Earnings Inequality in Production Networks*

Federico Huneeus† Kory Kroft‡ Kevin Lim §

Abstract

We develop a quantitative model in which heterogeneous firms hire heterogeneous workers in an imperfectly competitive labor market and source intermediates from heterogeneous suppliers in a production network. The model delivers an earnings equation with a firm-specific wage premium that depends endogenously on firm productivities and firm-to-firm network linkages. We establish identification of model parameters and estimate them using linked employer-employee and firm-to-firm transactions data from Chile. Counterfactual simulations show that heterogeneity in network linkages explains 21% of earnings variance. We also examine earnings volatility and a minimum wage policy, finding important roles for network linkages in both cases.

Keywords: Earnings inequality, production networks, monopsony power.

JEL Codes: J31, F16.

*We thank Lorenzo Caliendo, Arnaud Costinot, Dave Donaldson, Emmanuel Farhi, Cecilia Fieler, Gregor Jarosch, Pete Klenow, Sam Kortum, Andres Rodriguez-Clare, Bradley Setzler, Isaac Sorkin, Felix Tintelnot, Jose Vasquez, Ivan Werning and Daniel Xu for invaluable comments on our paper. We are also grateful for feedback from seminar participants at the University of Toronto, MIT, Universidad de Chile, Universidad Mayor, UT Austin, Harvard University, Georgetown, and UBC; and from conference participants at the Yale University Cowles Trade Day conference, the Central Bank of Chile Heterogeneity in Macroeconomics workshop, the Santiago-Macro Annual Conference, the West Coast Trade Conference, the Econometric Society World Congress, the European Economics Association Congress, the Stanford Institute for Theoretical Economics, and the ASSA 2021 Annual Meeting. We thank the Munk School of Global Affairs and Public Policy for financial support.

†Central Bank of Chile (fhuneeus@bcentral.cl).

‡University of Toronto and NBER (kory.kroft@utoronto.ca).

§University of Toronto (kvn.lim@utoronto.ca).

1 Introduction

There is growing evidence that firms matter for worker earnings. In a survey of the empirical literature concerned with estimation of worker and firm fixed effects on earnings, Card et al. (2018) summarize that firm effects explain around 20% of the variation in worker earnings.[1] A standard explanation for this proposed by the literature is that employers are heterogeneous in some innate characteristics – productivity and amenities, for example – with this heterogeneity then passing through into differences in earnings of otherwise similar workers. At the same time, a separate, emerging literature has documented that a substantial share of firm heterogeneity is explained by differences in the *connections* that firms form with each other in a production network. For instance, using firm-to-firm transactions data for Belgium, Bernard et al. (2019) show that variation in both the number and characteristics of a firm's customers and suppliers explains more than half of the variation in firm sales. Motivated by these two facts – that network heterogeneity matters for firm heterogeneity and firm heterogeneity matters for earnings heterogeneity – we investigate in this paper the importance of the production network structure for earnings inequality, both theoretically and empirically.

This research agenda is important for at least two reasons. First, many countries have witnessed secular increases in earnings inequality, largely driven by growing inequality between firms (Song et al. (2019)). At the same time, technological improvements have led to a greater role for value chains in production (Antràs and de Gortari (2020)). A quantitative economic framework capturing both earnings inequality and production network linkages is needed to study these developments in parallel. This will afford a better understanding of how changes in the structure of the production network affect both differences in earnings across workers in the cross-section and the volatility of these earnings over time. Second, many countries in both the developed and developing world have responded to growth in earnings inequality by adopting policies that aim to mitigate these trends, such as minimum wage requirements. Yet, little is known about how network linkages matter for the effects of such policy initiatives, despite the fact that these linkages are fundamental to the organization of production.

We shed light on these issues by studying a panel dataset from Chile that combines matched employer-employee records with firm-to-firm transactions data. This allows us to observe both the earnings for every employee at each firm in our data and the buyers and sellers of every firm. We use the data to structurally estimate the parameters of a quantitative general equilibrium model, which features imperfectly competitive labor markets and production network linkages between firms. Counterfactual simulations of the model then allow us to quantify the importance of the production network for earnings inequality, earnings volatility, and the effects of a

[1] Bonhomme et al. (2020) show that correcting for the well-known limited mobility bias in these estimates lowers the firm effect share of earnings variance, but increases the importance of the covariance between worker and firm effects (explaining 15% of earnings variance among US workers).

minimum wage policy. The outline of our approach is as follows.

First, in section 2, we develop the theoretical framework. We build on the model of labor markets in Lamadon et al. (2019), where workers are heterogeneous in ability and firms have wage-setting power arising from workers' idiosyncratic preferences for employment at different firms.[2] The model also allows for employer amenities that vary at the worker-firm level and complementarties in production between worker ability and firm technology, which drive the heterogeneous sorting of workers to firms. We enrich this by introducing a production structure in which firms produce output using both labor and materials (intermediate inputs), where the latter are sourced from suppliers in a production network as in Huneeus (2019) and Lim (2019). Firms are heterogeneous in the sets of customers and suppliers they are connected to in the network, as well as in total factor productivities (TFPs), labor productivities, and buyer-seller productivities (relationship capabilities).

This framework delivers an earnings equation that nests a well-known class of reduced-form earnings models, such as those studied by Abowd et al. (1999) and Bonhomme et al. (2019). Given our structure, the firm effects (wage premia) in this earnings equation depend endogenously on both firm productivities and the set of firm-to-firm linkages in the production network. In particular, we show that the relevance of the production network for the firm effect of a firm i at time t, W_{it}, can be summarized by two sufficient statistics. First, a demand shifter D_{it}, which captures both the number of customers that the firm is connected to and the demand received from each of these customers. Second, a materials price index Z_{it}, which captures both the number of suppliers that the firm is connected to and the price charged by each of these suppliers. Our model hence forges a direct link between the structure of the network and worker outcomes.[3]

To illuminate the key mechanisms in the model and the quantitative findings that follow from it, we then derive three comparative static results, which are presented in section 3. First, we characterize how changes in demand D_{it} and input price indices Z_{it} that firms face in the production network affect worker earnings through a combination of *scale* and *substitution* effects (Proposition 1). An increase in demand ($\Delta D_{it} > 0$) generates a positive scale effect on earnings by directly raising the marginal revenue product of labor (MRPL) for all workers at a firm. A fall in input costs ($\Delta Z_{it} < 0$), on the other hand, generates both a positive scale effect and a substitution effect, where the latter occurs as firms adjust the ratio of labor to materials used in production. Lower input costs can therefore lead to either higher or lower

[2] Employer differentiation as a source of labor market has been studied extensively in the literature. See Rosen (1986), Manning (2003), Sorkin (2018), Card et al. (2018), and Chan et al. (2019), just to name a few examples.

[3] Instead of employer differentiation, labor market power could also arise from concentration (Berger et al. (2019), Jarosch et al. (2019)) or search frictions (Burdett and Mortensen (1998), Postel-Vinay and Robin (2002), Taber and Vejlin (2018)). Like ours, most of these models imply that wages are a markdown below the marginal revenue product of labor (MRPL) at a firm, where the firm effect on earnings is the component of the MRPL that is common to all workers at a firm. Hence, the mechanisms that we highlight regarding the interaction between the production network and worker earnings are relevant for a broader class of models of the labor market.

earnings, depending on the comparison of two key elasticities: the price elasticity of demand σ (which determines the strength of scale effects) and the elasticity of substitution between labor and materials ϵ (which determines the strength of substitution effects). Second, we examine the passthrough from firm productivity shocks to worker earnings, showing that first-order changes in earnings can be characterized using only knowledge of $\{\sigma, \epsilon\}$, the labor supply elasticity γ, and observable network and factor cost shares (Proposition 2). Third, we establish a novel link connecting between-firm changes in earnings with within-firm changes in employment composition (Proposition 3). Through this mechanism, increases in between-firm earnings inequality lead to *reductions* in within-firm earnings inequality, due to general equilibrium effects in labor markets.

We then turn toward identification of the model parameters in section 4. We first state the assumptions needed for identification. These restrictions follow leading papers in the respective literatures on firm labor market power (Lamadon et al. (2019)), estimation of worker and firm effects on earnings (Bonhomme et al. (2019)), firm-to-firm networks (Bernard et al. (2019)), and production function estimation (Doraszelski and Jaumandreu (2018)). Under these assumptions, we formally establish identification results for each model parameter.

An important contribution of this paper is the development of a novel approach for the identification of the labor-materials substitution elasticity ϵ. It is well-known in the literature on production function estimation that ϵ can be identified from the relationship between firms' relative expenditures on labor versus materials and the relative prices of these inputs. Unlike standard models that assume homogeneous labor and material inputs, however, firms in our setting pay heterogeneous wages to different workers and heterogeneous prices to suppliers of different inputs. In this environment, we show that one should measure input costs using theoretical price indices for labor and materials that reflect the aggregation of these heterogeneous input costs. Our model further shows that the labor and material price indices correspond exactly to the firm effect on earnings W_{it} and the materials price index Z_{it}, respectively. In particular, W_{it} can be identified from the decomposition of earnings into worker and firm effects as in Lamadon et al. (2019) and Z_{it} can be identified from the decomposition of firm-to-firm transaction values into buyer and seller effects as in Bernard et al. (2019), where the seller effects reflect the marginal costs and hence output prices for every supplier of a firm. Given these theory-consistent price indices, one can then identify ϵ by applying the instrumental variables strategy of Doraszelski and Jaumandreu (2018), where the instruments correspond to, among other things, lagged input prices $W_{i,t-s}$ and $Z_{i,t-s}$ for $s > 0$. This identification strategy for ϵ hence leverages a key feature of our data, since it requires linked employer-employee and firm-to-firm transaction records.

The remaining model parameters are then identified as follows. First, the labor supply elasticity is identified from the passthrough of firm wage bill shocks to changes in worker earnings, since this parameter controls the extent of firm labor market power. Second, the price elasticity of demand is identified from an adjusted measure of profit-sales ratios, as this parameter governs the markups that firms charge. Third, worker abilities and production complementarities are identified from worker-firm effects on earnings, following the approach in Bonhomme et al. (2019). Fourth, firm amenities are identified from residual variation in employment shares of each worker ability type that is not explained by observed variation in wages. Fifth, firm relationship capability is identified from variation in the share of a firm's sales in the production network (as opposed to sales to final consumers), since this parameter affects only the productivity of relationships with other firms in the network. Sixth, labor productivity is identified from residual variation in relative expenditures on labor versus materials that is not explained by variation in the relative price indices of these inputs. Finally, TFP is identified from firm effects on earnings, since this parameter directly determines the MRPL of a worker type at a firm, given all other determinants of the MRPL described above.

In section 5, we then take the model to data and implement estimation of its parameters, building on our identification results. We begin by providing a detailed description of the data that we rely on for our analysis. The highlights of these data are two linked administrative datasets from the Chilean Internal Revenue Service. First, matched employer-employee records (2005-2018), providing information about labor income (wages and other measures) for every worker in the formal private sector. Second, value-added tax records (2005-2010), in which firms report transaction values with each of their customers and suppliers. To our knowledge, there are only three other papers that study linked employer-employee and firm-to-firm transactions data. Adao et al. (2020) use data from Ecuador to measure the effects of international trade on individual-level factor prices, while Demir et al. (2018) study the effects of trade-induced product quality upgrading on wages in Turkey. Both of these analyses assume a market price for skill and focus on the effects of trade shocks. In contrast, we allow for imperfect competition in labor markets and use our data to speak to the role of the production network itself in shaping earnings inequality. Finally, Alfaro-Ureña et al. (2020) adopt an event study research design to examine the effects on worker earnings in Costa Rica when a local firm starts interacting with multinationals. In contrast, we use our data to address both worker-level earnings and aggregate outcomes such as earnings inequality, which requires a general equilibrium model.

Next, we discuss our estimation results, the highlights of which can be summarized as follows. First, we estimate an elasticity of substitution between labor and materials of 1.5, indicating gross substitutability of these two inputs. Second, we estimate a labor supply elasticity of 5.5, which is consistent with other estimates in the literature (for example, see Staiger et al. (2010), Azar et al. (2019), Kline et al. (2019), Lamadon et al. (2019), Dube et al. (2019), and Kroft et al. (2019)). Third, we estimate a price elasticity of demand equal to 4.2. Combined with the preceding result, this implies that reductions in input costs lead to higher wages (as established in Proposition 1). Fourth, we find that the variance of firm effects on earnings explains around 11% of total log earnings variance across workers, while the covariance between firm and worker effects explains 20%. This is consistent with similar estimates in the literature and confirms an important role for firms in determining worker earnings in our data. Fifth, in decomposing firm-to-firm transactions into buyer and seller effects, we find that the former account for 12% of total variation in log transaction values while the latter account for 34%. These shares are similar to those reported by Bernard et al. (2019) using Belgian data (approximately 13% and 26% respectively).

Our empirical results also indicate substantial heterogeneity in firm-to-firm matching in the network. Two features in particular are important for understanding the effects of this heterogeneity on earnings inequality. On one hand, larger, higher-paying, and more productive firms tend to form more connections with buyers and suppliers. This heterogeneity in the extensive margin of the network therefore *amplifies* heterogeneity in own-firm characteristics. On the other hand, we observe negative assortative matching on firm sales, employment, TFP, and degree (number of customers or suppliers). For example, the average TFP of a firm's customers or suppliers is negatively correlated with the firm's own TFP. This heterogeneity in the intensive margin of the network therefore *dampens* heterogeneity in own-firm characteristics. Hence, heterogeneity in the production network overall does not necessarily make firms more different from each other and does not mechanically induce greater earnings inequality. Rather, how network heterogeneity affects earnings inequality is a quantitative question.

Finally, in section 6, we use the estimated model to investigate the importance of the production network for worker earnings through a series of counterfactual simulations. We examine three sets of exercises. First, we quantify the share of earnings variance that is attributable to heterogeneity in each set of primitives in our model: worker abilities, firm productivities, firm amenities, and production network linkages. This extends the usual earnings variance decomposition into worker and firm effects by accounting for the structural dependence of firm effects on underlying primitives. Our novel finding is that network heterogeneity accounts for 21% of earnings variance, with upstream heterogeneity in matching with suppliers accounting for 12% and downstream heterogeneity in matching with customers accounting for 9%. In contrast, own-firm productivities and amenities jointly account for 12% of earnings variance. Hence, we find that heterogeneity in the production network is in fact a key driver of earnings inequality.

Second, we quantify the importance of network linkages for earnings volatility. We do this by estimating the stochastic processes for firm TFP and labor productivity, simulating productivity shocks from these processes, and leveraging our analytic characterization of passthrough from firm productivity shocks to worker earnings. We find that indirect passthrough of shocks to a firm's customers and suppliers in the production network explains between 20-25% of earnings volatility, depending on worker ability type, with own-firm productivity fluctuations explaining the remainder. This result contributes to the empirical literature studying the relationship between firm shocks and worker earnings (for example, Guiso et al. (2005) and Chan et al. (2021)) by extending the analysis to account for passthrough via the network. It also builds on work by Dhyne et al. (2020), who show that the passthrough from foreign demand shocks to a firm's revenue depends not only on the firm's direct exposure to exports, but also on its indirect exposure through domestic network linkages.

Lastly, we examine a minimum wage (MW) application by simulating an equilibrium under a wage floor equal to the 20^{th} percentile of earnings in the baseline equilibrium. This exercise is similar to the study of MW policies by Haanwinckel (2020), but we extend the analysis to consider the role of the production network. Our counterfactual findings indicate that this policy induces a reallocation of low-ability workers from small to large firms, driven by the exit of small firms from the labor markets for these workers, consistent with findings in Dustmann et al. (2020). We also find that this policy reduces the variance of log earnings by 18.6%, with earnings for workers that are paid the MW at the smallest quintile of firms increasing by around 12%. Furthermore, we find important within-firm spillover effects on workers that earn above the MW, with earnings for these workers increasing by around 3.5% at the smallest quintile of firms.[4] We highlight two findings on the importance of the production network for these effects. The first is a positive one: a firm's ability to respond to the MW policy by substituting materials for labor explains 40% of the spillover effects on unconstrained worker earnings. The second result is a negative one accompanied by an important insight. Consistent with the empirical literature, we find that the MW policy increases producer prices at the most constrained firms by around 1.5%. However, propagation of these effects through network linkages is quantitatively negligible, because the firms that are most likely to be constrained by the wage floor also tend to be the least important buyers and sellers in the production network.[5]

2 Model

The economy is populated by a set of workers Ω^L and a set of firms Ω^F. Workers are heterogeneous in a characteristic that we refer to as *ability*, denoted by a, with an exogenous measure of each ability type denoted by $L(a)$ and the set of abilities denoted by $A \subset \mathbb{R}^d_+$. The theoretical results established in this section do not require restrictions on the dimension d of the worker ability space.[6] Firms are also heterogeneous in a variety of characteristics that we specify below. While the model allows for dynamics, all meaningful economic decisions can be analyzed statically. Nonetheless, estimation of the model will involve panel data and hence in anticipation of this, we index (discrete) time by t to make explicit the variables that are allowed to vary

[4] These spillovers are distinct from the spillovers emphasized by Engbom and Moser (2018), which occur *between* firms.

[5] In contrast, existing research has shown that network linkages can be important for propagating shocks that affect sectors or firms more broadly and not just those that are relatively unimportant buyers or sellers. See for example Caliendo et al. (2017), Baqaee and Farhi (2019, 2020), Lim (2019), and Huneeus (2019).

[6] When taking the model to data, we will assume $d = 2$ with worker ability comprised of a time-invariant and time-varying component.

temporally.

2.1 Labor market

Firms and workers interact in the labor market as follows. Each firm i posts a wage $w_{it}(a)$ that is conditional on worker ability a. We take the price of final consumption that workers face as the numeraire, hence wages should be interpreted in real terms. Each worker observes all wage offers for her ability type and chooses an employer to maximize utility, where the utility of a worker with ability a employed at firm i is given by:

$$u_{it}(a) = \log w_{it}(a) + \log \tau_t + \log g_i(a) + \beta^{-1}\epsilon_{it} \tag{2.1}$$

In addition to receiving labor income, workers are also residual claimants to firm profits, which are rebated through a lump-sum income transfer τ_t that is independent of employer. Workers also derive utility from amenities $g_i(a)$ offered by firm i and idiosyncratic preferences ϵ_{it} for employment firm i, with β an inverse measure of the preference dispersion across firms.

We highlight several important features of this utility specification. First, lump-sum transfers τ_t are paid to workers in proportion to their income. This is necessary to ensure that transfers do not affect the sorting of workers across firms. Second, firms have complete information about the ability of every worker but cannot observe idiosyncratic preferences ϵ_{it}. Hence, wages are conditioned only on ability, which will imply the existence of inframarginal workers at every firm who enjoy positive rents from their employment. Third, differences in amenities $g_i(\cdot)$ allow for *vertical* differentiation across potential employers, while differences in idiosyncratic preferences ϵ_{it} introduce *horizontal* differentiation. The former rationalizes heterogeneity in compensating differentials across firms for workers of a given ability, while the latter is the source of labor market power for firms. Fourth, we follow the literature on discrete choice and assume that idiosyncratic preferences are characterized as follows.

Assumption **2.1.** *The distribution of idiosyncratic preferences across workers, $\epsilon_t \equiv \{\epsilon_{it}\}_{i\in\Omega^F}$, is a multivariate Gumbel distribution with cumulative distribution function:*

$$F_\epsilon(\epsilon_t) = exp\left[-\left(\sum_{i\in\Omega^F} e^{-\frac{\epsilon_{it}}{\rho}}\right)^\rho\right] \tag{2.2}$$

where $\rho \in (0,1]$.

The parameter ρ controls the correlation of idiosyncratic preferences across firms: as ρ approaches zero, workers view all firms as perfect substitutes, whereas as ρ approaches one, idiosyncratic preferences across firms become independent random variables. Note also that Assumption 2.1 imposes structure on the cross-sectional distribution of ϵ_t but does not otherwise restrict its time-series properties.

Under Assumption 2.1, the probability that a worker with ability a chooses to work at firm i is given by:

$$\mathbb{P}_{it}(a) = \left[\frac{g_i(a) w_{it}(a)}{I_t(a)}\right]^{\gamma} \tag{2.3}$$

where $\gamma \equiv \beta/\rho$. Hence, labor supply is more elastic when preference shocks are less dispersed or more correlated. In what follows, only γ will be of interest and not β or ρ separately. The term $I_t(a)$ is an aggregate of wage and amenity values offered by all firms in the labor market for workers of ability a, which we henceforth refer to as the *labor market index* for these workers:

$$I_t(a) \equiv \left[\sum_{i \in \Omega^F} [g_i(a) w_{it}(a)]^{\gamma}\right]^{\frac{1}{\gamma}} \tag{2.4}$$

Appealing to a law of large numbers, the total supply of workers of ability a for firm i can then be written as:

$$L_{it}(a) = \kappa_{it}(a) w_{it}(a)^{\gamma} \tag{2.5}$$

where $\kappa_{it}(a)$ is a firm-specific labor supply shifter:

$$\kappa_{it}(a) \equiv L(a) \left[\frac{g_i(a)}{I_t(a)}\right]^{\gamma} \tag{2.6}$$

We further assume that the cardinality of the set of firms Ω^F is large enough such that each firm views itself as atomistic in the labor market. In choosing wages for workers of any ability $a \in A$, each firm thus views the labor market index $I_t(a)$ as invariant to its own choices. Hence, equation (2.5) implies that every firm behaves as though it faces an upward-sloping labor supply curve with a constant elasticity γ that is common to all firms and worker ability types.

2.2 Final demand

Workers use their income to finance consumption, with consumption utility derived from a constant elasticity of substitution (CES) aggregate of products produced by all firms in the economy. For a worker of ability a employed at firm i, this is given by:

$$v_{it}(a) = \left[\sum_{j \in \Omega^F} c_{ijt}(a)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{2.7}$$

where $c_{ijt}(a)$ denotes the worker's consumption of firm j's output and $\sigma > 1$ denotes the elasticity of substitution across products. Since we take the unit price of the final consumption aggregate in equation (2.7) as the numeraire, consumption utility can also be expressed as

$v_{it}(a) = w_{it}(a)\tau_t$, which corresponds to the first two terms in the utility specification (2.1).[7] Aggregate final demand for firm $i's$ output $C_{it} \equiv \sum_{j\in\Omega^F}\sum_{a\in A} c_{jit}(a) L_{jt}(a)$ is then:

$$C_{it} = E_t p_{Fit}^{-\sigma} \tag{2.8}$$

where p_{Fit} is the price of firm i's output for final sales and E_t is aggregate consumer income:

$$E_t = \sum_{i\in\Omega^F}\sum_{a\in A} w_{it}(a) L_{it}(a) + \sum_{i\in\Omega^F}\pi_{it} \tag{2.9}$$

with π_{it} denoting profit earned by firm i. Note that E_t is also equivalent to aggregate value-added in the economy.

2.3 Production technologies

Firms produce output using labor and materials. Combining $L_{it}(a)$ workers of ability a with $M_{it}(a)$ units of materials at firm i produces $f[\phi_{it}(a) L_{it}(a), M_{it}(a)]$ units of output, where $\phi_{it} : A \to \mathbb{R}_+$ maps worker ability into productivity. We allow the labor productivity function ϕ_{it} to vary by firm, which will capture potential worker-firm complementarities in production. Total output of firm i is then a linear aggregation of output produced by workers of all abilities:

$$X_{it} = T_{it}\sum_{a\in A} f[\phi_{it}(a) L_{it}(a), M_{it}(a)] \tag{2.10}$$

where T_{it} denotes total factor productivity (TFP) of firm i. The production function f satisfies the following standard properties.

Assumption **2.2.** *The production function f is: (i) strictly increasing in both arguments; (ii) strictly concave in both arguments; and (iii) homogeneous of degree one.*

Furthermore, we assume that labor productivity can be decomposed as follows.

Assumption **2.3.** *Productivity of ability a workers at firm i is of the form $\phi_{it}(a) = \phi_i(a)\omega_{it}$.*

In other words, we assume that any worker-firm complementarities are time-invariant, with all time variation in labor productivity accounted for by ω_{it}, which we henceforth refer to simply as labor productivity of firm i. This assumption will be important for identification of the productivity terms.

While firms hire workers in the labor market as described in section 2.1, materials are sourced through firm-to-firm trade in the production network. We denote the set of firm i's customers and suppliers by $\Omega_{it}^C \subset \Omega^F$ and $\Omega_{it}^S \subset \Omega^F$ respectively. Where convenient for exposition, we will

[7] In other words, the indirect consumption utility function is $v_{it}(a) = \frac{w_{it}(a)\tau_t}{P_t}$, where $P_t \equiv \left(\sum_{i\in\Omega^F} p_{Fit}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$ is the unit price of the final consumption bundle and we normalize $P_t = 1$.

also describe the production network in terms of a matching function m_{ijt}, which is equal to 1 if $j \in \Omega^S_{it}$ and 0 otherwise. Materials for firm i are then aggregated by combining inputs from all of its suppliers using a CES technology:

$$M_{it} = \left[\sum_{j \in \Omega^S_{it}} \psi_{ijt}^{\frac{1}{\sigma}} \left(x_{ijt} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{2.11}$$

where x_{ijt} denotes the quantity of inputs purchased by i from j and ψ_{ijt} is a relationship-specific productivity shifter. As is standard in the literature, we assume the same elasticity of substitution across products in equation (2.11) as in the consumption utility function (2.7). This simplifies the firm's profit maximization problem as it ensures that both final and intermediate demand have the same price elasticity. The total allocation of materials to a firm's workers must then be equal to total materials sourced by the firm:

$$\sum_{a \in A} M_{it} \left(a \right) da = M_{it} \tag{2.12}$$

Note that the production network is not restricted to be bipartite: firms can simultaneously be buyers and sellers. However, we treat the set of active buyer-seller relationships in the economy as exogenous. This allows us to focus on how labor market outcomes are determined conditional on a given production network structure. In the counterfactual simulations described below, we then study how changes in the production network structure alter these labor market outcomes.[8]

2.4 Output market structure and profit maximization

We assume a market structure of monopolistic competition in output markets: each firm in the economy produces a unique product and sets prices for each of its customers taking the prices set by all other firms as given. Demand by firm i for inputs from firm j then takes the standard form implied by the CES production technology (2.11):

$$x_{ijt} = \Delta_{it} \psi_{ijt} p_{ijt}^{-\sigma} \tag{2.13}$$

[8] The introduction of imperfect labor market competition complicates the modeling of endogenous production network formation because it implies that firms face *increasing* marginal costs of production. Hence, the incentive for a firm to sell to one customer depends on its existing set of customers. This violates the key assumptions needed for tractability in existing models of endogenous production network formation (e.g. Huneeus (2019) and Lim (2019)).

where p_{ijt} is the price charged by seller j to buyer i. The term Δ_{it} is a firm-specific intermediate demand shifter that we refer to as *network demand*:

$$\Delta_{it} = E_{it}^{M} \left(Z_{it}\right)^{\sigma-1} \tag{2.14}$$

where E_{it}^{M} is total material cost and Z_{it} is the unit cost of materials for firm i, both of which are characterized below.

As in the labor market, we assume that firms behave atomistically in the output market. Hence, in choosing output prices, firms take as given the network demands of each of its customers and perceive a constant price elasticity of demand equal to $-\sigma$. Note that the upward-sloping labor supply curves faced by each firm imply that marginal costs of production are increasing in output. Hence, a firm's relationships with each of its customers are inherently interlinked: a reduction in the price charged to one customer increases demand and hence raises both output and marginal cost, which in turn affects the choice of prices charged to other customers. However, even though we allow firms to charge different prices to different customers, the following result establishes that it is never optimal for them to do so.[9]

CLAIM **1.** *The profit-maximizing price charged by a firm i to each of its customers j (including final consumers) does not vary across customers:*

$$p_{jit} = p_{it}, \, \forall j \in \Omega_{it}^{C} \cup \{F\}$$

Intuitively, each firm maximizes profits by choosing prices such that marginal revenue from each customer is equal to marginal cost. Since demand features a constant and common price elasticity of $-\sigma$, marginal revenue is proportional to price. Furthermore, even though marginal cost is increasing, it depends only on total output of the firm and hence is common across customers. As a result, each firm optimally chooses to charge a common price to each of its customers in equilibrium.[10]

With this result, we can express total demand for firm i's output as:

$$X_{it} = D_{it} p_{it}^{-\sigma} \tag{2.15}$$

where D_{it} is a demand shifter for the firm given by the sum of final demand (common to all firms) and the network demands of the firm's customers:

$$D_{it} = E_t + \sum_{j \in \Omega_{it}^{C}} \Delta_{jt} \psi_{jit} \tag{2.16}$$

[9]Proofs of all claims and propositions are relegated to Section A of the appendix.

[10]In recent work, Kikkawaa et al. (2020) develop a model of firm-to-firm trade with oligopolistic competition in output markets that features buyer-seller specific markups, while abstracting from imperfect competition in factor markets.

Similarly, the unit cost of materials can be expressed as:

$$Z_{it} = \left[\sum_{j \in \Omega_{it}^S} \psi_{ijt} \Phi_{jt} \right]^{\frac{1}{1-\sigma}} \tag{2.17}$$

where Φ_{it} is the *network productivity* of firm i, an inverse measure of the firm's price:

$$\Phi_{it} \equiv p_{it}^{1-\sigma} \tag{2.18}$$

Finally, we can now write the profit-maximization problem for firm i concisely as a choice over its production inputs:

$$\pi_{it} = \max_{\{w_{it}(a), M_{it}(a)\}_{a \in A}} \left\{ D_{it}^{\frac{1}{\sigma}} X_{it}^{\frac{\sigma-1}{\sigma}} - \sum_{a \in A} w_{it}(a) L_{it}(a) - Z_{it} \sum_{a \in A} M_{it}(a) \right\} \tag{2.19}$$

subject to the labor supply curves (2.5) and production technology (2.10). Note that the unit material cost Z_{it} given by equation (2.17) already reflects the firm's optimal choice of inputs across its suppliers.

2.5 Wage determination

Wages are determined by the solution to the firm's profit maximization problem. First, note that since the price of materials is invariant with respect to worker ability, the marginal revenue product of materials must be equalized across worker ability types in equilibrium. Hence, materials are allocated to workers in proportion to their productivity:

$$M_{it}(a) = \nu_{it} \phi_i(a) \omega_{it} L_{it}(a) \tag{2.20}$$

where ν_{it} is an endogenous variable equal to materials per efficiency unit of labor at firm i.

The first-order condition for $w_{it}(a)$ in the firm's profit maximization problem then allows us to express equilibrium wages as:

$$w_{it}(a) = \eta \phi_i(a) W_{it} \tag{2.21}$$

where we have defined $\eta \equiv \frac{\gamma}{1+\gamma}$ for brevity. Equation (2.21) states the familiar result that wages are a constant markdown $\eta \in (0,1)$ over the marginal revenue product of labor (MRPL) of the respective worker types, $\phi_i(a) W_{it}$. Note that in the limit as labor supply becomes infinitely elastic ($\gamma \to 1$), the markdown η approaches unity as in the benchmark with perfectly competitive labor markets. The component of wages that is common to all workers employed at firm i, W_{it}, is given by:

$$W_{it} = \frac{1}{\mu} D_{it}^{\frac{1}{\sigma}} X_{it}^{-\frac{1}{\sigma}} \omega_{it} T_{it} f_L \left(1, \nu_{it}\right) \tag{2.22}$$

where f_L denotes the derivative of f with respect to its first argument and we have defined $\mu \equiv \frac{\sigma}{\sigma-1}$ for brevity. We henceforth refer to W_{it} as the *firm-level wage.*

Similarly, the first-order condition for materials implies:

$$Z_{it} = \frac{1}{\mu} D_{it}^{\frac{1}{\sigma}} X_{it}^{-\frac{1}{\sigma}} T_{it} f_M \left(1, \nu_{it}\right) \tag{2.23}$$

where f_M denotes the derivative of f with respect to its second argument. Equilibrium output for firm i can also be expressed as:

$$X_{it} = T_{it} f \left(1, \nu_{it}\right) \bar{L}_{it} \tag{2.24}$$

where $\bar{L}_{it} \equiv \sum_{a \in A} \phi_i \left(a\right) \omega_{it} L_{it} \left(a\right)$ is the total efficiency units of labor hired by the firm. This can be written as:

$$\bar{L}_{it} = \left(\eta W_{it}\right)^{\gamma} \omega_{it} \tilde{\phi}_{it} \tag{2.25}$$

where we define $\tilde{\phi}_{it}$ as the *sorting composite* for firm i:

$$\tilde{\phi}_{it} \equiv \sum_{a \in A} \kappa_{it} \left(a\right) \phi_i \left(a\right)^{1+\gamma} \tag{2.26}$$

since this varies across firms only due to primitives that affect differential sorting of worker types across firms ($g_i(\cdot)$ and $\phi_i(\cdot)$).

Given the labor supply shifters $\kappa_{it}(\cdot)$ (which are determined by equilibrium in labor markets) and the network statistics $\{D_{it}, Z_{it}\}$ (which are determined by equilibrium in output markets), equations (2.22)-(2.24) define a system of three equations in the three firm-level variables $\{W_{it}, \nu_{it}, X_{it}\}$. The solution to this system determines the firm-level wage W_{it} and hence the wage of every worker employed at firm i. This system of equations plays a key role in the model since it mediates the interaction between the production network and the labor market for each firm. We study the properties of this system in more detail below in section 3.

2.6 Sales, profits, and costs

Anticipating the empirical application of the theory, we also derive expressions in the model for observables in the data. In particular, sales, profits, labor costs, and material costs for firm i

are given respectively by:

$$R_{it} = D_{it}\Phi_{it} \tag{2.27}$$

$$\pi_{it} = D_{it}\Phi_{it}\left[\frac{1}{\sigma} + \frac{1-\eta}{\mu}\left[\frac{f_L\left(1,\nu_{it}\right)}{f\left(1,\nu_{it}\right)}\right]\right] \tag{2.28}$$

$$E_{it}^{L} = \eta W_{it}\bar{L}_{it} \tag{2.29}$$

$$E_{it}^{M} = Z_{it}\nu_{it}\bar{L}_{it} \tag{2.30}$$

while sales from firm j to firm i can be written concisely in terms of the buyer's network demand, the seller's network productivity, and relationship-specific productivity:

$$R_{ijt} = \Delta_{it}\Phi_{jt}\psi_{ijt} \tag{2.31}$$

Note from equations (2.27) and (2.28) that the ratio of firm profits to sales is variable, which is a departure from standard models with constant returns to scale in production (where $\eta = 1$).

2.7 Welfare

Finally, we present a measure of welfare. Aggregating utility (2.1) over workers of a given ability type and accounting for the optimal sorting of workers to firms gives us the following measure of welfare for workers at each firm.

Claim **2.** *The average utility of workers of ability a that are employed at firm i is:*

$$\bar{u}_{it}\left(a\right) = \log I_t\left(a\right) + \log\tau_t + \textit{const.} \tag{2.32}$$

where the value of the transfer τ_t is equal to the aggregate ratio of value-added to labor income:

$$\tau_t = \frac{E_t}{\sum_{i\in\Omega^F} E_{it}^{L}} \tag{2.33}$$

Note that expected utility for a given worker type is equalized across firms in equilibrium, which is a standard result that follows from the distributional assumption over workers' idiosyncratic preferences in Assumption 2.1.

3 Comparative Static Results

Before taking the model to data, we develop several comparative static results to highlight the mechanisms through which the production network and labor market interact. We present these results in two sections. Section 3.1 studies comparative static results treating wages $w_{it}\left(a\right)$ as the object of interest. Section 3.2 then discusses comparative statics focusing on the composition of employment within a firm.

3.1 Wages

First, we examine the determinants of the wages that are offered by a firm. To do so, we study the system of equations (2.22)-(2.24) highlighted in section 2.5. Recall that this determines the firm-level wage W_{it} and hence wages for every worker at the firm $w_{it}(a)$, given values for the firm's demand shifter D_{it}, material input cost Z_{it}, TFP T_{it}, labor productivity ω_{it}, and labor supply shifters $\kappa_{it}(\cdot)$. To describe our first result, let $\epsilon_{it} \equiv \left[\frac{\log(f_L(1,\nu_{it})/f_M(1,\nu_{it}))}{d\log\nu_{it}}\right]^{-1}$ denote the elasticity of substitution between labor and materials under the optimal choice of ν_{it}.[11] The following proposition then summarizes the main comparative static results in this system.

Proposition **1.** *The wage offered by firm i to workers of ability a, $w_{it}(a)$, is:*

(i) strictly increasing in D_{it} and T_{it};

(ii) strictly increasing in ω_{it} if $\epsilon_{it} > 1$;

(iii) strictly decreasing in Z_{it} if $\sigma > \epsilon_{it}$, strictly increasing in Z_{it} if $\sigma < \epsilon_{it}$, and independent of Z_{it} if $\sigma = \epsilon_{it}$; and

(iv) strictly decreasing in $\kappa_{it}(a')$ for all $a' \in A$.

Part (iv) of Proposition 1 is a standard supply-side effect, where an upward shift of the labor supply curve tends to reduce wages. The intuition for parts (i)-(iii) can then be understood in terms of scale and substitution effects. First, since firms face upward sloping labor supply curves, any increase in employment must be accompanied by higher wages. Hence, wages increase in response to shocks that induce the firm to increase its scale, such as an increase in the demand shifter D_{it} or TFP T_{it}.

Second, a fall in material cost Z_{it} also generates positive scale effects for the firm. However, there is an additional substitution effect, as firms respond by substituting away from labor toward materials. Intuitively, the scale effect is mediated by the elasticity of substitution across goods, σ: when products are more substitutable, demand is more sensitive to price, and hence reductions in input costs translate into larger increases in sales. On the other hand, the substitution effect is naturally mediated by the elasticity of substitution between labor and materials, ϵ_{it}. In particular, when $\epsilon_{it} > 1$ (labor and materials are substitutes), the substitution effect leads to a *reduction* in wages, whereas when $\epsilon_{it} < 1$ (labor and materials are complements), it leads to an

[11] For a general production function f, the elasticity of substitution between workers and materials may depend on the inputs chosen and hence may vary by firm. In the estimation below, we impose a restriction on the functional form of f that ensures ϵ is constant across all firms, although this is not required for the comparative static results discussed here.

increase in wages. Therefore, the condition $\sigma > \epsilon_{it}$ reflects the case in which the postiive scale effect dominates the potentially negative substitution effect. Since $\sigma > 1$, a sufficient condition for this to hold is $\epsilon_{it} < 1$, in which case the scale and substitution effects both operate in the same direction. In our estimation of the model's parameters described below, we find that the condition $\sigma > \epsilon_{it}$ in fact holds and hence lower material costs induce higher wages.

Third, changes in labor productivity ω_{it} exert similar scale and substitution effects on wages as changes in material costs. However, the two are not equivalent, since an improvement in labor productivity directly increases the MRPL, whereas a decline in material costs only affects the MRPL endogenously through changes in material inputs. Hence, a sufficient condition for wages to be increasing in ω_{it} is $\epsilon_{it} > 1$, which is weaker than the corresponding condition for wages to be declining in Z_{it} ($\epsilon_{it} > \sigma$). We find that this condition is satisfied empirically in our estimation results described below and hence improvements in labor productivity increase wages.

We next characterize how wages respond to exogenous shocks to TFP T_{it} and labor productivities ω_{it}, accounting for the effect of these shocks on $\{D_{it}, Z_{it}\}$ through the production network. We will restrict attention to shocks that do not affect aggregate income E_t and the labor market indices $I_t(\cdot)$ in order to highlight the main mechanisms of interest. To describe the result, let $\hat{Y}_{it} \equiv d\log Y_{it}$ denote the marginal log change in a firm-level variable Y_{it} and $\hat{Y}_t \equiv \left\{\hat{Y}_{it}\right\}_{i\in\Omega^F}$ the vector of these log changes for all firms. In addition, let Σ_t^C denote the sales share matrix, Σ_t^S the input share matrix, ϵ_t the vector of labor-material substitution elasticities, and s_t^L the vector of labor shares in production costs.[12]

Proposition **2.** *The passthrough to wages from shocks to productivities $y \in \{T, \omega\}$ can be decomposed as:*

$$\hat{W}_t = \left[H_{y,t}^{direct} + G_{y,t}^{down} + G_{y,t}^{up} + G_{y,t}^{int}\right]\hat{y}_t \tag{3.1}$$

where:

$$G_{y,t}^{down} \equiv H_t^{down}\sum_{d=1}^{\infty}\left(\Sigma_t^C H_t^C\right)^d H_{y,t}^{down}, \qquad G_{y,t}^{up} \equiv H_t^{up}\sum_{d=1}^{\infty}\left(\Sigma_t^S H_t^S\right)^d H_{y,t}^{up} \tag{3.2}$$

and $G_{y,t}^{int}$ is a matrix polynomial of interaction terms in $\left\{\Sigma_t^C, \Sigma_t^S\right\}$. All elements of the matrices H and coefficients of $G_{y,t}^{int}$ depend only on γ, σ, ϵ_t, and s_t^L.

This expression for passthrough can be interpreted as follows. The first term $H_{y,t}^{direct}$ captures direct passthrough from shocks to a firm's productivity to its own wages. The second term $G_{y,t}^{down}$

[12]The (i, j)-element of Σ_t^C is the share of firm i's sales accounted for by firm j: R_{jit}/R_{it}. The (i, j)-element of Σ_t^S is the share of firm i's materials expenditures accounted for by firm j: R_{ijt}/E_{it}^M. Labor shares of cost s_{it}^L are equal to $E_{it}^L/\left(E_{it}^L + E_{it}^M\right)$.

captures passthrough from downstream productivity shocks that affect a firm's customers, where the d^{th} term in the sum captures the effects of shocks d steps downstream. Intuitively, the importance of downstream shocks for a firm's wages depends on the matrix of sales shares in the network, Σ_t^C. Similarly, the third term $G_{y,t}^{up}$ captures passthrough from upstream productivity shocks that affect a firm's immediate and indirect suppliers, with the input share matrix Σ_t^S determining the importance of each upstream firm. The last term $G_{y,t}^{int}$ accounts for interactions between downstream and upstream shocks, which arise from the fact that marginal costs are increasing and hence are dependent on scale.[13]

The following remark characterizes a special case of the result in Proposition 2 to provide deeper insight into the determinants of the strength of passthrough. As shown in Proposition 1, upstream shocks that affect firm input costs Z_{it} have no effect on wages if $\sigma = \epsilon_{it}$. This eliminates the complex interaction effects that arise from increasing marginal costs and allows us to simplify the expression for passthrough as follows.

Remark **1.** *Suppose that $\sigma = \epsilon_{it}$ globally. Then the passthrough to wages from shocks to firm TFPs and labor productivities is given by:*

$$\hat{W}_t = \frac{\sigma - 1}{\gamma + \sigma}\left(\mathbb{I} - \Sigma_t^C\right)^{-1}\hat{T}_t + \frac{\sigma - 1}{\gamma + \sigma}\hat{\omega}_t \tag{3.3}$$

where $\mathbb{I}$ is the $\left|\Omega^F\right|$-dimensional identity matrix.

This result highlights the dependence of passthrough on the labor supply elasticity γ and the price elasticity of demand σ. In particular, passthrough is increasing in σ because the MRPL is more sensitive to changes in productivity when σ is large, whereas passthrough is decreasing in γ because firms have less market power in labor markets when γ is large. Furthermore, in this special case of the model with $\sigma = \epsilon_{it}$, the importance of downstream buyers in terms of passthrough can now be measured simply by the Leontief inverse of the sales share matrix Σ_t^C. On the other hand, since scale and substituton effects exactly offset each other when $\sigma = \epsilon_{it}$ as discussed above, there is no downstream or upstream passthrough of labor productivity shocks. Finally, when $\Sigma_t^C = 0$, equation (3.3) is identical to the passthrough of firm TFP shocks to wages in the Lamadon et al. (2019) model, which does not have intermediate inputs.

Note that Proposition 2 implies that one can estimate the passthrough from TFP or labor productivity shocks at one firm in the network to wages at any other firm, as long as one is able to observe $\left\{\Sigma_t^C, \Sigma_t^S, s_t^L\right\}$ and can estimate $\{\sigma, \gamma, \epsilon_{it}\}$. In particular, one does not require knowledge of relationship productivities $\{\psi_{ijt}\}$ or other features of the production function f to compute the passthrough matrices. We will leverage this result in the counterfactual analysis below and use the passthrough equation (3.1) to decompose the sources of earnings volatility.

[13] Section B.1 of the appendix provides a simple three-firm example that illustrates these interactions.

3.2 Within-firm employment composition

Note from equation (2.21) that the relative wage paid by a firm i to workers of abilities a' and a depends only on the ratio $\frac{\phi_i(a')}{\phi_i(a)}$. Since this ratio is exogenous, relative wages between different worker ability types within a firm are also exogenous. Hence, the results in Propositions 1 and 2 are mainly relevant for understanding the drivers of between-firm earnings inequality. Nonetheless, note from equations (2.5), (2.6), and (2.21) that relative employment $\frac{L_{it}(a')}{L_{it}(a)}$ depends endogenously on the relative labor market indices $\frac{I_t(a')}{I_t(a)}$, which in turn depend on the wages offered by all firms for workers of abilities a' and a. Therefore, the *composition* of workers within a firm varies endogenously, which gives rise to endogenous within-firm earnings inequality. The following proposition connects changes in employment composition within a firm to changes in the firm-level wages studied in section 3.1.

Proposition **3.** *Marginal changes in relative employment are related to marginal changes in firm-level wages as follows:*

$$\hat{L}_{it}(a') - \hat{L}_{it}(a) = -\gamma \left[\operatorname{cov}_j\left(\Lambda_{jt}(a'), \hat{W}_{jt}\right) - \operatorname{cov}_j\left(\Lambda_{jt}(a), \hat{W}_{jt}\right)\right] \tag{3.4}$$

where $\Lambda_{jt}(a)$ denotes the share of ability a workers hired by firm j and cov_j denotes the covariance operator across firms.

To illustrate the intuition for this result, consider the effect of an increase in the firm-level wage W_{it} for some firm i on the market indices $I_t(\cdot)$. Recall from equation (2.3) that the share of ability a workers hired by firm i is proportional to $[g_i(a)\, w_{it}(a)]^\gamma$, which is firm i's "contribution" to the market index $I_t(a)$. Furthermore, since an increase in W_{it} raises the wage offer at firm i by the same amount for all workers regardless of ability, the differential effects of this change on $I_t(a)$ for different abilities a depend on how many workers of each ability type are hired by the firm. Hence, a high covariance between labor market shares $\Lambda_{it}(a')$ and changes in firm-level wages $\hat{W}_{it}$ reflects a large increase in labor market competition for ability a' workers, which induces all firms to reduce employment of such workers relative to others.

Proposition 3 then implies the following prediction linking within- and between-firm earnings inequality. Suppose that, as we find in our data, firms with higher values of W_{it} account for larger shares of employment of higher ability workers. Then, any shock that leads to an increase in the dispersion of firm-level wages also leads to a relative increase in competition for high ability workers versus low ability workers, since the covariance term $\operatorname{cov}_j\left(\Lambda_{jt}(a), \hat{W}_{jt}\right)$ increases and more so when a is high. As Proposition 3 establishes, this leads all firms to reduce employment of higher ability workers relative to lower ability workers. Since relative wages are exogenous, this implies a reduction in within-firm wage dispersion as well. In sum, shocks that increase between-firm earnings inequality also tend to reduce within-firm earnings inequality.[14]

[14] We illustrate this further in appendix B.2 with a simple two-firm example.

4 Identification of Model Parameters

We now turn towards identification of the model's parameters. In section 4.1, we first impose additional assumptions needed for identification. We then discuss identification results for each model parameter of interest in section 4.2. As we move toward connecting the model with worker-level and firm-level data, we now explicitly index individual workers by m.

4.1 Assumptions for identification

We impose three additional sets of assumptions: (i) functional form assumptions (section 4.1.1); (ii) assumptions regarding stochastic processes and orthogonality of shocks (section 4.1.2); and (iii) a steady-state assumption (section 4.1.3).

4.1.1 Functional form assumptions

ASSUMPTION **4.1.** *The production function f takes the following CES form:*

$$f(\phi L, M) = \left[\lambda^{\frac{1}{\epsilon}} (\phi L)^{\frac{\epsilon-1}{\epsilon}} + (1-\lambda)^{\frac{1}{\epsilon}} M^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}} \tag{4.1}$$

where $\lambda \in [0,1]$ and $\epsilon \in (0,\infty)$.

As shown in Proposition 1, the elasticity of substitution between labor and material inputs plays a key role in mediating the effects of upstream shocks in the production network on worker earnings. The CES functional form in Assumption 4.1 allows us to parsimoniously account for this through a single parameter, ϵ. Furthermore, this functional form nests a model without intermediates: in the limit as $\lambda \to 1$, output is produced using labor alone and the model simplifies to a version of the model studied in Lamadon et al. (2019).

ASSUMPTION **4.2.** *The ability of worker m at time t, a_{mt}, is comprised of a permanent (time-invariant) component $\bar{a}_m$ and a transient (time-varying) component $\hat{a}_{mt}$. The labor productivity function takes the following form:*

$$\log \phi_i(a_{mt}) = \theta_i \log \bar{a}_m + \log \hat{a}_{mt} \tag{4.2}$$

and the firm amenity function depends only on permanent worker ability, $g_i(a_{mt}) = g_i(\bar{a}_m)$.

We refer to the parameter θ_i as the *production complementarity* of firm i. The distinction between permanent and transient worker ability follows Lamadon et al. (2019) and is important for the identification of worker-firm interaction effects on worker earnings. In the model, there are two sources of worker-firm sorting: workers of different abilities may have different productivity levels in different firms through $\phi_i(a_t)$ and may value the amenities of these firms differently through $g_i(a_t)$. Assumption 4.2 restricts these determinants of sorting to be time-invariant, which facilitates identification based on earnings at the worker-firm-time level.

Assumption **4.3.** *Log relationship-specific productivity between buyer i and seller j at time t is given by:*

$$\log\psi_{ijt} = \log\psi_{it} + \log\psi_{jt} + \log\tilde{\psi}_{ijt} \tag{4.3}$$

where ψ_{it} denotes the relationship capability of firm i and $\log\tilde{\psi}_{ijt}$ is a residual.

As discussed in section 4.2.6, Assumption 4.3 will be important for decomposing observed firm-to-firm transactions into buyer and seller effects, which we will then use to construct firm-specific material prices Z_{it} as specified in equation (2.17).

4.1.2 Stochastic processes and orthogonality conditions

The sources of heterogeneity in the model can now be summarized as follows: workers are heterogeneous with respect to $\chi^L_{mt} \equiv \{\bar{a}_m, \hat{a}_{mt}\}$, firms are heterogeneous with respect to $\chi^F_{it} \equiv \{T_{it}, \omega_{it}, \psi_{it}, \theta_i, g_i(\cdot)\}$, and buyer-seller matches are heterogeneous with respect to $\chi^M_{ijt} \equiv \left\{\tilde{\psi}_{ijt}\right\}$. We now specify stochastic processes for these variables and describe the orthogonality conditions that characterize them.

Assumption **4.4.** *Log transient worker ability,* $\log\hat{a}_{mt}$*, follows a stationary mean-zero stochastic process that is independent of permanent worker ability $\bar{a}_m$.*

Stationarity of transient ability implies that we can treat the supply of workers of each ability type, $\{L(a)\}_{a\in A}$, as time-invariant, which is consistent with the steady-state assumption that we impose below. Note that all mean differences in ability across workers are captured by differences in permanent ability.

Assumption **4.5.** *Time-varying firm productivities $\{T_{it}, \omega_{it}, \psi_{it}\}$ follow stationary first-order Markov processes:*

$$\log T_{it} = F^T(\log T_{i,t-1}) + \xi^T_{it} \tag{4.4}$$

$$\log\omega_{it} = F^\omega(\log\omega_{i,t-1}) + \xi^\omega_{it} \tag{4.5}$$

$$\log\psi_{it} = F^\psi(\log\psi_{i,t-1}) + \xi^\psi_{it} \tag{4.6}$$

where the Markov innovations $\left\{\xi^T_{it}, \xi^\omega_{it}, \xi^\psi_{it}\right\}$ are iid across both firms and time.

The Markov structure of firm productivities follows well-known papers in the literature on production function estimation such as Olley and Pakes (1996) and Doraszelski and Jaumandreu (2018). As described below, we adopt the approach proposed in the later paper to estimate parameters of the production function f and hence adopt this Markov structure. This will not be required otherwise for estimation of the productivity variables themselves. Stationarity of $\{T_{it}, \omega_{it}, \psi_{it}\}$ also implies that the cross-sectional distribution of firm characteristics χ_{it}^F is time-invariant, which is consistent with the steady-state assumption that we impose below.

Assumption **4.6.** *Relationship productivity residuals $\tilde{\psi}_{ijt}$ are iid across firm pairs and time.*

As with Assumption 4.3, this will be important for decomposing observed firm-to-firm transactions into buyer and seller effects. Note that this does not imply that relationship productivities ψ_{ijt} are serially uncorrelated. Instead, persistence of ψ_{ijt} is allowed for through the Markov structure of firm relationship capabilities ψ_{it} in Assumption 4.5.

Assumption **4.7.** *The stochastic processes for worker characteristics χ_{mt}^L, firm characteristics χ_{it}^F, and firm-to-firm characteristics χ_{ijt}^M are mutually independent.*

Together with the conditions imposed in Assumption 4.2, independence of the stochastic processes for worker and firm characteristics ensures that residual worker earnings due to transient ability shocks are uncorrelated with the characteristics of the worker's firm. This is the same as the orthogonality assumption imposed in Lamadon et al. (2019). Note also that independence of firm characteristics and relationship productivity residuals does not imply that firms match at random, only that they do not match based on the residual $\tilde{\psi}_{ijt}$.

4.1.3 Steady-state

The last assumption that we impose for identification is that the data are characterized by a steady-state of the model in which general equilibrium terms do not vary over time.

Assumption **4.8.** *Aggregate income, E_t, and the labor market indices, $I_t(\cdot)$, are time invariant.*

This is equivalent to the restriction that there are no aggregate (economy-wide) shocks in the model and implies that the labor supply shifters $\kappa_{it}(\cdot)$ are time-invariant.[15] As discussed in section 4.2.2, this assumption will be important for the identification of firm effects in earnings.

4.2 Identification results

The parameters of the model that we seek to identify – denoted by Θ – can now be summarized as: (i) the labor supply elasticity, γ; (ii) production function parameters, $\{\sigma, \epsilon\}$; (iii) worker abilities for every worker m, $\{\bar{a}_m, \hat{a}_{mt}\}$; (iv) firm productivity parameters for every firm i, $\{T_{it}, \omega_{it}, \psi_{it}, \theta_i\}$; (v) amenity values for every firm i and worker m, $g_i(\bar{a}_m)$; and (vi) relationship productivity residuals for every buyer-seller firm pair ij, $\tilde{\psi}_{ijt}$. We now describe identification of each of these parameters.

[15] The Lamadon et al. (2019) model allows for multiple regions of production with aggregate shocks within each region, but the identification strategy rules out economy-wide aggregate shocks across all regions. Since we abstract from multiple production regions, our steady-state assumption is equally restrictive.

4.2.1 Labor supply elasticity

We identify the labor supply elasticity γ from the passthrough of firm-level changes in wage bills E_{it}^L to worker-level wages w_{mt}. Here, we follow Lamadon et al. (2019) and allow for measurement error in firm wage bills, such that wage bills in the data $\ddot{E}_{it}^L$ are related to wage bills in the model E_{it}^L as follows:

$$\log E_{it}^L = \log \ddot{E}_{it}^L + e_{it}^L \tag{4.7}$$

where e_{it}^L denotes an MA(k) measurement error given by $e_{it}^L = \sum_{s=0}^k \delta^{L,s} u_{i,t-s}^L$ for some weights $\delta^{L,s}$ and mean-zero shocks u_{it}^L that are iid across firms and time. We allow the shocks u_{it}^L to be correlated with the firm productivity innovations specified in equations (4.4)-(4.6) only contemporaneously, so that $\mathbb{E}\left[\xi_{it}^x u_{is}^L\right] \neq 0$ only if $s = t$, for all $x \in \{T, \omega, \psi\}$ and $i \in \Omega^F$.

Combining equations (2.21), (2.29), and (4.2), we can then express the wage for worker m at firm i as:

$$\log w_{imt} = \theta_i \log \bar{a}_m - \frac{1}{1+\gamma} \log \tilde{\phi}_i + \frac{1}{1+\gamma} \log \ddot{E}_{it}^L + \frac{1}{1+\gamma} e_{it}^L + \log \hat{a}_{mt} \tag{4.8}$$

Under Assumption 4.2, the worker-firm productivity term $\theta_i \log \bar{a}_m$ is time-invariant, while under Assumption 4.8, the sorting composite $\tilde{\phi}_i$ is time-invariant. Restricting attention to workers that do not change employers between t and $t+1$ (stayers), we can then take first-differences of equation (4.8) and write:

$$\Delta \log w_{imt} = \frac{1}{1+\gamma} \Delta \log \ddot{E}_{it}^L + \Delta \log \hat{a}_{mt} + \frac{1}{1+\gamma} \Delta e_{it}^L \tag{4.9}$$

Note that the passthrough of changes in firm-level wage bills to worker-level wages is informative about the coefficient $\frac{1}{1+\gamma}$ and hence the labor supply elasticity γ. Intuitively, γ controls the extent of imperfect competition in the labor market and hence mediates the extent of rent-sharing between a firm and its employees. Furthermore, equation (4.8) makes clear why identification of γ should rely only on stayers: the change in earnings for a worker that switches employers between t and $t+1$ is driven not only by rent-sharing but also by changes in permanent firm characteristics $\left\{\theta_i, \tilde{\phi}_i\right\}$ and hence cannot be used to identify γ.

Note that under the structural assumptions of the model, the wage bill is a sufficient statistic for all firm-level shocks that matter for worker-level wages. This implies that in the absence of measurement error, the residual in equation (4.9) contains only worker-level shocks ($\Delta \log \hat{a}_{mt}$). However, with measurement error in wage bills, the unobserved error term in equation (4.9) contains a component that is potentially correlated with the observed wage bill since $\mathbb{E}\left[\Delta \log \ddot{E}_{it}^L \Delta \log e_{it}^L\right] \neq 0$. To address this, note that $\mathbb{E}\left[\log \ddot{E}_{is}^L \Delta e_{it}^L\right] = 0$ for all $s < t-k-1$ since Δe_{it}^L depends only on measurement error shocks u_{it}^L in periods $\{t-k-1, \cdots, t\}$. Hence, under Assumption 4.7, lagged changes in wage bills $\log \Delta \ddot{E}_{is}^L$ for any $s < t-k-1$ are valid instruments for $\log \Delta \ddot{E}^L_{it}$ in identifying γ from equation (4.9). This requires serial correlation in $\Delta \ddot{E}^L_{it}$ to be non-zero with at least $k+2$ lags, which is consistent with the Markov processes for firm productivities specified in Assumption 4.5.

To provide context, we point out that identification of γ from equation (4.9) resembles the passthrough analysis in Guiso et al. (2005) and Lamadon et al. (2019). The key difference is that both of these papers treat the firm-level shock of interest as value-added, whereas our model implies that the relevant variable is the firm wage bill.[16] In Guiso et al. (2005), the relevance of value-added is based on a reduced-form model of worker earnings and hence is not structurally derived. Lamadon et al. (2019) go one step further and model the structural relationship between firm and worker outcomes, but because intermediate inputs are absent from their model, the wage bill is a constant fraction of value-added for any given firm. Hence, the passthrough coefficient $\frac{1}{1+\gamma}$ in equation (4.8) can be identified from either changes in value-added or the wage bill. In contrast, with both imperfect competition in output markets and intermediate inputs in our model, the wage bill is no longer proportional to firm value-added and identification stems from changes in the former instead of the latter.[17]

4.2.2 Worker and firm wage effects

Next, we discuss identification of the worker and firm effects in the earnings equation (4.8). We first follow Lamadon et al. (2019) and rewrite this as:

$$\log \tilde{w}_{imt} = \underbrace{\theta_i \log \bar{a}_m}_{\text{worker-firm interaction}} + \underbrace{\log \bar{W}_i}_{\text{firm FE}} + \underbrace{\log \hat{a}_{mt}}_{\text{residual}} \tag{4.10}$$

where $\bar{W}_i \equiv \frac{1}{\eta}\left(\bar{E}^L_i/\tilde{\phi}_i\right)^{\frac{1}{1+\gamma}}$ is a time-invariant firm effect that depends on both $\tilde{\phi}_i$ and the firm's mean wage bill over time $\bar{E}^L_i$, while $\log \tilde{w}_{imt} \equiv \log w_{imt} - \frac{1}{1+\gamma}\left(\log E^L_{it} - \log \bar{E}^L_i\right)$ is worker earnings residualized by the innovation in its employer's wage bill. Equation (4.10) is of the same form as the reduced-form model of earnings in Bonhomme et al. (2019), who show that the model implies the following restriction:

$$\mathbb{E}\left[\frac{1}{\theta_j}\left(\tilde{w}_{jm,t+1} - \log \bar{W}_j\right) - \frac{1}{\theta_i}\left(\tilde{w}_{im,t} - \log \bar{W}_i\right) | m \in M^{i\to j}_{t,t+1}\right] = 0 \tag{4.11}$$

[16] There are also subtle differences in the assumptions placed on the stochastic processes for firm-level shocks. Guiso et al. (2005) assume that log value-added follows an AR(1) process with innovations comprised of a unit root process plus an MA(1) process. Lamadon et al. (2019) make the same assumptions as Guiso et al. (2005) but constrain the AR(1) coefficient to be zero. In contrast, we allow for non-linear first-order Markov processes in firm primitives that determine firm wage bills (Assumption 4.5) and MA(k) measurement errors in wage bills, but consider only stationary processes for firm and worker shocks (which is necessary for the steady-state described in Assumption 4.8 to exist).

[17] In appendix D.1, we document our estimates of the labor supply elasticity γ using value-added shocks instead of wage bill shocks and show that we obtain different results. Hence, the distinction is both theoretically and empirically relevant.

where the expectation is taken over the set of workers $M_{t,t+1}^{i\to j}$ that move from firm i at time t to firm j at time $t+1$.

In theory, equation (4.11) gives $\left|\Omega^F\right|^2$ moment conditions for identification of $2\left|\Omega^F\right|$ parameters. Intuitively, changes in earnings that accompany changes in employers are informative about the firm-specific determinants of earnings $\left\{\theta_i, \bar{W}_i\right\}$. In practice, however, we follow Bonhomme et al. (2019) and restrict the firm effects $\left\{\theta_i, \bar{W}_i\right\}$ to vary only by K clusters of firms, so that $\theta_i = \theta_{k(i)}$ and $\bar{W}_i = \bar{W}_{k(i)}$, where we refer to $k(i)$ as the *earnings cluster* of firm i. Although not strictly necessary for identification, this reduces the dimension of the parameter set that needs to be estimated and ameliorates the well-known limited mobility bias issue. We discuss the clustering procedure in more detail below.

Given identification of $\left\{\theta_{k(i)}, \bar{W}_{k(i)}\right\}$, permanent worker ability is then identified from:

$$\log \bar{a}_m = \mathbb{E}\left[\frac{\log \tilde{w}_{imt} - \log \bar{W}_{k(i)}}{\theta_{k(i)}}\right] \tag{4.12}$$

and transient worker ability is identified as the residual in earnings given identification of all other determinants of earnings. Furthermore, the time-varying firm effect W_{it} can be recovered from:

$$\log W_{it} = \log \bar{W}_{k(i)} + \frac{1}{1+\gamma}\left(\log E_{it}^L - \log \bar{E}_i^L\right) \tag{4.13}$$

Note that even though the time-invariant firm effect $\bar{W}_{k(i)}$ is restricted to vary only by cluster, the full time-varying firm effect W_{it} is firm-specific.

4.2.3 Amenities

Just as we restrict production complementarities θ_i to vary only by firm cluster for purposes of estimation, we impose a similar restriction on the firm amenity function:

$$g_i(a) = \tilde{g}_i \bar{g}_{k(i)}(\bar{a}) \tag{4.14}$$

This is the same decomposition of amenities as in Lamadon et al. (2019). The component $\bar{g}_{k(i)}(\bar{a})$ allows for worker-firm variation in amenities but restricts this to be the same for firms within a cluster, again for the purpose of reducing dimensionality. Variation in amenities across firms within a cluster is then accounted for by $\tilde{g}_i$.

In Appendix D.3, we show that the cluster-ability component of amenities can be identified from:

$$\bar{g}_k(\bar{a}) = (\bar{a})^{-\theta_k}\left[\Lambda_{kt}(\bar{a})\right]^{\frac{1}{\gamma}} \tag{4.15}$$

where $\Lambda_{kt}(\bar{a})$ is the share of workers of permanent ability $\bar{a}$ that are employed by firms in cluster k. Since a firm with a high value of amenities is able to attract a large share of workers at a lower wage, the amenities component $\bar{g}_k(\bar{a})$ is intuitively identified from cluster-ability level employment shares after controlling for the relevant determinants of earnings heterogeneity across cluster-ability groups, namely permanent worker abilities $\bar{a}$ and production complementarities θ_k. Similarly, the firm-specific component of amenities can be identified from:

$$\tilde{g}_i = \frac{1}{W_{it}} \left(\frac{\bar{\Lambda}_{it}}{\bar{\Lambda}_{k(i)t}} \right)^{\frac{1}{\gamma}} \tag{4.16}$$

where $\bar{\Lambda}_{it}$ and $\bar{\Lambda}_{k(i)t}$ denote the shares of employment across all worker types accounted for by firm i and cluster $k(i)$ respectively. The firm-specific component of amenities is hence intuitively identified from within-cluster employment shares $\frac{\bar{\Lambda}_{it}}{\bar{\Lambda}_{k(i)t}}$ controlling for the relevant determinant of earnings heterogeneity within clusters, namely the firm-level wage W_{it}.

4.2.4 Firm relationship capability and relationship-specific productivity

Using equations (2.31) and (4.3), we can write log sales from firm j to firm i as:

$$\log R_{ijt} = \gamma \log \eta + \underbrace{\log \tilde{\Delta}_{it}}_{\text{buyer effect}} + \underbrace{\log \tilde{\Phi}_{jt}}_{\text{seller effect}} + \log \tilde{\psi}_{ijt} \tag{4.17}$$

where $\tilde{\Delta}_{it} \equiv \Delta_{it}\psi_{it}$ and $\tilde{\Phi}_{it} \equiv \Phi_{it}\psi_{it}$. Under Assumption 4.3, the assignment of buyers to sellers is independent of $\tilde{\psi}_{ijt}$ and hence $\mathbb{E}\left[\log \tilde{\Delta}_{it} \log \tilde{\psi}_{ijt}\right] = \mathbb{E}\left[\log \tilde{\Phi}_{it} \log \tilde{\psi}_{ijt}\right] = 0$.[18] The buyer effect $\tilde{\Delta}_{it}$ is hence identified from purchases by firm i from all its suppliers controlling for total sales by these suppliers, while the seller effect $\tilde{\Phi}_{jt}$ is identified from sales by firm j to all its customers controlling for total expenditures by these customers. This follows Bernard et al. (2019). Note that this identification strategy only requires cross-sectional moments, in contrast with the worker and firm earnings effects in equation (4.10), which are identified based on movements of workers across firms over time. The difference in intermediate input markets is that matching occurs many-to-many: each seller can have several buyers at once and each buyer can have several sellers.

Next, from equations (2.8) and (2.13), the share of a firm's total sales that come from the network (i.e. excluding final sales) can be expressed as:

$$s_{it}^{net} = \frac{\psi_{it} \sum_{j \in \Omega_{it}^C} \tilde{\Delta}_{jt} \tilde{\psi}_{jit}}{E_t + \psi_{it} \sum_{j \in \Omega_{it}^C} \tilde{\Delta}_{jt} \tilde{\psi}_{jit}} \tag{4.18}$$

[18] Bernard et al. (2019) find strong evidence in support of this assumption using Belgian firm-to-firm transactions data.

Solving for ψ_{it}, we obtain:

$$\psi_{it} = E_t \left(\frac{s_{it}^{net}}{1 - s_{it}^{net}} \right) \frac{1}{\sum_{j \in \Omega_{it}^C} \tilde{\Delta}_{jt} \tilde{\psi}_{jit}} \tag{4.19}$$

Firm relationship capability ψ_{it} is therefore identified (up to a normalizing constant) from observable network sales shares s_{it}^{net} and terms $\left\{ \tilde{\Delta}_{jt}, \tilde{\psi}_{ijt} \right\}$ that are identified from equation (4.17). Intuitively, a higher value of ψ_{it} increases sales only within the network but not to final consumers. Hence, after controlling for total final expenditure E_t and characteristics of a firm's customers within the network through $\sum_{j \in \Omega_{it}^C} \tilde{\Delta}_{jt} \tilde{\psi}_{jit}$, variation in s_{it}^{net} is informative about ψ_{it}.

4.2.5 Demand price elasticity

In Appendix D.5, we show that the price elasticity of demand σ is identified from the following moment condition:

$$\sigma = \mathbb{E} \left[\frac{R_{it} \tilde{\eta}_{it}}{R_{it} \tilde{\eta}_{it} - E_{it}^L - E_{it}^M} \right] \tag{4.20}$$

where $\tilde{\eta}_{it} \equiv \eta \left(1 + \frac{E_{it}^M}{E_{it}^L} \right) \left(1 + \eta \frac{E_{it}^M}{E_{it}^L} \right)^{-1}$ is a firm-specific correction factor that accounts for imperfect competition in the labor market. Intuitively, σ controls the extent of product differentiation and hence is a key determinant of firm profits. We thus identify σ from firm sales R_{it} and input expenditures $\left\{ E_{it}^L, E_{it}^M \right\}$. Note that as $\gamma \to \infty$, the correction factor $\tilde{\eta}_{it}$ approaches one and σ is identified from the population average sales-profit ratio, as in standard CES production models with perfectly competitive labor markets.

4.2.6 Labor-materials substitution elasticity and labor productivity

In Appendix D.6, we show that a firm's relative expenditure on materials versus labor can be expressed as:

$$\log \frac{E_{it}^M}{E_{it}^L} = \log \left[\frac{1}{\eta} \left(\frac{1 - \lambda}{\lambda} \right) \right] + (\epsilon - 1) \log \frac{W_{it}}{Z_{it}} + (1 - \epsilon) F^{\omega} \left(\log \omega_{i,t-1} \right) + (1 - \epsilon) \xi_{it}^{\omega} \tag{4.21}$$

where W_{it} is identified from equation (4.13) and $Z_{it} = \left(\sum_{j \in \Omega_i^S} \tilde{\Phi}_{jt} \psi_{it} \tilde{\psi}_{ijt} \right)^{\frac{1}{1-\sigma}}$ depends only on identified terms discussed in section 4.2.4.[19] Equation (4.21) is the standard relationship between relative factor expenditures and relative factor prices implied by cost minimization under the CES technology in Assumption 4.1. With the first-order Markov structure of productivity

[19] Although we allow for measurement error in firm wage bills in section 4.2.1, we follow the production function estimation literature and assume that *relative* input costs are observed without error, so that $\frac{E_{it}^M}{E_{it}^L} = \frac{\ddot{E}_{it}^M}{\ddot{E}_{it}^L}$.

innovations in Assumption 4.5, identification of the substitution elasticity ϵ follows the strategy in Doraszelski and Jaumandreu (2018), which uses lagged values of input expenditures and factor prices as instruments for the regressor and a control function (in lagged factor prices and expenditures) to control for the lagged labor productivity term $F^{\omega}(\log \omega_{i,t-1})$. Given an estimate of ϵ, labor productivities are easily recovered as residuals in the relationship between relative input expenditures and prices, $\omega_{it} = \left[\frac{1}{\eta}\left(\frac{1-\lambda}{\lambda}\right)\left(\frac{E^L_{it}}{E^M_{it}}\right)\right]^{\frac{1}{\epsilon-1}}\left(\frac{W_{it}}{Z_{it}}\right)$.

Note that the relevant factor prices are firm-specific, instead of market-specific as is commonly assumed in the production function estimation literature: W_{it} due to imperfect competition in the labor market and Z_{it} due to heterogeneity in the production network. Furthermore, as pointed out in Doraszelski and Jaumandreu (2018), the weight on labor in the production function λ is not separately identified from the average level of labor productivity ω_{it} across firms. This is intuitive, since both λ and ω_{it} control the productivity of labor relative to materials. Hence, in what follows we set λ to an arbitrary constant in the interval $(0, 1)$. This is without loss of generality in the context of the results that we present below.[20]

4.2.7 Firm TFP

We have now established identification of all parameters of interest Θ except for firm TFPs, T_{it}. For this, we require (at least) $\left|\Omega^F\right|$ moment conditions. We construct these using the time-varying firm effects W_{it}, identification of which is given by equation (4.13). Note that one can write these firm effects in general as:

$$W_{it} = F_i\left(T_t|\Theta_{-T}\right) \tag{4.22}$$

where $T_t \equiv \{T_{it}\}_{i\in\Omega^F}$, $\Theta_{-T} \equiv \Theta \setminus T_t$, and $\{F_i\}_{i\in\Omega^F}$ is a set of *known* functions that depend on the structural relationships of the model. Given identification of all other parameters Θ_{-T}, equation (4.22) hence provides a set of moments for exact identification of firm TFPs.

We choose this approach because it ensures that the model replicates the firm effects on earnings that we estimate from the data, which in turn guarantees that the model matches observed earnings for a given worker conditional on also replicating the worker's observed choice of employer. This allows us to examine changes in earnings under various counterfactual scenarios with the confidence that the baseline model provides a good fit to observed earnings. Note that in the limit of our model without intermediates ($\lambda \to 1$), $\log W_{it}$ is linear in $\log T_{it}$ and hence identification is trivial. With intermediates, however, the functions F_i are in general defined

[20] This can be seen from the production function (2.10). Output for a given worker ability type (omitting firm, time, and ability indices) can be written as $X = \tilde{T}\left[(\phi\tilde{\omega}L)^{\frac{\epsilon-1}{\epsilon}} + M^{\frac{\epsilon}{\epsilon-1}}\right]^{\frac{\epsilon}{\epsilon-1}}$, where $\tilde{T} \equiv (1-\lambda)^{\frac{1}{\epsilon-1}}\,T$ and $\tilde{\omega} \equiv \left(\frac{\lambda}{1-\lambda}\right)^{\frac{1}{\epsilon-1}}\omega$. Hence, the production function is parameterized in terms of $\tilde{T}$ and $\tilde{\omega}$ instead of $\{\lambda, T, \omega\}$ separately.

implicitly and involve complex non-linearities, which hence requires a numerical solution for the TFP vector.[21]

5 Estimation of Model Parameters

We first describe the data that we use to estimate the model in section 5.1. Our estimation results are presented in section 5.2 and the fit of the model to data is assessed in section 5.3.

5.1 Data Sources and Sample Selection

To implement estimation of the model's parameters, we use four administrative datasets from the Internal Revenue Service (IRS, or SII for its acronym in Spanish) in Chile. These datasets cover the entire formal private sector in Chile. Below, we describe the data sources, sample selection, and key variables.

First, we use a matched employer-employee dataset (IRS tax affidavits 1887 and 1879) that reports annual earnings from each job that a worker has from 2005-2018. Earnings include wages, salaries, bonuses, tips, and other sources of labor income deemed taxable by the IRS. As earnings are reported net of social security payments, we adjust the earnings measure to include these payments. Second, we use a database from the civil registry that has the year of birth of each individual who is alive in 2018. We merge this dataset with the employer-employee dataset using workers' unique tax IDs to measure the age of every worker. Third, we use a firm-to-firm dataset (IRS tax form 3323) that is based on value-added tax (VAT) records from 2005-2010. Each firm in this dataset reports the full list of its intermediate buyers and suppliers, as well as the total gross value of transactions with each buyer and supplier. As reporting occurs semi-annually, we aggregate this data to the annual level to make it consistent with the other datasets. Since this dataset reports transactions gross of taxes, we measure transactions net of taxes by using the flat value-added tax rate of 19% that was in effect in Chile during the sample period. Finally, we use an administrative dataset (IRS tax form 29) that contains a set of firm balance sheet characteristics from 2005-2018. We use this dataset to measure total sales and material cost for each firm. Firms in each of the datasets above are assigned a unique tax ID that is consistent across datasets, which facilitates the merging of these datasets. In what follows, we define a firm as a tax ID.[22]

For the firm-to-firm dataset, we impose the following sample restrictions. We drop relationships involving firms that do not report value-added or employment, or firms that report

[21] Due to these features of the F_i functions, establishing a unique solution for T_t given a vector of firm effects W_t is not trivial. Nonetheless, we have explored the potential for multiplicity by varying the initial guess for the TFP vector and never find multiplicity to occur in practice.

[22] As all tax forms are reported at the headquarter-level, plant-level information is not available. Furthermore, while it is possible that a firm has several tax IDs, information that allows us to observe firm ownership is not available.

negative value-added, sales, or materials. Next, we implement an iterative procedure that drops firms that have only one relationship, as in Bernard et al. (2019), which is required for the decomposition of firm-to-firm transaction values into buyer and seller effects. After imposing these sample restrictions, the dataset includes 32 million firm-to-firm-year observations and 17 million observations of unique firm pairs. This corresponds to 593 (923) thousand supplier-year (buyer-year) observations and 195 (289) thousand unique suppliers (buyers). We refer to this restricted dataset as the *baseline firm-to-firm dataset*.

For the employer-employee dataset, we impose the following sample restrictions following the criteria of Lamadon et al. (2019). In each year, we start with all individuals aged 25-60 who are linked to at least one employer. We identify links using only information on labor contracts (tax affidavit 1887). Next, we drop firms that have missing or negative value-added, sales, or materials in the balance sheet data (tax form 29). Then, we keep for each worker the firm that pays the highest earnings in a given year. Since we do not have hours worked or a direct measure of full-time employment, we follow the literature by including workers for whom annual earnings are above a minimum threshold (Song et al., 2019). We set the threshold equal to 32.5% of the national average of earnings in order to make our estimates comparable to the cross-country study of earnings inequality in Bonhomme et al. (2020). After imposing these sample restrictions, the dataset includes 42 (2) million worker-year (firm-year) observations and 6,497 (488) thousand unique workers (firms). We refer to this restricted dataset as the *baseline employer-employee dataset*. Finally, for both the employer-employee and firm-to-firm datasets, we transform nominal variables to 2015 real dollars.

Starting from the baseline employer-employee dataset, we define two subsamples that we will use in different parts of the paper. The first, which we refer to as the *stayers* sample, restricts the baseline sample to workers observed with the same employer for at least 8 consecutive years. This restriction is needed to allow for a flexible specification of how worker's earnings evolve over time. We also omit the first and last years of these spells to avoid concerns over workers exiting and entering employment during the year, confounding the measure of annual earnings. The stayers sample is also restricted to firms with at least 10 stayers every year which helps to ensure sufficient sample size to perform the analyses at the firm level. The stayers sample includes 6,571 (603) thousand spells and 725 (6) thousand unique workers (firms).

The second, which we refer to as the *movers* sample, restricts the baseline to workers observed at multiple firms over time. In other words, the firm that pays a worker her greatest earnings in a given year is not the same firm in all years. Following previous work and motivated by concerns about limited mobility bias, we also restrict the movers sample to firms with at least two movers (Lamadon et al., 2019). Finally, as in the previous literature (Abowd et al., 1999; Lamadon et al., 2019), we restrict this sample to firms that belong to the largest connected set of firms, which in our dataset represents 99.9% of workers. The movers sample includes 40 (1.4) million worker-year (firm-year) observations and 6,184 (201) thousand unique workers (firms).

Finally, for the purpose of estimating the elasticity of substitution between labor and materials, we merge the baseline employer-employee and the baseline firm-to-firm dataset using the unique tax IDs discussed above. We implement this merge at the firm-year level and thus exclude in the merged dataset the set of firms that do not have information in either the employer-employee or the firm-to-firm dataset. The sample includes 126 thousand firm-year observations and 48 thousand unique firms. We refer to this merged dataset as the *baseline firm-level dataset*.

Appendix Table A.1 compares the size of the three employer-employee datasets, the firm-to-firm dataset and the firm dataset we use throughout the paper. Detailed summary statistics of these samples are provided in Appendix Table A.2. The samples are broadly similar. The most noticeable differences are that the stayers sample has older, higher-earning workers and higher labor share, as well as larger firms in terms of employment and degree (number of suppliers and buyers). Nonetheless, the firms in the stayers sample are broadly similar to the firms in the baseline employer-employee dataset in terms of value-added per worker, materials share of sales, and intermediate sales as a share of total sales.

5.2 Estimation results

In this section we present the estimation results for each of the parameters in the model. Henceforth, we follow Lamadon et al. (2019) in removing age and year effects from measured wages. Specifically, we assume that measured wages $\ddot{w}_{mt}$ are related to model wages w_{mt} through $\log \ddot{w}_{mt} = \beta'_{\Upsilon} \Upsilon_{mt} + \log w_{mt}$, where Υ_{mt} is a vector of year and cubic age effects. We estimate β_{Υ} via OLS and construct $\log w_{mt} = \log \ddot{w}_{mt} - \beta'_{\Upsilon} \Upsilon_{mt}$ as our measure of wages for subsequent steps of the estimation.

5.2.1 Labor supply elasticity

We estimate the labor supply elasticity γ using equation (4.9) applied to the stayers sample, with lagged values of $\Delta \log \ddot{E}^L_{it}$ as instruments. We assume an MA(1) process for the measurement errors e^L_{it}, which implies that lagged changes in wage bills $\Delta \log \ddot{E}^L_{it}$ for any $s < t - k - 1$ are valid instruments. Hence, we use a cubic polynomial of instruments with 3 to 5 lags of wage bill changes and choose the specification with the highest F-statistic. We do not use lags above 5 in order to avoid reducing the sample size available for implementing the estimation.

Our preferred specification based on the criterion above is shown in Column 1 of Table 1, which uses 3, 4 and 5 lags of wage bill changes as instruments. We find that the passthrough elasticity of changes in firm wage bills to changes in worker earnings is around 0.15, which implies a labor supply elasticity of $\gamma = 5.5$.[23] For comparison, we also report estimation results that

[23] For robustness, we also estimate γ using the difference-in-difference estimator proposed by Lamadon et al. (2019). Results obtained using this approach are discussed in Appendix D.1. We find the same estimate using this alternative approach.

we obtain from other specifications. In Column 2, we use wage bill changes with the minimum instruments allowed by the MA(1) process – 3 lags of wage bill changes (with a cubic polynomial) – and find that the passthrough elasticity increases to 0.18 ($\gamma = 4.6$). The estimates reported in Columns 1-2 are in line with estimates from the literature of the passthrough elasticity from firm shocks to worker earnings.[24] Finally, in Column 3, we report the OLS estimate that ignores measurement error in wage bills. We find that the passthrough elasticity is substantially larger at 0.27. This implies $\gamma = 2.7$, which is half of our preferred estimate.

Table 1: Estimation of labor supply elasticity (γ)

	$\Delta \log w_{imt}$		
	(1)	(2)	(3)
$\Delta \log \tilde{E}_{it}^{L}$	0.155	0.177	0.268
	(0.006)	(0.007)	(0.001)
γ	5.5	4.6	2.7
Strategy	GMM	GMM	OLS
Instruments Accumulated Lags	5	3	
First Stage F-Stat	2325	1426	
Number of Observations	2,507,868	2,507,868	2,507,868

Notes: This table presents results from the passthrough regression based on equation (4.9). All GMM strategies use different instruments of cubic polynomials of lags of wage bill and is implemented in two stages with a robust weighting matrix used to compute standard errors. Column 1 (our preferred specification) uses changes of wage bill lagged for 3, 4 and 5 periods as instruments. Column 2 uses changes of wage bill lagged for 3 periods as instruments. Column 3 estimates the model with OLS, which ignores measurement error on the wage bill.

5.2.2 Worker and firm effects on earnings

To estimate the worker and firm effects in the earnings equation (4.10), we use the movers sample. We first follow Bonhomme et al. (2019) and assign each firm in our data to one of ten earnings clusters via a K-means clustering algorithm that targets moments of the within-firm distribution of residualized earnings $\tilde{w}_{imt}$.[25] This groups together firms whose earnings distributions are the most similar, which is motivated by the restriction that the firm-level determinants of these earnings – the firm fixed effect $\log \bar{W}_i$ and production complementarity θ_i – do not vary within a cluster. With the cluster assignment in hand, we then estimate $\left\{\log \bar{W}_i, \theta_i\right\}$ by cluster via

[24] For example, in a review of the literature, Card et al. (2018) report values for this elasticity between 0.10 and 0.15. Lamadon et al. (2019) in particular estimate a passthrough elasticity of 0.15. Note that these estimates rely on different sources of variation. Whereas we use changes in wage bills (as justified by our model), Card et al. (2018) review estimates using value added per worker while Lamadon et al. (2019) use changes in value added.

[25] Appendix D.2 provides more details including diagnostics of the clustering procedure and robustness of our results with respect to the number of clusters.

limited information maximum likelihood, based on the moment condition (4.11).[26]

Our results are presented in Table 2, where clusters are sorted according to the firm fixed effect $\log \bar{W}_i$. We observe a positive correlation between $\log \bar{W}_i$ and θ_i, indicating that firms with higher wage premia are also those where workers of higher ability are more productive.[27] In addition, the estimates that we obtain for θ_i are indicative of strong production complementarities. For example, they imply that workers in the top 2% of the permanent ability distribution are around 40% more productive when employed at firms in the highest $\bar{W}_i$ earnings cluster than at firms in the lowest $\bar{W}_i$ cluster.

Table 2: Estimates of firm fixed effects and production complementarities

Cluster	1	2	3	4	5	6	7	8	9	10
$\log \bar{W}_i$	0	0.25	0.61	0.89	1.06	1.24	1.50	1.69	1.80	1.92
θ_i	1	1.13	1.42	1.66	1.77	1.91	2.19	2.37	2.44	2.26

Notes: This table presents estimates of the firm fixed effect $\log \bar{W}_i$ and production complementarities θ_i in the earnings Equation (4.10) using the movers sample. Clusters are sorted in ascending order of the firm fixed effect, $\log \bar{W}_i$. Note that $\log \bar{W}_i$ and θ_i are normalized to zero and one respectively for firms in the first earnings cluster.

We also use our estimates to perform a preliminary decomposition of the variance of log worker earnings, which will inform the counterfactual simulations that we examine below. We follow Lamadon et al. (2019) and base this exercise on the following transformation of components in the earnings equation (4.10):

$$\log w_{imt} = x_m + \bar{f}_i + \hat{f}_{it} + \rho_{ij} + \log \hat{a}_{mt} \tag{5.1}$$

where the transformed components are defined as follows:

$$\begin{aligned} x_m &\equiv \bar{\theta} \left(\log \bar{a}_m - \log \bar{a}\right), & \bar{f}_i &\equiv \log \eta \bar{W}_i + \theta_i \log \bar{a} \\ \rho_{im} &\equiv \left(\theta_i - \bar{\theta}\right) \left(\log \bar{a}_m - \log \bar{a}\right) & \hat{f}_{it} &= \log W_{it} - \log \bar{W}_i \end{aligned} \tag{5.2}$$

This transformation separates the worker-firm interaction effect $\theta_i \log \bar{a}_m$ into worker and firm effects, thus facilitating interpretation of the variance decomposition that follows. Here, $\log \bar{a}$ and $\bar{\theta}$ denote the average values of $\log \bar{a}_m$ and θ_i respectively, where both averages are calculated at the worker-level. Intuitively, x_m is a measure of productivity for worker m when employed at the average firm, $\bar{f}_i$ is the time-averaged firm effect on earnings when matched with the average worker, $\hat{f}_{it}$ accounts for time-variation in the firm effect, and ρ_{im} captures non-linear interactions

[26] We thank Lamadon et al. (2019) for providing the code for this step of the estimation procedure.

[27] This positive correlation is also document in Lamadon et al. (2019) using US data.

between worker and firm effects.[28] Hence, the variance of log earnings can be decomposed as:

$$\operatorname{var}(\log w_{imt}) = \underbrace{\operatorname{var}(x_m)}_{\text{1. worker effect var.}} + \underbrace{\operatorname{var}\left(\bar{f}_i\right)}_{\text{2. firm fixed effect var.}} + \underbrace{\operatorname{var}\left(\hat{f}_{it}\right)}_{\text{3. time-varying firm effect var.}} + \underbrace{2\operatorname{cov}(x_m, f_{it})}_{\text{4. sorting cov.}} + \underbrace{\operatorname{var}(\rho_{im}) + 2\operatorname{cov}(\rho_{im}, x_m + f_{it})}_{\text{5. interactions}} + \underbrace{\operatorname{var}(\log \hat{a}_{mt})}_{\text{6. residual}} \tag{5.3}$$

where $f_{it} \equiv \bar{f}_i + \hat{f}_{it}$ and all variances are computed at the worker-level.

Column 1 of Table 3 presents the shares of log earnings variance accounted for by each component in equation (5.3). Unsurprisingly, the variance of the worker effect accounts for the majority (57.0%) of earnings variance. However, we also find that firms play an important role. The variance of the total firm effect f_{it} accounts for 10.8% of the log earnings variance, with most of this share accounted for by cross-sectional heterogeneity rather than variation over time. The sorting of workers to firms is even more important, with the sorting covariance explaining 19.8% of the log earnings variance. The interaction term, on the other hand, explains little.[29] Note that in the context of the structural model, the firm effect is endogenously determined by firm primitives, including network connections. In the counterfactual simulations below, we expand on this in detail to quantify the contributions of each primitive to earnings inequality.

For comparison, Table 3 also presents variance decomposition results obtained under three alternative approaches. In Column 2, we estimate the earnings equation (4.10) without residualizing worker earnings by firm innovations in wage bills, which eliminates the time-varying component of the firm effect $\hat{f}_{it}$. This has a negligible effect on the variance shares of the remaining components. In Column 3, we repeat the approach in Column 2 but set $\theta_i = 1$ for all firms, which eliminates production complementarities and hence the interaction term in equation (5.3). This leads to a slightly smaller share for the firm effect. In Column 4, we repeat the approach in Column 3 but estimate the earnings equation without grouping firms into earnings clusters, which is equivalent to the approach in Abowd et al. (1999). This increases the share of the firm effect and decreases the share of the sorting covariance, which is qualitatively similar to the effects of firm clustering reported in Bonhomme et al. (2019, 2020).

[28] In what follows, we abuse terminology somewhat by referring to both $\bar{W}_i$ and f_i as "firm fixed effects" on earnings even though these correspond to slightly different concepts.

[29] For comparison, Lamadon et al. (2019) find the following variance shares using US data: worker effect variance, 71.6%; fixed firm effect variance, 4.3%; time-varying firm effect variance, 0.3%; sorting covariance, 13.0%; interactions, 0.9%; and residual, 10.0%. Hence, we find a slightly larger role for firm effects and sorting.

Table 3: Earnings variance decomposition results

share of earnings variance explained by:		(1)	(2)	(3)	(4)
1.	worker effect variance	57.0	56.6	56.8	58.7
2.	firm fixed effect variance	10.3	10.2	7.8	12.3
3.	time-varying firm effect variance	0.5	-	-	-
4.	sorting covariance	19.8	20.5	19.9	14.4
5.	interactions	-2.0	-2.1	-	-
6.	residual	14.4	14.8	15.5	14.6
	time-varying firm effects	yes	no	no	no
	production complementarities	yes	yes	no	no
	firm clustering	yes	yes	yes	no

Notes: This table presents variance decomposition results for worker earnings using the movers sample. Column 1 is based on the model-consistent specification in Equation (5.3). Columns 2-4 successively remove time-varying firm effects $\hat{f}_{it}$, production complementarities θ_i, and firm clustering respectively from the estimation procedure. Thus, Column 4 corresponds to estimates following the approach from Abowd et al. (1999), whereas Column 1-3 corresponds to estimate following the approach from Lamadon et al. (2019) and Bonhomme et al. (2019).

5.2.3 Amenities

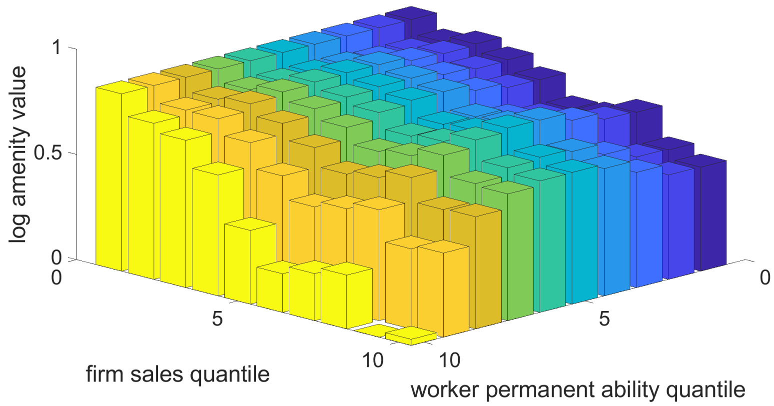
Amenities $g_i(\bar{a})$ are identified for each value of permanent worker ability $\bar{a}$ from equations (4.15) and (4.16). However, to reduce the dimension of the parameters that we estimate, we average log amenity values across workers in each of 50 quantiles of $\bar{a}$. These estimates are shown in Figure 1, where we further average log amenity values by deciles of firm sales and worker permanent ability for presentation purposes. We highlight two observations. First, for a given worker type, we find lower amenity values at larger firms. Second, this negative relationship is stronger for workers of higher ability.

Note that amenities can also be interpreted as residuals in employment shares that are not explained by observed wages. Therefore, our procedure for estimating amenities allows the model to fit the observed share of each worker type employed at each firm earnings cluster. The fit to shares constructed at the cluster level is shown in Figure 2, from which we observe the sorting of high-ability workers to firms with high wage premia (large values of $\bar{W}_i$).

5.2.4 Firm relationship capability and relationship-specific productivity

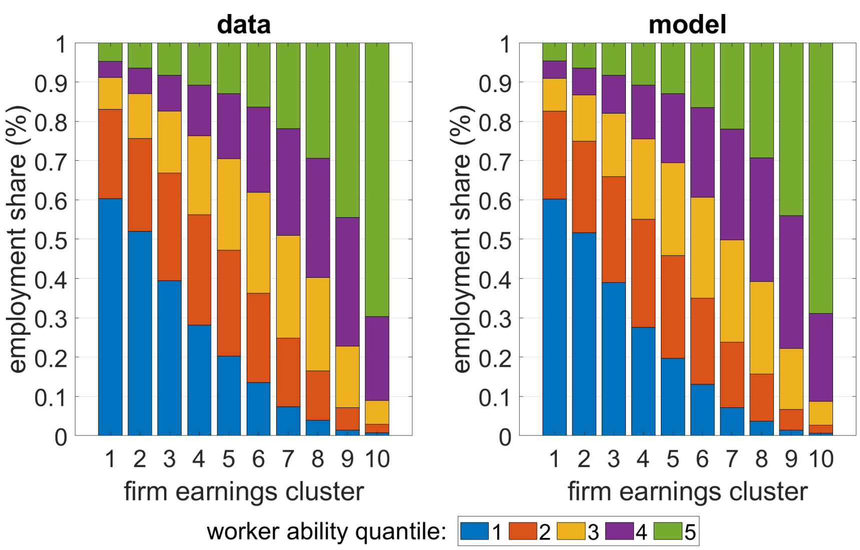
We estimate firm buyer effects $\tilde{\Delta}_{it}$, firm seller effects $\tilde{\Phi}_{it}$, and residual relationship productivities $\tilde{\psi}_{ijt}$ via a two-way fixed effects OLS regression based on equation (4.17). Details of the implementation are discussed in Appendix D.4. Of the total variance in log transaction values across all relationships, we find that 11.8% is explained by the buyer effect, 33.6% by the seller effect, -0.5% by the covariance of the seller and buyer effect and the remaining 55.1% by residual relationship productivity. Therefore, both firm-specific and relationship-specific characteristics

Figure 1: Distribution of Amenities



Notes: This figure shows the joint distribution of amenity estimates $\log g_i(\bar{a})$ by deciles of firm sales and worker permanent ability. Values are normalized for presentation purposes such that: (i) average log amenities within the smallest decile of firm sales are equal across deciles of worker permanent ability, and (ii) the smallest value of mean log amenities across sales-ability quantiles is equal to zero.

Figure 2: Model fit to employment shares by firm earnings cluster and worker ability



Notes: Firm earnings clusters are sorted in ascending order of the time-invariant firm earnings effect, $\bar{W}_i$.

are important determinants of variation in firm-to-firm sales.

With estimates of these effects in hand, we then recover firm relationship capabilities ψ_{it} using equation (4.19) and data on network sales shares s_{it}^{net}. As described in section 4.2.4, this approach only identifies ψ_{it} up to a constant. Hence, we calibrate the overall level of ψ_{it} to match the aggregate ratio of gross output to value-added in the sample.

5.2.5 Demand price elasticity

We estimate σ using the sample moment analog of the adjusted sales-profit ratio population moment in equation (4.20). Pooling observations across years, we find an average value of $\sigma = 4.2$, which implies an output markup of around 31%. We also investigate an alternative method for estimating σ, by choosing this parameter to match the aggregate profit share of sales directly in the model simulations. Using this approach, we find a similar estimate of $\sigma = 3.6$. These estimates are in line but on the low end of the range of typical values estimated in the literature, which is to be expected given that we constrain the product substitution elasticity to be same across all goods in the economy.[30]

5.2.6 Labor-materials substitution elasticity and labor productivity

We implement the approach in Doraszelski and Jaumandreu (2018) to estimate the labor-materials substitution elasticity ϵ and labor productivities ω_{it} from equation (4.21), which we repeat here for convenience:

$$\log \frac{E_{it}^{M}}{E_{it}^{L}} = \log \left[\frac{1}{\eta} \left(\frac{1-\lambda}{\lambda} \right) \right] + (\epsilon - 1) \log \frac{W_{it}}{Z_{it}} + (1-\epsilon)\, F^{\omega} \left(\log \omega_{i,t-1} \right) + (1-\epsilon)\, \xi_{it}^{\omega} \tag{5.4}$$

Following the production function estimation literature, we adopt a control function approach to control for $F^{\omega}(\cdot)$, approximating this non-parametrically using a cubic polynomial in $\log \frac{E_{it-1}^{M}}{E_{it-1}^{L}}$ and $\log \frac{W_{it-1}}{Z_{it-1}}$. To instrument for $\log \frac{W_{it}}{Z_{it}}$, we use polynomials of one-period lagged input expenditures and factor prices of labor and materials, $\{E_{it-1}^{M}, E_{it-1}^{L}, W_{it-1}, Z_{it-1}\}$. For all factor prices, we use the estimated values of W_{it} and Z_{it}, as described in sections 4.2.2 and 4.2.6. Since there are many potential instruments available, we implement estimation using all possible combinations of the instruments and vary the order of the polynomials used. We then choose the specification that delivers a first-stage F-statistic greater than 10 and a p-value of the Hansen J test above 0.1. If there is more than one specification that satisfies these criteria, we choose the one with the highest F-statistic.

Table 4 presents our results. Our preferred specification based on the criteria above is shown

[30] For example, Broda and Weinstein (2006) find an average value of $\sigma = 4$ across SITC-3 product categories, estimated using trade data for the US between 1990 and 2001, and report that estimates of σ increase when using data at higher levels of disaggregation.

in Column 1. This specification uses quadratic polynomials in $\{E^M_{it-1}, E^L_{it-1}\}$ as instruments and delivers an estimate of $\epsilon = 1.5$ (s.e.=0.058). This implies that labor and materials are substitutes in the production function ($\epsilon > 1$), a result that holds with statistical significance. For comparison, we also present the estimates of ϵ that we obtain under other specifications. In Column 2, we use cubic polynomials in $\{E^M_{it-1}, E^L_{it-1}, W_{it-1}, Z_{it-1}\}$ as instruments instead of applying the instrument selection criteria discussed above. With this specification, we find that $\epsilon = 1.0$ (s.e.=0.027). In Column 3, we estimate W_{it} using the wage model and estimation strategy in Abowd et al. (1999), which does not address the issue of limited mobility bias. Applying the instrument selection criteria above, we use a linear polynomial in $\{E^M_{it-1}, E^L_{it-1}, W_{it-1}, Z_{it-1}\}$ as instruments and find $\epsilon = 1.6$ (s.e.=0.094), which is not statistically different from our preferred estimate in Column 1. Finally, in Column 4 we follow the standard approach in the literature of using average firm wages instead of the model-consistent firm-level wage W_{it}. Our instrument set in this case is comprised of quadratic polynomials in $\{W_{it-1}, Z_{it-1}\}$. We find $\epsilon = 1.05$ (s.e.=0.043), which is not statistically different from one. Note that this specification is not consistent with our theory since the model-consistent price index of labor is W_{it} and not the average wage.[31]

Despite differences in the estimates of ϵ across the four specifications in Table 4, we find in all cases that $\sigma > \epsilon$, which holds with statistical significance (recall from section 5.2.5 that we estimate $\sigma = 4.2$). This result has important implications for the counterfactual exercises that we study below, since from Proposition 1, it implies that reductions in material input costs Z_{it} have *positive* effects on wages.

[31]The previous literature has estimated a value of ϵ below one (Doraszelski and Jaumandreu, 2018; Oberfield and Raval, 2019). Doraszelski and Jaumandreu (2018) use the ratio of the average cost of labor to the average input cost as the main right-hand side variable. Thus, our estimates, which are based on constructed price indices, are not strictly comparable. Nevertheless, in Column 4, we move closer to the empirical specification in Doraszelski and Jaumandreu (2018) by using the average wage instead of our model-based labor price index. Our estimate of ϵ falls and becomes more similar to their estimates. However, some important differences remain. In particular, we do not observe the average intermediate input price in our production network dataset (since we only observe transaction values) and as such, we cannot replicate their exact specification. A further difference is that their sample is restricted to the manufacturing sector, whereas our sample spans all sectors.

Table 4: Estimation of Elasticity of Substitution between Materials and Labor (ϵ)

	$\log E^M/E^L$			
	(1)	(2)	(3)	(4)
$\log W/Z$	0.553	0.023	0.623	
	(0.058)	(0.027)	(0.094)	
$\log \bar{w}/Z$				0.052
				(0.043)
ϵ	1.55	1.02	1.62	1.05
Model for Wage Component	BLM	BLM	AKM	Average
Instruments	$\{E^M_{it-1}, E^L_{it-1}\}$	$\{E^M_{it-1}, E^L_{it-1}, W_{it-1}, Z_{it-1}\}$	$\{E^M_{it-1}, E^L_{it-1}, W_{it-1}, Z_{it-1}\}$	$\{W_{it-1}, Z_{it-1}\}$
Instrument Polynomial	Quadratic	Cubic	Linear	Quadratic
First Stage F-Stat	130	45	84	186
Hansen's J Test	0.121	0.000	0.379	0.003
Number of Observations	44,967	44,967	44,967	44,967

Notes: This table presents the results of the estimation of ϵ outlined in Section 4.2.6 by using the baseline firm-level dataset described in section 5.1. In particular, the table presents estimates of equation (5.4). Column 1, which is our preferred result because it is consistent with our theory and also satisfy the selection criteria of instruments that delivered an F-stat above 10 and a p-value of Hansen's J test above 0.1, presents estimates of equation (5.4). Column 2 presents estimates of equation (5.4) including all available instruments: expenditure and input prices with a cubic polynomial. Column 3 and 4 presents estimates from equation (5.4) but using different measures of labor price. Column 3 uses the AKM wage model to estimate W_{it} and Column 4 uses the average wage of the firm, as in Doraszelski and Jaumandreu (2018). Note that Column 4 presents a Hansen's J test p-value lower than 0.1. That happens because across all specifications using the average wage as the price index of labor, that is the highest Hansen's J test p-value achieved. All specifications are estimated with a two-stage GMM. A robust weighting matrix is used.

5.2.7 Firm TFPs

We estimate firm TFPs by fitting the reduced-form estimates of firm-level wages W_{it} as specified in equation (4.22). Since this requires numerical solution of a non-linear set of equations, we first perform a secondary clustering procedure to further reduce the dimensionality of the parameter space. We do this also in anticipation of the counterfactual simulations discussed below, which will require numerical solution of the general equilibrium model and hence necessitates a reduction in dimensionality from the 29 thousand firms in the baseline firm-level sample. Therefore, within each earnings cluster k, we again cluster firms into k' subclusters via a K-means clustering algorithm targeting the other primitives $\{\omega_{it}, \psi_{it}, \tilde{g}_i\}$ that have been estimated at the firm-level. For our baseline results, we use $k' = 10$ subclusters for a total of 100 firm cluster-subcluster pairs that we henceforth simply refer to as firm *groups*.

Table 5 shows the correlation matrix of our TFP estimates, other estimated firm primitives, and observed sales. These are computed at the firm group level, weighted by the number of firms in each group. We highlight four observations. First, we find a negative correlation between TFP and labor productivity, although TFP is positively correlated with the product $T\omega$.[32]

[32] We estimate similar standard deviations for log TFP and labor productivity, both around 1.8.

Note that one can interpret this product as a total measure of labor productivity and T as the productivity of material inputs alone. Hence, firms that use labor more efficiently also tend to tend to be firms that use materials more efficiently. Second, we find positive correlation between production complementarities θ and both TFP and labor productivity. Therefore, higher-ability workers tend to be more productive at firms that are also inherently more productive. Third, we find negative correlation between relationship productivity ψ and TFP, but positive correlation between ψ and ω. This implies that firms that are more productive do not necessarily have more productive relationships. Fourth, firm-level amenities are negatively correlated with all productivity primitives except ψ, which is consistent with Figure 1. Finally, T, $T\omega$, and θ are all positively correlated with sales, which is reassuring.

Table 5: Correlation Matrix of Firm Characteristics

	$\log \omega$	$\log T\omega$	$\log \psi$	θ	$\log \tilde{g}$	$\log R$
$\log T$	-0.59	0.30	-0.82	0.13	-0.08	0.46
$\log \omega$		0.60	0.41	0.36	-0.15	-0.03
$\log T\omega$			-0.32	0.56	-0.25	0.42
$\log \psi$				-0.18	0.05	-0.59
θ					-0.87	0.77
$\log \tilde{g}$						-0.60

Notes: Correlations are computed at the firm group level weighted by the number of firms in each group.

5.2.8 The network

To quantify the importance of the production network for worker earnings, we measure the fraction of potential buyer-seller linkages between each pair of 100×100 firm groups in each year. We then take the average of this across years as our measure of the network $\{m_{ijt}\}_{i,j \in \Omega^F}$ in the counterfactual simulations that we study below. Figure 3 shows the patterns of matching that we observe in the network in terms of sales, employment, average wages, and TFP, where we further bin firms by decile on each variable for presentation purposes. We highlight two features of the network that are particularly important for understanding the effects of network heterogeneity on earnings inequality.

First, firms with larger sales, employment, average wages, and estimated TFPs tend to have more customers and suppliers. This dimension of heterogeneity in network connections is substantial. For example, firms in the largest TFP decile have around 3 times the number of suppliers as firms in the smallest TFP decile and 14 times the number of customers. These differences in the extensive margin of the network therefore *amplify* differences in own-firm characteristics, suggesting that network heterogeneity contributes positively to differences in mean earnings across firms and therefore to earnings inequality overall. Second, we observe negative assortative matching on sales, employment, TFP, and degree (number of customers or suppliers). For example, the average TFP of a firm's customers or suppliers is negatively correlated with the firm's own TFP. This stems from the fact that more productive firms sell to firms with both high and low TFP, whereas less productive firms sell mainly to high TFP firms. This heterogeneity in the intensive margin of the network therefore *dampens* heterogeneity in own-firm characteristics and suggests that network heterogeneity has a negative effect on earnings inequality. Hence, heterogeneity in the production network overall does not necessarily make firms more different from each other and does not mechanically induce greater earnings inequality. Rather, how network heterogeneity affects earnings inequality is a quantitative question.[33]

Figure 3: Firm-to-firm matching in the production network

Notes: Each subfigure shows the fraction of all potential buyer-seller relationships that are formed between each buyer-seller firm decile pair, where deciles are computed in terms of the indicated firm-level outcome.

5.3 Model fit

Table 6 shows the fit of the estimated model to key moments in the data. There are several important takeaways. First, the model matches observed aggregate factor shares well (panel (a)). This is by construction: the value-added share of sales is targeted through the mean of ψ_{ijt} (section 5.2.4), the aggregate labor share of value-added is targeted through σ (section 5.2.5),

[33]Negative matching on sales, employment, and degree has been documented in various other firm-to-firm datasets. See for example Bernard et al. (2018, 2019), Huneeus (2019), and Lim (2019). We also find weakly *positive* assortative matching on average wages in the data, which is consistent with findings reported by Demir et al. (2018).

and the labor-material cost ratio is targeted through the labor-materials substitution elasticity ϵ and labor productivities ω_{it} (section 5.2.6).

Second, since we estimate amenities by fitting observed employment shares (section 5.2.3) and firm TFP by fitting the firm effects in the earnings equation (section 5.2.7), the model closely replicates the observed earnings distribution (panel (b)). This is despite the fact that our amenities estimates are at a level coarser than the worker-firm level, which implies that the model does not perfectly replicate the observed assignment of individual workers to firms. In the counterfactual exercises below, we also show that the simulated model provides a good fit to the components of earnings variance in equation (5.3).

Finally, the model provides a reasonable fit to the dispersion in firm-level outcomes such as sales and wage bills, although the predicted moments have slightly lower variance (panel (c)). This is to be expected given that we restrict the estimates of firm production complementarities and TFP to vary only at the earnings cluster and group level respectively, thereby effectively abstracting from within-cluster variation. The model also slightly overpredicts the firm size-wage premium, which as discussed above is due to mismatch on sales rather than earnings. The correlation between sales and network statistics such as out-degree (number of customers) and in-degree (number of suppliers), on the other hand, matches more closely with the data.

Table 6: Fit of the model to aggregate, worker, and firm moments

(a) Aggregate			**(b) Worker earnings**			**(c) Firm-level**		
	data	model		data	model		data	model
labor share of VA	0.24	0.24	s.d.	0.75	0.80	sales, s.d.	1.60	1.27
VA share of sales	0.39	0.39	Gini coeff.	0.48	0.49	wage bill, s.d.	1.66	1.30
labor/material cost	0.13	0.15	90/10 ratio	7.10	7.37	mean wage, s.d.	0.57	0.50
			75/25 ratio	2.91	2.76	corr(sales, avg. wage)	0.53	0.73
			75/50 ratio	1.81	1.79	corr(sales, out-degree)	0.53	0.70
			50/25 ratio	1.61	1.54	corr(sales, in-degree)	0.78	0.75

Notes: Empirical moments are averages over years in 2005-2010. All variables are in logs except those expressed in shares or ratios.

6 Model Counterfactuals

We now use the estimated model to assess the role of production networks in determining worker outcomes by studying a series of counterfactual exercises. First, we perform counterfactual simulations to quantify the sources of earnings inequality (section 6.1). This extends the preliminary variance decomposition of earnings in Table 3 by relying on the structure of the model to further decompose the firm effects f_{it} in the earnings equation (5.1). These effects are endogenously determined by firm primitives, including network connections, and matter for earnings inequality through both $\text{var}(f_{it})$ and the sorting covariance $\text{cov}(x_m, f_{it})$. Second, we leverage the theoretical expressions for the passthrough of firm productivity shocks to worker earnings described in Proposition 2 to decompose the sources of earnings volatility (section 6.2). Finally, we present a minimum wage application (section 6.3), highlighting the role of networks in mediating the effects of such policies on earnings inequality.

These exercises will rely on simulations of the model that require an iterative numerical solution algorithm. Since this is not computationally feasible at the level of individual workers and firms, we group workers into 50×50 quantiles by permanent and transient ability and aggregate firms to the group level as described in section 5.2.7.

6.1 The sources of earnings inequality

The variance decomposition of earnings presented in Table 3 provides evidence that firms are quantitatively important in shaping earnings inequality. Since the firm effect f_{it} in this decomposition is determined endogenously by model primitives, however, the question of *why* firms matter for earnings inequality requires further analysis. In particular, what drives the variance of the firm effect (which accounts for 10.8% of earnings variance) and the covariance between the worker and firm effects (which accounts for 19.8%)? We now use our structural model to shed light on these questions.

To begin, note that all heterogeneity in worker earnings w_{imt} is accounted for in the model by heterogeneity in the following primitives: (i) the extensive and intensive margins of the production network, $\{m_{ijt}, \psi_{ijt}\}$; (ii) firm productivities, $\{T_{it}, \omega_{it}\}$; (iii) production complementarities, θ_i; (iv) amenities, $g_i(\cdot)$; (v) permanent worker ability, $\bar{a}_m$; and (vi) transient worker ability, $\hat{a}_{mt}$. To quantify the contribution of each set of primitives to earnings inequality, we then simulate counterfactual equilibria of the model in which each dimension of heterogeneity is eliminated by replacing the relevant parameters by their corresponding means across firms or workers. For example, to quantify the importance of heterogeneity in supplier matching for earnings inequality, we replace the observed network $\{m_{ijt}, \psi_{ijt}\}$ with a counterfactual network $\left\{\hat{m}^S_{ijt}, \hat{\psi}^S_{ijt}\right\}$ that is randomized across suppliers while holding constant the total supplier count and mean relationship productivity across suppliers for each firm:

$$\hat{m}^S_{ijt} = \frac{1}{|\Omega^F|}\sum_{j\in\Omega^F} m_{ijt}, \qquad \log\hat{\psi}_{ijt} = \frac{1}{|\Omega^F|}\sum_{j\in\Omega^F}\log\psi_{ijt} \tag{6.1}$$

We follow an analogous procedure to quantify the importance of heterogeneity in customer matching and in the other model primitives listed above.[34]

Note that eliminating heterogeneity in a given set of primitives Θ not only removes the contribution to earnings variance arising from $\text{var}(\Theta)$, but also from the covariance between

[34] In each counterfactual simulation, we also hold constant the aggregate ratio of gross output to value-added by recalibrating the grand mean of relationship productivity ψ_{ijt} (see section 5.2.4). This keeps the overall importance of materials relative to labor constant as various dimensions of primitive heterogeneity are eliminated.

Θ and all other sets of primitives. Therefore, the reduction in earnings variance that arises from eliminating heterogeneity in Θ cannot be attributed to Θ alone. To address this, we simulate counterfactuals by eliminating *all* possible combinations of heterogeneity in the primitives listed above and compute the Shapley value for each primitive in terms of its effect on earnings variance.[35] This procedure and the formal definition of the Shapley value is described in Appendix E. Intuitively, this provides an average measure of the reduction in earnings variance when heterogeneity in one set of primitives is eliminated under all possible combinations of heterogeneity in the remaining primitives. This procedure can hence be viewed as a generalization of the reduced-form variance decomposition exercise described in section 4.2.2 that allows us to quantify the share of variance accounted for by high-dimensional objects such as the production network.

To provide context, we point out that this exercise is similar in spirit to the structural counterfactual exercise examined by Lamadon et al. (2019), in which they examine the dependence of the firm effect variance and sorting covariance on TFP and amenities. The key difference is that we expand the set of firm primitives to allow for heterogeneity in the production network. Hence, we are able to precisely quantify the contribution of this dimension to earnings variance and its individual components relative to other dimensions of heterogeneity such as TFP. The inclusion of the network in the set of model primitives introduces a technical challenge: unlike the model in Lamadon et al. (2019), which is log-linear in TFP T_{it} and the sorting composite $\tilde{\phi}_i$, the dependence of the firm effect f_{it} on its underlying primitive determinants in our model does not afford an additively separable representation. This arises from the non-linearities in the input-output structure of our model. Therefore, we require an iterative numerical solution procedure, the details of which are provided in Appendix F.

Table 7 presents our findings. The first two rows compare the results of the earnings variance decomposition in equation (5.3) from the raw data (as documented in Column 1 of Table 3) and in the baseline of our simulated model. There are two reasons for potential discrepancies between these two sets of values. First, the empirical decomposition uses data at the level of individual workers and only aggregates firms only by earnings cluster, whereas the model simulations aggregate both workers and firms into groups (as described in the preface to this section). Second, given the restriction on amenities $g_i(\cdot)$ described in section 4.2.3, the model does not perfectly rationalize employment shares of different worker ability types within an

[35] To illustrate, consider two univariate primitives, Θ_A and Θ_B, and suppose the variance of earnings in the baseline can be expressed as $\text{var}(\Theta_A) + \text{var}(\Theta_B) + 2\text{cov}(\Theta_A, \Theta_B)$. The change in earnings variance from eliminating heterogeneity in Θ_A relative to the baseline is $\delta_{A1} = \text{var}(\Theta_A) + 2\text{cov}(\Theta_A, \Theta_B)$. The change in earnings variance from eliminating heterogeneity in Θ_A relative to the equilibrium in which heterogeneity in Θ_B has already been eliminated is $\delta_{A2} = \text{var}(\Theta_A)$. The Shapley contribution of Θ_A to earnings variance is then $\frac{\delta_{A1}+\delta_{A2}}{2} = \text{var}(\Theta_A) + \text{cov}(\Theta_A, \Theta_B)$. The Shapley approach is therefore equivalent to splitting the covariance equally between Θ_A and Θ_B in this univariate linear case, but generalizes the variance decomposition of the form in equation (5.3) to cases where Θ_A is high-dimensional (for example, the production network) and where the dependence of earnings is not linear in primitives.

earnings cluster. Nonetheless, we see that the model provides a good fit to the empirical variance decomposition shares.

Column 1 shows the share of total earnings variance accounted for by each set of model primitives (rows a-g). Similar to the results from the preliminary decomposition in Table 3, we find that permanent worker ability (row a) accounts for just over half of log earnings variance, while transient worker ability (row b) accounts for 13.8%. Firm-specific primitives (rows c-g) account for the remaining 32.4% of earnings variance. In particular, we find network heterogeneity to be a key driver of earnings inequality: heterogeneity in upstream connections with suppliers accounts for 11.9% of log earnings variance (row c), while heterogeneity in downstream connections with customers explains 8.6% (row d). Network heterogeneity overall therefore explains approximately 60% of the firm primitive share of earnings variance (20.5% out of 32.4%). The remaining firm-specific primitives (rows e-g) jointly account for 11.9% of log earnings variance, with more important roles for TFP, labor productivity, and production complementarity than for amenities.

To provide context for these findings, we note that the result that network heterogeneity explains the majority of the firm primitive share of earnings variance mirrors the findings in Bernard et al. (2019), who report that more than half of the variance in log firm sales is explained by heterogeneity in network connections. Hence, our findings are consistent with evidence that network connections matter for firm-level outcomes in general. These findings also show that it is important to account for production networks, as not doing so will load heterogeneity in networks onto other factors like TFP.

To provide further insight into the role of each set of primitives in shaping earnings inequality, we also report in Columns 2-5 of Table 7 the share of each component of earnings (as defined in equation (5.1)) accounted for by various model primitives.[36] Naturally, worker ability accounts for almost all of the variance in the worker effect (Column 2).[37] Interestingly, our results reveal that network heterogeneity explains almost all of the variance of the firm effect (Column 3). However, this heterogeneity matters less for sorting (Column 4). This indicates that good network connections are important for how much a firm pays its workers overall, but less so for determining the types of workers that sort to a firm. Finally, we see that network heterogeneity is not important for interactions (Column 5), although this component contributes little to overall earnings variance to begin with.

[36] Since the model-based variance decomposition is purely cross-sectional, we do not distinguish between the fixed component $\bar{f}_i$ and the innovation component $\hat{f}_{it}$ of the firm effect f_{it}.

[37] Given the definition of the worker effect in (5.2), changes in firm primitives such as TFP and labor productivity also affect the worker effect through the term $\bar{\theta}$, which is computed as an average across *workers* rather than *firms*. Hence, changes in firm primitives that affect the allocation of workers across firms also directly change the worker effect.

Table 7: Earnings variance decomposition results

	(1) earnings variance	(2) worker effect variance	(3) firm effect variance	(4) sorting covariance	(5) interactions
share of earnings variance (data)	100	57.0	10.8	19.8	-2.0
share of earnings variance (model)	100	52.5	9.8	20.8	3.1
of which:					
a. worker permanent ability, $\bar{a}_m$	53.8	48.6	-1.5	4.1	2.6
b. worker transient ability, $\hat{a}_{mt}$	13.8	-	-	-	-
c. supplier network, $\{m_{ijt}, \psi_{ijt}\}_{j\in\Omega^S_{it}}$	11.9	0.9	7.9	2.7	0.4
d. customer network, $\{m_{jit}, \psi_{jit}\}_{j\in\Omega^C_{it}}$	8.6	-0.1	6.7	1.5	0.4
e. firm productivities, $\{T_{it}, \omega_{it}\}$	6.1	7.5	-4.3	3.3	-0.5
f. production complementarities, θ_i	4.6	-4.0	-2.7	8.6	2.6
g. amenities, $g_i(\cdot)$	1.2	-0.4	3.6	0.5	-2.6

Notes: The first two rows show results from the earnings variance decomposition of equation (5.3) in the data and baseline model simulation. Values in the first row are the same as in Column 1 of Table 3. Subsequent rows show the share of earnings variance (Column 1) and each component of earnings variance (Columns 2-5) that are accounted for by each set of model primitives. Variance shares are computed using the Shapley approach described in Appendix E. Units are in percentage points.

6.2 The sources of earnings volatility

Firm-to-firm network connections may matter not only for the cross-sectional variance of worker earnings, but also for the *volatility* of earnings over time. Recall that Proposition 2 establishes this theoretically, showing how the production network mediates the passthrough of firm productivity shocks $\left\{\hat{T}_t, \hat{\omega}_t\right\}$ to changes in worker earnings via the passthrough matrices H^{direct}, G^{down}, G^{up}, and G^{int}, which account for direct, downstream, upstream, and interaction effects respectively. We now leverage this result to examine empirically the role of the production network in shaping earnings volatility. We proceed as follows.

First, we use our estimates of model parameters $\{\gamma, \sigma, \epsilon\}$ and observed network and labor cost shares $\left\{\Sigma^C_t, \Sigma^S_t, s^L_t\right\}$ to compute the passthrough matrices $\left\{H^{direct}, G^{down}, G^{up}, G^{int}\right\}$. We further split the network passthrough matrices G^{down} and G^{up} into direct and indirect components based on the representation of these terms in equation (3.2):

$$G^{down,direct}_{y,t} \equiv H^{down}_t \left(\Sigma^C_t H^C_t\right) H^{down}_{y,t}, \qquad G^{up,direct}_{y,t} \equiv H^{up}_t \left(\Sigma^S_t H^S_t\right) H^{up}_{y,t} \tag{6.2}$$

$$G^{down,indirect}_{y,t} \equiv H^{down}_t \sum_{d=2}^{\infty} \left(\Sigma^C_t H^C_t\right)^d H^{down}_{y,t}, \qquad G^{up,indirect}_{y,t} \equiv H^{up}_t \sum_{d=2} \left(\Sigma^S_t H^S_t\right)^d H^{up}_{y,t} \tag{6.3}$$

The direct components therefore account for passthrough from shocks to immediate customers and suppliers, whereas the indirect components account for passthrough from further away in the network.

Next, we estimate the stochastic processes for TFP and labor productivity. To do this, we model the joint process for $\{\log T_{it}, \log \omega_{it}\}$ as a vector autoregression (VAR) with two lags.[38] Given our estimates of $\{T_{it}, \omega_{it}\}$ for each of the six years in our sample, we then estimate the VAR model and simulate a large number of draws ($t_{sim} = 10,000$) for $\{T_{it}, \omega_{it}\}$ from its stationary distribution. We treat changes in these simulated draws as the shocks $\left\{\hat{T}_t, \hat{\omega}_t\right\}$ described in Proposition 2. This approach allows us to characterize earnings volatility using shocks drawn from the entire estimated distribution of productivities instead of using only estimated draws from the six years in our sample.

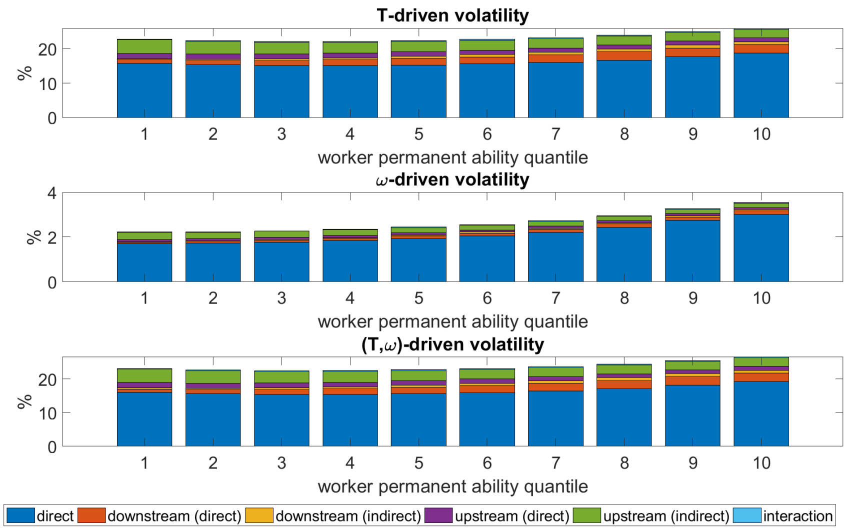
With the passthrough matrices and productivity shocks in hand, we then simulate counterfactual changes in worker earnings using equation (3.1). This provides a first-order approximation to the changes in earnings that arise from firm productivity fluctuations. This approximation is useful because it allows us to decompose earnings volatility into economically-meaningful components using Proposition 2. Following the same procedure for the variance analysis described in section 6.1, we measure the importance of each component of passthrough for earnings volatility in terms of its Shapley contribution to the absolute percentage change in worker earnings over consecutive years averaged over the t_{sim} periods. We repeat this for workers in each quantile of the permanent ability distribution to examine heterogeneous effects across worker types.

Figure 4 presents our findings, with each bar showing the mean absolute percentage change in earnings driven by the simulated productivity fluctuations for workers in each ability decile, and the shares of each bar representing the Shapley contributions of each component of passthrough to the earnings volatility measure. We highlight four key observations. First, the effects of own-employer productivity shocks account for the majority (75-80%) of worker earnings volatility, as expected. However, the contribution of passthrough from the production network is substantial, accounting for the remainder (20-25%) of simulated earnings volatility. We also observe that the network contributions are more important for the passthrough of TFP shocks compared with labor productivity shocks. This follows from the fact that the scale effects of labor productivity shocks are partially offset by substitution effects, as discussed in section 3.1, whereas TFP shocks exert only scale effects.

Second, as a share of total earnings volatility, network contributions are relatively more important for low-ability workers than for high-ability workers. This is driven less by differences in the extent of network passthrough across worker types, however, and more by the fact that own-employer shocks are more important for high-ability workers. This in turn stems from the disproportionate sorting of high-ability workers to large firms that tend to experience more volatile productivity fluctuations. For example, the estimated standard deviations of the productivity processes for employers of workers in the top ability decile are around 30% higher than for employers of workers in the lowest ability decile.

[38] This generalizes the first-order Markov structure in Assumption 4.5 to allow for second-order dependence but imposes linearity on the transition functions F^T and F^ω.

Figure 4: Components of earnings volatility generated by productivity fluctuations



Notes: The vertical axis of each figure shows the mean absolute percentage change in worker earnings across 10,000 simulated periods. These changes are generated using estimates of the passthrough matrices and productivity shocks described in Proposition 2. The shares of each bar represent the Shapley value contribution to earnings volatility of different forms of passthrough, where these values are computed following the procedure described in Appendix E.

Third, the importance of upstream shocks relative to downstream shocks is greater for low-ability workers than for high-ability workers. For instance, upstream shocks are four times as important as downstream shocks for workers in the lowest ability decile, whereas upstream and downstream shocks contribute almost equally for workers in the highest ability decile. This suggests that low-ability workers tend to be employed at firms that rely on more volatile supply chains, whereas high-ability workers tend to be employed at firms that face slightly more volatile demand conditions.

Finally, indirect passthrough of productivity shocks via the network matters as much if not more than passthrough from directly connected customers and suppliers. For example, for workers in the highest decile of permanent ability, indirect upstream passthrough accounts for around half of the total network passthrough share. This indicates that it is important for the analysis of earnings volatility to account for the full structure of the production network, rather than considering only direct customers and suppliers of each firm.

6.3 A minimum wage application

We conclude our counterfactual exercises by examining the role of production networks in shaping the effects of minimum wage (MW) policies, which are often used for the purpose of mitigating earnings inequality. We introduce this by adding to the model a constraint that requires each firm to pay any worker that it hires a wage no less than some wage floor w_{min}.[39] In our simulations below, we set the value of w_{min} equal to the 20^{th} percentile of earnings in the baseline equilibrium without a wage floor.

Recall that without the MW constraint, the optimal wage offered by firm i to workers of ability a is $w^*_{it}(a) = \eta\phi_i(a) W_{it}$, where $\phi_i(a) W_{it}$ is the MRPL of workers of ability a at firm i. The optimal wage schedule under a wage floor of w_{min} can then be expressed as:

$$w_{it}(a) = \begin{cases} w^*_{it}(a), & \text{if } w_{min} \leq w^*_{it}(a) \\ w_{min}, & \text{if } w^*_{it}(a) < w_{min} \leq \phi_i(a) W_i \\ 0, & \text{if } w_{min} > \phi_i(a) W_{it} \end{cases} \tag{6.4}$$

Wages for workers in $A^*_{it} \equiv \{a \in A : \; w_{min} \leq w^*_{it}(a)\}$ are not constrained by the MW policy, since the optimal wage that the firm would like to offer exceeds the wage floor. In contrast, the wage floor is binding for workers in $A^{min}_{it} \equiv \{w^*_{it}(a) < w_{min} \leq \phi_i(a) W_i\}$, but the firm still prefers to hire these workers because the wage floor does not exceed their MRPL at the firm. However, the firm rejects all workers in $A^0_{it} \equiv \{w_{min} > \phi_i(a) W_{it}\}$ because the lowest wage it can offer these workers exceeds their MRPL at the firm. Hence, it is now possible that some firms exit the market for certain types of workers. In the discussion below, we refer to A^*_{it}, A^{min}_{it}, and A^0_{it} as *unconstrained*, *MW*, and *rejected* worker types respectively for firm i.

Panels (a)-(d) of Table 8 show the reallocation of employment across firms induced by the MW policy, where we aggregate employment changes by quantiles of firm size (F#) and worker permanent ability (W#).[40] Panel (a) shows changes in employment of all worker types (as percentages of aggregate employment), while panels (b)-(d) show changes in employment separately for unconstrained, MW, and rejected worker types respectively. Similarly, panels (e)-(g) show changes in mean log worker earnings induced by the MW policy (excluding rejected worker types, which have zero earnings under the MW policy). We observe that reallocation of employment occurs mainly among low-ability workers, who move from small firms to large firms, with smaller reallocation of high-ability workers in the opposite direction (panel (a)). Furthermore, the MW policy increases earnings for all workers except those of the highest ability that are also employed at the largest firms (panel (e)). These changes are a composite of three sets of effects.

First, employment of MW workers increases for all firms (panel (b)). This occurs because firms are constrained to raise wages for these workers to the wage floor, with earnings for MW workers increasing by around 12% at the smallest firms and by around 8% at the largest firms

[39] See Haanwinckel (2020) for a formal discussion of the problem and equilibrium solution.

[40] Let $Q^{firm,R}_q$ and $Q^{worker,\bar{a}}_q$ denote the set of firms and worker types in the q^{th} quantiles of the firm size and worker permanent ability distributions respectively, where firm size is computed under the baseline equilibrium. Cell (q, q') of panel (a) of Table 8 shows $\sum_{i \in Q^{firm,R}_q} \int_{A \cap Q^{worker,\bar{a}}_{q'}} \left[\hat{L}_i(a) - L_i(a)\right] da$. Panels (b), (c), and (d) show the same thing but replacing A with $\hat{A}^*_i$, $\hat{A}^{min}_i$, and $\hat{A}^0_i$ respectively. Note that the values in panels (b)-(d) for a given cell (q, q') aggregate to the corresponding value in panel (a).

(panel (f)).[41] Note that the relative increase in earnings for MW workers is smaller at larger firms, since MW workers employed at larger firms are typically paid more than at small firms before the MW policy. Hence, overall log earnings variance falls by 18.6% under the policy.

Second, there are spillover effects on both employment and earnings of unconstrained workers. Notice that the spillover effects on employment largely account for the reallocation of high-ability workers from large to small firms (panel (c)). Furthermore, even though the earnings of unconstrained workers are not directly affected by the MW policy by definition, earnings increase by around 3.5% for unconstrained worker types in the smallest quintile of firms (panel (g)), which is around a quarter of the direct effect on MW workers at the same firms. In contrast, earnings for unconstrained workers at the largest firms fall. We examine the determinants of these spillover effects when considering the role of the production network below.

Finally, firms reject low-ability workers whose MRPLs fall below the wage floor (panel (d)).[42] Worker rejection is more prevalent at smaller firms, where productivity and hence MRPLs tend to be lower. Furthermore, worker rejection dominates the positive employment effects for MW workers at small firms, reducing employment of low-ability workers overall. Large firms, on the other hand, are less constrained by the MW policy as they have higher MRPLs to begin with. These firms absorb the excess supply of low-ability workers generated by exit of smaller firms from the labor market for these workers. Hence, employment of low-ability workers – earning both unconstrained and minimum wages – increases for large firms.

What is the relevance of the production network for the effects of the MW policy documented above? We highlight two key results. First, the ability of a firm to respond to the MW policy by adjusting the material inputs that it sources from the network is quantitatively important in explaining the spillover effects on unconstrained workers. Formally, note that the change in log earnings for unconstrained workers of ability a at firm i can be expressed using equations (2.22)-(2.24) as:

$$\Delta \log w_{it}(a) = -\underbrace{\frac{1}{\sigma}\Delta \log \bar{L}_{it}}_{\text{L-scale effect}} - \underbrace{\frac{1}{\sigma}\Delta \log \tilde{\nu}_{it}}_{\text{M-scale effect}} + \underbrace{\frac{1}{\epsilon}\Delta \log \tilde{\nu}_{it}}_{\text{M-substitution effect}} + \underbrace{\frac{1}{\sigma}\Delta \log D_{it}}_{\text{GE demand effect}} \tag{6.5}$$

where $\tilde{\nu}_{it} \equiv \left[\lambda^{\frac{1}{\epsilon}} + (1-\lambda)^{\frac{1}{\epsilon}} \nu_{it}^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}}$ is a monotonic transformation of the optimally chosen ratio of materials per efficiency unit of labor ν_{it} and $\Delta X \equiv \hat{X} - X$ denotes the change in outcome X.

The first term, which we label the *L-scale* effect, can be considered the direct effect of the MW policy: a firm that responds by reducing the total efficiency units of labor that it hires contracts

[41] Note that the sets of MW worker types A_{it}^{min} need not be identical across firms i and hence it is possible for employment of these worker types to increase across all firms despite the fact that supply of a given ability type a is exogenously fixed at $L(a)$.

[42] There are also some worker types that are rejected by all firms, leading to an aggregate unemployment rate of 2.05% as shown in the bottom row of panel (a).

Table 8: Changes in employment and earnings under a 20% minimum wage

	change in employment						change in mean log earnings					
	(a)	W1	W2	W3	W4	W5	(e)	W1	W2	W3	W4	W5
all worker types, A	F1	-1.53	-0.06	0.06	0.06	0.07	F1	8.99	7.52	5.38	4.24	3.71
	F2	-1.86	-0.20	0.02	0.07	0.09	F2	6.53	3.73	2.35	1.82	1.65
	F3	-1.08	-0.30	-0.08	0.02	0.06	F3	5.33	2.04	1.03	0.67	0.57
	F4	-0.98	-0.39	-0.14	-0.01	0.06	F4	5.16	1.75	0.83	0.47	0.39
	F5	3.86	0.54	0.09	-0.13	-0.29	F5	3.99	1.74	0.53	0.08	-0.24
	unemp.	1.59	0.41	0.05	0.00	0.00						
	(b)	W1	W2	W3	W4	W5	(f)	W1	W2	W3	W4	W5
min. wage worker types, A_{it}^{min}	F1	0.40	0.29	0.08	0.02	0.00	F1	12.26	11.67	11.22	11.26	10.86
	F2	0.60	0.31	0.09	0.02	0.01	F2	10.19	9.32	9.23	9.48	9.76
	F3	0.71	0.29	0.09	0.02	0.00	F3	9.43	8.39	8.27	8.32	8.50
	F4	1.04	0.39	0.13	0.03	0.00	F4	9.40	8.53	8.22	7.67	9.32
	F5	2.97	0.93	0.46	0.13	0.00	F5	8.43	8.59	7.66	8.42	6.66
	(c)	W1	W2	W3	W4	W5	(g)	W1	W2	W3	W4	W5
unconstrained worker types, A_{it}^{*}	F1	0.07	0.09	0.08	0.07	0.08	F1	3.61	3.41	3.38	3.44	3.58
	F2	-0.02	0.06	0.10	0.09	0.09	F2	1.73	1.51	1.47	1.48	1.56
	F3	-0.14	-0.09	0.01	0.04	0.06	F3	0.75	0.67	0.58	0.52	0.54
	F4	-0.22	-0.18	-0.04	0.03	0.06	F4	0.57	0.47	0.40	0.36	0.37
	F5	1.85	0.02	-0.14	-0.20	-0.29	F5	-0.53	0.05	0.02	-0.06	-0.24
	(d)	W1	W2	W3	W4	W5						
rejected worker types, A_{it}^{0}	F1	-2.00	-0.43	-0.10	-0.03	-0.01						
	F2	-2.43	-0.57	-0.16	-0.04	-0.01						
	F3	-1.65	-0.51	-0.18	-0.05	0.00						
	F4	-1.80	-0.60	-0.22	-0.07	0.00						
	F5	-0.96	-0.41	-0.23	-0.05	0.00						

Notes: Rows (F#) and columns (W#) indicate quantiles of firm size and permanent worker ability respectively. Employment changes are shown in units of percentages of aggregate employment. Earnings changes are shown in units of log differences multiplied by 100.

in scale and increases its output price, which then raises MRPLs for all unconstrained workers at the firm. The second and third terms account for the readjustment of material inputs sourced from the production network. On one hand, an increase in ν_{it} holding employment constant increases firm output, which implies a lower output price and a lower MRPL for all workers at the firm. We refer to this as the *M-scale effect*. On the other hand, an increase in ν_{it} also makes each worker at the firm more productive, hence raising MRPLs directly. We refer to this as the *M-substitution effect*. Note that the relative strength of the M-scale and M-substitution effects are determined by the parameters σ and ϵ respectively. Since we estimate $\sigma > \epsilon$, an increase in ν_{it} *increases* MRPLs on net and hence increases worker earnings. Finally, the last term captures changes in demand shifters that occur through general equilibrium effects, where any fall in demand lowers the output price that a firm is able to charge and hence reduces MRPLs for all workers.

Table A.6 in the appendix provides detailed results of the decomposition in equation (6.5) by various worker-firm groups. We summarize the key observations pertaining to the role of the production network here by noting that the readjustment of material inputs *amplifies* the positive spillover effects on earnings of unconstrained workers at small firms, explaining around 40% of the total spillover effect at the smallest quintile of firms (approximately 1.5% out of 3.5% increase in earnings). This occurs because the MW policy raises the effective cost of labor for small firms that are likely to be constrained by the wage floor, inducing them to substitute towards material inputs and *increasing* the optimally chosen ratio of materials per effective unit of labor. In contrast, for the largest firms, the excess supply of low-ability workers generated by exit of small firms from the labor market for these workers lowers the effective cost of labor, inducing substitution away from materials toward labor. Hence, the readjustment mechanism *lowers* earnings for unconstrained workers on net.

Our second result on the role of the production network in shaping the effects of the MW policy is a negative one. Existing studies have found that firms increase prices in response to MW policies (see Lemos (2008) for a survey of these effects). This evidence suggests the potential for *propagation* of the MW policy effects through the network, affecting firms who sell to and buy from other firms in the network that are heavily constrained by a wage floor. These effects enter into equation (6.5) in two ways. First, changes in $\tilde{\nu}_{it}$ may stem from changes in material costs Z_{it} and hence from changes in the output price p_{jt} charged by each supplier j to firm i. Second, changes in demand D_{it} can arise from changes in the network demand Δ_{jt} for each customer j of firm i, which are also dependent on the output prices of these customers. Therefore, our model allows us to quantitatively assess the importance of network propagation for MW policy effects.

In doing so, we arrive at an important insight: propagation of MW effects through the network is quantitatively negligible because the firms that tend to be the most constrained by a wage floor also tend to be the least important buyers and sellers in the network. For example, we find that half of all workers that earn the MW in our counterfactual scenario are employed by firms that account for only 6% of total sales and 3% of total expenditures in the network. This is consistent with the firm-to-firm matching patterns documented in Figure 3, which shows that lower-paying firms are substantially less connected in the network than higher-paying firms. Therefore, although we find that the MW policy raises output prices by as much as 1.5% for the most affected firms, the passthrough of this to changes in material costs and demand for the customers and suppliers of these constrained firms is minimal.

In sum, in evaluating the effects of MW policies on worker earnings, we find that it is important to account for readjustment of sourcing behavior in the network by constrained firms, but also that one can safely abstract from propagation of the MW policy effects through the network linkages themselves.

7 Conclusion

Matched employer-employee and firm-to-firm transactions datasets have attracted substantial interest from researchers in recent years, but these have largely been studied in isolation from each other. We have argued in this paper that the ability to link these two rich sources of data offer novel and important insights into a fundamental set of economic questions, and have provided a unified quantitative framework linking a theory of labor markets with a theory of production networks to uncover these insights.

Our analysis establishes three important takeaways. First, heterogeneity in network linkages matters for earnings inequality, explaining 21% of log earnings variance in total, with upstream heterogeneity accounting for 12% and downstream heterogeneity accounting for 9%. Second, passthrough of firm productivity shocks via network linkages matters for earnings volatility, explaining 20-25% of these fluctuations in earnings. Third, how firms substitute material inputs for labor in response to minimum wage policies matters for the spillover effects of these policies on earnings for above-minimum wage workers, explaining 40% of these effects, although propagation of the policy's effects through network linkages is quantitatively negligible.

We conclude with four potential directions for future research on the interaction between workers and production networks. First, there is growing evidence that worker outsourcing is a key driver of increases in earnings inequality (Goldschmidt and Schmieder (2017)). However, there are as yet no studies documenting such evidence where both worker flows between firms and firm-to-firm linkages are simultaneously observed. The ability to observe these jointly will allow for a refinement of the definition of outsourcing and hence of the study of its effects on worker earnings. For example, worker transitions between linked buyers and sellers may differ fundamentally in both cause and effect from worker transitions between unrelated firms.

Second, there is growing interest among both policymakers and researchers in understanding the effects of automation on worker outcomes. It is natural to view these effects as arising from the substitution of labor by inputs such as industrial robots. For example, Acemoglu and Restrepo (2020) estimate the effects of increased robot usage on employment and wages in US labor markets, finding robust negative effects. More recent theoretical work by Jackson and Kanik (2020) develops a model of robot-labor substitution that accounts for production network linkages between firms. A quantitative study of the mechanisms highlighted by this literature using matched employer-employee and firm-to-firm transactions data is therefore likely to yield important insights.

Third, an emerging literature has emphasized the importance of production network linkages for determining optimal industrial policy (Liu (2019)). However, this literature has largely focused on outcomes in product markets such as sales and aggregate output, while abstracting from labor market frictions. The framework that we have developed in this paper offers a natural starting point for the extension of such policy analyses to consider implications for heterogeneous workers, in a context with imperfect competition in labor markets and production network linkages.

Finally, while we consider in this paper how changes in the production network structure affect worker earnings, there is also nascent evidence that worker flows between firms shape the formation of network linkages. For example, Patault and Lenoir (2020) document using French data that movements of sales managers across firms induce the formation of new buyer-seller relationships. This evidence points toward the need for a better understanding of the economic determinants of both worker transitions and firm-to-firm relationship formation, which linked employer-employee and firm-to-firm transactions data are well-suited to examine.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. Econometrica 67(2), 251–333.

Acemoglu, D. and P. Restrepo (2020). Robots and jobs: Evidence from us labor markets. Journal of Political Economy 128(6).

Adao, R., P. Carrillo, A. Costinot, D. Donaldson, and D. Pomeranz (2020). International trade and earnings inequality: A new factor content approach. Working paper.

Alfaro-Ureña, A., I. Manelici, and J. P. Vasquez (2020). The Effects of Multinationals on Workers: Evidence from Costa Rica. Working paper.

Antràs, P. and A. de Gortari (2020). On the geography of global value chains. Econometrica 84(4), 1553–1598.

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

Baqaee, D. R. and E. Farhi (2019). The macroeconomic impact of microeconomic shocks: Beyond hulten's theorem. Econometrica 87(4), 1155–1203.

Baqaee, D. R. and E. Farhi (2020). Productivity and Misallocation in General Equilibrium. Quarterly Journal of Economics 135(1), 105–163.

Berger, D., K. Herkenhoff, and S. Mongey (2019). Labor market power. Working paper.

Bernard, A., E. Dhyne, G. Magerman, A. Moxnes, and K. Manova (2019). The Origins of Firm Heterogeneity: A Production Network Approach. Working paper.

Bernard, A. B., A. Moxnes, and K. H. Ulltveit-Moe (2018). Two-sided heterogeneity and trade. Review of Economics and Statistics 100(3), 424–439.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2020). How Much Should we Trust Estimates of Firm Effects and Worker Sorting? Working paper.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A distributional framework for matched employer employee data. Econometrica 87(3), 699–739.

Broda, C. and D. E. Weinstein (2006). Globalization and the gains from variety. Quarterly Journal of Economics 121(2), 541–585.

Burdett, K. and D. T. Mortensen (1998). Wage differentials, employer size, and unemployment. International Economic Review 39(2), 257–273.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2017, 12). The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy. The Review of Economic Studies 85(4), 2042–2096.

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

Chan, M., K. Kroft, and I. Mourifié (2019). An empirical framework for matching with imperfect competition. Working paper.

Chan, M., S. Salgado, and M. Xu (2021). Heterogeneous passthrough from tfp to wages. Working paper.

Demir, B., A. C. Fieler, D. Y. Xu, and K. K. Yang (2018). O-Ring Production Network. Working paper.

Dhyne, E., A. K. Kikkawa, M. Mogstad, and F. Tintelnot (2020). Trade and domestic production networks. Review of Economic Studies, forthcoming.

Doraszelski, U. and J. Jaumandreu (2018). Measuring the bias of technological change. Journal of Political Economy 126(3), 1027–1084.

Dube, A., J. Jacobs, S. Naidu, and S. Suri (2019). Monopsony in online labor markets. Working paper.

Dustmann, C., A. Lindner, U. Schönberg, M. Umkehrer, and P. vom Berge (2020). Reallocation effects of the minimum wage. CReAM Discussion Paper Series (2007).

Engbom, N. and C. Moser (2018). Earnings inequality and the minimum wage: Evidence from brazil. Working paper.

Goldschmidt, D. and J. F. Schmieder (2017). The rise of domestic outsourcing and the evolution of the german wage structure. Quarterly Journal of Economics 132(3), 1165–1217.

Guiso, L., L. Pistaferri, and F. Schivardi (2005). Insurance within the firm. Journal of Political Economy 113(5), 1054–1087.

Haanwinckel, D. (2020). Supply, demand, institutions, and firms: A theory of labor market sorting and the wage distribution. Working paper.

Huneeus, F. (2019). Production Network Dynamics and the Propagation of Shocks. Working paper.

Jackson, M. O. and Z. Kanik (2020). How automation that substitutes for labor affects production networks, growth, and income inequality. Working paper.

Jarosch, G., J. Nimczik, and I. Sorkin (2019). Granular search, market structure, and wages. Working paper.

Kikkawaa, A. K., G. Magerman, and E. Dhyne (2020). Imperfect competition in firm-to-firm trade. Working paper.

Kline, P., N. Petkova, H. Williams, and O. Zidar (2019). Who profits from patents? rent sharing at innovative firms. Quarterly Journal of Economics 134(3), 1343–1404.

Kroft, K., Y. Luo, M. Mogstad, and B. Setzler (2019). Rent sharing and imperfect competition: Quasi-experimental evidence from procurement auctions. Working paper.

Lamadon, T., M. Mogstad, and B. Setzler (2019). Imperfect Competition, Compensating Differentials and Rent Sharing in the U.S. Labor Market. Working paper.

Lemos, S. (2008). A survey of the effects of the minimum wage on prices. Journal of Economic Surveys 22(1), 187–212.

Lim, K. (2019). Production Networks and the Business Cycle. Working paper.

Liu, E. (2019). Industrial policies in production networks. Quarterly Journal of Economics 134(4), 1883–1948.

Manning, A. (2003). Monopsony in motion: Imperfect competition in labor markets. Princeton University Press.

Oberfield, E. and D. Raval (2019). Micro Data and Macro Technology. Working paper.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. Econometrica 64(6), 1263–1297.

Patault, B. and C. Lenoir (2020). How valuable are business networks? evidence from sales managers in international markets. Working paper.

Postel-Vinay, F. and J.-M. Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. Econometrica 70(6), 2295–2350.

Rosen, S. (1986). The theory of equalizing differences. Handbook of Labor Economics 1, 641–692.

Song, J., D. J. Price, N. Bloom, and T. von Watcher (2019). Firming up inequality. Quarterly Journal of Economics 134(1), 1–50.

Sorkin, I. (2018). Ranking firms using revealed preference. Quarterly Journal of Economics 133(3), 1331–1393.

Staiger, D., J. Spetz, and C. Phibbs (2010). Is there monopsony in the labor market? evidence from a natural experiment. Journal of Labor Economics 28(2), 211–236.

Taber, C. and R. Vejlin (2018). Estimation of a roy/search/compensating differentials model of the labor market. Working paper.

A Proofs of Claims and Propositions

A.1 Proof of Claim 1

Omitting time subscripts for brevity, the profit-maximization problem for a firm i can be written generally as:

$$\max_{\{p_{ji}\}_{j\in\Omega_i^C\cup\{F\}}} \left\{ \sum_{j\in\Omega_i^C\cup\{F\}_i} p_{ji}x_{ji} - C\left[X_i|l_i\left(\cdot\right), Z_i\right] \right\} \tag{A.1}$$

$$\text{s.t. } x_{ji} = \Delta_j \psi_{ji} p_{ji}^{-\sigma} \tag{A.2}$$

$$X_i = \sum_{j\in\Omega_i^C\cup\{F\}} x_{ji} \tag{A.3}$$

where $\psi_{Fi} = 1$. Here, $C\left[X_i|l_i\left(\cdot\right), Z_i\right]$ denotes the total cost of producing X_i units of output given the labor supply functions $l_i\left(\cdot\right)$ and material input cost Z_i. The latter depends on the prices charged by suppliers of firm i, which firm i takes as given in the problem above. Importantly, the total production cost for firm i depends only on total output of the firm X_i and not on how this output is allocated to each customer.

The first-order condition for the profit-maximization problem with respect to p_{ji} is then:

$$\left(1-\sigma\right)\Delta_j\psi_{ji}p_{ji}^{-\sigma} = -\sigma C'\left[X_i|l_i\left(\cdot\right), Z_i\right]\Delta_j\psi_{ji}p_{ji}^{-\sigma-1} \tag{A.4}$$

Solving for the optimal price yields:

$$p_{ji} = \frac{\sigma}{\sigma-1}C'\left[X_i|l_i\left(\cdot\right), Z_i\right] \tag{A.5}$$

Note that the right-hand side of (A.5) does not vary by customer j. Hence, the optimal prices set by firm i do not vary by customer and are equal to the standard CES markup over the firm's marginal cost. The existence of imperfect competition in the labor market implies that marginal cost is not constant, but this does not break the standard CES markup result.

A.2 Proof of Claim 2

Let $F_{\epsilon i}\left[\epsilon_i, \{\epsilon_j\}_{j\neq i}\right]$ denote the partial derivative of the idiosyncratic preference shock CDF (2.2) with respect to the shock for firm i:

$$F_{\epsilon i}\left[\epsilon_i, \{\epsilon_j\}_{j\neq i}\right] = exp\left[-\left(\sum_{i\in\Omega^F} e^{-\frac{\epsilon_{it}}{\rho}}\right)^{\rho}\right]\left(\sum_{i\in\Omega^F} e^{-\frac{\epsilon_{it}}{\rho}}\right)^{\rho-1} e^{-\frac{\epsilon_i}{\rho}} \tag{A.6}$$

First, consider the problem of a worker of a given ability who is choosing which firm to work for. For brevity, we omit time and ability indices. The probability that firm i is chosen by this worker is:

$$
\begin{aligned}
\mathbb{P}_i &= \mathrm{P}[\log\left(w_i g_i\right) + \tfrac{1}{\beta}\epsilon_i \geq \log\left(w_j g_j\right) + \tfrac{1}{\beta}\epsilon_j, \forall j \neq i] && \text{(A.7)}\\
&= \mathrm{P}[\epsilon_j \leq \log\left(\tfrac{w_i g_i}{w_j g_j}\right)^{\beta} + \epsilon_i, \forall j \neq i] && \text{(A.8)}\\
&= \int_{-\infty}^{\infty} F_{\epsilon i}\left[\epsilon_i, \left\{\log\left(\tfrac{w_i g_i}{w_j g_j}\right)^{\beta} + \epsilon_i\right\}_{j\neq i}\right] d\epsilon_i && \text{(A.9)}\\
&= \int_{-\infty}^{\infty} e^{-\Psi_i^{\rho} e^{-\epsilon_i}} \Psi_i^{\rho-1} e^{-\epsilon_i} d\epsilon_i && \text{(A.10)}
\end{aligned}
$$

where $\Psi_i \equiv \sum_{j\in\Omega^F}\left(\frac{w_j g_j}{w_i g_i}\right)^{\beta/\rho}$. Using the change of variables $z = \Psi_i^{\rho} e^{-\epsilon_i}$, we can write this as:

$$
\begin{aligned}
\mathbb{P}_i &= \frac{1}{\Psi_i}\int_0^{\infty} e^{-z} dz && \text{(A.11)}\\
&= \frac{\left(w_i g_i\right)^{\beta/\rho}}{\sum_{j\in\Omega^F}\left(w_j g_j\right)^{\beta/\rho}} && \text{(A.12)}
\end{aligned}
$$

which gives the firm choice probability (2.3).

Next, let $G_{\epsilon i}$ denote the CDF of the idiosyncratic preference shock for employment at firm i, ϵ_i, conditional on firm i being chosen. This is given by:

$$
\begin{aligned}
G_{\epsilon_i}\left(\epsilon\right) &= \frac{1}{\mathbb{P}_i}\int_{-\infty}^{\epsilon} F_{\epsilon i}\left[\epsilon_i, \left\{\log\left(\tfrac{w_i g_i}{w_j g_j}\right)^{\beta} + \epsilon_i\right\}_{j\neq i}\right] d\epsilon_i && \text{(A.13)}\\
&= \frac{1}{\mathbb{P}_i}\int_{-\infty}^{\epsilon} e^{-\Psi_i^{\rho} e^{-\epsilon_i}} \Psi_i^{\rho-1} e^{-\epsilon_i} d\epsilon_i && \text{(A.14)}
\end{aligned}
$$

Using the change of variables $z = \Psi_i^{\rho} e^{-\epsilon_i}$, we can write this as:

$$
\begin{aligned}
G_{\epsilon_i}\left(\epsilon\right) &= \frac{1}{\mathbb{P}_i \Psi_i}\int_{\Psi^{\rho} e^{-\epsilon}}^{\infty} e^{-z} dz && \text{(A.15)}\\
&= exp\left(-\Psi_i^{\rho} e^{-\epsilon}\right) && \text{(A.16)}
\end{aligned}
$$

which is a univariate Gumbel distribution with mean $\log\Psi_i^{\rho} + \Gamma$, where Γ is the Euler-Mascheroni constant. Average utility of workers employed at firm i is then given by:

$$
\begin{aligned}
\bar{u}_i &= \log w_i g_i + \frac{1}{\beta}\int_{-\infty}^{\infty} \epsilon_i dG_{\epsilon_i}\left(\epsilon_i\right) + \log\tau && \text{(A.17)}\\
&= \log w_i g_i \Psi_i^{\rho/\beta} + \log\tau + \frac{\Gamma}{\beta} && \text{(A.18)}\\
&= \log\left[\sum_{j\in\Omega^F}\left(w_j g_j\right)^{\beta/\rho}\right]^{\rho/\beta} + \log\tau + \frac{\Gamma}{\beta} && \text{(A.19)}
\end{aligned}
$$

To determine the value of the transfer τ, note that the aggregate value of transfers to workers must equal aggregate profits in the economy:

$$(\tau - 1) \sum_{i \in \Omega^F} \sum_{a \in A} w_i(a) L_i(a) = \sum_{i \in \Omega^F} \pi_i \tag{A.20}$$

Solving for τ:

$$\tau = \frac{\sum_{i \in \Omega^F} \pi_i + \sum_{i \in \Omega^F} \sum_{a \in A} w_i(a) L_i(a)}{\sum_{i \in \Omega^F} \sum_{a \in A} w_i(a) L_i(a)} \tag{A.21}$$

where the numerator in (A.21) is aggregate value-added and the denominator is aggregate labor income.

A.3 Proof of Proposition 1

In what follows, we omit firm and time subscripts for brevity and all derivatives of the production function f are evaluated at $\{\phi L, M\} = \{1, \nu\}$. Totally differentiating (2.22)-(2.24) for a given firm, we obtain:

$$\hat{W} + \frac{1}{\sigma}\hat{X} - \left(\frac{f_{LM}\nu}{f_L}\right)\hat{\nu} = \frac{1}{\sigma}\hat{D} + \hat{T} + \hat{\omega} \tag{A.22}$$

$$\frac{1}{\sigma}\hat{X} - \left(\frac{f_{MM}\nu}{f_M}\right)\hat{\nu} = \frac{1}{\sigma}\hat{D} + \hat{T} - \hat{Z} \tag{A.23}$$

$$-\gamma\hat{W} + \hat{X} - \left(\frac{f_M v}{f}\right)\hat{\nu} = \hat{T} + \hat{\omega} + \sum_{a \in A} s^{\phi}(a)\,\hat{\kappa}(a) \tag{A.24}$$

where $s^{\phi}(a) \equiv \frac{\kappa(a)\phi(a)^{1+\gamma}}{\int_A \kappa(a')\phi(a')^{1+\gamma} da'}$ is the share of the firm's efficiency units of labor that are derived from workers of ability a. Solving for $\left\{\hat{W}, \hat{X}, \hat{\nu}\right\}$, we obtain:

$$\begin{aligned}\hat{W} &= \Gamma\hat{D} + (\sigma - 1)\Gamma\hat{T} - (\sigma - \epsilon)\epsilon_m\Gamma\hat{Z} + [\sigma - 1 - (\sigma - \epsilon)\epsilon_M]\Gamma\hat{\omega} \\ &\quad - \Gamma\sum_{a \in A} s^{\phi}(a)\,\hat{\kappa}(a)\end{aligned} \tag{A.25}$$

$$\begin{aligned}\hat{X} &= (\gamma + \epsilon\varepsilon_M)\Gamma\hat{D} + \sigma(\gamma + \epsilon\epsilon_m + 1 - \varepsilon_M)\Gamma\hat{T} - \sigma(\gamma + \epsilon)\varepsilon_M\Gamma\hat{Z} \\ &\quad + \sigma(1 - \varepsilon_M)(1 + \gamma)\Gamma\hat{\omega} + \sigma(1 - \varepsilon_M)\Gamma\sum_{a \in A} s^{\phi}(a)\,\hat{\kappa}(a)\end{aligned} \tag{A.26}$$

$$\begin{aligned}\hat{v} &= \epsilon\Gamma\hat{D} + \epsilon(\sigma - 1)\Gamma\hat{T} - \epsilon(\gamma + \sigma)\Gamma\hat{Z} - \epsilon(1 + \gamma)\Gamma\hat{\omega} \\ &\quad - \epsilon\Gamma\sum_{a \in A} s^{\phi}(a)\,\hat{\kappa}(a)\end{aligned} \tag{A.27}$$

where $\varepsilon_M \equiv \frac{f_M \nu}{f}$ denotes the elasticity of f with respect to materials, $\epsilon \equiv \left(\frac{f_{LM}\nu}{f_L} - \frac{f_{MM}v}{f_M}\right)^{-1}$ denotes the elasticity of substitution between labor and materials, and $\Gamma \equiv [\gamma + \sigma - (\sigma - \epsilon)\varepsilon_M]^{-1}$.

Note that the coefficients on $\hat{D}$ and $\hat{T}$ on the right-hand side of equation (A.25) are strictly positive, the sign of the coefficient on $\hat{Z}$ depends only on the sign of $\sigma - \epsilon$, and the coefficient on $\hat{\omega}$ is strictly positive whenever $\epsilon > 1$. This establishes the comparative static results described in Proposition 1.

A.4 Proof of Proposition 2

In what follows, we omit time subscripts for brevity and denote by $\hat{Y}$ the vector of changes in a firm-specific variable $\hat{Y}_i$ for all firms. We first establish the following relationship between the production function elasticity $\varepsilon_M \equiv \frac{f_M \nu}{f}$ and the labor share of cost s^L at a given firm.

Lemma **1.** *The elasticity of the production function with respect to materials at firm i satisfies $\varepsilon_{M,i} = 1 - \tilde{s}_i^L$, where $\tilde{s}_i^L \equiv \frac{s_i^L}{s_i^L + \eta\left(1 - s_i^L\right)}$ is an adjusted measure of the labor cost share.*

Proof. From equations (2.29) and (2.30), we can express the labor share of cost for firm i as:

$$s_i^L = \frac{\eta}{\eta + \nu_i \left(Z_i / W_i\right)} \tag{A.28}$$

where recall $\eta \equiv \frac{\gamma}{1+\gamma}$. Then, from the first-order conditions (2.22) and (2.23), relative factor prices can be expressed as:

$$\frac{Z_i}{W_i} = \frac{f_M\left(1, \nu_i\right)}{f_L\left(1, \nu_i\right)} \tag{A.29}$$

Combining (A.28) and (A.29) and using the result that $f = f_M \nu + f_L$ for a homogeneous of degree one function f (see Assumption 2.2) establishes the result. ∎

Next, we derive an expression for marginal changes in demand shifters, $\hat{D}$. Totally differentiating equation (2.16) gives:

$$\hat{D} = \Sigma^C \hat{\Delta} \tag{A.30}$$

where we have used the result that the share of firm j's sales accounted for by firm i can be expressed using (2.8), (2.16) and (2.31) as:

$$\Sigma_{ijt}^C \equiv \frac{R_{ijt}}{\sum_{k \in \Omega_{it}^C \cup \{F\}} R_{kit}} = \frac{\Delta_j}{D_i} \tag{A.31}$$

Recall also that we are assuming no changes in general equilibrium variables and hence $\hat{\Delta}_{Ft} = 0$. Totally differentiating (2.14) and using (2.30), we obtain:

$$\hat{\Delta} = \gamma \hat{W} + \sigma \hat{Z} + \hat{\nu} \tag{A.32}$$

Then, taking the ratio of the first-order conditions for the profit-maximization problem (2.22)-(2.23) and totally differentiating gives:

$$\hat{W} - \hat{Z} = \epsilon^{-1} \hat{\nu} \tag{A.33}$$

where here ϵ denotes a $\left|\Omega^F\right| \times \left|\Omega^F\right|$ diagonal matrix with i^{th}-diagonal element equal to the elasticity of substitution between labor and materials for firm i, $\epsilon_i \equiv \left[\frac{f_{LM}(1,\nu_i)\nu_i}{f_L(1,\nu_i)} - \frac{f_{MM}(1,\nu_i)v_i}{f_M(1,\nu_i)}\right]^{-1}$. Combining (A.30), (A.32), and (A.33), we then obtain the following expression for marginal changes in demand shifters:

$$\hat{D} = \Sigma^C \left[\left(\gamma + \epsilon\right) \hat{W} + \left(\sigma - \epsilon\right) \hat{Z}\right] \tag{A.34}$$

Next, we derive an expression for marginal changes in material costs, $\hat{Z}$. Totally differentiating equation (2.17) gives:

$$\hat{Z} = -\frac{1}{\sigma - 1}\Sigma^S \hat{\Phi} \tag{A.35}$$

where we have used the result that the share of firm i's input expenditures accounted for by firm j can be expressed using (2.17) and (2.31) as:

$$\Sigma^S_{ijt} \equiv \frac{R_{ijt}}{\sum_{k\in\Omega^S_{it}} R_{ikt}} = \frac{\Phi_{jt}\psi_{ijt}}{Z^{1-\sigma}_{it}} \tag{A.36}$$

Then, from (2.15) and (2.18), we can express marginal changes in network productivities as:

$$\hat{\Phi} = \frac{\sigma - 1}{\sigma}\left(\hat{X} - \hat{D}\right) \tag{A.37}$$

Hence, combining (A.35) and (A.37), we obtain the following expression for marginal changes in material costs:

$$\hat{Z} = \frac{1}{\sigma}\Sigma^S\left(\hat{D} - \hat{X}\right) \tag{A.38}$$

Now equations (A.25)-(A.27), (A.34), and (A.38) define a linear system in $\left\{\hat{W}, \hat{X}, \hat{\nu}, \hat{D}, \hat{Z}\right\}$, given changes in TFP $\hat{T}$ and labor productivity $\hat{\omega}$. Recall that we are assuming no changes in general equilibrium variables and hence $\hat{\kappa}(\cdot) = 0$. Eliminating $\hat{X}$ and $\hat{\nu}$ from this system, we can write the remaining equations as:

$$\hat{W} = H^{WT}\hat{T} + H^{W\omega}\hat{\omega} + H^{WD}\hat{D} + H^{WZ}\hat{Z} \tag{A.39}$$

$$\hat{D} = \Sigma^C\left[H^{DT}\hat{T} + H^{D\omega}\hat{\omega} + H^{DD}\hat{D} + H^{DZ}\hat{Z}\right] \tag{A.40}$$

$$\hat{Z} = \Sigma^S\left[H^{ZT}\hat{T} + H^{Z\omega}\hat{\omega} + H^{ZZ}\hat{Z} + H^{ZD}\hat{D}\right] \tag{A.41}$$

where the H matrices are all $\left|\Omega^F\right| \times \left|\Omega^F\right|$ diagonal matrices. Using Lemma 1, the matrices summarizing the dependence of $\left\{\hat{W}, \hat{D}, \hat{Z}\right\}$ on productivity shocks $\left\{\hat{T}, \hat{\omega}\right\}$ have i^{th}-diagonal elements given by:

$$\begin{array}{ll} H^{WT}_i = (\sigma - 1)\Gamma_i & H^{W\omega}_i = \left[(\sigma - 1) - (\sigma - \epsilon)\left(1 - \tilde{s}^L_i\right)\right]\Gamma_i \\ H^{DT}_i = (\gamma + \epsilon)(\sigma - 1)\Gamma_i & H^{D\omega}_i = (1 + \gamma)(\sigma - \epsilon)\tilde{s}^L_i\Gamma_i \\ H^{ZT}_i = -\left[\gamma + \epsilon\left(1 - \tilde{s}^L_i\right) + \tilde{s}^L_i\right]\Gamma_i & H^{Z\omega}_i = -(1 + \gamma)\tilde{s}^L_i\Gamma_i \end{array} \tag{A.42}$$

while the matrices summarizing the interrelation between $\left\{\hat{W}, \hat{D}, \hat{Z}\right\}$ have i^{th}-diagonal elements given by:

$$\begin{array}{lll} H^{WT}_i = (\sigma - 1)\Gamma_i & H^{WD}_i = \Gamma_i & H^{WZ}_i = -(\sigma - \epsilon)\left(1 - \tilde{s}^L_i\right)\Gamma_i \\ H^{DT}_i = (\gamma + \epsilon)(\sigma - 1)\Gamma_i & H^{DD}_i = (\gamma + \epsilon)\Gamma_i & H^{DZ}_i = (\sigma - \epsilon)(\gamma + \sigma)\tilde{s}^L_i\Gamma_i \\ H^{ZT}_i = -\left[\gamma + \epsilon\left(1 - \tilde{s}^L_i\right) + \tilde{s}^L_i\right]\Gamma_i & H^{ZD}_i = \tilde{s}^L_i\Gamma_i & H^{ZZ}_i = (\gamma + \epsilon)\left(1 - \tilde{s}^L_i\right)\Gamma_i \end{array} \tag{A.43}$$

Note that all the coefficients in equations (A.39)-(A.41) depend only on $\{\gamma, \sigma, \epsilon\}$, network shares $\left\{\Sigma^C, \Sigma^S\right\}$, and labor shares of cost s^L. Hence, the dependence of $\hat{W}$ on productivity shocks $\hat{T}$

and $\hat{\omega}$ depends only on these terms, as claimed in Proposition 2.

The interpretation of the system (A.39)-(A.41) is as follows. Equation (A.39) says that changes in earnings depend on (i) own TFP and labor productivity shocks (H^{WT} and $H^{W\omega}$); (ii) shocks to downstream firms (H^{WD}); and (iii) shocks to upstream firms (H^{WZ}). We hence first define the *direct* passthrough matrices in Proposition 2 as:

$$H_y^{direct} \equiv H^{Wy}, \ \ y \in \{T, \omega\} \tag{A.44}$$

Next, consider equation (A.40). This says that the change in the network demand for a firm depends on (i) shocks to TFP and labor productivity of the firm's customers (H^{DT} and $H^{D\omega}$); (ii) changes in the network demand for the firm's customers (H^{DD}), and (iii) changes in the network input cost of the firm's customers (H^{DZ}). The first term captures direct downstream passthrough, the second term captures higher-order downstream passthrough, and the third term captures interactions between downstream and upstream shocks due to the fact that marginal costs are increasing (and hence are dependent on scale). All of these are averaged across the firm's customers weighted by sales shares through Σ^C. Note that we can rewrite equation (A.40) as:

$$\hat{D} = \left[I - \Sigma^C H^{DD}\right]^{-1} \Sigma^C \left[H^{DT}\hat{T} + H^{D\omega}\hat{\omega} + H^{DZ}\hat{Z}\right] \tag{A.45}$$

Let us first ignore the interaction term involving H^{DZ} for now and return to this later. We define the *downstream* passthrough matrices in Proposition 2 as:

$$G_y^{down} \equiv H^{WD}\left[I - \Sigma^C H^{DD}\right]^{-1} \Sigma^C H^{Dy}, \ \ y \in \{T, \omega\} \tag{A.46}$$

By expanding the Leontief inverse $\left[I - \Sigma^C H^{DD}\right]^{-1}$, we can again rewrite this as:

$$G_y^{down} \equiv H^{down} \sum_{d=1}^{\infty} \left(\Sigma^C H^C\right)^d H_y^{down} \tag{A.47}$$

where $H^{down} \equiv H^{WD}$, $H^C \equiv H^{DD}$, and $H_y^{down} \equiv \left(H^{DD}\right)^{-1} H^{Dy}$. This is of the form in Proposition 2.

Next, consider equation (A.41). This says that that the change in input cost for a firm depends on (i) shocks to TFP and labor productivity of the firm's suppliers (H^{ZT} and $H^{Z\omega}$); (ii) changes in the network input cost of the firm's suppliers (H^{ZZ}), and (iii) changes in the network demand for the firm's suppliers (H^{ZD}). As with the downstream channel, the first effect captures direct upstream passthrough, the second term captures higher-order upstream passthrough, and the third term captures interactions between downstream and upstream shocks. All of these are averaged across the firm's suppliers weighted by input cost shares through Σ^S. Note that we can rewrite equation (A.41) as:

$$\hat{Z} = \left[I - \Sigma^S H^{ZZ}\right]^{-1} \Sigma^S \left[H^{ZT}\hat{T} + H^{ZD}\hat{D}\right] \tag{A.48}$$

Again, let us first ignore the interaction term involving H^{ZD} for now and return to this later.

We define the *upstream* passthrough matrices in Proposition 2 as:

$$G_y^{up} \equiv H^{WZ} \left[I - \Sigma^S H^{ZZ}\right]^{-1} \Sigma^S H^{Zy}, \;\; y \in \{T, \omega\} \tag{A.49}$$

By expanding the Leontief inverse $\left[I - \Sigma^S H^{ZZ}\right]^{-1}$, we can again rewrite this as:

$$G_y^{up} \equiv H^{up} \sum_{d=1}^{\infty} \left(\Sigma^S H^S\right)^d H_y^{up} \tag{A.50}$$

where $H^{up} \equiv H^{WZ}$, $H^S \equiv H^{ZZ}$, and $H_y^{up} \equiv \left(H^{ZZ}\right)^{-1} H^{Zy}$. This is of the form in Proposition 2.

Finally, we can write the passthrough from productivity shocks to firm wages as:

$$\hat{W} = \left[H_T^{direct} + G_T^{down} + G_T^{up} + G_T^{int}\right]\hat{T} + \left[H_\omega^{direct} + G_\omega^{down} + G_\omega^{up} + G_\omega^{int}\right]\hat{\omega} \tag{A.51}$$

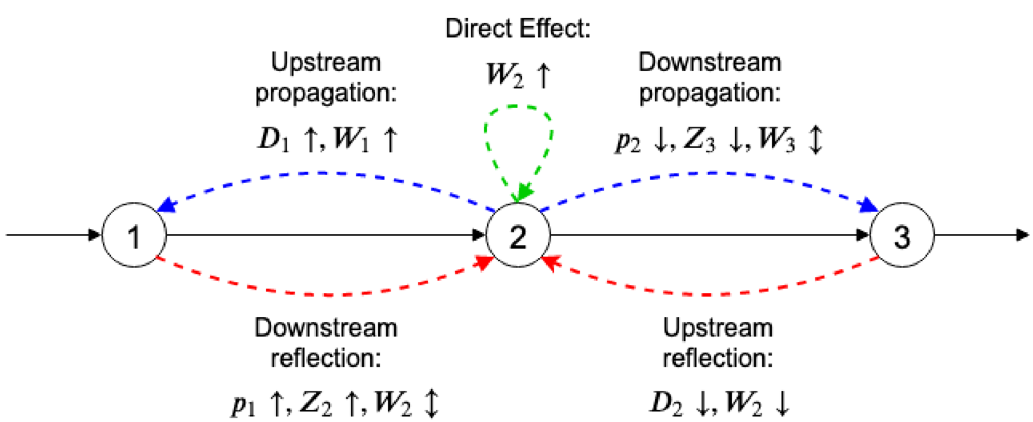
where the terms G_y^{int} for $y \in \{T, \omega\}$ account for the remaining H^{DZ} and H^{ZD} interaction terms that we ignored in equations (A.40) and (A.41). The expressions for passthrough in Remark 1 follow from setting $\sigma = \epsilon_{it}$ in the definitions in (A.42) and (A.43).

B Comparative Static Examples

B.1 Downstream and upstream interaction effects of TFP shocks

We provide a simple example to illustrate the interactions between downstream and upstream effects of TFP shocks on firm-level wages described in section 3.1. Consider three firms that are linked in a supply chain as shown in Figure A.1 and suppose that there is a positive shock to TFP for the firm in the middle of the sub-chain ($\hat{T}_2 > 0$). This shock has several effects on wages in the supply chain.

Figure A.1: Example of interaction between upstream and downstream effects of TFP on wages



First, as shown in Proposition 1, the direct effect of the shock is an increase in wages at the firm hit by the shock ($\hat{W}_2 > 0$). Second, there is upstream propagation of the shock to firm 1: it leads to higher demand for firm 1's output ($\hat{D}_1 > 0$) due to an increase in scale for firm 2, which increases wages at firm 1 ($\hat{W}_1 > 0$). Third, there is downstream *reflection* of the shock back from firm 1 to firm 2: due to increasing marginal costs, higher demand for firm 1 raises its price of output ($\hat{p}_1 > 0$), which increases firm 2's input cost ($\hat{Z}_2 > 0$) and has non-neutral effects on wages at firm 2 if $\sigma \neq \epsilon_{it}$ ($\hat{W}_2 \gtrless 0$). Fourth, there is downstream propagation of the shock to firm 3: it reduces the price of firm 2's output ($\hat{p}_2 < 0$), which reduces firm 3's input cost ($\hat{Z}_3 < 0$) and has non-neutral effects on wages at firm 3 if $\sigma \neq \epsilon_{it}$ ($\hat{W}_3 \gtrless 0$). Finally, there is upstream *reflection* of the shock back from firm 3 to firm 2: changes in input cost and wages at firm 3 lead to lower demand for firm 2 ($\hat{D}_2 < 0$) which leads to lower wages at firm 2 ($\hat{W}_2 < 0$).

B.2 Firm dispersion and within-firm employment composition

To illustrate the intuition behind Proposition 3, consider a simple example in which the economy consists of two firms, 1 and 2. For brevity, omit time subscripts and let $h_i(a) \equiv g_i(a)\phi_i(a)$ denote the composite of amenities and labor productivity. Then the relative market indices for two worker types a' and a is:

$$\left[\frac{I_t(a')}{I_t(a)}\right]^\gamma = \frac{[h_1(a')W_1]^\gamma + [h_2(a')W_2]^\gamma}{[h_1(a)W_1]^\gamma + [h_2(a)W_2]^\gamma} \tag{B.1}$$

Note that the right-hand side of equation (B.1) is strictly increasing in the relative firm wage $\frac{W_2}{W_1}$ if and only if the following condition holds:

$$\frac{h_2(a')}{h_2(a)} > \frac{h_1(a')}{h_1(a)} \tag{B.2}$$

Intuitively, if the *comparative advantage* (in terms of amenities and labor productivity) of a'-ability workers versus a-ability workers is higher at firm 2 than firm 1, then any shock to the economy that increases the firm-level wage of firm 2 relative to firm 1 will increase relative competition for a'-ability workers relative to a-ability workers.

What is the effect of an increase in $\frac{W_2}{W_1}$ on the within-firm earnings distribution in this example? Suppose for concreteness that $W_2 > W_1$ and $\phi_i(a') > \phi_i(a)$ for all $i \in \Omega^F$, so that a'-ability workers are more productive than a-ability workers in all firms and hence earn higher wages. Suppose also that equation (B.2) holds, which implies that high-ability workers have a comparative advantage in high-wage firms. Then, an increase in $\frac{W_2}{W_1}$ raises labor market competition for high-ability workers relative to low-ability workers and thus reduces employment of the former relative to the latter within *all firms*. This implies that an increase in between-firm wage dispersion *reduces* within-firm employment dispersion if the comparative advantage condition holds. Since relative wages are exogenous, this implies a reduction in within-firm wage dispersion as well.

C Data Details

Table A.1: Overview of Sample Sizes

Panel A: Firm-to-Firm Dataset	**Links**		**Suppliers**		**Buyers**	
Sample	**Unique**	**Observation-Years**	**Unique**	**Observation-Years**	**Unique**	**Observation-Years**
Baseline	16,831,546	31,743,495	194,615	592,622	289,344	923,155
Panel B: Employer-Employee Dataset	**Workers**		**Firms**			
Sample	**Unique**	**Observation-Years**	**Unique**	**Observations-Years**		
Baseline	6,496,849	41,954,008	487,504	2,315,927		
Movers	6,183,692	40,130,960	200,592	1,378,554		
Stayers: Complete Spells	953,865	8,472,302	64,670	602,622		
Stayers: 10 Stayers per Firm	724,957	6,571,483	5,726	61,823		
Panel C: Firm Dataset	**Firms**					
Sample	**Unique**	**Observations-Years**				
Baseline	47,685	125,726				

Notes: This table provides an overview of the samples used throughout the paper.

Table A.2: Descriptive Satistics of Datasets

Dataset	**Employer-Employee**			**Firm**	**Firm-to-Firm**
Panel A: Worker Characteristics	**Baseline**	**Movers**	**Stayers**	**Baseline**	**Baseline**
Mean Log Worker Earnings (Log US)	9.36	9.38	9.74	9.17	9.22
Median Log Worker Earnings (Log US)	9.25	9.27	9.66	9.02	9.10
Mean Worker Age	40.2	40.1	42.6	39.3	39.8
Median Worker Age	39.4	39.4	42.6	38.5	39.0
Panel B: Firm Characteristics	**Baseline**	**Movers**	**Stayers**	**Baseline**	**Baseline**
Mean Number of Workers	9	20	281	27	12
Median Number of Workers	2	4	94	7	2
Mean Wage Bill per Worker (US)	10,199	11,145	7,833	9,440	8,306
Median Wage Bill per Worker (US)	6,943	8,323	6,672	7,103	5,490
Mean Value Added per Worker (US)	56,315	58,610	50,077	49,604	50,091
Median Value Added per Worker (US)	23,424	25,659	26,583	23,389	18,771
Mean Log Value Added (Log US)	11.0	11.8	14.6	12.2	10.9
Median Log Value Added (Log US)	11.0	11.7	14.8	12.1	10.9
Mean Labor Share	0.49	0.45	0.70	0.42	0.49
Median Labor Share	0.32	0.34	0.21	0.34	0.32
Panel C: Production Network Characteristics	**Baseline**	**Movers**	**Stayers**	**Baseline**	**Baseline**
Mean Number of Suppliers	67	67	306	67	35
Median Number of Suppliers	36	36	208	36	19
Mean Number of Buyers	80	80	580	80	34
Median Number of Buyers	8	8	59	8	4
Mean Materials Share of Sales	0.58	0.58	0.55	0.58	0.57
Median Materials Share of Sales	0.61	0.61	0.60	0.61	0.61
Mean Intermediate Share of Sales	0.40	0.40	0.45	0.40	0.38
Median Intermediate Share of Sales	0.38	0.38	0.50	0.38	0.33

Notes: This table provides descriptive statistics of all the samples used in the paper.

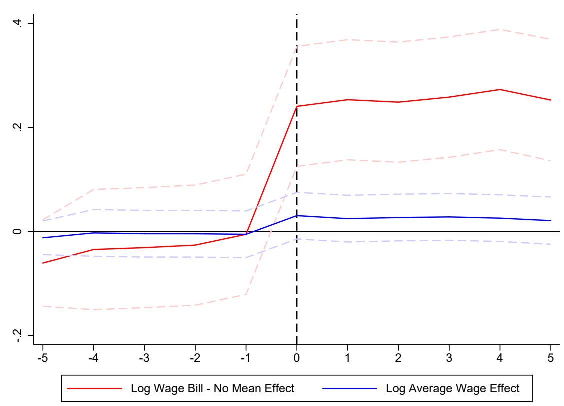
D Estimation Details

D.1 Labor supply elasticity

For robustness, we also follow Lamadon et al. (2019) and estimate γ using a difference-in-difference approach (DiD). For this, we follow a three step procedure. First, for each year, we order firms according to log changes of the wage bill of the firm. Second, we identify the treatment when firms have log changes of their wage bill above the median of log changes of wage bill across firms each year. Finally, we plot difference in wage bill of treated and control firms both at each year ($t = 0$) and years before ($t < 0$) and after ($t > 0$). We perform this step for each calendar year and weight firms by the number of workers.

Results are presented Figure A.2. By construction, the treatment and control groups differ in the wage bill from period $t = -1$ to $t = 0$. On average, firms in the treatment group face an increase of 21 log points growth in their wage bill relative to firms in the control group. The effect of the treatment appears to be permanent in levels up to 5 years after the treatment. Figure A.2 also shows the effect on the average earnings of firms. On average, firms in the treatment group face an increase of 3.25 log points of their average earnings relative to firms in the control group. Once again, the effect of the treatment appears to be permanent in levels up to 5 years after the treatment. Finally, firms in the treatment and control group do not experience statistically significant differences up to 5 years before the treatment, for both the wage bill and the average earnings. Through the lens of a DiD design, these results imply a passthrough rate of firms shocks of around 0.155 ($= 0.0325/0.21$). From equation (4.8), this implies a labor supply elasticity of $\hat{\gamma} = 5.5$, which is the same as our preferred estimate documented in the main text.

Figure A.2: Difference-in-difference Estimate of passthrough of Firm Shocks to Worker Earnings



Notes: This figure presents the results from the Lamadon et al. (2019) difference-in-difference approach to estimating passthrough of wage bill shocks to worker wages.

D.2 Worker and firm wage effects

To estimate the Bonhomme et al. (2019) decomposition of worker earnings from equation (4.10), we first cluster firms using a k-means clustering algorithm into $K = 10$ groups. We use a weighted K-means algorithm with 100 randomly generated starting values. We use firms' empirical distributions of log earnings on a grid of 10 percentiles of the overall log-earnings distribution. Second, we use these K groups as the relevant firm identifier in the Bonhomme et al. (2019) estimation approach. This procedure yields estimates of the firm fixed effect $\bar{W}_i$ and the worker-firm production complementarity θ_i for every firm $i \in \Omega^F$, as well as the permanent and transient components of ability for every worker.

To assess robustness of our results to the number of clusters used, Table A.3 documents the share of variance of wages accounted for by the firm fixed effect $\bar{W}_i$. We implement this for the basic model of Abowd et al. (1999) and also the basic version of the model of Bonhomme et al. (2019) with only firm and worker fixed effects for different levels of K (thus, excluding interactions and time-varying firm effects). First, one can see that the basic version of the model of Bonhomme et al. (2019) implies a role for the firm fixed effect that is significantly lower than the model of Abowd et al. (1999), consistent with previous literature that has found that addressing the limited mobility bias inherent in estimates of Abowd et al. (1999) decreases the share of the variance accounted for by the firm fixed effect (Bonhomme et al., 2020). Second, as one increases K from 10 to 50, the share of the variance of wages accounted for the firm fixed effects increases only 0.7 percentage points from 7.8 to 8.5%. At least with this piece of evidence, this implies that the limited mobility bias does not represent a substantially bigger problem for $K = 50$ than what it represents for $K = 10$.

Table A.3: Share of Log Earnings Variance Accounted for by the Firm Fixed Effect

Estimation Strategy	Number of Clusters	Firm Fixed EffectShare
AKM		12.3
BLM	10	7.8
BLM	50	8.5

Notes: This table documents the share of the log of earnings variance accounted for by the firm fixed effect. It is documented for the estimation strategy of Abowd et al. (1999) (row 1), for the estimation strategy of Bonhomme et al. (2019) with $K = 10$ clusters (row 2) and the estimation strategy of Bonhomme et al. (2019) with $K = 50$ clusters (row 3). Note that this table documents the version of the wage models following the estimation strategy of Lamadon et al. (2019) documented in Section 5.2.2 without interactions and without time-varying firm effects. Thus, the share documented in row 2 corresponds to the same one document in row 2 and column 3 of Table 3.

To further assess whether clustering with $K = 10$ or $K = 50$ makes a difference, we document how much clusters account for the variance of firm-level characteristics. Tables A.4-A.5 document the share of the variance of variables accounted for by within-cluster variation. Table A.4 shows the within-cluster share of variance of variables in levels, whereas Table A.5 shows the same evidence for variables in ratios. Although there is substantial heterogeneity across firms that the clustering procedure of Bonhomme et al. (2019) does not account for, this result does not vary significantly if one uses $K = 10$ or $K = 50$ clusters.

Table A.4: Within Clusters Share of Total Variance of Variables in Levels

Number Clusters	Total Sales	Materials	Wage Bill	Employment	Number of Buyers	Number of Suppliers	Firm-to-Firm Sales
10	79	90	67	88	90	85	95
50	74	86	62	84	88	81	92

Notes: This table documents the share of the variance of each variable accounted for by the within cluster variance. It is implemented for $K = 10$ and $K = 50$ and for variables in levels.

Table A.5: Within Clusters Share of Total Variance of Variables in Ratios

Number Clusters	Wage Bill/Sales	Materials/Sales	Materials/Wage Bill	Sales/Employment	Wage Bill/Employment	Materials/Employment
10	96	97	95	92	26	99
50	95	97	95	90	21	98

Notes: This table documents the share of the variance of each variable accounted for by the within cluster variance. It is implemented for $K = 10$ and $K = 50$ and for variables in ratios.

D.3 Amenities

To estimate firm amenities, we begin with the labor supply equation (2.5). It will be useful for the exposition to write this explicitly in terms of permanent and transient worker abilities:

$$\frac{L_{it}\left(\bar{a},\hat{a}\right)}{L\left(\bar{a},\hat{a}\right)}=\frac{\left[g_i\left(\bar{a}\right)w_{it}\left(\bar{a},\hat{a}\right)\right]^{\gamma}}{\sum_{j\in\Omega^F}\left[g_j\left(\bar{a}\right)w_{jt}\left(\bar{a},\hat{a}\right)\right]^{\gamma}} \tag{D.1}$$

where note that under Assumption 4.2, amenity values only vary across workers in relation to permanent ability $\bar{a}$. Next, consider the equilibrium wage equation (2.21). Under assumption 4.2, we can write this as:

$$w_{it}\left(\bar{a},\hat{a}\right)=\eta\bar{a}^{\theta_i}\hat{a}W_{it} \tag{D.2}$$

The average wage paid by firm i to workers with permanent ability $\bar{a}$ is hence:

$$\bar{w}_{it}\left(\bar{a}\right)=\eta\bar{a}^{\theta_i}\mathbb{E}\left[\hat{a}\right]W_{it} \tag{D.3}$$

where $\mathbb{E}\left[\hat{a}\right]$ denotes the average value of transient ability. Under Assumptions 4.2 and 4.4, this mean does not depend on permanent ability of the worker or the identity of the firm. Combining (D.2) and (D.3), we then have:

$$w_{it}\left(\bar{a},\hat{a}\right)=\bar{w}_{it}\left(\bar{a}\right)\frac{\hat{a}}{\mathbb{E}\left[\hat{a}\right]} \tag{D.4}$$

Substituting this into (D.1) and using the decomposition of amenities in equation (4.14), we obtain:

$$\frac{L_{it}\left(\bar{a},\hat{a}\right)}{L\left(\bar{a},\hat{a}\right)}=\frac{\left[\tilde{g}_i\bar{g}_{k(i)}\left(\bar{a}\right)\bar{a}^{\theta_{k(i)}}W_{it}\right]^{\gamma}}{\sum_j\left[\tilde{g}_j\bar{g}_{k(j)}\left(\bar{a}\right)\bar{a}^{\theta_{k(j)}}W_{jt}\right]^{\gamma}} \tag{D.5}$$

Now notice that the employment share of workers of ability $\{\bar{a},\hat{a}\}$ varies across firms only in relation to permanent ability $\bar{a}$. This is a direct implication of Assumption 4.2, which implies that workers do not sort to firms based on transient ability $\hat{a}$. Therefore, the share of workers of permanent ability $\bar{a}$ employed by firm i is also given by equation (D.5). Summing this (D.5) across all firms within cluster k, we can similarly express the share of workers of permanent ability $\bar{a}$ that are employed by firms in cluster k as:

$$\Lambda_{kt}\left(\bar{a}\right)=\frac{\sum_{i\in k}\left[\tilde{g}_i\bar{g}_k\left(\bar{a}\right)\bar{a}^{\theta_k}W_{it}\right]^{\gamma}}{\sum_{j}\left[\tilde{g}_j\bar{g}_{k(j)}\left(\bar{a}\right)\bar{a}^{\theta_{k(j)}}W_{jt}\right]^{\gamma}} \tag{D.6}$$

Next, note that for each value of permanent ability $\bar{a}$, equilibrium outcomes are invariant to scaling $g_i\left(\bar{a}\right)$ by a constant for all firms i. Therefore, we are allowed to choose one normalization of amenity values for each permanent worker ability type $\bar{a}$. For this, we choose $\sum_{j}\left[\tilde{g}_j\bar{g}_{k(j)}\left(\bar{a}\right)\bar{a}^{\theta_{k(j)}}W_{jt}\right]^{\gamma}=1$. Furthermore, mean differences in amenity values can be loaded onto either $\tilde{g}_i$ or $\bar{g}_{k(i)}\left(\bar{a}\right)$. Hence, we are allowed to choose one normalization of the values for $\tilde{g}_i$ for each firm cluster. For this, we choose $\sum_{i\in k}\left[\tilde{g}_iW_{it}\right]^{\gamma}=1$. With these normalizations, equations (4.15) and (4.16) follow immediately.

D.4 Firm relationship capability and relationship-specific productivity

To estimate equation (4.17), firms must have multiple connections. To identify seller fixed effects, each seller needs to have at least two buyers. Similarly, to identify buyer fixed effects, each buyer needs to have at least two sellers. In the data, some firms have either one supplier or one seller. Hence, we implement the aforementioned restriction using an iterative approach known as “avalanching”. Specifically, we first drop firms with one supplier or seller. Doing this may result in additional firms that have one supplier or seller, hence in the next step, we drop these firms as well. We continue this process until firms are no longer dropped from the sample. The algorithm takes three iterations to converge in practice and reduces the sample size of firm-to-firm linkages from a total of 32 million transactions to 31.7 million transactions, that is, a reduction of 1% of transactions. Hence, the avalanching algorithm has little impact on our sample size. Bernard et al. (2019) report that avalanching also eliminates around 1% of firm-to-firm links in the production network for Belgium.

D.5 Product substitution elasticity

To derive equation (4.20), first note that from equations (2.27) and (2.28), the share of firm profits in total sales is:

$$\frac{\pi_{it}}{R_{it}}=\frac{1}{\sigma}\left[1+\left(\sigma-1\right)\left(1-\eta\right)\frac{f_L\left(1,\nu_{it}\right)}{f\left(1,\nu_{it}\right)}\right] \tag{D.7}$$

Under the CES functional form for the production function in Assumption 4.1, we can write this as:

$$\frac{\pi_{it}}{R_{it}}=\frac{1}{\sigma}\left[1+\frac{\left(\sigma-1\right)\left(1-\eta\right)}{1+\eta\frac{E_{it}^M}{E_{it}^L}}\right] \tag{D.8}$$

Solving for σ and using the fact that $\pi_{it} = R_{it} - E_{it}^L - E_{it}^M$, we obtain:

$$\sigma = \frac{R_{it}\tilde{\eta}_{it}}{R_{it}\tilde{\eta}_{it} - E_{it}^L - E_{it}^M} \tag{D.9}$$

where $\tilde{\eta}_{it} \equiv \eta \left(1 + \frac{E_{it}^M}{E_{it}^L}\right)\left(1 + \eta \frac{E_{it}^M}{E_{it}^L}\right)^{-1}$. Hence, we estimate σ using the sample average of the right-hand side of (D.9), which is observable given our estimate of the labor supply elasticity γ and data on firm sales, labor costs, and material costs.

D.6 Labor-materials substitution elasticity and labor productivity

To derive equation (4.21), first note that under the CES production function specified in Assumption 4.1, a firm's relative expenditure on materials versus labor inputs can be expressed using equations (2.22), (2.23), (2.29), and (2.30) as:

$$\log \frac{E_{it}^M}{E_{it}^L} = \log \left[\frac{1}{\eta}\left(\frac{1-\lambda}{\lambda}\right)\right] + (\epsilon - 1)\log \frac{W_{it}}{Z_{it}} + (1-\epsilon)\log \omega_{it} \tag{D.10}$$

Using the Markov process for labor productivity (4.5) to substitute for ω_{it} in (D.10) then gives equation (4.21).

For estimation of ϵ using equation (4.21), we follow the approach in Doraszelski and Jaumandreu (2018). To control for $F^{\omega}(\omega_{i,t-1})$, we first rearrange the $t-1$ version of equation (D.10) to write:

$$\log \omega_{i,t-1} = \frac{1}{\epsilon - 1}\log \left[\frac{1}{\eta}\left(\frac{1-\lambda}{\lambda}\right)\right] - \frac{1}{\epsilon-1}\log \frac{E_{i,t-1}^M}{E_{i,t-1}^L} + \log \frac{W_{i,t-1}}{Z_{i,t-1}} \tag{D.11}$$

$$\equiv G\left(\log \frac{E_{i,t-1}^M}{E_{i,t-1}^L}, \log \frac{W_{i,t-1}}{Z_{i,t-1}}\right) \tag{D.12}$$

Substituting this into (4.21), we obtain:

$$\log \frac{E_{it}^M}{E_{it}^L} = \log \left[\frac{1}{\eta}\left(\frac{1-\lambda}{\lambda}\right)\right] + (\epsilon - 1)\log \frac{W_{it}}{Z_{it}} + H\left(\log \frac{E_{i,t-1}^M}{E_{i,t-1}^L}, \log \frac{W_{i,t-1}}{Z_{i,t-1}}\right) \tag{D.13}$$

$$+ (1-\epsilon)\,\xi_{it}^{\omega} \tag{D.14}$$

where $H(\cdot,\cdot) \equiv (1-\epsilon) F^{\omega}[G(\cdot,\cdot)]$. Hence, we control for the term H using a third-degree polynomial in lagged relative expenditures $\log \frac{\tilde{E}_{i,t-1}^M}{\tilde{E}_{i,t-1}^L}$ and lagged relative input prices $\log \frac{\tilde{W}_{i,t-1}}{Z_{i,t-1}}$. We then instrument for relative input prices at date t using all available lags of logged input expenditures and constructed prices from dates $t-1$ and earlier.

D.7 Firm TFP

We choose values for TFP T_{it} to fit the estimated firm-level wages W_{it} as specified in equation (4.22). We do this using an iterative numerical procedure that is similar in spirit to the equilibrium solution algorithm described in section F:

1. Compute $\left\{\tilde{\phi}_{it}\right\}_{i\in\Omega^F}$ from (2.26), using (2.4), (2.6), and the estimated firm-level wages $\{W_{it}\}_{i\in\Omega^F}$.

2. Guess E_t.

   (a) Guess $\{D_{it}, Z_{it}\}_{i\in\Omega^F}$.

   (b) Compute the values of $\{T_{it}\}_{i\in\Omega^F}$ implied by equation (F.1), given the estimated firm-level wages $\{W_{it}\}_{i\in\Omega^F}$.

   (c) Compute new guesses of $\{D_{it}\}_{i\in\Omega^F}$ from (2.16) and $\{Z_{it}\}_{i\in\Omega^F}$ from (2.17).

   (d) Iterate on steps (a)-(c) until convergence.

3. Compute a new guess of E_t from (2.9), using (2.5), (2.19), and (2.21).

4. Iterate on steps 1-2 until convergence.

In practice, we also add as an outer loop iterations over guesses of mean relationship productivity (as described in section 5.2.4) and the product substitution elasticity σ (as described in section 5.2.5), targeting the aggregate value-added share of gross output and labor share of value-added respectively.

E A Shapley value approach for model counterfactuals

In the counterfactual exercises studied in sections 6.1 and 6.2, we deal with interdependencies between model primitives in shaping outcomes of interest using the following approach. Let Θ denote the estimated vector of values for all model primitives and let $X(\Theta)$ denote the value of some equilibrium outcome X under this parameter vector. Now, define some N subsets of the parameter vector $\{\theta_n\}_{n=1}^N$ such that $\Theta = \cup_{n=1}^N \theta_n$ and denote $\mathcal{N} \equiv \{1, \cdots, N\}$. We are interested in computing values of outcome X under known counterfactual values $\hat{\theta}_n$ for each subset of the parameter vector. Therefore, let $\hat{\Theta}_S \equiv \left\{\cup_{n\in S}\hat{\theta}_n\right\} \cup \{\cup_{n\notin S}\theta_n\}$ denote the parameter vector under counterfactual values for parameter subsets in S for some $S \subseteq \mathcal{N}$. We define the Shapley value X_n for parameter subset n in relation to outcome X as follows:

$$X_n = \sum_{S\subseteq\mathcal{N}\setminus\{n\}} \frac{|S|!\,(N! - |S|! - 1)}{N!} \left[X\left(\hat{\Theta}_{S\cup\{n\}}\right) - X\left(\hat{\Theta}_S\right)\right] \tag{E.1}$$

For example, suppose that X is the variance of log earnings across all workers, θ_n is the estimated vector of firm TFPs, and $\hat{\theta}_n$ is a counterfactual vector of firm TFPs with each value equal to the mean of θ_n across firms. Then, we measure the contribution of TFP heterogeneity to earnings variance as $-\frac{X_n}{X(\Theta)}$. By construction of the Shapley value, these measures sum to one across all $n \in \mathcal{N}$.

F Solution Algorithm

We solve numerically for an equilibrium of the model using the following solution algorithm.

1. Guess E_t.

(a) Guess $\left\{\Delta_{it}, \Phi_{it}, \tilde{\phi}_{it}\right\}_{i\in\Omega^F}$.

(b) Compute $\{D_{it}\}_{i\in\Omega^F}$ from (2.16) and $\{Z_{it}\}_{i\in\Omega^F}$ from (2.17).

(c) Solve for $\{W_{it}, \nu_{it}, X_{it}\}_{i\in\Omega^F}$ from (2.22), (2.23), and (2.24).

(d) Compute new guesses of $\{\Delta_{it}\}_{i\in\Omega^F}$ from (2.14), $\{\Phi_{it}\}_{i\in\Omega^F}$ from (2.18), and $\left\{\tilde{\phi}_{it}\right\}_{i\in\Omega^F}$ from (2.26).

(e) Iterate on steps (a)-(d) until convergence.

2. Compute a new guess of E_t from (2.9), using (2.5), (2.19), and (2.21).

3. Iterate on steps 1-2 until convergence.

Note that step 1(c) involves numerical solution of a system in $\{W_{it}, \nu_{it}, X_{it}\}$. This system can be reduced to one in firm-level wages alone:

$$W_{it}^{\gamma+\epsilon}\left[\lambda W_{it}^{1-\epsilon} + (1-\lambda) Z_{it}^{1-\epsilon}\right]^{\frac{\sigma-\epsilon}{1-\epsilon}} \tilde{\phi}_{it}^{1+\gamma} = \frac{\lambda}{\mu^\sigma \eta^\gamma} D_{it} \left(T_{it}\omega_{it}\right)^{\sigma-1} \tag{F.1}$$

which has a unique solution for W_{it} given $\left\{D_{it}, Z_{it}, \tilde{\phi}_{it}\right\}$. Solutions for ν_{it} and X_{it} are then easy to recover given W_{it}.

G Detailed results for a minimum wage counterfactual

Table A.6 shows the results of the decomposition in equation (6.5), which decomposes the spillover effects on earnings for unconstrained workers under a 20% minimum wage policy.

Panel (a) shows the L-scale effects, which we see are typically positive for all worker and firm types. The MW policy induces a *reduction* in total efficiency units of labor for most firms, especially small firms that are likely to reject many low-ability workers. Hence, this reduces the scale of these firms, increases their output prices, and raises the MRPLs of unconstrained workers.

Panels (b) and (c) show the M-scale and M-substitution effects. Here, we see opposite patterns for small versus large firms. Since small firms are likely to be constrained by the wage floor, the MW policy raises the effective cost of labor for these firms, inducing them to substitute towards material inputs and *increasing* the optimally chosen ratio of materials per effective unit of labor. As discussed above, the M-substitution effect dominates the M-scale effect and readjustment of material inputs *amplifies* the positive spillover effects on unconstrained workers at small firms. Quantitatively, we find that this amplification is important: for example, it accounts for around one-third of the total spillover effect on earnings for unconstrained workers at the smallest quintile of firms. For the largest firms, the excess supply of low-ability workers generated by exit of small firms from the labor market for these workers lowers the effective cost of labor, inducing substitution away from materials toward labor. Hence, the M-scale and M-substitution effects have signs that are opposite to the effects for smaller firms, with the readjustment mechanism *lowering* earnings for unconstrained workers on net.

Finally, panel (d) shows the GE demand effect. This is negative for all workers at all firms, which is largely due to a reduction in consumer income generated by the MW policy, which lowers demand for all firms and reduces MRPLs.

Table A.6: Decomposition of spillover earnings effects under a 20% minimum wage

(a) L-Scale Effect

	W1	W2	W3	W4	W5
F1	2.52	2.41	2.39	2.43	2.51
F2	1.41	1.31	1.30	1.30	1.33
F3	0.82	0.76	0.71	0.67	0.68
F4	0.68	0.63	0.59	0.57	0.57
F5	-0.02	0.36	0.35	0.29	0.18

(b) M-Scale Effect

	W1	W2	W3	W4	W5
F1	-1.03	-0.96	-0.96	-0.97	-1.01
F2	-0.48	-0.40	-0.38	-0.39	-0.42
F3	-0.20	-0.17	-0.14	-0.12	-0.13
F4	-0.17	-0.13	-0.11	-0.09	-0.10
F5	0.14	-0.02	-0.01	0.01	0.06

(c) M-Substitution Effect

	W1	W2	W3	W4	W5
F1	2.43	2.29	2.27	2.31	2.40
F2	1.14	0.94	0.91	0.92	0.99
F3	0.47	0.40	0.34	0.30	0.31
F4	0.40	0.31	0.25	0.22	0.23
F5	-0.34	0.06	0.03	-0.03	-0.15

(d) GE Demand Effect

	W1	W2	W3	W4	W5
F1	-0.32	-0.32	-0.32	-0.32	-0.32
F2	-0.34	-0.35	-0.35	-0.35	-0.35
F3	-0.34	-0.33	-0.33	-0.32	-0.32
F4	-0.34	-0.34	-0.34	-0.34	-0.34
F5	-0.31	-0.34	-0.34	-0.34	-0.33

Notes: Rows (F#) and columns (W#) indicate quantiles of firm size and permanent worker ability respectively. The effects shown in each panel correspond to the terms in equation (6.5). Units are in log changes multiplied by 100.