} = \frac{\mu_{ft}^{\mathrm{l}} - 1}{\mu_{ft}^{\mathrm{l}}}.$$

Similarly, the coal price markup is defined as the ratio of the coal price over marginal costs, $\mu_{ft} \equiv \partial L_{ft}/\mathrm{MC}_{ft}$.

C. Employer Behavior

We assume that firms choose variable input quantities to minimize a combination of their own and their competitors' costs, as specified in equation (4). The collusion weights λ_{fgt} parametrize the weight that each firm f puts on the costs of every other firm g within the same input market $i(f)$, with the set of firms in market i being denoted $\mathcal{F}_{i(f)t}$. This is the cost minimization equivalent of the objective functions in empirical collusion models such as Bresnahan (1987). The shadow value parameter MC_{ft} captures the marginal cost of increasing output by 1 unit at firm f:

$$\min_{L_{ft}, M_{ft}} \left(\sum_{g \in \mathcal{F}_{i(f)t}} (\lambda_{fgt}(L_{gt} W_{gt}^{\mathrm{l}} + M_{gt} W_{gt}^{\mathrm{m}})) - \mathrm{MC}_{ft}(Q(L_{ft}, M_{ft}, K_{ft}, \Omega_{ft}; \boldsymbol{\beta}) - Q_{ft}) \right), \tag{4}$$

with $\lambda_{fgt} = 1$ if $f = g$ and $0 \leq \lambda_{fgt} \leq 1$ if $f \neq g$.

The collusion weights λ_{fgt} indicate the extent to which firms internalize only their own costs when choosing inputs or the costs of their competitors as well. If firms choose variable inputs to minimize only their own costs, this implies that the matrix of λ_{fgt} weights, Λ_t, is the identity matrix, in which case our model collapses to the one in De Loecker and Warzynski (2012).

If firms are colluding perfectly, they are choosing inputs to minimize joint costs, as if they would be a single firm, and Λ_t becomes a matrix of ones. This general formulation nests different kinds of collusive practices: for instance, firms can agree to a nonpoaching agreement or they can outright collude on their employment quantities (or wages). All these forms of collusive behavior are captured by the collusion parameter λ_{fgt}. We note that collusion on output quantities or prices is also picked up in terms of the collusion parameter λ_{fgt}: firms do not internalize each other's revenues and costs differently.[18]

We quantify the bounds of the wage markdown μ^{l}_{ft} under two different employer conduct assumptions: noncooperative employment choices and perfect collusion.[19] In appendix section A.1, we generalize the aforementioned model and identification approach to a broader class of models that does not rely on the homogeneous employers assumption, and that allows for different noncooperative baseline conduct than Cournot competition.

1. No Collusion

In the absence of collusion, firms choose inputs to minimize their own costs without internalizing their rivals' costs. Hence, the objective function in equation (5) assumes that firms choose their variable inputs L and M in every time period to minimize their current variable costs:

$$\min_{L_{ft}, M_{ft}} \Big((L_{ft} W^{\mathrm{l}}_{ft} + M_{ft} W^{\mathrm{m}}_{ft}) - \mathrm{MC}_{ft} (Q_{ft} - Q(L_{ft}, M_{ft}, K_{ft}, \Omega_{ft}; \boldsymbol{\beta})) \Big). \qquad (5)$$

Given that employers are assumed to be homogeneous to their workers, this implies a model of Cournot competition. Taking the first-order condition with respect to labor, adjusting wage subscripts to the fact that wages are market-specific, and rewriting marginal costs MC_{ft} as the coal price over the markup $\mu_{ft} \equiv P_{ft}/\mathrm{MC}_{ft}$ results in

$$L_{ft} \frac{\partial W^{\mathrm{l}}_{it}}{\partial L_{it}} + W^{\mathrm{l}}_{it} = \frac{\partial Q_{ft}}{\partial L_{ft}} \frac{P_{ft}}{\mu_{ft}}. \qquad (6)$$

The right-hand side of equation (6) is equal to the marginal revenue product of labor, its left-hand side is the marginal cost of labor. Denoting the labor market share of firm f in market i as $s^{\mathrm{l}}_{ft} \equiv L_{ft}/L_{it}$, it follows that the ratio of the marginal revenue product of labor over the wage is equal to the market-level inverse labor supply elasticity weighted by the labor

[18] In theory, one could distinguish different collusion weights on competitor sales and costs, but to separately identify these, one would need to impose a model of competition both downstream and upstream, whereas we only do the latter.

[19] Under perfect labor market competition, wages are equal to the marginal revenue product of labor, so $\mu^{\mathrm{l}}_{ft} = 1$ and $\delta^{\mathrm{l}}_{ft} = 0$.

market share, as shown in equation (7). We denote this markdown in the absence of collusion as $\underline{\mu}^{\mathrm{l}}_{ft}$:

$$\underline{\mu}^{\mathrm{l}}_{ft} = 1 + s^{\mathrm{l}}_{ft}\Psi^{\mathrm{l}}. \tag{7}$$

2. Collusion

Under perfect collusion, all firms in market i form a cartel that collectively chooses the input bundle that minimizes joint costs of all the firms, as defined by the set $\mathcal{F}_{i(f)t}$. This implies the objective function (8):

$$\min_{L_{ft}, M_{ft}} \left(\sum_{g \in \mathcal{F}_{i(f)t}} (L_{gt} W^{\mathrm{l}}_{gt} + M_{gt} W^{\mathrm{m}}_{gt}) - \mathrm{MC}_{ft}(Q_{ft} - Q(L_{ft}, M_{ft}, K_{ft}, \Omega_{ft}; \boldsymbol{\beta})) \right). \tag{8}$$

The first-order condition becomes equation (9). In contrast to the first-order condition in the Cournot case, equation (6), the firms do not optimize individually over their residual labor supply curve but jointly, treating the entire market-level labor supply curve as endogenous:

$$L_{it} \frac{\partial W^{\mathrm{l}}_{it}}{\partial L_{it}} + W^{\mathrm{l}}_{it} = \frac{\partial Q_{ft}}{\partial L_{ft}} \frac{P_{ft}}{\mu_{ft}}. \tag{9}$$

The resulting collusive markdown, which we denote $\bar{\mu}^{\mathrm{l}}_{ft}$, is equal to the market-level inverse labor supply elasticity, as is expressed in equation (10). As firms choose inputs jointly, their collective labor market share is equal to 1, which rationalizes the collusive markdown (10) in comparison with the Cournot markdown (7):

$$\bar{\mu}^{\mathrm{l}}_{ft} = 1 + \Psi^{\mathrm{l}}. \tag{10}$$

3. General Formulation

To nest these two extreme cases of conduct into one specification, we rewrite the first-order conditions from equations (6) and (9) more generally as equation (11). We introduce a conduct parameter $\tilde{\lambda}_{ft}$ that parameterizes the extent of collusion in the market. If firms do not collude, the conduct parameter is equal to the labor market share, $\tilde{\lambda}_{ft} = s^{\mathrm{l}}_{ft}$, and the first-order condition collapses to the Cournot model. In contrast, if firms collude perfectly, the conduct parameter is one, $\tilde{\lambda}_{ft} = 1$. The conduct parameter $\tilde{\lambda}_{ft}$ is a firm-level aggregate of the bilateral conduct parameters λ_{fgt} from equation (4):

$$W^{\mathrm{l}}_{it} + \tilde{\lambda}_{ft} \frac{\partial W^{\mathrm{l}}_{it}}{\partial L_{it}} L_{it} = \frac{\partial Q_{ft}}{\partial L_{ft}} \frac{P_{ft}}{\mu_{ft}}. \tag{11}$$

Working out this first-order condition results in the markdown expression in equation (12):

$$\mu_{ft}^{l} = 1 + \tilde{\lambda}_{ft}\Psi^{l}. \tag{12}$$

This expression nests the markdown bounds under no collusion, equation (7), and under perfect collusion, equation (10). In the next section, we will compare these markdown bounds to cost-side markdown estimates to identify collusion.

4. Timing of Choices

In accordance with the assumptions made above, the timing of choices is as follows. At time $t - 1$, prior to observing productivity shocks v_{ft}, firms choose their capital investment and decide whether to collude.[20] At time t, after the productivity shock materializes, they choose labor and intermediate inputs.

One caveat related to the model is that, as was mentioned earlier, there is anecdotal evidence for both wage fixing and employment coordination. The cartel restricted output and hence employment, which is more in line with the Cournot model. For the employers' associations, we find anecdotal evidence for both wage coordination, as was mentioned earlier, but also for various types of employment coordination. Given that we are mainly interested in the labor market effects of the cartel, we will rely on a model in which firms collude in their employment choices in the next section. We refer to appendix section C.2.3 for an alternative model with collusive wage setting, rather than employment setting.

D. *Quantifying Employer Collusion*

The model above shows that the labor supply elasticity allows us to identify the wage markdown μ_{ft}^{l} only under a specific assumption about employer conduct, as parameterized by the conduct parameter $\tilde{\lambda}_{ft}$. In this section, we show that the wage markdown can also be written independently of the conduct parameter but relying on the production function parameters instead.[21] Substituting the output elasticity of labor β^{l} and the revenue share of labor $\alpha_{ft}^{l} \equiv (W_{ft}^{l}L_{ft})/(P_{ft}Q_{ft})$ into the first-order condition for labor in the cost minimization problem, (11), results in the following markup expression, which is an extension of the markup expression in De Loecker and Warzynski (2012):

$$\mu_{ft} = \frac{\beta^{l}}{\alpha_{ft}^{l}(1 + \tilde{\lambda}_{ft}\Psi^{l})}.$$

[20] We do not formally model the underlying collusion decisions, which are likely dynamic.

[21] Again, we refer to app. sec. A.1 for the more general version of this argument beyond the homogeneous firms model.

The first-order condition for materials is identical to the markup derivation in De Loecker and Warzynski (2012). Given that intermediate input prices are exogenous to firms, we can write the following equation for markups:

$$\mu_{ft} = \frac{\beta^{\mathrm{m}}}{\alpha_{ft}^{\mathrm{m}}}. \tag{13}$$

Following previous work (Morlacco 2020; Brooks et al. 2021; Yeh, Macaluso, and Hershbein 2022) but now allowing for collusion, we divide the markup derived from labor by the markup derived from intermediate inputs to obtain the markdown equation (14). The right-hand side of this equation, $(\beta^{\mathrm{l}}\alpha_{ft}^{\mathrm{m}})/(\beta^{\mathrm{m}}\alpha_{ft}^{\mathrm{l}})$, is the cost-side markdown estimate, which does not depend on the conduct parameter $\tilde{\lambda}_{ft}$. The left-hand-side term, $1 + \tilde{\lambda}_{ft}\Psi^{\mathrm{l}}$, is the labor supply side markdown from the generalized Cournot model, which does depend on the conduct parameter:

$$\mu_{ft}^{\mathrm{l}} = 1 + \tilde{\lambda}_{ft}\Psi^{\mathrm{l}} = \frac{\beta^{\mathrm{l}}\alpha_{ft}^{\mathrm{m}}}{\beta^{\mathrm{m}}\alpha_{ft}^{\mathrm{l}}}. \tag{14}$$

Equation (14) captures the core of our empirical strategy. If we have an estimate of the market-level inverse labor supply elasticity Ψ^{l}, the wage markdown is known up to the conduct parameter $\tilde{\lambda}_{ft}$. The wage markdown is also known if the production function parameters are identified. Hence, identification of both the labor supply function and the production function permits identification of the conduct parameter $\tilde{\lambda}_{ft}$ by equating the two terms in equation (14).

Rather than estimating the conduct parameter $\tilde{\lambda}_{ft} \in [s_{ft}^{\mathrm{l}}, 1]$, we estimate a slightly altered conduct parameter $\hat{\lambda}_{ft} \in [0, 1]$ as defined in equation (15), which is more easily interpretable as it ranges from 0 to 1.[22] In the absence of collusion, $\hat{\lambda}_{ft} = 0$, whereas in a fully collusive market, $\hat{\lambda}_{ft} = 1$.

$$\hat{\lambda}_{ft} \equiv \frac{\mu_{ft}^{\mathrm{l}} - \underline{\mu}_{ft}^{\mathrm{l}}}{\bar{\mu}_{ft}^{\mathrm{l}} - \underline{\mu}_{ft}^{\mathrm{l}}}. \tag{15}$$

We operationalize this approach by following a stepwise approach. First, section IV.A presents the estimation of the production and labor supply functions. Second, in section IV.B, we estimate and discuss the evolution of wage markdowns. Third, in section IV.C, we quantify collusion and examine how it changed in response to the cartel. Finally, in section IV.D, we carry out counterfactual exercises to examine the employment and wage effects of collusion.

[22] It is easy to show that $\hat{\lambda}_{ft} = (\tilde{\lambda}_{ft} - s_{ft}^{\mathrm{l}})/(1 - s_{ft}^{\mathrm{l}})$.

IV. Identification, Estimation, and Results

A. *Labor Demand and Supply*

1. Production Function

We start by estimating the production function, equation (1). As is usual in the literature, we rely on timing assumptions on firms' input choices for identification, in the spirit of Olley and Pakes (1996). However, we combine these timing assumptions with a labor supply shifter to achieve overidentification, and we also test the timing assumptions, as will be explained further below. As labor and materials were assumed to be static and variable inputs, they are chosen after the productivity shock v_{ft} is observed by the firm at time t, while capital is fixed and dynamic, so investment is chosen before the productivity shock is observed at time $t-1$. Second, we rely on agricultural wage shocks as an additional instrument. It is a well-established fact in Belgian economic history that the Walloon coal belt attracted a large surplus of agricultural labor, predominantly from Flanders, the northern area of Belgium (Segers 2003, 334; Buyst, forthcoming). Hence, negative shocks to agricultural wages should have acted as positive labor supply shocks to coal mines. We include lagged agricultural wages in Belgium, as measured by Segers (2003, 622–23), in the instruments vector. The assumption here is that changes in agricultural wages in the previous year, w_{t-1}^{agri}, affected labor supply to the mines but did not affect coal mining productivity directly.[23] In table D.4 (tables C.1–D.4 are available online), we provide evidence on the first stage by regressing the annual change in log total mining employment in the Liège and Namur coal basin on the annual change in log agricultural wages in Belgium. Negative agricultural wage shocks were indeed followed by increased coal mining employment. Following these assumptions, we can now write the moment conditions to estimate the mining production function as

$$\mathbb{E}[v_{ft} | (l_{fr-1}, m_{fr-1}, k_{fr}, w_{r-1}^{\text{agri}})]_{r\in[2,..,t]} = 0. \tag{16}$$

The usual approach in the literature is to invert the intermediate input demand function to recover the latent productivity level ω_{ft}, which can be used to construct the productivity shock v_{ft} using the productivity law of motion (Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg, Caves, and Frazer 2015). This approach hinges on productivity being the only latent, serially correlated input demand shifter. However, input

[23] We include lagged agricultural wages rather than current wages because we also include lagged labor rather than current labor, due to our variable labor assumption. We further examine our instrumental variables assumptions in app. sec. C.1.12, where we also estimate a version of the model that does not rely on the agricultural price instrument and in which we also test other instruments using agricultural demand and supply shocks.

demand varies due to markup and markdown variation as well. The approach with input inversion can still be used when making additional parametric assumptions about the distribution of markups and markdowns. Another possibility is to impose more structure on the productivity transition process. Following Blundell and Bond (2000), we rely on the first-order autoregressive (AR[1]) structure of the productivity transition (2).[24] By taking ρ differences of equation (2), one can express the productivity shock v_{ft} as a function of estimable coefficients without having to invert the input demand function.[25]

We pursue this approach so as to avoid having to impose additional structure on the distribution of markups and markdowns across firms and over time. This comes at the cost of not allowing entry and exit of mines to be endogenous to their productivity level, contrary to Olley and Pakes (1996). However, as is often noted in the literature, the use of an unbalanced panel, in which one does not select negatively on market exit, already alleviates most concerns of selection bias.[26]

Rewriting the moment conditions from equation (16) and using lags of at most 1 year, the moment conditions are given by equation (17):[27]

$$\begin{aligned}\mathbb{E}[q_{ft} - \rho q_{ft-1} - \beta^0(1-\rho) - \beta^{\mathrm{l}}(l_{ft} - \rho l_{ft-1}) - \beta^{\mathrm{m}}(m_{ft} - \rho m_{ft-1}) \\ - \beta^{\mathrm{k}}(k_{ft} - \rho k_{ft-1})|(l_{ft-1}, m_{ft-1}, k_{ft}, k_{ft-1}, w_{t-1}^{\mathrm{agri}})] = 0.\end{aligned} \tag{17}$$

We measure q_{ft} as the logarithm of annual coal production in metric tons at mine f during year t. Similarly, labor l_{ft} is measured as the logarithm of the average number of workers employed throughout the year, multiplied by the average number of days worked. Materials m_{ft} are measured as the logarithm of the "ordinary expenses" variable, which is reported in the data. The logarithm of the capital stock k_{ft} is constructed by using the perpetual inventory method on the "extraordinary expenses" category, which we describe in more detail in appendix section B.3. To estimate the production function using ordinary least squares (OLS), the logs of output, employment, material usage, and capital need to be observed. This reduces the sample size from 8,779 to 4,480 observations, as also explained in table B.6. For the generalized method of moments (GMM) estimator, the lagged values of these variables need to be observed as well. This additional sample restriction further decreases the number of observations

[24] In app. sec. C.1.7, we do a robustness check in which we set $\rho = 1$, rather than estimating ρ. In app. sec. C.1.9, we test for serial correlation in the productivity shocks v and also estimate a second-order autoregressive model (AR[2]) as an extension.

[25] An alternative is to estimate the output elasticities using a cost shares approach, rather than estimating the production function. Appendix sec. C.1.8 contains the estimates using this approach, which are of similar magnitudes as those in the main text.

[26] See Olley and Pakes (1996) and De Loecker et al. (2016).

[27] In theory, one could use more lags, but this further reduces the dataset, which is already small.