ONLINE APPENDIX
"Monopsony in the U.S. Labor Market"

Chen Yeh, Claudia Macaluso and Brad Hershbein

O.1 Additional results on markdowns

O.1.1 Unionization

In the following, we provide some external validity on our markdown measures by correlating them with measures of unionization. The ASM and CM unfortunately do not contain measures of unionization at the plant level. Instead, we leverage the Current Population Survey (CPS), which since 1984 has asked about unionization and collective bargaining status in outgoing rotation months, to construct measures of unionization at the 3-digit NAICS-state-year level. To do so, we convert the Census industry codes for manufacturing in the CPS to 21 consistent, 3-digit, 2012-vintage NAICS codes using crosswalks provided by IPUMS.[1] We then run a logit regression of union coverage (member or covered by a union) on a vector of state indicators, NAICS3 indicators, and year indicators. After collapsing the data to 3-digit NAICS–state–year cells, we fit values of union coverage based on the estimated logit coefficients. This simulated instrument adjusts for small cells (including missings) and mitigates endogeneity, although it still contains measurement error.

Due to data limitations, we can construct these measures only from 1984 onward. Hence, our sample to correlate markdowns with unionization will be somewhat smaller than our baseline sample (which starts in 1976). There are only a limited amount of observations available at this narrow cell level in the CPS, so our correlations with labor market power could be noisy. To avoid this, we create a binary variable which categorizes a plant's level of unionization either above or below the median of the unionization distribution for a given year. Our results are displayed in the table below.

As expected, markdowns are negatively correlated with unionization; albeit the correlation is noisily estimated. A plant operating in a 3-digit NAICS-state cell that is in the upper half

[1] See `https://cps.ipums.org/cps-action/variables/IND#codes_section`.

Table 1: Plant-level markdowns are negatively correlated with unionization.†

Dependent variable: PLANT-LEVEL TRANSLOG MARKDOWNS		
UNIONIZATION	-0.07463 (0.04760)	-0.07628 (0.04731)
Fixed effects		
YEAR	N	Y
Weights	`empwt`	`empwt`
Observations (in millions)	10.91	10.91

†Markdowns are estimated under the assumption of a translog specification for gross output. Each industry group in manufacturing corresponds to the manufacturing categorization of the BEA which approximately follows a 3-digit NAICS specification. Standard errors are clustered at the industry-state level and denoted between parentheses. Regressions are weighted by the product of employment count and ASM sampling weights. Source: Authors' calculations from ASM/CM data in 1984–2014.

of the unionization distribution has a markdown that is about 7.5 percent lower on average. This is intuitive since plants can extract less rents in those environments in which workers are more likely to be affiliated to a union.

O.1.2 Below-unity markdowns

Our baseline estimates on markdowns in section II indicate that most plants operate in a monopsonistic environment since markdowns are above unity. However, a relatively small fraction of our sample (approximately 11 percent) features markdowns below unity. We have verified that our core results are robust to dropping establishment-years with below-unity markdowns, but while these types of markdowns could partly be the result of statistical noise, they could also be real, especially when temporary. In the following, we rationalize why below-unity markdowns can occur under the production approach.

First, we deal with measurement error in output, but we do not account for measurement error in *inputs*. This type of measurement error can obviously impact the estimated production function coefficients. Whenever we allow for a translog specification, it is not unlikely that some of the higher order (cross- and second-order) terms are negative which pulls estimated output elasticities below their revenue shares. Given that the overwhelming majority of observations with below-unity markdowns are between 0.75 and 1 (see table below), we believe moderate measurement error in inputs can likely account for some markdowns to be estimated below unity.

Table 2: Estimated plant-level markdowns in U.S. manufacturing (below-unity sample).†

BELOW-UNITY SAMPLE	Median	Mean	25%	75%	SD
	0.864	**0.816**	**0.748**	**0.942**	**0.173**
Sample size	$1.56 \cdot 10^5$				

†Markdowns are estimated under the assumption of a translog specification for gross output. The flexible input is materials. Each industry group in manufacturing corresponds to the manufacturing categorization of the BEA which approximately follows a 3-digit NAICS specification. The sample is restricted to those plant-year observations with markdowns strictly below unity. Source: Authors' calculations from ASM/CM data in 1976–2014.

Second, our baseline results are relying on the assumption that material inputs are not subject to any monopsony forces. However, it is not unlikely that this specific assumption does not apply equally to all plants in a given industry. Think about monopolistic competition across space in the spirit of Hotelling (1929) and Salop (1979). Whenever this is the case, we are identifying monopsony for labor *relative to material inputs*. If the latter is larger than the former, then we expect to see below-unity labor markdowns.

Third, we are also assuming that labor is chosen statically. Whenever this is the case, our markdown formula based on static first-order conditions applies. Even though we show in Online Appendix O.4 that labor adjustment costs are unlikely to change our estimates, we did not rule out other dynamic considerations. It might be the case that some plants in our sample are subject to a, for example, "customer capital" mechanism. Under this narrative, a plant's future demand directly depends on the amount of quantity currently sold. As a result, some plants are willing to make losses (i.e., set below-unity markdowns and/or markups) in order to sell more in the future. This reflects "investing-harvesting" incentives that are present in models of the customer base. Even though our baseline estimates do not capture these dynamic considerations, we do think they describe the data in a reasonable fashion.

Fourth and last, the estimated wedges for labor cannot be interpreted as labor market power under the classical monopsony framework whenever these wedges are below unity. However, Dobbelaere and Mairesse (2013) show that these below-unity wedges can be interpreted as labor market imperfections in a setting where risk-neutral workers and firms efficiently bargain over wages in the spirit of McDonald and Solow (1981). In fact, the estimated wedges can be used to retrieve the relevant bargaining parameters. Let $\gamma_{it} \in (0, 1)$ denote workers' bargaining power (also referred to as the "absolute extent of rent sharing" by Dobbelaere and Mairesse, 2013), then it can be shown that:

$$\frac{\theta^M_{it}}{\alpha^M_{it}} - \frac{\theta^\ell_{it}}{\alpha^\ell_{it}} = \mu_{it} \cdot \frac{\gamma_{it}}{1-\gamma_{it}} \cdot \left[\frac{1-\alpha^\ell_{it}-\alpha^M_{it}}{\alpha^\ell_{it}}\right]$$

which we can rearrange as:

$$1 - \frac{\theta^\ell_{it}}{\alpha^\ell_{it}} \Big/ \frac{\theta^M_{it}}{\alpha^M_{it}} = \frac{\gamma_{it}}{1-\gamma_{it}} \cdot \left[\frac{1-\alpha^\ell_{it}-\alpha^M_{it}}{\alpha^\ell_{it}}\right] \tag{1}$$

Obviously, the interpretation for γ_{it} is only valid whenever relative labor wedges $\frac{\theta^\ell_{it}}{\alpha^\ell_{it}} \Big/ \frac{\theta^M_{it}}{\alpha^M_{it}}$ are below unity. Following Dobbelaere and Mairesse (2013), below-unity markdowns in the classical monopsony setting can also be reinterpreted as a different labor market "regime" in which there are labor market imperfections under efficient bargaining.

O.1.3 Markdowns with energy as flexible input

In our baseline estimates, we assumed that material inputs were flexible and used these inputs to identify markups. We argued in section 5 that material inputs are more suitable than energy because (a) Davis et al. (2013) document that a large fraction of the cross-sectional dispersion in electricity prices is due to variation in purchase quantities contradicting the required "no monopsony" assumption **III**, and (b) revenue shares for energy are much smaller when compared to material inputs, thus measurement error in energy inputs gets amplified when estimating markdowns due to division bias. In this section, we provide some additional evidence supporting these claims.

We start by recalculating markdowns with energy as the flexible input. If there is indeed a substantial amount of monopsony in energy markets, then our estimates do not necessarily reflect labor market power alone but labor markdowns *relative to energy markdowns*, say ν_ℓ/ν_E. The evidence in Davis et al. (2013) indicates that $\nu_E > 1$ is likely, so we expect our markdown results with energy inputs to be lower when compared to our baseline. If monopsony in energy markets is so prevalent, in fact, it is also possible that our estimates

Table 3: Estimated plant-level markdowns in U.S. manufacturing (energy as a flexible input).†

INDUSTRY GROUP	Median	Mean	SD
Food and Kindred Products	0.559	0.758	0.825
Textile Mill Products	1.871	2.998	3.085
Apparel and Leather	0.473	0.727	0.970
Lumber	0.681	1.032	1.369
Furniture and Fixtures	0.634	0.889	1.007
Paper and Allied Products	1.118	1.553	1.632
Printing and Publishing	1.396	2.287	2.450
Chemicals	0.980	1.870	2.380
Petroleum Refining	1.963	2.258	1.781
Plastics and Rubber	1.023	1.264	1.135
Non-metallic Minerals	0.389	0.531	0.606
Primary Metals	1.218	1.603	1.501
Fabricated Metal Products	0.656	0.846	0.889
Non-electrical Machinery	0.310	0.376	0.276
Electrical Machinery	0.494	0.914	1.366
Motor Vehicles	0.387	0.492	0.457
Computer and Electronics	0.986	2.084	2.77
Miscellaneous Manufacturing	0.518	0.691	0.765
Whole sample	**0.618**	**0.957**	**1.350**
Sample size	$1.018 \cdot 10^6$		

†Markdowns are estimated under the assumption of a translog specification for gross output. The flexible input is energy. Each industry group in manufacturing corresponds to the manufacturing categorization of the U.S. Bureau of Economic Analysis (BEA) which approximately follows a 3-digit NAICS specification. Source: authors' calculations from ASM/CM data in 1976–2014.

fall below unity most of the time. This is the case whenever $\nu_E > \nu_\ell$. This is exactly what we observe in table 3. For many industries, the median markdown is smaller than unity.

Furthermore, energy shares in U.S. manufacturing are small. The NBER-CES Manufacturing Database indicates that revenue shares average at around 2 percent.[2] In addition, the dispersion in energy shares is substantial: its 10th and 90th percentile equal 0.59 percent and 4.26 percent, respectively. Note, however, that energy is not only more dispersed across plants, but it is also more volatile for a given plant. To show this, we have calculated

[2] The median revenue share for energy is even smaller at 1.18 percent.

the standard deviation of log inputs for each plant's life cycle. Inputs are normalized by the mean of its log level over time. The results are displayed in table 4. Because of its modest and volatile revenue share, we conjectured that markdowns estimated with energy as the flexible input would be much less accurate. Indeed, due to the volatility of expenditure on energy inputs, measurement error in energy shares is amplified by division bias. This is reflected in the within-industry standard deviations of markdowns when estimated with energy inputs, which are significantly higher compared to our baseline estimates.

Table 4: Variability of inputs.†

INPUT	Median	Mean	25%	75%	SD
Capital	0.0154	0.0280	0.0080	0.0339	0.0341
Labor	0.0307	0.0401	0.0171	0.0518	0.0349
Materials	0.0391	0.0493	0.0222	0.0648	0.0394
Energy	0.0625	0.0954	0.0335	0.1158	0.1331

†For each plant, we calculate the standard deviation of its log normalized inputs over time. Each plant's input is normalized by the mean of its log level over time. The sample is restricted to those plants that have at least three observations over their life cycle. Source: Authors' calculations from ASM/CM data in 1976–2014.

As expected, we see that energy usage is much more volatile for the average plant when compared to other inputs. Hence, it is not surprising that our markdown estimates with energy are much more volatile when compared to our baseline estimates.

O.1.4 Markups

In the following, we report our estimates for markups. Summary statistics are provided for each industry group. The results clearly indicate that there is market power in output markets: the median (and mean) markup at the plant-year level equals about 20 percent. Similar to markdowns, there is a substantial amount of variation across industry groups, though the within-industry variation of markups is substantially more limited when compared to markdowns. The IQR for markups is about 16.5 percent whereas its standard deviation for the whole sample is 18.8 percent.

While these estimates are informative for markups, it should be noted that our estimates for markups *in isolation* are faced with a bias. This is because we proxied physical output with deflated revenues which causes a downward bias in markups (see Klette and Griliches, 1996). This has recently been reiterated by Bond et al. (2021). Thus, in a conservative sense, our estimates for markups can also be interpreted as lower bounds for market power in output markets. Note, however, that these estimates for markups are still valid when they are used in order to obtain estimates for markdowns. This is a point we emphasize in Online Appendix O.6.

Table 5: Estimated plant-level markups in U.S. manufacturing.†

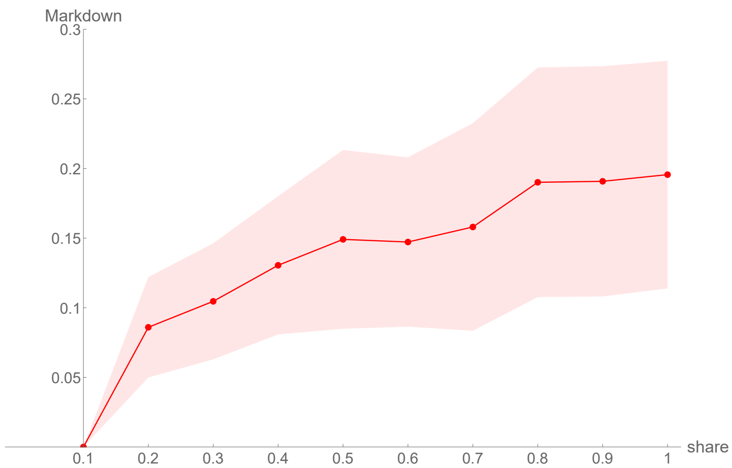
INDUSTRY GROUP	Median	Mean	IQR_{75-25}	SD
Food and Kindred Products	1.145	1.165	0.139	0.123
Textile Mill Products	1.218	1.220	0.136	0.122
Apparel and Leather	1.286	1.293	0.152	0.193
Lumber	1.056	1.055	0.115	0.107
Furniture and Fixtures	1.227	1.226	0.143	0.122
Paper and Allied Products	1.081	1.084	0.129	0.106
Printing and Publishing	1.249	1.234	0.136	0.183
Chemicals	1.330	1.368	0.243	0.214
Petroleum Refining	1.119	1.160	0.194	0.192
Plastics and Rubber	1.107	1.105	0.147	0.131
Non-metallic Minerals	1.219	1.218	0.104	0.135
Primary Metals	1.129	1.142	0.116	0.096
Fabricated Metal Products	1.194	1.198	0.073	0.058
Non-electrical Machinery	1.449	1.488	0.278	0.193
Electrical Machinery	1.286	1.294	0.105	0.083
Motor Vehicles	1.170	1.178	0.082	0.071
Computer and Electronics	1.023	1.018	0.197	0.180
Miscellaneous Manufacturing	1.255	1.263	0.071	0.068
Whole sample	**1.205**	**1.214**	**0.165**	**0.188**
Sample size	$1.393 \cdot 10^6$			

†Markups are estimated under the assumption of a translog specification for gross output. The flexible input is materials. Each industry group in manufacturing corresponds to the manufacturing categorization of the BEA which approximately follows a 3-digit NAICS specification. Source: Authors' calculations from ASM/CM data in 1976–2014.

O.1.5 Size, age and productivity effects

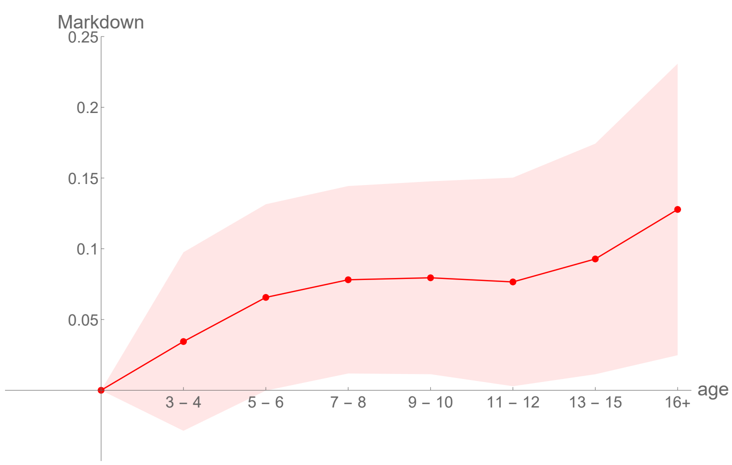
O.1.5.1 Size and age regressions without controls

Figure 1: Markdowns increase with establishment size.



Note: The figure shows point estimates and 95-percent confidence intervals of plant-specific markdowns on size (as measured by employment share) indicators, controlling for state, industry and year fixed effects. The omitted group is the smallest size indicator, so coefficients reflect deviations relative to this baseline. The indicator labeled “0.1” is equal to unity for those plants with employment shares $s \in (0, 0.1]$. Other indicators are defined similarly. Standard errors are clustered at the industry level. Source: Authors’ own calculations from ASM/CM data in 1976–2014.

Figure 2: Markdowns increase with establishment age, but this result only holds when not controlling for establishment size.



Note: The figure shows point estimates and 95-percent confidence intervals of plant-specific markdowns on age category indicators, controlling for state, industry and year fixed effects. The omitted group is the smallest age category, less than three years, so coefficients reflect deviations relative to this baseline. Standard errors are clustered at the industry level. Source: Authors’ own calculations from ASM/CM data in 1976–2014.

O.1.5.2 Baseline results in tabular form

Table 6: Non-parametric estimates of markdowns on size, age and productivity†

Dependent variable: MARKDOWNS					
	SIZE		AGE		TFPR
Share bin		Age bin		TFPR %	
0.1 – 0.2	0.0849 (0.0181)	3 – 4	0.0242 (0.0308)	1% – 5%	−0.8088 (0.2842)
0.2 – 0.3	0.1030 (0.0212)	5 – 6	0.0536 (0.0327)	5% – 10%	−0.8162 (0.3920)
0.3 – 0.4	0.1286 (0.0254)	7 – 8	0.0637 (0.0326)	10% – 25%	−0.7629 (0.4198)
0.4 – 0.5	0.1471 (0.0326)	9 – 10	0.0557 (0.0333)	25% – 50%	−0.6257 (0.4360)
0.5 – 0.6	0.1452 (0.0308)	11 – 12	0.0586 (0.0365)	50% – 75%	−0.5020 (0.4383)
0.6 – 0.7	0.1560 (0.0377)	13 – 15	0.0709 (0.0401)	75% – 90%	−0.4031 (0.4486)
0.7 – 0.8	0.1880 (0.0419)	16+	0.0978 (0.0514)	90% – 95%	−0.2453 (0.4747)
0.8 – 0.9	0.1882 (0.0420)			95% – 99%	0.1084 (0.5182)
0.9 – 1	0.1934 (0.0420)			99%+	0.8046 (0.5321)
Observations (in millions)	1.393		1.393		1.393
R^2	0.2579		0.2579		0.3385

†All regression specifications contain fixed effects at the state, industry and year level, and are weighted by the product of employment and the ASM sampling weights. The results are almost identical whenever only ASM sampling weights are used instead. The specifications for size and age control for age and size, respectively. The omitted categories for the size, age and productivity specifications are 0 – 0.1, 1 – 2 and $<1\%$, respectively. Hence, the regression coefficients reflect deviations relative to these baselines. The indicator labeled "0.1 – 0.2" is equal to unity for those plants with employment shares $s \in (0.1, 0.2]$. Other indicators for the size specification are defined similarly. Standard errors, in parentheses, are clustered at the industry level. Source: Authors' calculations from ASM/CM data in 1976–2014.

O.2 Additional results on the aggregate markdown

O.2.1 Aggregate markdowns and employment concentration

We calculate the cross-sectional correlation (across local labor markets) between the aggregate markdown $\mathcal{V}_{jlt}$ and employment concentration HHI_{jlt}. The results for each Census year can be found in table 7.

Table 7: The correlation between employment HHIs and aggregate markdown across local labor markets is close to zero.†

Specification: TRANSLOG MARKDOWNS		
YEAR	$\rho(\mathcal{V}_{jlt}, \mathrm{HHI}_{jlt})$	$\rho(\mathcal{V}_{jlt}^{\mathrm{RHST}}, \mathrm{HHI}_{jlt})$
1977	0.01656	0.00017
1982	0.00779	0.03593
1987	-0.00164	0.03528
1992	-0.01491	0.03305
1997	0.00097	0.01567
2002	0.00385	0.01444
2007	0.00440	0.00425
2012	-0.01964	0.01108
AVERAGE	-0.00033	0.01873

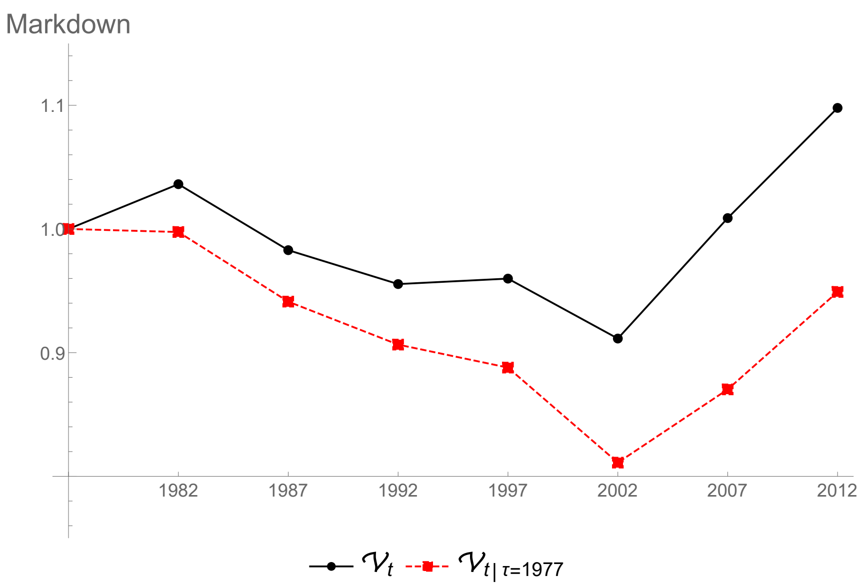
†Markdowns are estimated under the assumption of a translog specification for gross output. Cross-market correlations are calculated at the 3-digit NAICS-county level for each Census year. Aggregate markdowns are calculated according to formulas (12) and (14) whereas HHI_{jlt} denotes a market's employment Herfindahl-Hirschman index. Source: Authors' own calculations from quinquennial CM data from 1977–2012.

Our results indicate that the correlations between labor market power and employment concentration are low. In fact, these correlations are close to zero and for some Census years even negative. When we take the average cross-market correlation across Census years, we basically find a value of zero. Our conclusions do not change whenever we base our results on rank correlations (e.g., Spearman's ρ or Kendall's τ) instead. In section III, we found that the aggregate markdown in the spirit of Rossi-Hansberg, Sarte and Trachter (2020), calculated with equation (14), displayed a relatively strong correlation over time with local concentration LOCAL_t. However, our results in table 7 indicate that the cross-sectional correlations are also fairly weak under this specification.

O.2.2 Compositional effects and benefits

In this section, we provide several robustness checks on the aggregate markdown $\mathcal{V}_t$. First, we verify that the distinct time evolution of the aggregate markdown is not purely driven by compositional changes across local labor markets. To do so, we recalculate the aggregate markdown but fix its weights across local labor markets at their 1977 level. That is, we construct $\mathcal{V}_{t|\tau} \equiv \sum_{j\in J}\sum_{l\in L}\omega_{jl\tau}\mathcal{V}_{jlt}$ with $\tau = 1977$. The results can be found in figure 3.

Figure 3: The qualitative nature of the time evolution for the aggregate markdown cannot be explained by compositional changes across local labor markets.

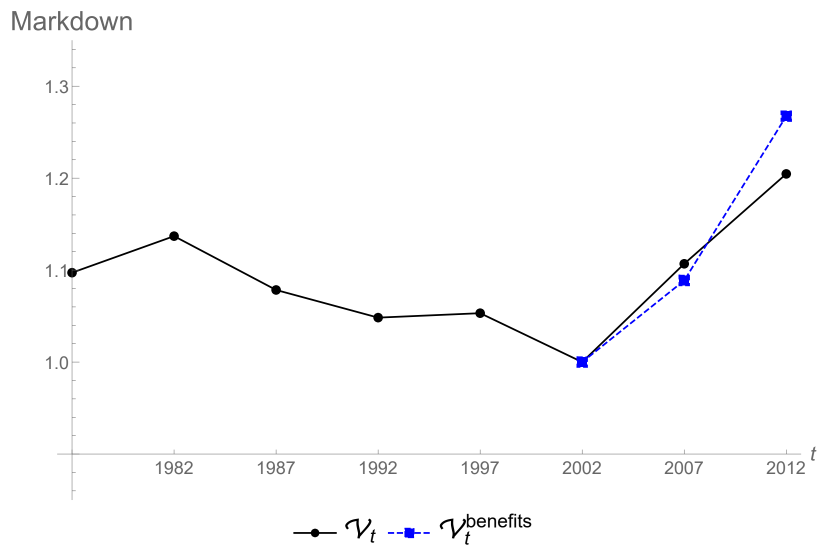


Markdowns are constructed under the assumption of translog production and aggregated according to equation (12). Our baseline measure $\mathcal{V}_t$ is depicted by the solid black line. The aggregate markdown $\mathcal{V}_{t|\tau=1977}$ (dashed red) is calculated by fixing the employment weights for local labor markets at their 1977 values. All measures are normalized relative to their initial value in 1977. Source: Authors' own calculations from quinquennial CM data from 1977–2012.

We find that the qualitative nature of the aggregate markdown is preserved. When employment weights across local labor markets are fixed at their 1977 values, the aggregate markdown also decreases until 2002 and increases afterward. However, its decrease from 1977 to 2002 is a bit stronger than in our baseline specification. Nevertheless, we conclude that the evolution of the aggregate markdown $\mathcal{V}_t$ cannot be accounted for by changes in the employment composition across local labor markets.

Second, our baseline specification of the aggregate markdown does not include health and pension benefits. However, these benefits are available from 2002 onward. We verify that the aggregate markdown also starkly increases whenever benefits are taken into consideration. Given that benefits are available from only 2002 onward, we normalize our series to unity in 2002. As shown in figure 4, the aggregate markdown also increases from 2002 onward whenever benefits are included.

Figure 4: The stark increase of the aggregate markdown $\mathcal{V}_t$ (solid black) from 2002 onward is preserved whenever benefits (dashed blue) are also taken into account.

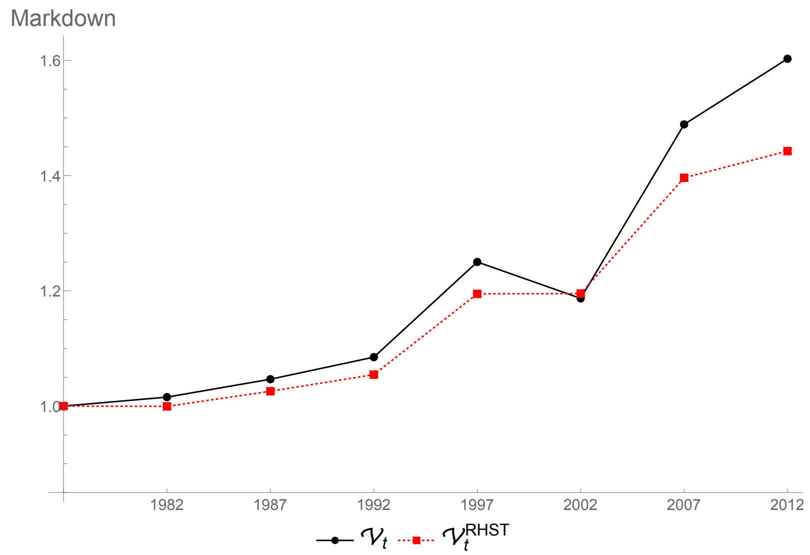


Markdowns are constructed under the assumption of translog production and aggregated according to equation (12). The aggregate markdown $\mathcal{V}_t^{\text{benefits}}$ is calculated by including health and pension benefits. All measures are normalized relative to their values in 2002. Source: Authors' own calculations from quinquennial CM data from 1977–2012.

O.2.3 Secular trend in markdowns: Cobb-Douglas

In our baseline estimates, we specified production functions to be translog. By construction, the translog specification allows output elasticities to vary with the level of inputs. As a result, these output elasticities can vary over time as well. Under a Cobb-Douglas specification, output elasticities are constant and markdowns can only vary over time due to changes in revenue shares. In the following, we show that allowing for time-varying output elasticities is important for several measures of the aggregate markdown.

Figure 5: Time evolution of aggregate markdowns across U.S. manufacturing plants from 1977 to 2012 (Cobb-Douglas case). Unlike the baseline estimation using translog, these measures are increasing over time (cfr. figure 4 in main text).



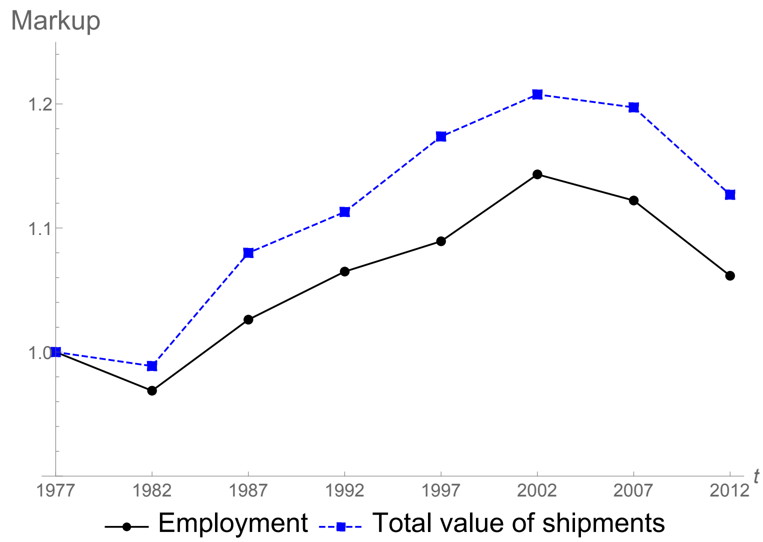
Markdowns are constructed under the assumption of Cobb-Douglas production and aggregated according to equations (12) and (14), respectively. All measures are normalized relative to their initial value in 1977. Source: Authors' own calculations from quinquennial CM data from 1977–2012.

We start by calculating the aggregate measures $\mathcal{V}_t$ and $\mathcal{V}_t^{\text{RHST}}$ whenever production technologies are assumed to be Cobb-Douglas. While these measures are decreasing over time (at least before 2002) under a translog specification, the opposite is true whenever markdowns are estimated under Cobb-Douglas technologies. This is illustrated in figure 5. These differences underline that Cobb-Douglas specifications can be quite restrictive. By construction, the Cobb-Douglas specification assumes that output elasticities are constant and, hence, ignores any time variation in a plant's output elasticities. Conversely, a translog specification allows precisely for this. Our results favor the translog specification since they indicate that this time variation is quantitatively important.

O.2.4 Secular trend in markups

In this section, we present the time series for the aggregate markup. The aggregate markup at the market level is calculated according to equation (13). Then, we aggregate markups across markets through either employment or revenue weights.

Figure 6: Time evolution of revenue- and employment-weighted markups (the black and blue line, respectively) across U.S. manufacturing plants from 1977 to 2012.



Markups are constructed under the assumption of translog production and aggregated according to equation (13). Source: Authors' own calculations from ASM/CM data in 1976–2014.

As emphasized by Bond et al. (2021), the estimation of micro-level markups with deflated revenues, instead of physical output, leads to biases that make interpretation challenging (see Online Appendix O.6). In turn, bias in the level of markups at the micro level will lead to bias in the aggregate markup. Note however, as we show formally in Online Appendix O.6, this concern does *not* apply to our estimation of markdowns.

Consequently, we feel that using our methodology to present markups should—at the very least—be treated cautiously by other researchers. However, presenting a trend of aggregate markups could still be useful to others even when bias is present—perhaps in comparison to markup trends created under different approaches and different biases. This trend is depicted in figure 6.

O.2.5 Decomposition of aggregate markdowns

In the following, we will apply the decomposition by Foster, Haltiwanger and Krizan (2001) to aggregate markdowns in order to understand what was driving its changes. However, this is not straightforward because the accounting decomposition by Foster, Haltiwanger and Krizan (2001) applies to arithmetic (weighted) averages only. In the discussion below, we will present some accounting identities that will allow us to apply the decomposition by Foster, Haltiwanger and Krizan (2001) to harmonic (weighted) averages. To do so, we start with the following lemma.

LEMMA 1. For any aggregate variable X_t, we have:

$$\Delta X_t = -\frac{\Delta X_t^{-1}}{1+\Delta X_t^{-1}} \tag{2}$$

Proof. By definition, we have:

$$\begin{aligned}\Delta X_t^{-1} &= \frac{X_t^{-1} - X_{t-1}^{-1}}{X_{t-1}^{-1}} \\ &= -\frac{X_{t-1}}{X_t}\left(\frac{X_t - X_{t-1}}{X_{t-1}}\right) \\ &= -\frac{\Delta X_t}{1+\Delta X_t}\end{aligned}$$

Then, the lemma follows directly by solving for ΔX_t. □

This is useful since our definition of the aggregate markdown consists of a ratio of two sales-weighted harmonic averages. That is, we have $\mathcal{V}_{jlt} \equiv \frac{V_{jlt}}{\mathcal{M}_{jlt}}$ with:

$$V_{jlt} = \left(\sum_{i\in F_t(j,l)} s_{it} \cdot \frac{\theta_{it}^L}{\theta_{jlt}^L} \cdot (\nu_{it}\mu_{it})^{-1}\right)^{-1} \tag{3}$$

$$\mathcal{M}_{jlt} = \left(\sum_{i\in F_t(j,l)} s_{it} \cdot \frac{\theta_{it}^M}{\theta_{jlt}^M} \cdot \mu_{it}^{-1}\right)^{-1} \tag{4}$$

Note that for any weighted harmonic average X_t, we can write:

$$\widetilde{X}_t \equiv X_t^{-1} = \sum_{i\in F_t} s_{it} x_{it}^{-1} \equiv \sum_{i\in F_t} s_{it}\tilde{x}_{it}$$

The latter is just a simple (i.e., arithmetic) weighted average. For these types of averages, we can apply the decomposition of Foster, Haltiwanger and Krizan (2001):

$$\begin{aligned}\Delta \widetilde{X}_t = & \sum_{i \in C_t} s_{it-1} \Delta \tilde{x}_t + \sum_{i \in C_t} \left(\tilde{x}_{it-1} - \widetilde{X}_{t-1}\right) \Delta s_{it} + \sum_{i \in C_t} \Delta \tilde{x}_{it} \Delta s_{it} \\ & + \sum_{i \in N_t} s_{it} \left(\tilde{x}_{it} - \widetilde{X}_{t-1}\right) - \sum_{i \in X_t} s_{it-1} \left(\tilde{x}_{it-1} - \widetilde{X}_{t-1}\right) \qquad (5) \\ & \equiv \text{WITHIN}_t + \text{BTWN}_t + \text{COV}_t + \text{ENTRY}_t - \text{EXIT}_t \qquad (6)\end{aligned}$$

where the growth rate of $\widetilde{X}_t$ can be decomposed into within-firm, between-firm, covariance, entry and exit components, respectively. Note that the first three components can only be applied to incumbent firms (i.e., firms active in periods t and $t-1$). By definition of the aggregate markdown, we have:

$$\begin{aligned}\widetilde{V}_{jlt} \equiv V_{jlt}^{-1} &= \sum_{i \in F_t(j,l)} s_{it} \cdot \frac{\theta_{it}^L}{\theta_{jlt}^L} \cdot (\nu_{it}\mu_{it})^{-1} \\ &\equiv \sum_{i \in F_t(j,l)} s_{it} \cdot \tilde{v}_{it} \\ \widetilde{\mathcal{M}}_{jlt} \equiv \mathcal{M}_{jlt}^{-1} &= \sum_{i \in F_t(j,l)} s_{it} \cdot \frac{\theta_{it}^M}{\theta_{jlt}^M} \cdot \mu_{it}^{-1} \\ &\equiv \sum_{i \in F_t(j,l)} s_{it} \cdot \tilde{\mu}_{it}\end{aligned}$$

Thus, we can apply the insight of Foster, Haltiwanger and Krizan (2001) in (5) to $\widetilde{V}_{jlt}$ and $\widetilde{\mathcal{M}}_{jlt}$ to obtain decompositions for $\Delta\widetilde{V}_{jlt}$ and $\Delta\widetilde{\mathcal{M}}_{jlt}$. This will aid us in understanding growth in the aggregate markdown since we have:

$$\begin{aligned}\Delta \mathcal{V}_{jlt} &= \Delta V_{jlt} - \Delta \mathcal{M}_{jlt} \\ &= -\frac{\Delta \widetilde{V}_{jlt}}{1 + \Delta \widetilde{V}_{jlt}} + \frac{\Delta \widetilde{\mathcal{M}}_{jlt}}{1 + \Delta \widetilde{\mathcal{M}}_{jlt}} \qquad (7) \\ &\approx \Delta \widetilde{\mathcal{M}}_{jlt} - \Delta \widetilde{V}_{jlt} \qquad (8)\end{aligned}$$

where the last approximation follows from the fact that we have $-\frac{x}{1+x} \simeq -x$ up to a first order for small values of x. This seems appropriate in our setting given the observed movements in aggregate markdowns. Thus, growth in the aggregate markdown, for a given local labor market, is primarily led by those components that are more important for the growth rate of the inverse aggregate markup, i.e. $\Delta\widetilde{\mathcal{M}}$, whereas it is slowed down by those components that determine the growth rate of the inverse aggregate labor wedge, i.e. $\Delta\widetilde{V}$.

Table 8: Decomposition of $\Delta\widetilde{\mathcal{M}}$ and $\Delta\widetilde{V}$ (cfr. equation 8).† Movements in the aggregate markdown are not clearly driven by one specific type of reallocation.

YEAR		WITHIN_t	BTWN_t	COV_t	ENTRY_t	EXIT_t
1977 – 1982	$\Delta\widetilde{\mathcal{M}}$	0.3618	0.1370	0.1547	0.2042	0.1423
1977 – 1982	$\Delta\widetilde{V}$	0.3997	0.1231	0.1120	0.2162	0.1490
1982 – 1987	$\Delta\widetilde{\mathcal{M}}$	0.3724	0.1125	0.1140	0.2317	0.1694
1982 – 1987	$\Delta\widetilde{V}$	0.3386	0.1271	0.1553	0.2261	0.1528
1987 – 1992	$\Delta\widetilde{\mathcal{M}}$	0.3782	0.1131	0.1218	0.2244	0.1625
1987 – 1992	$\Delta\widetilde{V}$	0.3537	0.1236	0.1585	0.2190	0.1453
1992 – 1997	$\Delta\widetilde{\mathcal{M}}$	0.3903	0.1250	0.1164	0.2113	0.1570
1992 – 1997	$\Delta\widetilde{V}$	0.3452	0.1281	0.1753	0.2119	0.1395
1997 – 2002	$\Delta\widetilde{\mathcal{M}}$	0.3555	0.1189	0.1193	0.2408	0.1655
1997 – 2002	$\Delta\widetilde{V}$	0.3358	0.1262	0.1583	0.2307	0.1491
2002 – 2007	$\Delta\widetilde{\mathcal{M}}$	0.3777	0.1273	0.1244	0.2172	0.1534
2002 – 2007	$\Delta\widetilde{V}$	0.3363	0.1384	0.1819	0.1966	0.1469
2007 – 2012	$\Delta\widetilde{\mathcal{M}}$	0.3979	0.1281	0.1280	0.2033	0.1426
2007 – 2012	$\Delta\widetilde{V}$	0.3441	0.1449	0.1767	0.190	0.1444

†Markdowns are estimated under the assumption of a translog specification for gross output. The flexible input is materials. Each industry group in manufacturing corresponds to the manufacturing categorization of the BEA which approximately follows a 3-digit NAICS specification. Each component is denoted in absolute values and normalized by the sum of absolute values for each component. The table reports the employment-weighted mean across local labor markets. Source: Authors' calculations from ASM/CM data in 1976–2014.

We follow Foster, Haltiwanger and Krizan (2001) and calculate the employment-weighted average across local labor markets of the *absolute* contribution for each component. By construction, we can write $\Delta\widetilde{V} = \text{WITHIN} + \text{BTWN} + \text{COV} + \text{ENTRY} - \text{EXIT}$. Then, for each local labor market, we calculate each component's absolute contribution by taking its absolute value and dividing it by the sum of absolute values for each component. That is:

$$\widehat{x} = \frac{|x|}{|\text{WITHIN}| + |\text{BTWN}| + |\text{COV}| + |\text{ENTRY}| + |\text{EXIT}|}$$

for $x \in \{\text{WITHIN}, \text{BTWN}, \text{COV}, \text{ENTRY}, \text{EXIT}\}$.[3] Then, we report averages across local labor markets using employment weights. This is appropriate in our setting since we aggregate markdowns across local labor markets by taking employment-weighted averages in order to obtain $\mathcal{V}_t$.

Our decomposition in equation (8) indicates that movements in the aggregate markdown are primarily determined by those components that are relatively important for $\Delta\widetilde{\mathcal{M}}$ but not for $\Delta\widetilde{V}$. However, our results in table 8 indicate that each component is about equally important for $\Delta\widetilde{\mathcal{M}}$ and $\Delta\widetilde{V}$. As a result, we conclude that movements in the aggregate markdown are not clearly driven by one specific type of reallocation.

[3]We report absolute contributions for each component since the patterns over time for each raw component are difficult to interpret: they can switch signs over time and are also quite volatile. This is similar to Foster, Haltiwanger and Krizan (2001) who apply the decomposition to aggregate productivity in U.S. manufacturing sectors (see their table 8.7). In fact, they mention that their results can be quite "erratic" under the used accounting decomposition.

O.2.6 Aggregate markdowns and local concentration in tabular form

Table 9: Measures of the aggregate markdown and local concentration.†

Specification: TRANSLOG MARKDOWNS				
YEAR	$\mathcal{V}_t$	$\mathcal{V}_t^{\text{RHST}}$	$\mathcal{V}_t^{\text{dLEU}}$	LOCAL_t
1977	1.000	1.000	1.000	1.000
1982	1.0362	0.9653	0.9495	0.9640
1987	0.9829	0.9515	0.9392	0.9841
1992	0.9555	0.9460	0.9289	0.9707
1997	0.9599	0.9344	0.9330	0.9224
2002	0.9114	0.9322	0.9310	0.9269
2007	1.0088	0.9366	0.9815	0.9297
2012	1.0979	0.9272	1.016	0.9646

†Markdowns are estimated under the assumption of a translog specification for gross output. Aggregate markdowns $\mathcal{V}_t$, $\mathcal{V}_t^{\text{dLEU}}$ and $\mathcal{V}_t^{\text{RHST}}$ are calculated according to formulas (12), (15) and (14), respectively, whereas LOCAL_t denotes local concentration as calculated according to equation (18). All values are normalized with respect to 1977. Source: Authors' own calculations from quinquennial CM data from 1977–2012.

O.3 Details on GMM-IV estimation procedure

O.3.1 Implementation of De Loecker and Warzynski (2012)

In the following, we will provide more details on how we obtain output elasticities. To do so, we will follow the "proxy variable" literature on production function estimation (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; De Loecker and Warzynski, 2012; Ackerberg, Caves and Frazer, 2015).

Let the production function be given by:

$$Q_{it} = F(\mathbf{V}_{it}, \mathbf{K}_{it}; \omega_{it})$$

where we categorize inputs as flexible or non-flexible inputs, i.e. $\mathbf{X}'_{it} = (\mathbf{V}'_{it}, \mathbf{K}'_{it})$. In particular, we have:

$$\begin{aligned}\mathbf{V}_{it} &= (X^1_{it}, \dots, X^V_{it})' \\ \mathbf{K}_{it} &= (X^{V+1}_{it}, \dots, X^K_{it})'\end{aligned}$$

where the first $V \geq 1$ inputs are flexible and the latter $K - V$ inputs are not fully flexible. In particular, $\mathbf{K}_{it}$ is a state variable when choosing the inputs $\mathbf{V}_{it}$. Furthermore, ω_{it} denotes a firm's productivity. In particular, suppose that $X^1_{it} = M_{it}$ are material inputs.

To account for measurement error, we assume that *observed* logged output satisfies $y_{it} = \ln(Q_{it}) + \varepsilon_{it}$, i.e. measurement error enters production in a multiplicative fashion. Note that the error term ε_{it} is *not* observed by firms when they have to make their optimal input decisions. Given our econometric assumptions 1 – 5, we can write:

$$y_{it} = f(\mathbf{v}_{it}, \mathbf{k}_{it}; \boldsymbol{\beta}) + \omega_{it} + \varepsilon_{it}$$

where $f(\mathbf{v}_{it}, \mathbf{k}_{it}; \boldsymbol{\beta}) = \ln(F(\mathbf{V}_{it}, \mathbf{K}_{it}; \boldsymbol{\beta}))$, and $\mathbf{v}_{it}$ and $\mathbf{k}_{it}$ denote componentwise natural log transformations of $\mathbf{V}_{it}$ and $\mathbf{K}_{it}$, respectively. Firm-level productivities ω_{it} are not observed by the econometrician, but are observable for firms themselves.

Unobservable productivity is the main cause of endogeneity concerns in our estimation procedure. To deal with this, we use the insight of Levinsohn and Petrin (2003). Under assumptions 4 and 5, material demand $\ln(X_{it}^1) = m_{it}$ can be used to proxy for productivity. Note that firms choose flexible inputs given the state $\mathbf{K}_{it}$, idiosyncratic productivity ω_{it} and some controls that can influence their decisions $\mathbf{c}_{it}$ (e.g., input prices):

$$m_{it} = m_t(\omega_{it}; \mathbf{k}_{it}, \mathbf{c}_{it})$$

where the vector $\mathbf{c}_{it}$ denotes any additional, observable variables that can affect a plant's optimal demand for material inputs.[4] The above mapping for materials is invertible in productivity ω_{it} by assumption 4. Following Levinsohn and Petrin (2003), a sufficient condition for invertibility is:

$$D^M = \begin{vmatrix} \frac{\partial \mathbf{V}_{it}(\mathbf{K}_{it},\omega_{it})}{\partial \omega_{it}} & \mathbf{H}_{2,V}^F(\mathbf{K}_{it}, \omega_{it}) & \dots & \mathbf{H}_{V,V}^F(\mathbf{K}_{it}, \omega_{it}) \end{vmatrix} > 0$$

where $\frac{\partial \mathbf{V}_{it}(\mathbf{K}_{it},\omega_{it})}{\partial \omega_{it}} = \left(\frac{\partial X_{it}^1(\mathbf{K}_{it},\omega_{it})}{\partial \omega_{it}}, \dots, \frac{\partial X_{it}^V(\mathbf{K}_{it},\omega_{it})}{\partial \omega_{it}}\right)'$ and
$\mathbf{H}_{r,V}^F(\mathbf{K}_{it}, \omega_{it}) = \left(\frac{\partial F(\mathbf{V}_{it},\mathbf{K}_{it},\omega_{it})}{\partial X_{it}^r \partial X_{it}^1}, \dots, \frac{\partial F(\mathbf{V}_{it},\mathbf{K}_{it},\omega_{it})}{\partial X_{it}^r \partial X_{it}^V}\right)'$ is the r^{th} column of the Hessian matrix for $F(\cdot, \mathbf{K}_{it}; \omega_{it})$ evaluated at $\mathbf{V}_{it} \in \mathbb{R}_+^V$.

Under this assumption, the material input demand function is monotonic in productivity ω_{it}.[5] Then, there exists some function $h_t(\cdot; \mathbf{k}_{it}, \mathbf{c}_{it})$ such that:

$$\omega_{it} = h_t(m_{it}; \mathbf{k}_{it}, \mathbf{c}_{it})$$

As a result, production y_{it} can be written in terms of observables only:

$$\begin{aligned} y_{it} &= f(\mathbf{v}_{it}, \mathbf{k}_{it}; \boldsymbol{\beta}) + h_t(m_{it}; \mathbf{k}_{it}, \mathbf{c}_{it}) + \varepsilon_{it} \\ &= \phi_t(\mathbf{v}_{it}, \mathbf{k}_{it}, \mathbf{c}_{it}) + \varepsilon_{it} \\ &= \varphi_{it} + \varepsilon_{it} \end{aligned}$$

[4]In the empirical implementation, $\mathbf{c}_{it}$ only contains a set of year fixed effects. Industry fixed effects are not required whenever production technology parameters are estimated industry-by-industry. However, the used methodology is flexible enough to account for other observables.

[5]This follows from standard arguments for comparative statics under multiple inputs. We then apply Cramer's rule to arrive at the stated condition. Levinsohn and Petrin (2003) show a similar result for $V = 2$ in their appendix A. In a nutshell, assumption 5 imposes a set of regularity conditions on the cross-derivatives of the production function in $\mathbf{V}_{it}$ which are fairly mild.

Estimating the production technology parameters $\boldsymbol{\beta}$ is done in a three stage fashion which is in a similar spirit to Ackerberg, Caves and Frazer (2015). To implement our estimation procedure, we set $\mathbf{v}_{it} = m_{it}$, $\mathbf{k}_{it} = (k_{it}, \ell_{it}, e_{it})'$ and $\mathbf{c}_{it} = (d_{i,1}, \ldots, d_{i,T})'$ where $d_{i,t}$ is a fixed effect for a specific year t. Even though we will mainly focus on translog production functions, we also occasionally report results for Cobb-Douglas specifications.

Step 1. Non-parametric estimation of φ_{it} and ε_{it}.

First, we estimate φ_{it} and ε_{it} non-parametrically by approximating y_{it} with a third degree polynomial in $\tilde{\mathbf{x}}_{it} = (k_{it}, \ell_{it}, m_{it}, e_{it})'$ with interaction terms. In the case of translog production, we have:

$$\mathbf{x}_{it} = (k_{it}, \ell_{it}, m_{it}, e_{it}, k_{it}\ell_{it}, k_{it}m_{it}, k_{it}e_{it}, \ell_{it}m_{it}, \ell_{it}e_{it}, m_{it}e_{it}, k_{it}^2, \ell_{it}^2, m_{it}^2, e_{it}^2)'$$

Let its fitted values and residuals be denoted by $\widehat{\varphi}_{it}$ and $\widehat{\varepsilon}_{it}$ respectively. These residuals are then interpreted as measurement error in observed output.

Step 2. Construction of innovations ξ_{it} to productivity ω_{it}.

By assumption 3, idiosyncratic productivity ω_{it} is Markovian, thus its expected value is only a function of its lagged value. As a result, we have $\omega_{it} = g_t(\omega_{it-1}) + \xi_{it}$. Then, productivity is approximated in the data by:

$$\omega_{it}(\boldsymbol{\beta}) = \widehat{\varphi}_{it} - f(\mathbf{x}_{it}; \boldsymbol{\beta})$$

Then, we approximate $g_t(.)$ with a $\mathcal{P}^{\text{th}}$ order polynomial in its argument:

$$\begin{aligned}\omega_{it}(\boldsymbol{\beta}) &= \Omega_{it-1}(\boldsymbol{\beta})'\rho(\boldsymbol{\beta}) + \xi_{it} \\ &= \sum_{p=0}^{\mathcal{P}} \rho_p \omega_{it-1}^p(\boldsymbol{\beta}) + \xi_{it}\end{aligned}$$

where we follow De Loecker and Warzynski (2012) and set $\mathcal{P} = 3$. Thus, the innovations to productivity can be constructed as a function of $\boldsymbol{\beta}$ through:

$$\xi_{it}(\boldsymbol{\beta}) = \omega_{it}(\boldsymbol{\beta}) - \Omega_{it-1}(\boldsymbol{\beta})'\widehat{\rho}(\boldsymbol{\beta})$$

The estimates $\widehat{\rho}(\boldsymbol{\beta}) = (\{\widehat{\rho}_p\}_{p=1}^{\mathcal{P}})'$ are simply obtained by running a least squares regression of $\Omega_{it-1}(\boldsymbol{\beta})$ on $\omega_{it}(\boldsymbol{\beta})$.

STEP 3. GMM-IV ESTIMATION OF $\boldsymbol{\beta}$.
By assumption 2, capital is predetermined at time t as a firm chooses it one period ahead. As a result, it is safe to assume that k_{it} is orthogonal to the innovation $\xi_{it}(\boldsymbol{\beta})$. Similarly, firms cannot observe the string of future innovations to their productivity. As a result, current input decisions (with the exception of investment in capital) must be orthogonal to shocks to their idiosyncratic productivity in the future. Define the instrument $\mathbf{z}_{it} \in \mathbb{R}^Z$ as the vector that contains one-period lagged values of every polynomial term containing ℓ_{it}, m_{it} and e_{it} in the production technology $f(\mathbf{x}_{it}; \boldsymbol{\beta})$ but with capital preserved at its current value k_{it}. Then, we define the following system of moment conditions to identify $\boldsymbol{\beta} \in \mathbb{R}^Z$:

$$\mathbb{E}\left(\xi_{it}(\boldsymbol{\beta})\mathbf{z}_{it}\right) = \mathbf{0}_{Z\times 1} \tag{9}$$

By construction, this system of equations defines a set of exogeneity conditions. Lagged inputs are used to instrument for current period inputs. To validate this identification strategy, we need to argue that the moment conditions in (9) also satisfy rank conditions. Our focus lies on material inputs, so we will pay particular attention for this specific input. For lagged material inputs to be a valid instrument for current material inputs, m_{it} and m_{it-1} need to be correlated. A sufficient condition would be that input prices for material inputs are persistent over time. In fact, Atalay (2014) finds empirical evidence for this using data from the Census of Manufactures.

To obtain $\boldsymbol{\beta}$, we rely on the minimization of a quadratic loss function which is standard in GMM estimation.[6] Thus, we get:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^Z}{\arg\min} \sum_{m=1}^{Z} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \xi_{it}(\boldsymbol{\beta}) z_{it}^m \right)^2$$

where we have $\mathbf{z}_{it} = (z_{it}^1, \ldots, z_{it}^Z)'$.

CONSTRUCTING MARKUPS AFTER OBTAINING ESTIMATES $\widehat{\boldsymbol{\beta}}$. In general, output elasticities with respect to material inputs can depend on the level of *all* inputs; be it flexible or

[6]By construction, the number of parameters in $\boldsymbol{\beta}$ is equal to the amount of identifying moments. This case of "just identification" renders the specification of a weighting matrix useless.

predetermined. This implies that $\theta_{it}^M = \theta_M^{j(i)}(\tilde{\mathbf{x}}_{it}; \boldsymbol{\beta})$. Following the estimation procedure by De Loecker and Warzynski (2012), we can furthermore correct for measurement error ε_{it} in logged output. This is particularly important for data in the ASM and CM. Output prices are not available at the firm level, so output levels are obtained by deflating revenues adjusted for inventories. Unfortunately, the deflators used in the NBER-CES Manufacturing database are only available at the industry level. This causes an unavoidable bias in measuring real output.

However, De Loecker and Warzynski (2012) mention that some concern of this bias can be taken care of with the correction term ε_{it}. By construction, any unobserved variation in output prices orthogonal to a firm's inputs will be absorbed by the measurement error correction term. In addition, if pricing decisions are correlated with a plant's productivity, then this specific variation will be controlled for as well through the use of a proxy for productivity. Then, markups are constructed as:

$$\begin{aligned} \widehat{\mu}_{it} &= \widehat{\theta}_{it}^M \left(\frac{\mathtt{vm}_{it}}{\mathtt{tvs}_{it}/\widehat{\varepsilon}_{it}} \right)^{-1} \\ &= \theta_M^{j(i)}(\tilde{\mathbf{x}}_{it}; \widehat{\boldsymbol{\beta}}) \left(\frac{\mathtt{vm}_{it}}{\mathtt{tvs}_{it}/\exp(\widehat{\varepsilon}_{it})} \right)^{-1} \end{aligned} \tag{10}$$

where $\mathtt{vm}_{it}$ and $\mathtt{tvs}_{it}$ denote a plant i's total expenditure on intermediate inputs and total value of shipments in year t. Production technologies do not differ over time but are allowed to vary across industries by assumption 3.[7]

To construct output elasticities explicitly, we need to take a stance on the production function. In the following, we demonstrate how to obtain output elasticities in the case of translog production.[8] Our preferred specification assumes that production is translog, for two reasons. First, the translog specification is a second-order log approximation to *any* arbitrary, differentiable production function. In fact, the Cobb-Douglas setup is nested within our translog specification. Second, output elasticities are allowed to vary with the level of any input under the translog specification. This implies that markups and markdowns have two sources of time variation: time-varying output elasticities and input rev-

[7]Note that this assumption can be relaxed by estimating, for example, time-varying Cobb-Douglas parameters. This is easily done by restricting the estimation sample to repeated cross-sections in a subset of years. Theoretically, this should be possible for the translog case as well, but the amount of cross-sectional variation in these subsamples might not be sufficient to identify all parameters properly.

[8]Under Cobb-Douglas production, output elasticities are equal to their respective production coefficients.

enue shares.

TRANSLOG PRODUCTION. Assumption 3 under translog production implies:

$$
\begin{aligned}
f(\mathbf{x}_{it};\boldsymbol{\beta}) = {} & \beta_K k_{it} + \beta_L \ell_{it} + \beta_M m_{it} + \beta_E e_{it} \\
& + \beta_{KL} k_{it}\ell_{it} + \beta_{KM} k_{it} m_{it} + \beta_{KE} k_{it} e_{it} + \beta_{LM} \ell_{it} m_{it} + \beta_{LE} \ell_{it} e_{it} + \beta_{ME} m_{it} e_{it} \\
& + \beta_{KK} k_{it}^2 + \beta_{LL} \ell_{it}^2 + \beta_{MM} m_{it}^2 + \beta_{EE} e_{it}^2
\end{aligned}
$$

Assuming that capital is chosen one period ahead, the instrument vector becomes:

$$
\begin{aligned}
\mathbf{z}_{it} = \Big(& k_{it}, \ell_{it-1}, m_{it-1}, e_{it-1}, k_{it}\ell_{it-1}, k_{it} m_{it-1}, k_{it} e_{it-1}, \ell_{it-1} m_{it-1}, \ell_{it-1} e_{it-1}, m_{it-1} e_{it-1}, \\
& k_{it}^2, \ell_{it-1}^2, m_{it-1}^2, e_{it-1}^2 \Big)'
\end{aligned}
$$

where $\boldsymbol{\beta} \in \mathbb{R}^{14}$ is estimated for each industry j. Note that the number of parameters increases exponentially whenever more inputs are considered.[9] Markdowns are then empirically implemented through:

$$
\widehat{\nu}_{it}^{\text{TL}} = \widehat{\theta}_{\ell}^{j(i)}(\tilde{\mathbf{x}}_{it};\widehat{\boldsymbol{\beta}}) \left(\frac{\texttt{sw}_{it}}{\texttt{tvs}_{it}}\right)^{-1} \left[\widehat{\theta}_{M}^{j(i)}(\tilde{\mathbf{x}}_{it};\widehat{\boldsymbol{\beta}}) \left(\frac{\texttt{vm}_{it}}{\texttt{tvs}_{it}/\exp(\widehat{\epsilon}_{it})}\right)^{-1}\right]^{-1}
$$

s.t.

$$
\begin{aligned}
\widehat{\theta}_{\ell}^{j(i)}(\tilde{\mathbf{x}}_{it};\widehat{\boldsymbol{\beta}}) &= \widehat{\beta}_{L}^{j(i)} + \widehat{\beta}_{KL}^{j(i)} k_{it} + \widehat{\beta}_{LM}^{j(i)} m_{it} + \widehat{\beta}_{LE}^{j(i)} e_{it} + 2\widehat{\beta}_{LL}^{j(i)} \ell_{it} \\
\widehat{\theta}_{M}^{j(i)}(\tilde{\mathbf{x}}_{it};\widehat{\boldsymbol{\beta}}) &= \widehat{\beta}_{M}^{j(i)} + \widehat{\beta}_{KM}^{j(i)} k_{it} + \widehat{\beta}_{LM}^{j(i)} \ell_{it} + \widehat{\beta}_{ME}^{j(i)} e_{it} + 2\widehat{\beta}_{MM}^{j(i)} m_{it}
\end{aligned}
$$

O.3.2 Implementation of constant returns to scale restriction

We implement the “production approach” for obtaining markdowns by relying on proxy variable methods. While the induced moment conditions are easily derived and understood, Gandhi, Navarro and Rivers (2020) emphasize that point identification is not achieved when applying the methodology by De Loecker and Warzynski (2012), for example. To address this criticism, we apply the solution suggested in Flynn, Gandhi and Traina (2019). They

[9] With a translog production function with K inputs, there are K linear terms, K quadratic components and $\binom{K}{2}$ unique input pairs. Thus, there are a total of $2K + \binom{K}{2} = \frac{K(K+3)}{2}$ terms.

show that the non-identification problem can be resolved whenever a production function's return to scale is ex-ante specified. Similar to their work, we show the robustness of our markdown estimates whenever we impose a constant returns to scale restriction.[10] Assuming constant returns to scale seems reasonable since a substantial body of previous work (e.g., Basu and Fernald, 1997; Syverson, 2004a; Syverson, 2004b) has shown that constant returns to scale is a good approximation for manufacturing plants.

In the following, we will briefly describe how our estimation procedure is adjusted (for the translog case) when imposing constant returns to scale. In fact, this requires minor adjustments only. Steps 1 and 2 are unchanged whereas we only need to add some moment conditions to step 3. To do so, define a firm's returns to scale as follows:

$$\Sigma_{it}(\boldsymbol{\beta}) = \sum_{\iota \in \{k,\ell,m,e\}} \frac{\partial f(\mathbf{x}_{it};\boldsymbol{\beta})}{\partial \iota_{it}} \tag{11}$$

Also, define the vector $\boldsymbol{\chi}_{it} = (1, \tilde{\mathbf{x}}_{it}')' = (1, k_{it}, \ell_{it}, m_{it}, e_{it})' \in \mathbb{R}^{K+1}$, then the new set of moment conditions can be compactly written as:

$$\mathbb{E}\begin{pmatrix} \xi_{it}(\boldsymbol{\beta})\mathbf{z}_{it} \\ \Sigma_{it}(\boldsymbol{\beta}) - 1 \end{pmatrix} = \mathbf{0}_{(Z+1)\times 1} \tag{12}$$

In the case of a translog production function, we can write the constant returns to scale restriction as a linear operator:

$$\Sigma_{it}(\boldsymbol{\beta}) - 1 = (R\boldsymbol{\beta})' \boldsymbol{\chi}_{it}$$

where R is a $5 \times Z$ matrix defined as:

$$R = \begin{bmatrix} -1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 2 \end{bmatrix}$$

[10]We draw similar conclusions whenever we allow for deviations around constant returns to scale.

Our estimation results are displayed in column 2 (panel B) of table 5 in the main text (see section IV).

O.3.3 Bootstrapping procedure

The GMM-IV estimator of the proxy variable approach does not have a closed-form solution for its standard errors. Furthermore, even if we did have these standard errors for the production function coefficients, it is difficult to derive standard errors for the aggregate markdown due to its non-linear structure. As a result, we resort to bootstrapping methods; similar to De Loecker and Warzynski (2012). In the following, we describe the bootstrap algorithm that we implemented with the Census data.

Initiate bootstrap round parameter at $b = 1$.

I. For each industry group $j \in \{1, \ldots, \mathcal{J}\}$, draw a random sample with replacement from the unbalanced ASM panel containing $N_j^{[b]} = 0.9 \times N_j$ observations.

II. For each plant that has been sampled, select its entire life cycle, i.e. we engage in *panel bootstrapping* (or block-bootstrapping at the plant level). This generates the unbalanced sample $S_j^{[b]}$.

III. Obtain the estimated production function parameters $\widehat{\boldsymbol{\beta}}_j^{[b]}$ (with the two-step GMM-IV estimator from De Loecker and Warzynski, 2012) for each industry j using data from sample $S_j^{[b]}$.

IV. For each Census year τ, calculate the aggregate markdown $\widehat{\mathcal{V}}_\tau^{[b]}$ (normalized to unity in 1977) with the universe of manufacturing plants from the CM using the production function parameters $\widehat{\boldsymbol{\beta}}^{[b]} = (\widehat{\boldsymbol{\beta}}_1^{[b]\prime}, \ldots, \widehat{\boldsymbol{\beta}}_{\mathcal{J}}^{[b]\prime})'$.

V. Define $b := b + 1$ and repeat step **I**. Stop the algorithm whenever $b > B$.

Confidence interval bounds at the α-significance level for the aggregate markdown $\mathcal{V}_\tau$ can then be constructed by taking the $100 \cdot \frac{\alpha}{2}$ and $100 \cdot (1 - \frac{\alpha}{2})$ percentile of the set $\{\widehat{\mathcal{V}}_\tau^{[b]}\}_{b=1}^{B}$. We construct 95 percent confidence intervals through 500 simulations, i.e. $\alpha = 0.05$ and $B = 500$.

Note that the constructed confidence interval for the normalized aggregate markdown does not necessarily have to be symmetric around the estimated (normalized) aggregate markdown $\mathcal{V}_t$. This is because of the non-linear structure of markdowns at the firm level and how firm-level markdowns enter the aggregate markdown in a non-linear fashion. Note that we only sample with replacement in the ASM to estimate the production function parameters $\boldsymbol{\beta}$. However, markdowns at the firm level and the aggregate markdown are always calculated using the full sample of the CM for every Census year $\tau \in \{1977, \dots, 2012\}$. By construction, there is no confidence interval for the aggregate markdown in 1977 since this value is always normalized to unity.

Using these block-bootstrap methods, we have verified that the production function parameters $\boldsymbol{\beta}$ are statistically significant for every industry group. In particular, we find that the cross- and second-order terms of our production function specification are statistically significant; indicating the importance of the translog specification.

O.4 Labor adjustment costs

In this appendix, we show that the wedge between the marginal revenue product of labor and the wage is no longer reflective of only labor market power whenever labor adjustment costs are present. This is not a trivial result since a firm's profit maximization problem becomes dynamic when labor is subject to costly adjustments. Intuitively, this is because labor adjustment costs depend on the level of labor in the previous period. If these adjustment costs take a quadratic form however, it is possible to "correct" our initial estimates for markdowns. When we apply these correction terms to our estimates, we obtain measures for markdowns that are only reflective of monopsony forces and not of labor adjustment costs. In the end, we find that these correction terms are quantitatively small.

The proposition below shows that labor adjustment costs can also drive a wedge between marginal revenue products of labor and wages. Nevertheless, we can identify the "monopsony" component whenever these adjustment costs take a quadratic form.

PROPOSITION 1. Let $\mathbf{z}$ denote a firm's set of stochastic state variables and suppose revenue, labor adjustment cost and wage schedule functions are differentiable. Then, a firm's wedge between its MRPL and wage satisfies:

$$\frac{R'(\ell^*)}{w(\ell^*)} = \left(\varepsilon_S^{-1} + 1\right) + \mathcal{A}(\ell^*, \ell_{-1})$$

where $\mathcal{A}(\ell^*, \ell_{-1})$ equals zero whenever labor adjustment costs are absent. If, in addition, a firm is subject to convex labor adjustment costs of the form $\Phi(\ell, \ell_{-1}) = \frac{\gamma}{2}\ell\left(\frac{\ell-\ell_{-1}}{\ell_{-1}}\right)^2$ for $\gamma \geq 0$ and it discounts future profits at the rate $\beta \in [0, 1]$, then a firm's monopsony power can be characterized as:

$$\varepsilon_S^{-1} + 1 = \frac{\frac{R'(\ell^*)}{w(\ell^*)} - \gamma \cdot \left(g_\ell(1+g_\ell) - \beta\mathbb{E}_{\mathbf{z}'}\left[g_{\ell'}(1+g_{\ell'})(1+g_{\text{sw}'})|\mathbf{z}\right]\right)}{1 + \frac{\gamma}{2}g_\ell^2} \tag{13}$$

where g_ℓ, $g_{\ell'}$ and $g_{\text{sw}'}$ denote current and future labor growth, and future wage bill growth, respectively.

Proof. We will consider environments in which revenue, labor adjustment costs and wage schedules are continuously differentiable (at least in labor). Furthermore, we will restrict our attention to convex adjustment costs in labor, but we do allow for dynamic considerations (i.e., adjustment costs in labor are allowed to depend on the stock of labor in the previous period; denoted by ℓ_{-1}). Then, consider a firm's *dynamic* profit maximization problem:

$$v(\ell_{-1};\mathbf{z}) = \max_{\ell \geq 0} R(\ell;\mathbf{z}) - w(\ell)\cdot\ell - w(\ell)\cdot\Phi(\ell,\ell_{-1}) + \beta\cdot\mathbb{E}_{\mathbf{z}'}\left[v(\ell;\mathbf{z}')|\mathbf{z}\right] \tag{14}$$

where $\Phi(\ell,\ell_{-1})$ denotes a firm's adjustment cost (in real terms) whenever it wants to change its stock of labor to $\ell \neq \ell_{-1}$ and $\beta \in [0,1]$ is its discount factor. We will assume that the adjustment cost function is homogeneous of degree one and continuously differentiable in both arguments. Furthermore, we have that $\Phi(\ell,\ell_{-1}) > 0$ for $\ell \neq \ell_{-1}$ and zero otherwise. Similar to before, we denote the revenue function by $R(\ell;\mathbf{z}) \equiv \text{rev}(\ell;\mathbf{X}^*_{-\ell}(\ell),\mathbf{z})$ where $\mathbf{z}$ denotes a firm's (possibly stochastic) state variable, e.g. productivity. Given this setup, a firm's optimal choice is characterized by its first order condition:

$$\begin{aligned}
R'(\ell) &= w'(\ell)\ell + w(\ell) + w(\ell)\cdot\Phi_1(\ell,\ell_{-1}) + w'(\ell)\cdot\Phi(\ell,\ell_{-1}) - \beta\cdot\mathbb{E}_{\mathbf{z}'}\left[v'(\ell)|\mathbf{z}\right] \\
&= w'(\ell)\ell + w(\ell) + w(\ell)\cdot\Phi_1(\ell,\ell_{-1}) + w'(\ell)\cdot\Phi(\ell,\ell_{-1}) + \beta\cdot\mathbb{E}_{\mathbf{z}'}\left[\Phi_2(\ell',\ell)w(\ell')|\mathbf{z}\right]
\end{aligned}$$

where we applied the envelope theorem in the last equality. This can be rearranged to end up with an expression for a firm's markdown:

$$\begin{aligned}
\nu &\equiv \frac{R'(\ell)}{w(\ell)} \\
&= \varepsilon_S^{-1} + 1 + \Phi_1(\ell,\ell_{-1}) + \frac{\Phi(\ell,\ell_{-1})}{\ell}\varepsilon_S^{-1} + \beta\cdot\mathbb{E}_{\mathbf{z}'}\left[\Phi_2(\ell',\ell)\frac{w(\ell')}{w(\ell)}\middle|\mathbf{z}\right] \\
&\equiv \varepsilon_S^{-1} + 1 + \mathcal{A}(\ell,\ell_{-1})
\end{aligned} \tag{15}$$

where $\mathcal{A}(\ell,\ell_{-1})$ reflects a firm's expected continuation value of adjustment cost relative to its wage level.

Without specifying the shape of the real labor adjustment cost function further, it is hard to assess the magnitude of the bias (i.e., $\mathcal{A}(\ell,\ell_{-1})$) that we are dealing with. For illustrative purposes, we use a commonly specified labor adjustment cost function $\Phi(\ell,\ell_{-1}) = \frac{\gamma}{2}\ell\left(\frac{\ell-\ell_{-1}}{\ell_{-1}}\right)^2$ (Hall, 2004; Cooper, Haltiwanger and Willis, 2007). Given this specification and after some algebra, we can simplify equation (15) to:

$$\nu = \left(1 + \frac{\gamma}{2} g_\ell^2\right)(\varepsilon_S^{-1} + 1) + \gamma g_\ell (1 + g_\ell) - \beta\gamma \mathbb{E}_{\mathbf{z}'}\left[g_{\ell'}(1 + g_{\ell'})(1 + g_{\mathtt{sw}'})|\mathbf{z}\right] \tag{16}$$

where we defined labor growth rates as $g_\ell = \frac{\ell - \ell_{-1}}{\ell_{-1}}$ and $g_{\ell'} = \frac{\ell' - \ell}{\ell}$, respectively. Furthermore, we have a firm's future growth rate in its wage bill which equals $g_{\mathtt{sw}'} = \frac{w(\ell')\ell'}{w(\ell)\ell} - 1$. If our estimates for markdowns do not only reflect monopsony, then we can obtain "unbiased" estimates for labor market power (i.e., percentage wedges between marginal revenue products of labor and wages corrected for labor adjustment costs as reflected by $\varepsilon_S^{-1} + 1$ alone) by using equation (16) instead. To do so, we solve for $\varepsilon_S^{-1} + 1$ and obtain:

$$\varepsilon_S^{-1} + 1 = \frac{\frac{R'(\ell^*)}{w(\ell^*)} - \gamma \cdot \left(g_\ell(1 + g_\ell) - \beta \mathbb{E}_{\mathbf{z}'}\left[g_{\ell'}(1 + g_{\ell'})(1 + g_{\mathtt{sw}'})|\mathbf{z}\right]\right)}{1 + \frac{\gamma}{2} g_\ell^2}$$

which is exactly what we wanted to show. □

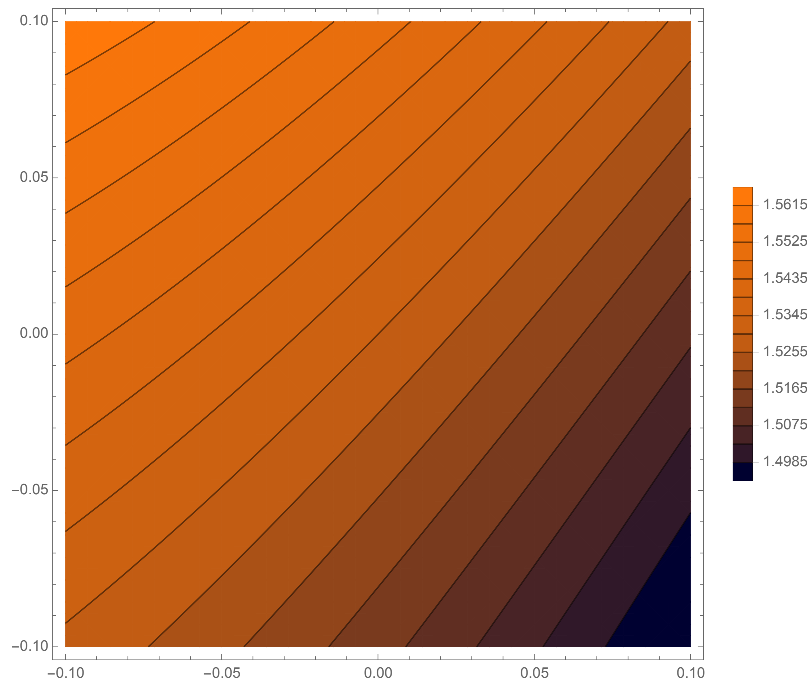
We apply the above proposition by substituting out expected growth rates with their realized counterparts. In particular, our estimates for markdowns $\widehat{\nu}$ can be adjusted for labor adjustment costs as follows:

$$\frac{\varepsilon_S + 1}{\varepsilon_S} = \frac{\widehat{\nu} - \gamma \cdot \left[g_\ell(1 + g_\ell) - \beta g_{\ell'}(1 + g_{\ell'})(1 + g_{\mathtt{sw}'})\right]}{1 + \frac{\gamma}{2} g_\ell^2} \tag{17}$$

The proposition above shows that the wedge between a firm's MRPL and the wage it pays its workers no longer only reflects monopsony power in the presence of convex labor adjustment costs. In other words, labor adjustment costs can also drive a wedge between MRPL and wages. Hence, one could be worried that our measured markdowns do not only reflect monopsony forces but also capture labor adjustment costs.

If labor adjustment costs are quadratic, then the second part of the above proposition demonstrates that we can correct our measured markdowns such that they only reflect forces of monopsony power. This can be done if we observe a plant's growth in labor and its wage bill, and know the parameters β and γ. Obviously, quadratic adjustment costs are not without loss of generality, but it is a specification that is often employed (see Hall,

Figure 7: Corrections to markdowns from convex labor adjustment costs are quantitatively small.



Wage bill growth $g_{\mathrm{sw}'}$ is set at 2.19 percent which is the average level of wage bill growth in U.S. manufacturing from 1987–2017 (BEA GDP by Industry accounts). Horizontal and vertical axes denote current and future labor growth g_{ℓ} and $g_{\ell'}$, respectively. The adjustment cost parameter γ is set at 0.185 (Hall, 2004).

2004; Cooper, Haltiwanger and Willis, 2007). Another advantage of this functional form is that is governed by only one parameter. Obviously, we are back to our baseline in the absence of adjustment costs when $\gamma = 0$ holds as can be seen from equation (13).

To be conservative, we choose the highest estimate for γ in Hall (2004) that is estimated with reasonable precision. This results in $\gamma = 0.185$.[11] In figure 7, we set $\beta = 1$ and show that our measured markdowns only have to be adjusted by a maximum of 3.15 percent for a broad range of labor growth rates (varying from -10 to 10 percent). We conclude that labor adjustment costs only play a minor quantitative role and, hence, our baseline estimates must reflect labor market power.

[11] See the estimation results in table II of Hall (2004).

O.5 Benefits

O.5.1 Measures of compensation

In our baseline estimation procedure, we use a plant's total wage bill (or "payroll") as its total variable expenditure on labor. Following the instructions of form MA-10000, payroll is an overall measure of wages and salaries paid to a plant's employee(s). An employee is defined according to Internal Revenue Service Form 941, Employer's Quarterly Federal Tax Return. This includes:

- All persons on paid sick leave, paid holidays, and paid vacation during these pay periods
- Officers at this establishment, if a corporation
- Spread on stock options that are taxable to employees as wages

An employer's wage bill is defined as its payroll before deductions excluding an employer's cost for fringe benefits. In particular, it includes:

- Employee's Social Security contributions, withholding taxes, group insurance premiums, union dues, and savings bonds
- In gross earnings: commissions, dismissal pay, paid bonuses, employee contributions to pension plans such as 401(k), vacation and sick leave pay, and the cash equivalent of compensation paid in kind
- Spread on stock options that are taxable to employees as wages
- Salaries of officers of this establishment, if a corporation
- Paid holiday, personal, funeral, jury duty, military and family leave
- Non-production bonuses
  - Cash profit-sharing
  - Employee recognition
  - End-of-year
  - Holiday
  - Payment in lieu of benefits - Referral
  - Other

By construction, the wage bill does not include benefits. Fortunately, the ASM/CM does include a measure of these benefits from 2002 onward. Benefits cover health insurance, pension plans and other employer paid benefits. The latter includes legally-required benefits (e.g., Social Security, workers' compensation insurance, unemployment tax, state disability insurance programs, Medicare), benefits for life insurance, "quality of life" benefits (e.g., childcare assistance, subsidized commuting, etc.), employer contributions to pre-tax benefit accounts (e.g., health savings accounts), education assistance, and other benefits. In the end, our results on markdowns are not qualitatively changed whenever we use a measure for labor that includes benefits.

O.5.2 Understanding markdowns with benefits

In one of our robustness exercises, we calculated micro-level markdowns whenever benefits were also included as a part of workers' compensation. We saw from table 5 that median markdowns at the industry group level slightly declined relative to our baseline results from table 1.

In this section, we verify that the differences between our baseline estimates for markdowns and those with benefits included can be rationalized by the fraction of benefits in total compensation.

Given that benefits are not included in our baseline estimates, we expect that they are biased upwards. This is intuitive since we are including only wage payments in the denominator of the markdown. As a result, the bias of our baseline estimates should increase more for those plants whose compensation to workers relies more on benefits. We measure the latter by the "benefit fraction", i.e. total benefits relative to the sum of total benefits and wage payments.

Our hypothesis is confirmed by table 10. Our baseline estimates, but more importantly the *difference* between our baseline estimates and those including benefits, are increasing in the benefit fraction. Our conclusions are not affected much whenever we take absolute differences instead. This is as expected since our baseline estimates are larger than those estimates including benefits for the overwhelming fraction of our sample anyway.

However, it could also be argued that the sign of the benefit fraction coefficient may at first

Table 10: The fraction of benefits in total compensation accounts for the difference between baseline and markdowns with benefits.†

Dependent variable	ν_{it}	$\nu_{it} - \nu_{it}^{\text{benefit}}$	$\lvert\nu_{it} - \nu_{it}^{\text{benefit}}\rvert$
BENEFIT FRACTION	1.682 (0.3153)	1.299 (0.2057)	1.0360 (0.1642)
Fixed effects			
INDUSTRY	Y	Y	Y
STATE	Y	Y	Y
YEAR	Y	Y	Y
Weights	`empwt`	`empwt`	`empwt`
Observations	$4.02 \cdot 10^5$	$4.02 \cdot 10^5$	$4.02 \cdot 10^5$

†Markdowns are estimated under the assumption of a translog specification for gross output. Each industry group in manufacturing corresponds to the manufacturing categorization of the BEA which approximately follows a 3-digit NAICS specification. Baseline markdowns are denoted by ν whereas markdowns with benefits are denoted by ν^{benefit}. A plant's benefit fraction is defined as benefit payments divided by the sum of wage and benefit payments to workers. All regressions contain size and age controls at the plant level. Furthermore, all regressions include average earnings (i.e., total wage bill divided by employment count) as a control. Standard errors are clustered at the industry group level and denoted between parentheses. Regressions are weighted by the product of employment count and ASM sampling weights. Source: Authors' calculations from ASM/CM data in 2002–2014.

be surprising, as one might associate larger benefit shares of compensation with stronger employee bargaining power and thus expect a *lower* markdown. However, because we control for plant-level average earnings, the results in table 10 show how markdown estimates change as the benefit share changes, holding average earnings constant. To the extent that benefit shares are higher in lower-wage plants, on average, our regressions control for this mechanical relationship.

Finally, note that this sample is smaller than our base sample since we can estimate markdowns with benefits from 2002 onward only.

O.6 Critique by Bond *et al.* (2021)

O.6.1 Deflated revenues

Unfortunately, most firm-level data sets do not have physical output available. As an alternative, physical output is typically approximated by deflating revenues with some industry-level deflator. While it could be argued that revenues are more easily comparable across firms, it does not align with the theory behind production function estimation. In fact, Klette and Griliches (1996) show that estimated production function coefficients in an imperfectly competitive environment with price heterogeneity are downward biased whenever physical output is approximated with deflated revenues. This immediately implies that markups are also downward biased under the production approach.

Bond et al. (2021) demonstrate that the problem is even more severe: using deflated revenues does not only induce a downward bias, but it must result in ratio estimators, such as the one we employ in equation (3), to be equal to unity. To see why this is the case, consider the analog of (3) using revenue elasticities:

$$\begin{aligned}\frac{\theta_{it}^{k',\text{rev}}}{\alpha_{it}^{k'}} &= \frac{\theta_{it}^{k',\text{Q}}}{\alpha_{it}^{k'}} \cdot \left(1 + \frac{\mathrm{d}P(Q_{it})}{\mathrm{d}Q_{it}}\frac{Q_{it}}{P_{it}}\right) \\ &\equiv \frac{\theta_{it}^{k',\text{Q}}}{\alpha_{it}^{k'}} \cdot (1 + \varepsilon_{P,Q,it}) \\ &\equiv \mu_{it} \cdot (1 + \varepsilon_{P,Q,it}) \\ &= 1 \end{aligned} \tag{18}$$

where the last equality follows directly from Lerner's monopoly pricing rule, i.e. $\mu_{it} = (1+\varepsilon_{P,Q,it})^{-1}$. Based on this result, Bond et al. (2021) conclude that it is basically hopeless to retrieve markups through the production approach whenever data on physical output is not available. Estimates of markups using deflated revenues that do not equal unity then indicate that assumptions **I** – **VI** and/or 1 – 5 must be violated. While this is an issue for the estimation of mark*ups*, we argue that it does not pose any problems when estimating mark*downs*. This can be shown most clearly through the following proposition:

PROPOSITION 2. Let $\theta_{it}^{j,Q} \equiv \frac{\partial \ln(Q_{it})}{\partial \ln(X_j)}$ and $\theta_{it}^{j,\text{rev}} \equiv \frac{\partial \ln(P(Q_{it})\cdot Q_{it})}{\partial \ln(X_j)}$ denote the output and revenue elasticities with respect to some differentiable input j, respectively. Furthermore, let $\alpha^j_{it} \equiv \frac{V^j_{it} \cdot X^j_{it}}{P_{it}Q_{it}}$ denote the revenue share of input j. Then, we have:

$$\frac{\theta^{\ell,\text{rev}}_{it}}{\alpha^{\ell}_{it}} \Big/ \frac{\theta^{M,\text{rev}}_{it}}{\alpha^{M}_{it}} = \frac{\theta^{\ell,Q}_{it}}{\alpha^{\ell}_{it}} \Big/ \frac{\theta^{M,Q}_{it}}{\alpha^{M}_{it}} \tag{19}$$

That is, it is sufficient to estimate *revenue* elasticities in order to construct markdowns on labor inputs.

Proof. We drop firm and time subscripts to ease notation. To prove the proposition, it is sufficient to show that $\frac{\theta^{\text{rev}}_{\ell}}{\theta^{\text{rev}}_{M}} = \frac{\theta^{Q}_{\ell}}{\theta^{Q}_{M}}$ is true. To do so, note that we have:

$$\begin{aligned}
\theta^{\text{rev}}_j &\equiv \frac{\partial\left[P(Q)\cdot Q\right]}{\partial X_j} \cdot \frac{X_j}{P(Q)Q} \\
&= \left[P'(Q)Q + P(Q)\right] \cdot \frac{\partial Q}{\partial X_j} \cdot \frac{X_j}{P(Q)Q}
\end{aligned}$$

Then, with some abuse of notation, it immediately follows that:

$$\begin{aligned}
\frac{\theta^{\text{rev}}_{\ell}}{\theta^{\text{rev}}_{M}} &= \frac{\left[P'(Q)Q + P(Q)\right] \cdot \frac{\partial Q}{\partial \ell} \cdot \frac{\ell}{P(Q)Q}}{\left[P'(Q)Q + P(Q)\right] \cdot \frac{\partial Q}{\partial M} \cdot \frac{M}{P(Q)Q}} \\
&= \frac{\frac{\partial Q}{\partial \ell} \cdot \frac{\ell}{Q}}{\frac{\partial Q}{\partial M} \cdot \frac{M}{Q}} \\
&\equiv \frac{\theta^{Q}_{\ell}}{\theta^{Q}_{M}}
\end{aligned}$$

which is exactly what we wanted to show. □

The proposition shows that the bias occurring from proxying physical output with deflated revenues cancels out since it appears in both the numerator and denominator (i.e., markup) of the markdown expression in a multiplicative manner. Thus, the lack of data availability on physical output would only affect our results if we were interested in estimating markups separately. As a result, the main point of critique by Bond et al. (2021) does not apply to markdowns.

O.6.2 Demand shifters

Another point of critique on the production approach in Bond et al. (2021) revolves around the assumption of inputs being solely used for the production of output (i.e., assumption **VI**). However, in reality, some inputs can also be used for activities to shift demand such as marketing and/or advertising. When inputs are also used to shift (or "influence" using the terminology of Bond et al., 2021) demand, then the markup formula in equation (3) is no longer correct.

To see this, consider an environment in which each input X_{it}^{k} can be used for either the production of output $X_{it}^{k,Q}$ or to shift demand $X_{it}^{k,D}$. Then, assume that a firm's inverse demand function is of the following form:

$$P(Q_{it}, D_{it}) \text{ s.t. } D_{it} = \mathscr{D}(\mathbf{X}_{it}^{D}) \tag{20}$$

where all functions are differentiable in their arguments and $\mathbf{X}_{it}^{D} = (X_{it}^{1,D}, \ldots, X_{it}^{K,D})'$ are those parts of each input that are used for shifting demand. Hence, by construction, we have $\mathbf{X}_{it} = \mathbf{X}_{it}^{D} + \mathbf{X}_{it}^{Q}$.

Let k' be some flexible input, then Hall's (1988) formula only holds for that part dedicated to production, i.e. we have:

$$\mu_{it} = \frac{\theta_{it}^{k',Q}}{\alpha_{it}^{k',Q}} \tag{21}$$

Bond et al. (2021) argue that, for most data sets, we can only observe $X_{it}^{k'}$ and its expenditure but not its components $X_{it}^{k',Q}$ and $X_{it}^{k',D}$ separately. If one would apply formula (3) to $X_{it}^{k'}$ rather than $X_{it}^{k',Q}$, we would obtain a biased estimate of the markup:

$$\mu_{it} \cdot \frac{\varepsilon_{X^{k',Q},X^{k'}}}{1 + \frac{X_{it}^{k',D}}{X_{it}^{k',Q}}} \tag{22}$$

where $\varepsilon_{X^{k',Q},X^{k'}}$ denotes by what percentage the usage of input k' for production purposes increases if total expenditure on input k' is raised by one percent. If assumption **VI** holds, then we must have $\varepsilon_{X^{k',Q},X^{k'}} = 1$ and $X_{it}^{k',D} = 0$.

In our baseline estimates, we adopt the definition for material inputs as used by the Census Bureau, which includes contract work. It is not unlikely that some of this contracted labor is used for activities such as marketing; even though it is less likely for manufacturers. However, our results are robust to using an alternative definition for materials as proposed by Kehrig (2015) in which contract work is disregarded and information on inventories for materials is used instead. Under this definition, material inputs only consist of materials and parts. Its exact definition can be taken from section 16A1 of form MA-10000 which we documented below for convenience.

Table 11: Description of what constitutes "material inputs" from section 16A1 in form MA-10000 of ASM.

MATERIALS		PARTS	CONTAINERS	SUPPLIES	
Lumber	Cement	Pumps	Pails	Bolts, screw and nuts	Cleaning supplies
Plywood	Clay	Wheels	Drums and barrels	Drills, tools, dies, jigs and fixtures which are charged to current accounts	Stationary and office supplies
Paper	Glass	Bearings	Tubes	Welding rods, electrodes and acetylene	First aid and safety supplies
Resins	Steel sheet	Engines	Boxes and bags	Lubricating oils	Dunnage
Sulfuric acid	Steel scrap	Gears	Crates		Water
Alcohols	Copper rods	Motors			
Rubber	Iron castings	Hardware			
Coking coal	Metal stampings	Compressors			
Crude petroleum	Wire				

If we impose the assumption that none of the expenditures on materials and parts are used to shift demand, which we believe to be reasonable given the table above, then there are no issues with the denominator of our markdown definition. On the other hand, the numerator of our markdown definition consists of the "labor markup." If some fraction of total labor is used to shift demand, then our markdown estimates are biased. This is formally shown in the proposition below.

PROPOSITION 3. Let there exist some input $k' \neq \ell$ that satisfies assumptions **I – VI**. If labor ℓ does not satisfy assumption **VI** and firms possess monopsony power but cannot discriminate between different workers, then the ratio estimator for the markdown in equation (3) retrieves:

$$\hat{\nu} = \left[\varepsilon_S^{-1}\frac{\ell^Q}{\ell} + 1\right] \cdot \frac{\varepsilon_{\ell^Q,\ell}}{1 + \frac{\ell^D}{\ell^Q}} \tag{23}$$

where total labor $\ell \equiv \ell^Q + \ell^D$ is the sum of labor used for production and shifting demand, respectively. Furthermore, $\varepsilon_{\ell_Q,\ell}$ denotes the elasticity of labor used for output with respect to total labor. If labor ℓ does not satisfy assumption **VI** but firms are allowed to discriminate between different workers, then the ratio estimator for the markdown in equation (3) retrieves instead:

$$\hat{\nu} = \nu_{\ell^Q} \cdot \frac{\varepsilon_{\ell^Q,\ell}}{1 + \frac{\ell^D}{\ell^Q}} \tag{24}$$

where ν_{ℓ^Q} denotes the markdown a firm charges on its production workers.

Proof. We follow the proof of Bond et al. (2021) closely. For notational convenience, we drop firm and time subscripts. A firm's profit maximization problem reads as:

$$\max_{Q,D\geq 0} P(Q,D) \cdot Q - C_Q(Q) - C_D(D) \tag{25}$$

where $C_D(D)$ denotes the cost of reaching a level D for the demand shifter. This results in the two first order conditions:

$$(1 + \varepsilon_{P,Q})^{-1} = \mu \tag{26}$$

$$\varepsilon_{P,D} = \frac{\frac{\mathrm{d}C_D(D)}{\mathrm{d}D} \cdot D}{P(Q)Q} \tag{27}$$

Assuming that a firm has monopsony power, but faces a residual labor supply curve for its total stock of workers only, the first order conditions for ℓ^Q and ℓ^D for the cost minimization problem give us:

$$\left[\varepsilon_S^{-1} \frac{\ell^Q}{\ell} + 1\right] \cdot \mu = \frac{\varepsilon_{Q,\ell^Q}}{\alpha_{\ell^Q}} \tag{28}$$

$$\left[\varepsilon_S^{-1} \frac{\ell^D}{\ell} + 1\right] = \frac{\frac{\mathrm{d}C_D(D)}{\mathrm{d}D} \cdot D}{P(Q)Q} \cdot \frac{\varepsilon_{D,\ell^D}}{\alpha_{\ell^D}} \tag{29}$$

where we defined $\varepsilon_{D,\ell^D} = \frac{\partial \mathscr{D}(\mathbf{X}^D)}{\partial \ell^D} \frac{\ell^D}{\mathscr{D}(\mathbf{X}^D)}$. Then, we get:

$$\begin{aligned} \alpha_\ell &= \alpha_{\ell^Q} + \alpha_{\ell^D} \\ &= (1 + \varepsilon_{P,Q})\varepsilon_{Q,\ell^Q} \left[\varepsilon_S^{-1} \frac{\ell^Q}{\ell} + 1\right]^{-1} + \varepsilon_{P,D}\varepsilon_{D,\ell^D} \left[\varepsilon_S^{-1} \frac{\ell^D}{\ell} + 1\right]^{-1} \end{aligned} \tag{30}$$

where $\varepsilon_{Q,\ell^Q} = \theta^{\ell^Q,Q}$ is the output elasticity with respect to labor for production purposes. Similarly, we define ε_{D,ℓ^D} as the demand shifter elasticity with respect to labor for "influencing" purposes. Then, the numerator for our markdown expression in equation (3) using *total* labor ℓ is equal to:

$$
\begin{aligned}
\frac{\theta^{\ell,Q}}{\alpha_\ell} &= \frac{\varepsilon_{Q,\ell^Q} \cdot \varepsilon_{\ell^Q,\ell}}{\alpha_\ell} \\
&= \frac{\varepsilon_{Q,\ell^Q} \cdot \varepsilon_{\ell^Q,\ell}}{(1+\varepsilon_{P,Q})\varepsilon_{Q,\ell^Q}\left[\varepsilon_S^{-1}\frac{\ell^Q}{\ell}+1\right]^{-1} + \varepsilon_{P,D}\varepsilon_{D,\ell^D}\left[\varepsilon_S^{-1}\frac{\ell^D}{\ell}+1\right]^{-1}} \\
&= \frac{\varepsilon_{\ell^Q,\ell}}{\mu^{-1}\left[\varepsilon_S^{-1}\frac{\ell^Q}{\ell}+1\right]^{-1} + \frac{\varepsilon_{P,D}\varepsilon_{D,\ell^D}}{\varepsilon_{Q,\ell^Q}}\left[\varepsilon_S^{-1}\frac{\ell^D}{\ell}+1\right]^{-1}} \\
&= \mu \cdot \left[\varepsilon_S^{-1}\frac{\ell^Q}{\ell}+1\right] \cdot \frac{\varepsilon_{\ell^Q,\ell}}{1+\frac{\alpha_{\ell^D}}{\alpha_{\ell^Q}}} \\
&= \mu \cdot \left[\varepsilon_S^{-1}\frac{\ell^Q}{\ell}+1\right] \cdot \frac{\varepsilon_{\ell^Q,\ell}}{1+\frac{\ell^D}{\ell^Q}}
\end{aligned} \tag{31}
$$

If there exists some input $k' \neq \ell$ that satisfies assumptions **I – VI**, then we get an unbiased estimate for markups. As a result, we must have:

$$
\begin{aligned}
\hat{\nu} &= \frac{\theta^{\ell,Q}}{\alpha_\ell}\left(\frac{\theta^{k',Q}}{\alpha_{k'}}\right)^{-1} \\
&= \left[\varepsilon_S^{-1}\frac{\ell^Q}{\ell}+1\right] \cdot \frac{\varepsilon_{\ell^Q,\ell}}{1+\frac{\ell^D}{\ell^Q}}
\end{aligned} \tag{32}
$$

which covers the case whenever a firm faces a residual labor supply curve as function of only its *total* stock of workers. This is similar to the case in Bond et al. (2021) in which it is assumed that production and non-production workers are compensated at an identical wage rate. The derivation for the case in which a firm faces different residual labor supply curves for its production and non-production workers is almost identical. Note that a firm can then charge different markdowns for different workers. We only need to replace $\left[\varepsilon_S^{-1}\frac{\ell^Q}{\ell}+1\right]$ by $\left[\varepsilon_{S,\ell^Q}^{-1}+1\right] \equiv \nu_{\ell^Q}$ and $\left[\varepsilon_S^{-1}\frac{\ell^D}{\ell}+1\right]$ by $\left[\varepsilon_{S,\ell^D}^{-1}+1\right] \equiv \nu_{\ell^D}$. Then, expression (62) becomes:

$$
\hat{\nu} = \nu_{\ell^Q} \cdot \frac{\varepsilon_{\ell^Q,\ell}}{1+\frac{\ell^D}{\ell^Q}} \tag{33}
$$

which is exactly what we wanted to show. $\square$

If labor was used for production only, then we must have $\varepsilon_{\ell_Q,\ell} = 1$, $\ell = \ell^Q$ and $\ell^D = 0$ and our markdown estimates would feature no bias(es) since $\hat{\nu} = \nu$. Bond et al. (2021) point out that bias-free estimates can be obtained if labor inputs used for production and "influencing demand" were observed separately. Even though our baseline estimates are somewhat subject to this point of critique in Bond et al. (2021), our markdown results for production and non-production workers (which are estimated separately) corroborate our baseline results. It supports the observation it is unlikely that manufacturers spend a large fraction of their workforce for non-production purposes (see Dey, Houseman and Polivka, 2012). As a result, it is reasonable in our setting to have $\varepsilon_{\ell^Q,\ell} \approx 1$ and $\frac{\ell^D}{\ell^Q} \approx 0$.

O.6.3 Scalar unobservable assumption

The last point of critique in Bond et al. (2021) relates to the scalar unobservable assumption of the proxy variable methodology. Bond et al. (2021) argue that this assumption cannot be satisfied whenever firms possess market power. Whenever this is the case, the econometrician also needs to observe a firm's marginal cost of production. This point is formally illustrated below through a simple example.

PROPOSITION 4. If a monopolist is faced with some differentiable, downward-sloping demand curve and is endowed with a Cobb-Douglas production technology, then there exist parameters $\boldsymbol{\alpha} = (\alpha_0, \alpha_\omega, \alpha_k, \alpha'_{\mathbf{p}}, \alpha_{\mathrm{MC}})'$ such that its optimal input demand schedule for materials under market power satisfies:

$$\begin{aligned} m_{it}(k_{it}, \omega_{it}, mc^*_{it}) &= \alpha_0 + \alpha_\omega \cdot \omega_{it} + \alpha_k \cdot k_{it} + \alpha'_{\mathbf{p}}\mathbf{p}_t + \alpha_{\mathrm{MC}}(p^*_{it} - \ln(\mu^*_{it})) \\ &= \alpha_0 + \alpha_\omega \cdot \omega_{it} + \alpha_k \cdot k_{it} + \alpha'_{\mathbf{p}}\mathbf{p}_t + \alpha_{\mathrm{MC}} \cdot mc^*_{it} \end{aligned} \tag{34}$$

That is, the optimal input demand schedule for materials depends on idiosyncratic productivity and a firm's marginal cost of production.

Proof. The monopolist's profit maximization problem becomes:

$$\max_{K_{it}, L_{it}, M_{it} \geq 0} P_t(Q_{it})Q_{it} - C(Q_{it}) \;\text{ s.t. }\; Q_{it} = \exp(\omega_{it}) K_{it}^{\beta_K} L_{it}^{\beta_L} M_{it}^{\beta_M} \tag{35}$$

It is easy to show that the firm's optimal input demand schedule for materials is:

$$M_{it} = \left(\frac{W_t}{\beta_L}\right)^{\frac{\beta_L}{\beta_L+\beta_M}} \cdot \left(\frac{P_t^M}{\beta_M}\right)^{-\frac{\beta_L}{\beta_L+\beta_M}} \cdot \left(\frac{Q_{it}}{\exp(\omega_{it})K_{it}^{\beta_K}}\right)^{\frac{1}{\beta_L+\beta_M}} \tag{36}$$

which leads to the Cobb-Douglas cost function (conditional on a given level of output and capital):

$$C(Q_{it}) = (\beta_L + \beta_M) \cdot \left(\frac{W_t}{\beta_L}\right)^{\frac{\beta_L}{\beta_L+\beta_M}} \left(\frac{P_t^M}{\beta_M}\right)^{\frac{\beta_M}{\beta_L+\beta_M}} \cdot \left(\frac{Q_{it}}{\exp(\omega_{it})K_{it}^{\beta_K}}\right)^{\frac{1}{\beta_L+\beta_M}} \tag{37}$$

Following the Lerner index pricing formula, a firm's optimal output is pinned down by:

$$\begin{aligned} \mu_{it}^* &\equiv \frac{\varepsilon_D(Q_{it}^*)}{\varepsilon_D(Q_{it}^*) - 1} \\ &= \frac{P_t(Q_{it}^*)}{C'(Q_{it}^*)} \end{aligned} \tag{38}$$

which, using (37) for the marginal cost of production, we can rearrange as:

$$\begin{aligned} C'(Q_{it}) &= \left(\frac{W_t}{\beta_L}\right)^{\frac{\beta_L}{\beta_L+\beta_M}} \left(\frac{P_t^M}{\beta_M}\right)^{\frac{\beta_M}{\beta_L+\beta_M}} \cdot \left(\frac{1}{\exp(\omega_{it})K_{it}^{\beta_K}}\right)^{\frac{1}{\beta_L+\beta_M}} Q_{it}^{\frac{1-\beta_L-\beta_M}{\beta_L+\beta_M}} \\ &= P_t(Q_{it})\mu_{it}^{-1} \end{aligned} \tag{39}$$

Using (39), we solve for the optimal level of output Q_{it}^*:

$$Q_{it}^* = \frac{P_{it}^*}{\mu_{it}^*} \cdot \left(\frac{W_t}{\beta_L}\right)^{-\frac{\beta_L}{1-\beta_L-\beta_M}} \left(\frac{P_t^M}{\beta_M}\right)^{-\frac{\beta_M}{1-\beta_L-\beta_M}} \left(\exp(\omega_{it})K_{it}^{\beta_K}\right)^{\frac{1}{1-\beta_L-\beta_M}} \tag{40}$$

Plugging (40) into (36) and taking natural logs, there exist values for α_0, α_ω, α_k, $\alpha_{\mathbf{p}}$ and α_{MC} such that the optimal input demand schedule for materials *under market power* satisfies:

$$\begin{aligned} m_{it}(k_{it}, \omega_{it}, mc_{it}^*) &= \alpha_0 + \alpha_\omega \cdot \omega_{it} + \alpha_k \cdot k_{it} + \alpha_{\mathbf{p}}' \mathbf{p}_t + \alpha_{\text{MC}}(p_{it}^* - \ln(\mu_{it}^*)) \\ &= \alpha_0 + \alpha_\omega \cdot \omega_{it} + \alpha_k \cdot k_{it} + \alpha_{\mathbf{p}}' \mathbf{p}_t + \alpha_{\text{MC}} \cdot mc_{it}^* \end{aligned} \tag{41}$$

As a result, a firm's input demand schedule for materials becomes a direct function of its marginal cost of production whenever it has pricing power. □

The proposition illustrates that the econometrician needs to observe both firm-level productivity and its marginal cost of production, contradicting the scalar unobservable assumption. As a result, Bond et al. (2021) argue that other estimators, in particular those that do *not* rely on the scalar unobservable assumption, should be used in order to estimate production function parameters. In particular, they refer to the estimator from Blundell and Bond (2000).

In the following, we evaluate a set of proxy variable estimators and two estimators that do not rely on the scalar unobservable assumption. Regarding the latter two, we use the dynamic panel IV estimator from Blundell and Bond (2000) and the estimator from Hu, Huang and Sasaki (2020). To evaluate the performance of all estimators, we apply them to simulated data. In particular, we adopt the third data generating process (DGP) from Ackerberg, Caves and Frazer (2015) (or "ACF – DGP3") which is least favorable to the family of proxy variable estimators. The latter paper only allows for gross output specifications in which materials enter in a Leontief fashion. We replicate ACF to the letter, but we also look at the performance of production function estimators whenever gross output is also Cobb-Douglas in materials. This requires us to specify a process for material prices. We follow ACF and assume it follows an AR(1) process in natural logs, i.e. $\ln(P_t^M) = \varphi_M \cdot \ln(P_{t-1}^M) + \varepsilon_{it}^M$. Online Appendix O.6.4 contains more details on what changes whenever we allow material inputs to enter production in a Cobb-Douglas fashion. We use the same parameter values as ACF unless otherwise specified.

ACF – DGP3: FULL COBB-DOUGLAS PRODUCTION. We start out with the case in which material inputs enter the production function in a non-Leontief fashion, i.e. $Y_{it} = \exp(\omega_{it}) K_{it}^{\beta_k} L_{it}^{\beta_\ell} M_{it}^{\beta_m}$ with $\beta_m \in (0, 1)$. Similar to the wage process in ACF – DGP3, we assume that prices for material inputs are idiosyncratic and follow an AR(1) process. This introduces two additional parameters compared to Ackerberg, Caves and Frazer (2015). We set all of the parameters in an identical fashion to the latter paper unless otherwise noted. Obviously, we cannot do this for the parameters φ_M and σ_M^2.

To solve this issue, we set $\varphi_M = 0.799$ based on evidence from Atalay (2014) and $\sigma_M^2 =$

$\sigma_W^2 = 0.1^2$.[12] Furthermore, we have to adjust the production function parameters to reflect a gross output (rather than a value added) specification. We choose $\beta_k = 0.1$, $\beta_\ell = 0.25$ and $\beta_m = 0.65$ which reflect data from the ASM/CM. To stay close to Ackerberg, Caves and Frazer (2015), we allow for optimization errors in labor by setting $\sigma_{\xi^\ell} = 0.37$. The results are not affected qualitatively by this choice though. The simulation results of the production function estimation procedure can be found in table 12.

The translog specification approximates the Cobb-Douglas production function in the best manner. Each cross and second-order term is not statistically significant (at the 5 percent level). Note that the parameters are estimated with some bias, but this is to be expected since the scalar unobservable assumption is violated. Furthermore, the production function is not of the Leontief form: Ackerberg, Caves and Frazer (2015) have pointed out that the family of proxy variable estimators then generates biased results. Nevertheless, the estimated parameters are very close to their true values. In fact, the true parameters are contained within the 95 percent confidence intervals generated with the Monte Carlo simulations. Somewhat surprisingly, the estimation results for the Cobb-Douglas specification are less precisely estimated when compared to its translog counterpart.

Our simulation results also indicate that other estimators from the proxy variable family do not perform as well. In particular, the coefficient for material inputs is always heavily underestimated. To assess the importance of the scalar unobservable assumption, we test the performance of the estimators mentioned in Blundell and Bond (2000) (DPD-IV) and, Hu, Huang and Sasaki (2020) (HHS).

As can be seen from the table below, the DPD-IV estimator from Blundell and Bond (2000) performs quite poorly; even when allowing for measurement error in output. In particular, capital coefficients are estimated to be implausibly large. This estimator is predicated upon several layers of differencing, and we suspect this approach eliminates the variation that is necessary for identification.

[12] Higher values for σ_M^2 will increase the standard errors of our estimates, but do not affect the point estimates themselves by much. All of the remaining parameters are set to their identical values in the appendix section of Ackerberg, Caves and Frazer (2015). There are two exceptions. First, we set ρ and σ_ω^2 at 0.9 and 0.2^2 instead of 0.7 and 0.3^2. We believe this reflects the U.S. data in a better fashion. Second, we leave adjustment cost parameters to be static, i.e. they do not evolve dynamically over time. However, the latter does not affect our results much and is without much loss of generality.

Table 12: Monte Carlo results with ACF – DGP3 under non-trivial Cobb-Douglas specification.† Our preferred estimator, DLW-TL, outperforms alternative estimators.

	β_k	β_ℓ	β_m	θ_ℓ/θ_m
$\boldsymbol{\beta}_0$	0.10	0.25	0.65	0.3846
DLW-TL	0.1097	0.2212	0.6231	0.2922
	(0.0381)	(0.0553)	(0.0566)	(0.4605)
	$\beta_{k\ell}$	β_{km}	$\beta_{\ell m}$	
	0.0428	-0.0156	0.0154	
	(0.0371)	(0.0386)	(0.0374)	
	β_{k^2}	β_{ℓ^2}	β_{m^2}	
	0.0237	-0.0167	0.0493	
	(0.0501)	(0.0205)	(0.0270)	
	β_k	β_ℓ	β_m	θ_ℓ/θ_m
DLW-CD	0.0394	0.1013	0.5682	0.1817
	(0.0386)	(0.01887)	(0.0453)	(0.0450)
LP-CD	0.1078	0.2214	0.1438	1.6303
	(0.0382)	(0.0074)	(0.0317)	(0.5031)
ACF-CD	0.0689	0.2219	0.1005	2.2255
	(0.0072)	(0.0075)	(0.0077)	(0.2478)
BB-CD: MA(0)	0.3538	0.2137	0.1069	1.9722
	(0.1775)	(0.0649)	(0.0678)	(29.3905)
BB-CD: MA(1)	0.2976	0.2254	0.1061	0.3362
	(0.2358)	(0.0749)	(0.0751)	(35.5410)
HHS-CD (capital only)	0.1414	0.1103	0.4587	-0.0214
	(0.5750)	(0.5617)	(0.7696)	(6.1008)
HHS-CD (capital and labor)	0.0981	0.2329	0.6675	0.1931
	(0.3302)	(0.3200)	(0.5247)	(2.4675)

†We estimate production function parameters through the two-step proxy variable estimator of De Loecker and Warzynski (2012) (denoted by DLW-CD and DLW-TL), the two-step proxy variable estimator of Levinsohn and Petrin (2003) (LP-CD), the two-step proxy variable estimator of Ackerberg, Caves and Frazer (2015) (ACF-CD), the dynamic panel estimator of Blundell and Bond (2000) (BB-CD) and the two-step GMM-IV estimator of Hu, Huang and Sasaki (2020) (HHS-CD). Starting values of the GMM-IV minimization processes for the proxy variable estimators are based on the true parameters of the DGP. Samples are generated based on DGP3 in Ackerberg, Caves and Frazer (2015) in which input prices are serially correlated, labor is chosen before materials and investment, and labor is subject to optimization error. However, production is generated through a Cobb-Douglas specification in capital, labor and material inputs. Furthermore, capital adjustment costs are heterogeneous but static. The table displays the mean of each estimated parameter across $S = 1000$ simulations. Standard errors, which are displayed in parentheses, are based on the standard deviation of each estimated parameter across the simulations.

We also focus on the estimator of Hu, Huang and Sasaki (2020). It is commonly assumed that labor is chosen simultaneously with material inputs. As a result, the policy function for material inputs should only contain capital as a state variable. Whenever we impose this in our moment conditions, we see from table 12 that the estimator from Hu, Huang and Sasaki (2020) is quite biased; performing worse than the Cobb-Douglas specification of De Loecker and Warzynski (2012). Note however that labor for production in period t is chosen at $t-b$ in DGP3 of Ackerberg, Caves and Frazer (2015). Thus, the model is correctly specified whenever labor is included as a state variable. The table below shows that the methodology of Hu, Huang and Sasaki (2020) does produce consistent estimates under this scenario. However, its standard errors are an order of magnitude larger than our preferred estimator.

Last, note that output elasticities are an explicit function of inputs under translog production. Thus, it could be argued that output elasticities are incorrectly estimated despite the small estimates for cross- and higher-order terms under the translog specification. It appears that this is not the case as can be seen from the last column in table 12. In fact, output elasticities are also most accurately estimated under the translog estimator from De Loecker and Warzynski (2012).

ACF – DGP3: LEONTIEF PRODUCTION. For completeness, we assess the reliability of the translog specification under the *exact same* DGP3 of Ackerberg, Caves and Frazer (2015). Material inputs enter production in a Leontief fashion instead, i.e. we have $Y_{it} = \min\left\{\exp\left(\omega_{it}\right)\beta_0 K_{it}^{\beta_k} L_{it}^{\beta_\ell}, \beta_m M_{it}\right\}$. Thus, we replicate the simulated data from Ackerberg, Caves and Frazer (2015) to the letter in this case. Most contributions in the production function literature run their Monte Carlo simulations on value added specifications; rather than gross output specifications as in the previous section.

To assess the reliability of the estimator used in our paper, we adapt it to estimate value added production functions instead which allows us to directly compare it to other production function estimation methodologies in the literature. We compare the Cobb-Douglas and translog estimators of De Loecker and Warzynski (2012) with the Cobb-Douglas estimators in Blundell and Bond (2000), Levinsohn and Petrin (2003), Ackerberg, Caves and Frazer (2015), and Hu, Huang and Sasaki (2020). The results can be found in table 13.

Table 13: Monte-Carlo results with ACF – DGP3 under Leontief specification: value added estimation.[2]

	β_k	β_ℓ	$\beta_{k\ell}$	β_{k^2}	β_{ℓ^2}
$\boldsymbol{\beta}_0$	0.40	0.60			
DLW-TL	0.4040 (0.0060)	0.6109 (0.0099)	0.0013 (0.0017)	-0.0028 (0.0028)	-0.0010 (0.0001)
DLW-CD	0.3878 (0.0170)	0.6048 (0.0077)			
LP-CD	0.5839 (0.0194)	0.4732 (0.0076)			
ACF-CD	0.4063 (0.0166)	0.5953 (0.0079)			
BB-CD: MA(0)	0.2277 (0.1008)	0.8974 (0.0675)			
BB-CD: MA(1)	0.1501 (0.1902)	0.8339 (0.0737)			
HHS-CD (capital only)	0.3161 (0.1186)	0.3634 (0.2028)			
HHS-CD (capital and labor)	0.4144 (0.0803)	0.6142 (0.1126)			

[2]We estimate production function parameters through the two-step proxy variable estimator of De Loecker and Warzynski (2012) (denoted by DLW-CD and DLW-TL), the two-step proxy variable estimator of Levinsohn and Petrin (2003) (LP-CD), the two-step proxy variable estimator of Ackerberg, Caves and Frazer (2015) (ACF-CD), the dynamic panel estimator of Blundell and Bond (2000) (BB-CD) and the two-step estimator of Hu, Huang and Sasaki (2020) (HHS-CD). Starting values of the GMM-IV minimization processes for the proxy variable estimators are based on the true parameters of the DGP. Samples are generated based on DGP3 in Ackerberg, Caves and Frazer (2015) in which input prices are serially correlated, labor is chosen before materials and investment, and labor is subject to optimization error. Furthermore, capital adjustment costs are heterogeneous but static. The table displays the mean of each estimated parameter across $S = 1000$ simulations. Standard errors, which are displayed in parentheses, are based on the standard deviation of each estimated parameter across the simulations.

Unlike the results for a gross output specification, the whole family of proxy variable estimators (with the exception of Levinsohn and Petrin, 2003) produces consistent estimates. Similar to the previous section, we see that the DPD-IV estimator from Blundell and Bond (2000) still performs poorly; its bias is less severe than before though. Moreover, we see that the estimator from Hu, Huang and Sasaki (2020) does produce consistent estimates, but it is crucial that the model (in particular, the state variables of the policy function for material inputs) is correctly specified. Hence, it appears that the estimator from Hu, Huang and Sasaki (2020) is quite sensitive to model misspecification. Also, its standard errors are an order of magnitude larger than our preferred translog estimator by De Loecker and Warzynski (2012).

O.6.4 Derivation of ACF – DGP3 process

In the following, we adapt the DGP in Ackerberg, Caves and Frazer (2015) to allow for material inputs to enter production in a Cobb-Douglas fashion. Conceptually, this does not change much, but the expressions, in particular the investment function, become more complicated. To ensure the validity of our results, we verify that the limits of our expressions (in which $\beta_m \to 0$ holds) coincide with those presented in Ackerberg, Caves and Frazer (2015) and Collard-Wexler and De Loecker (2020). Furthermore, we will also run our Monte Carlo experiments with the exact same DGP in Ackerberg, Caves and Frazer (2015).

We adapt the third data generating process (DGP3) in Ackerberg, Caves and Frazer (2015) and allow for material inputs to enter the production function through a Cobb-Douglas specification. Hence, production Y_{it} is generated through:

$$Y_{it} = \exp(\omega_{it})\beta_0 K_{it}^{\beta_k} L_{it}^{\beta_\ell} M_{it}^{\beta_m} \tag{42}$$

In the remainder of this section, we will set $\beta_0 = 1$. In the following, we will focus on DGP3 of Ackerberg, Caves and Frazer (2015): labor is chosen before material inputs without full knowledge of productivity ω_{it}. Instead, the firm observes some intermediate level of productivity ω_{it-b} between time periods $t-1$ and t.

Wages are idiosyncratic and stochastic. In particular, we assume that (natural log) productivity, wages and prices for material inputs follow AR(1) processes:

$$\omega_{it} = \rho \cdot \omega_{it-1} + \varepsilon_{it}^{\omega} \tag{43}$$

$$\ln(W_{it}) = \varphi_W \cdot \ln(W_{it-1}) + \varepsilon_{it}^{W} \tag{44}$$

$$\ln(P_{it}^{M}) = \varphi_M \cdot \ln(P_{it-1}^{M}) + \varepsilon_{it}^{M} \tag{45}$$

where $\varepsilon_{it}^{e} \sim N(0, \sigma_e^2)$ for $e \in \{\omega, W, M\}$ and all shocks are independent across firms and time. To avoid the functional dependence problem, labor is chosen at time $t - b$ for some $b \in (0, 1)$ when the firm observes only some intermediate productivity ω_{it-b}. This level of productivity evolves smoothly, i.e. it satisfies:

$$\omega_{it-b} = \rho^{1-b}\omega_{it-1} + \xi_{it}^{A} \tag{46}$$

$$\omega_{it} = \rho^{b}\omega_{it-b} + \xi_{it}^{B} \tag{47}$$

By construction, the variances of these innovations satisfy $V(\rho^b \xi_{it}^A + \xi_{it}^B) = V(\varepsilon_{it}^{\omega}) = \sigma_{\omega}^2$. We assume that investment and material inputs are chosen at time t. To solve the firm's problem, we use a backward induction strategy. At time t, given a level of capital K_{it} and labor L_{it}, a firm i chooses its optimal level of material inputs:

$$\max_{M_{it} \geq 0} P_{it}\text{exp}(\omega_{it}) K_{it}^{\beta_k} L_{it}^{\beta_\ell} M_{it}^{\beta_m} - P_{it}^{M} M_{it}$$

Assuming that output and input markets are perfectly competitive, the first order condition for M_{it} characterizes its optimal level:

$$\beta_m P_{it}\text{exp}(\omega_{it}) K_{it}^{\beta_k} L_{it}^{\beta_\ell} M_{it}^{\beta_m - 1} = P_{it}^{M}$$

Thus, we get:

$$\begin{aligned} M_{it}^{*} &\equiv \mathscr{M}_{it}(\omega_{it}, K_{it}; L_{it}) \\ &= \beta_m^{\frac{1}{1-\beta_m}} \text{exp}\left(\frac{\omega_{it}}{1-\beta_m}\right) P_{it}^{\frac{1}{1-\beta_m}} K_{it}^{\frac{\beta_k}{1-\beta_m}} L_{it}^{\frac{\beta_\ell}{1-\beta_m}} (P_{it}^{M})^{-\frac{1}{1-\beta_m}} \end{aligned} \tag{48}$$

At time $t - b$, a firm i takes ω_{it-b} (and *not* the level of productivity ω_{it}) as given and internalizes that its labor decision affects its choice for material inputs at time t. Hence, its maximization problem is given by:

$$\max_{L_{it}\geq 0} P_{it}\mathbb{E}_{it-b}\left[\exp(\omega_{it})K_{it}^{\beta_k}L_{it}^{\beta_\ell}\mathscr{M}_{it}(\omega_{it},K_{it};L_{it})^{\beta_m}\middle|\omega_{it-b}\right]-W_{it}L_{it}$$
$$=\max_{L_{it}\geq 0} P_{it}^{\frac{1}{1-\beta_m}}\mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\middle|\omega_{it-b}\right]K_{it}^{\frac{\beta_k}{1-\beta_m}}L_{it}^{\frac{\beta_\ell}{1-\beta_m}}\beta_m^{\frac{\beta_m}{1-\beta_m}}(P_{it}^M)^{-\frac{\beta_m}{1-\beta_m}}-W_{it}L_{it}$$

The first order condition for labor is then characterized by:

$$\frac{\beta_\ell}{1-\beta_m}P_{it}^{\frac{1}{1-\beta_m}}\mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\middle|\omega_{it-b}\right]K_{it}^{\frac{\beta_k}{1-\beta_m}}L_{it}^{\frac{\beta_\ell-1+\beta_m}{1-\beta_m}}\beta_m^{\frac{\beta_m}{1-\beta_m}}(P_{it}^M)^{-\frac{\beta_m}{1-\beta_m}}=W_{it}$$

Then, optimal labor L_{it}^* satisfies:

$$\begin{aligned}L_{it}^* &\equiv \mathscr{L}_{it}(\omega_{it-b},K_{it})\\ &=\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}}\beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}}\left(\mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\middle|\omega_{it-b}\right]\right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}}\times\\ &\quad K_{it}^{\frac{\beta_k}{1-\beta_m-\beta_\ell}}P_{it}^{\frac{1}{1-\beta_m-\beta_\ell}}(P_{it}^M)^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}}W_{it}^{-\frac{(1-\beta_m)}{1-\beta_m-\beta_\ell}}\end{aligned}\tag{49}$$

Note that the expression for $\lim_{\beta_m\to 0} L_{it}^*$ equals:

$$\beta_\ell^{\frac{1}{1-\beta_\ell}}P_{it}^{\frac{1}{1-\beta_\ell}}\left(\mathbb{E}_{it-b}\left[\exp(\omega_{it})\middle|\omega_{it-b}\right]\right)^{\frac{1}{1-\beta_\ell}}W_{it}^{-\frac{1}{1-\beta_\ell}}K_{it}^{\frac{\beta_k}{1-\beta_\ell}}$$

which coincides with the term for labor on p. 2443 in Ackerberg, Caves and Frazer (2015). To see this, note that $\lim_{\beta_m\to 0}\beta_m^{\beta_m}=1$. To simplify, we can also write the expression for labor as:

$$\begin{aligned}L_{it}^* &\equiv \mathscr{L}_{it}(\omega_{it-b},K_{it})\\ &=\left(\frac{\left(\frac{\beta_\ell}{1-\beta_m}\right)^{1-\beta_m}\cdot\beta_m^{\beta_m}\cdot P_{it}\cdot\left(\mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\middle|\omega_{it-b}\right]\right)^{1-\beta_m}}{(P_{it}^M)^{\beta_m}W_{it}^{1-\beta_m}}\right)^{\frac{1}{1-\beta_m-\beta_\ell}}K_{it}^{\frac{\beta_k}{1-\beta_m-\beta_\ell}}\end{aligned}\tag{50}$$

Given these optimal choices, we can define the following lemmas.

LEMMA 2. Under DGP1 of ACF and $\beta_k + \beta_\ell + \beta_m = 1$, revenues at the optimum can be written as:

$$P_{it}Y_{it}^* = \left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\omega_{it}}{1-\beta_m}\right) \mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\Big|\omega_{it-b}\right]^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} K_{it}$$
$$\times (P_{it}^M)^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}} \cdot P_{it}^{\frac{1}{1-\beta_m-\beta_\ell}} \cdot W_{it}^{-\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \tag{51}$$

Proof. We plug the optimal choices for material inputs and labor at time t and $t-b$, respectively, in the revenue function. Then, we get:

$$\begin{aligned}
P_{it}Y_{it}^* &= P_{it}\exp(\omega_{it})K_{it}^{\beta_k}\mathscr{L}_{it}(\omega_{it-b},K_{it})^{\beta_\ell}\mathscr{M}_{it}(\omega_{it},K_{it};\mathscr{L}_{it}(\omega_{it-b},K_{it}))^{\beta_m} \\
&= P_{it}^{\frac{1}{1-\beta_m}}\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)K_{it}^{\frac{\beta_k}{1-\beta_m}}\mathscr{L}_{it}(\omega_{it-b},K_{it})^{\frac{\beta_\ell}{1-\beta_m}}\beta_m^{\frac{\beta_m}{1-\beta_m}}(P_{it}^M)^{-\frac{\beta_m}{1-\beta_m}} \\
&= P_{it}^{\frac{1}{1-\beta_m}}\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\beta_m^{\frac{\beta_m}{1-\beta_m}}(P_{it}^M)^{-\frac{\beta_m}{1-\beta_m}} \\
&\times K_{it}^{\frac{\beta_k}{1-\beta_m}+\frac{\beta_k}{1-\beta_m-\beta_\ell}\frac{\beta_\ell}{1-\beta_m}}\left(\frac{\left(\frac{\beta_\ell}{1-\beta_m}\right)^{1-\beta_m}\cdot\beta_m^{\beta_m}\cdot P_{it}\cdot\left(\mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\Big|\omega_{it-b}\right]\right)^{1-\beta_m}}{(P_{it}^M)^{\beta_m}W_{it}^{1-\beta_m}}\right)^{\frac{1}{1-\beta_m-\beta_\ell}\frac{\beta_\ell}{1-\beta_m}} \\
&= \exp\left(\frac{\omega_{it}}{1-\beta_m}\right)K_{it}\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}}\beta_m^{\frac{\beta_m}{1-\beta_m}\left[1+\frac{\beta_\ell}{1-\beta_m-\beta_\ell}\right]}(P_{it}^M)^{-\frac{\beta_m}{1-\beta_m}\left[1+\frac{\beta_\ell}{1-\beta_m-\beta_\ell}\right]} \\
&\times P_{it}^{\frac{1}{1-\beta_m}\left[1+\frac{\beta_\ell}{1-\beta_m-\beta_\ell}\right]}\mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\Big|\omega_{it-b}\right]^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}}W_{it}^{-\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \\
&= \left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}}\beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}}\mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\Big|\omega_{it-b}\right]^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}}\cdot K_{it} \\
&\times \exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\cdot(P_{it}^M)^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}}\cdot P_{it}^{\frac{1}{1-\beta_m-\beta_\ell}}\cdot W_{it}^{-\frac{\beta_\ell}{1-\beta_m-\beta_\ell}}
\end{aligned} \tag{52}$$

which is exactly what we wanted to show. □

LEMMA 3. Under DGP1 of ACF and $\beta_k + \beta_\ell + \beta_m = 1$, revenues net of payments to labor at the optimum can be written as:

$$
\begin{aligned}
P_{it}Y^*_{it} - W_{it}L^*_{it} = & (P^M_{it})^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}} \cdot P_{it}^{\frac{1}{1-\beta_m-\beta_\ell}} \cdot W_{it}^{-\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\Big|\omega_{it-b}\right]^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} K_{it} \\
& \times \exp\left(\frac{\rho\omega_{it-1}}{1-\beta_m}\right)\exp\left(\frac{\rho^b\xi^A_{it}}{1-\beta_m}\right)\Bigg\{\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\xi^B_{it}}{1-\beta_m}\right) \\
& - \left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\left[\frac{1}{1-\beta_m}\right]^2\frac{\sigma^2_{\xi^B}}{2}\right)\Bigg\} \qquad (53)
\end{aligned}
$$

Proof. By applying lemma 1 and the optimal equation for labor (49), we get:

$$
\begin{aligned}
P_{it}Y^*_{it} - W_{it}L^*_{it} = & P_{it}^{\frac{1}{1-\beta_m-\beta_\ell}} (P^M_{it})^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}} W_{it}^{-\frac{(1-\beta_m)}{1-\beta_m-\beta_\ell}} K_{it}\left(\mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\Big|\omega_{it-b}\right]\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \\
& \times\Bigg\{\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\omega_{it}}{1-\beta_m}\right) \\
& - \left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\Big|\omega_{it-b}\right]\Bigg\} \\
= & P_{it}^{\frac{1}{1-\beta_m-\beta_\ell}} (P^M_{it})^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}} W_{it}^{-\frac{(1-\beta_m)}{1-\beta_m-\beta_\ell}} K_{it}\left(\mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\Big|\omega_{it-b}\right]\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \\
& \times \exp\left(\frac{\rho\omega_{it-1}}{1-\beta_m}\right)\Bigg\{\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\varepsilon^\omega_{it}}{1-\beta_m}\right) \\
& - \left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\rho^b\xi^A_{it}}{1-\beta_m}\right)\exp\left(\left[\frac{1}{1-\beta_m}\right]^2\frac{\sigma^2_{\xi^B}}{2}\right)\Bigg\} \\
= & P_{it}^{\frac{1}{1-\beta_m-\beta_\ell}} (P^M_{it})^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}} W_{it}^{-\frac{(1-\beta_m)}{1-\beta_m-\beta_\ell}} \left(\mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\Big|\omega_{it-b}\right]\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} K_{it} \\
& \times \exp\left(\frac{\rho\omega_{it-1}}{1-\beta_m}\right)\exp\left(\frac{\rho^b\xi^A_{it}}{1-\beta_m}\right)\Bigg\{\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\xi^B_{it}}{1-\beta_m}\right) \\
& - \left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\left[\frac{1}{1-\beta_m}\right]^2\frac{\sigma^2_{\xi^B}}{2}\right)\Bigg\} \qquad (54)
\end{aligned}
$$

where we exploited the fact that $\varepsilon^\omega_{it} = \rho^b\xi^A_{it} + \xi^B_{it}$. Then, we have showed exactly what we wanted. □

These lemmas will be extremely useful for characterizing the optimal investment function.

This is shown in the proposition below.

PROPOSITION 5. Let the environment of DGP1 in ACF hold with $Y_{it} = \exp(\omega_{it})\beta_0 K_{it}^{\beta_k} L_{it}^{\beta_\ell} M_{it}^{\beta_m}$ and $\beta_k + \beta_\ell + \beta_m = 1$. Whenever the price for output is the numéraire, the optimal investment function equals:

$$\begin{aligned}
I_{it}^* = \frac{\beta}{\varphi_{it}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} & \left[\left(\frac{\beta_\ell}{1-\beta_m} \right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} - \left(\frac{\beta_\ell}{1-\beta_m} \right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}} \right] \cdot \sum_{\tau=0}^{\infty} \Bigg\{ [\beta(1-\delta)]^\tau \\
& \times \exp \Bigg[\frac{\rho^{\tau+1}\omega_{it}}{1-\beta_m-\beta_\ell} - \frac{\beta_\ell \cdot \varphi_W^{\tau+1}}{1-\beta_m-\beta_\ell} \ln(W_{it}) - \frac{\beta_m \cdot \varphi_M^{\tau+1}}{1-\beta_m-\beta_\ell} \ln\left(P_{it}^M\right) \\
& + \frac{1}{2} \left[\frac{\beta_\ell}{1-\beta_m-\beta_\ell} \right]^2 \sigma_W^2 \sum_{s=0}^{\tau} \varphi_W^{2(\tau-s)} + \frac{1}{2} \left[\frac{\beta_m}{1-\beta_m-\beta_\ell} \right]^2 \sigma_M^2 \sum_{s=0}^{\tau} \varphi_M^{2(\tau-s)} \\
& + \frac{1}{2} \left[\frac{1}{1-\beta_m-\beta_\ell} \right]^2 \rho^2 \sum_{s=1}^{\tau} \rho^{2(\tau-s)} \sigma_\omega^2 + \frac{\sigma_{\xi^A}^2}{2} \cdot \rho^{2b} \cdot \left(\frac{1}{1-\beta_m-\beta_\ell} \right)^2 \\
& + \left(\frac{1}{1-\beta_m} \right) \left(\frac{1}{1-\beta_m-\beta_\ell} \right) \frac{1}{2} \sigma_{\xi^B}^2 \Bigg] \Bigg\}
\end{aligned} \tag{55}$$

Proof. By definition, a firm i's optimal level of investment I_{it}^* solves the following problem:

$$V(\mathbf{x}_{it}) = \max_{I_{it}, M_{it} \geq 0} \Bigg\{ P_{it} Y_{it}^* - W_{it} L_{it}^* - P_{it}^M M_{it} - \frac{\varphi_{it} I_{it}^2}{2} + \beta \mathbb{E}_{it} \left[V(\mathbf{x}_{it+1}) | \mathbf{x}_{it} \right]$$
$$\text{s.t. } K_{it+1} = (1-\delta) K_{it} + I_{it} \Bigg\}$$

Investment I_{it} is characterized by its first order condition:

$$\varphi_{it} I_{it} = \beta \mathbb{E}_{it} \left[\frac{\partial V(\mathbf{x}_{it+1})}{\partial K_{it+1}} \middle| \mathbf{x}_{it} \right] \tag{56}$$

We exploit the envelope condition to characterize the partial derivative $\frac{\partial V(\mathbf{x}_{it+1})}{\partial K_{it+1}}$. More precisely, we have:

$$\frac{\partial V(\mathbf{x}_{it})}{\partial K_{it}} = \frac{\partial}{\partial K_{it}} \left[P_{it} Y_{it}^* - W_{it} L_{it}^* \right] + (1-\delta) \beta \mathbb{E}_{it} \left[\frac{\partial V(\mathbf{x}_{it+1})}{\partial K_{it+1}} \middle| \mathbf{x}_{it} \right]$$

$$
\begin{aligned}
&= \frac{P_{it}}{K_{it}} \exp(\omega_{it}) K_{it}^{\beta_k} \mathscr{L}_{it}(\omega_{it-b}, K_{it})^{\beta_\ell} \mathscr{M}_{it}(\omega_{it}, K_{it}; \mathscr{L}_{it}(\omega_{it-b}, K_{it}))^{\beta_m} - \frac{W_{it}}{K_{it}} \mathscr{L}_{it}(\omega_{it-b}, K_{it}) \\
&+ (1-\delta)\beta \mathbb{E}_{it}\left[\frac{\partial V(\mathbf{x}_{it+1})}{\partial K_{it+1}}\middle|\mathbf{x}_{it}\right] \\
&= (P_{it}^M)^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}} \cdot W_{it}^{-\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \mathbb{E}_{it-b}\left[\exp\left(\frac{\omega_{it}}{1-\beta_m}\right)\middle|\omega_{it-b}\right]^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\rho\omega_{it-1}}{1-\beta_m}\right) \\
&\times \exp\left(\left[\frac{1}{1-\beta_m}\right]\left[\frac{\beta_\ell}{1-\beta_m-\beta_\ell}+1\right]\rho^b \xi_{it}^A\right)\Bigg\{\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\xi_{it}^B}{1-\beta_m}\right) \\
&- \left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\left[\frac{1}{1-\beta_m}\right]^2 \frac{\sigma_{\xi^B}^2}{2}\right)\Bigg\} + (1-\delta)\beta\mathbb{E}_{it}\left[\frac{\partial V(\mathbf{x}_{it+1})}{\partial K_{it+1}}\middle|\mathbf{x}_{it}\right] \\
&= (P_{it}^M)^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}} \cdot W_{it}^{-\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \exp\left(\left[\frac{1}{1-\beta_m}\right]^2\left[\frac{\beta_\ell}{1-\beta_m-\beta_\ell}\right]\sigma_{\xi^B}^2\right) \\
&\times \exp\left(\frac{\rho\omega_{it-1}}{1-\beta_m-\beta_\ell}\right)\exp\left(\frac{\rho^b\xi_{it}^A}{1-\beta_m-\beta_\ell}\right)\Bigg\{\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\xi_{it}^B}{1-\beta_m}\right) \\
&- \left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\left[\frac{1}{1-\beta_m}\right]^2 \frac{\sigma_{\xi^B}^2}{2}\right)\Bigg\} + (1-\delta)\beta\mathbb{E}_{it}\left[\frac{\partial V(\mathbf{x}_{it+1})}{\partial K_{it+1}}\middle|\mathbf{x}_{it}\right]
\end{aligned} \tag{57}
$$

where the second to the third equality follows from lemma 2 and applying P_{it} as the numéraire. We go from the third to the last equality by expanding the conditional expectation and collecting common terms. Note that, by assumption, investment and material inputs are chosen at time t after labor was determined in period $t-b$. Hence, we must take revenues *net of labor payments* (which are a function of physical capital K_{it}) when applying the envelope condition. Combining expresssions (56) and (57), we obtain:

$$
\begin{aligned}
\varphi_{it} I_{it} = {} & (P_{it}^M)^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}} \cdot W_{it}^{-\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \exp\left(\left[\frac{1}{1-\beta_m}\right]^2\left[\frac{\beta_\ell}{1-\beta_m-\beta_\ell}\right]\sigma_{\xi^B}^2\right) \\
& \times \exp\left(\frac{\rho\omega_{it-1}}{1-\beta_m-\beta_\ell}\right)\exp\left(\frac{\rho^b\xi_{it}^A}{1-\beta_m-\beta_\ell}\right)\Bigg\{\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\xi_{it}^B}{1-\beta_m}\right) \\
& - \left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\left[\frac{1}{1-\beta_m}\right]^2 \frac{\sigma_{\xi^B}^2}{2}\right)\Bigg\} + \beta(1-\delta)\mathbb{E}_{it}\left(\varphi_{it+1} I_{it+1} | \mathbf{x}_{it}\right)
\end{aligned} \tag{58}
$$

Iterating expression (58) forward, we get:

$$I_{it}^{*} = \frac{\beta}{\varphi_{it}} \times \mathbb{E}_{it}\Bigg[\sum_{\tau=0}^{\infty} [\beta(1-\delta)]^{\tau}\, W_{it+1+\tau}^{-\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \left(P_{it+1+\tau}^{M}\right)^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\left[\frac{1}{1-\beta_m}\right]^{2}\left[\frac{\beta_\ell}{1-\beta_m-\beta_\ell}\right]\sigma_{\xi^B}^{2}\right)$$
$$\times \exp\left(\frac{\rho\cdot\omega_{it+\tau}}{1-\beta_m-\beta_\ell}\right)\exp\left(\frac{\rho^{b}\xi_{it+1+\tau}^{A}}{1-\beta_m-\beta_\ell}\right)\Bigg\{\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}}\beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}}\exp\left(\frac{\xi_{it+1+\tau}^{B}}{1-\beta_m}\right)$$
$$-\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}}\beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}}\exp\left(\left[\frac{1}{1-\beta_m}\right]^{2}\frac{\sigma_{\xi^B}^{2}}{2}\right)\Bigg\}\Bigg|\mathbf{x}_{it}\Bigg] \qquad (59)$$

Apply the expectations operators on the productivity shocks, then we have:

$$I_{it}^{*} = \frac{\beta}{\varphi_{it}} \times \sum_{\tau=0}^{\infty}[\beta(1-\delta)]^{\tau}\,\mathbb{E}_{it}\left[W_{it+1+\tau}^{-\frac{\beta_\ell}{1-\beta_m-\beta_\ell}}\left(P_{it+1+\tau}^{M}\right)^{-\frac{\beta_m}{1-\beta_m-\beta_\ell}}\exp\left(\frac{\rho\omega_{it+\tau}}{1-\beta_m-\beta_\ell}\right)\Bigg|\mathbf{x}_{it}\right]$$
$$\times\exp\left(\frac{1}{1-\beta_m}\frac{1}{1-\beta_m-\beta_\ell}\frac{\sigma_{\xi^B}^{2}}{2}\right)\exp\left(\frac{\rho^{2b}}{(1-\beta_m-\beta_\ell)^{2}}\frac{\sigma_{\xi^A}^{2}}{2}\right)\Bigg\{\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}}\beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}}$$
$$-\left(\frac{\beta_\ell}{1-\beta_m}\right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}}\beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}}\Bigg\} \qquad (60)$$

The latter step is valid since it is assumed that productivity shocks are orthogonal to shocks to input prices; across time and firms. By assumption, productivity ω_{it} follows an AR(1) process. Hence, we must have:

$$\begin{aligned}\mathbb{E}_{it}\left[\exp(\omega_{it+\tau})|\mathbf{x}_{it}\right] &= \mathbb{E}_{it}\left[\exp\left(\rho^{\tau}\omega_{it}+\sum_{s=1}^{\tau}\rho^{\tau-s}\varepsilon_{it+s}^{\omega}\right)\Bigg|\mathbf{x}_{it}\right]\\ &= \mathbb{E}_{it}\left[\exp\left(\sum_{s=1}^{\tau}\rho^{\tau-s}\varepsilon_{it+s}^{\omega}\right)\Bigg|\mathbf{x}_{it}\right]\exp\left(\rho^{\tau}\omega_{it}\right)\end{aligned}$$

We can apply the same logic to input prices, so we can rewrite expression (60) as:

$$
\begin{aligned}
I_{it}^{*} = \frac{\beta}{\varphi_{it}} & \left[\left(\frac{\beta_\ell}{1-\beta_m} \right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} - \left(\frac{\beta_\ell}{1-\beta_m} \right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} \right] \\
& \times \sum_{\tau=0}^{\infty} \Bigg\{ [\beta(1-\delta)]^{\tau} \, W_{it}^{-\frac{\beta_\ell \cdot \varphi_W^{\tau+1}}{1-\beta_m-\beta_\ell}} \cdot \prod_{s=0}^{\tau} \exp\left(\frac{\sigma_W^2}{2} \left[\frac{\beta_\ell}{1-\beta_m-\beta_\ell} \right]^2 \cdot \varphi_W^{2s} \right) \\
& \times \left(P_{it}^{M} \right)^{-\frac{\beta_m \cdot \varphi_M^{\tau+1}}{1-\beta_m-\beta_\ell}} \cdot \prod_{s=0}^{\tau} \exp\left(\frac{\sigma_M^2}{2} \left[\frac{\beta_m}{1-\beta_m-\beta_\ell} \right]^2 \cdot \varphi_M^{2s} \right) \\
& \times \exp\left(\frac{\rho^{\tau+1}\omega_{it}}{1-\beta_m-\beta_\ell} \right) \cdot \prod_{s=1}^{\tau} \exp\left(\frac{\sigma_\omega^2}{2} \left[\frac{\rho}{1-\beta_m-\beta_\ell} \right]^2 \cdot \rho^{2(\tau-s)} \right) \\
& \times \exp\left(\frac{1}{1-\beta_m} \frac{1}{1-\beta_m-\beta_\ell} \frac{\sigma_{\xi^B}^2}{2} \right) \exp\left(\frac{\rho^{2b}}{(1-\beta_m-\beta_\ell)^2} \frac{\sigma_{\xi^A}^2}{2} \right) \Bigg\}
\end{aligned}
$$

Collecting terms, the above expression can be rewritten as:

$$
\begin{aligned}
\hat{I}_{it}^{*}(\beta_m) = \frac{\beta}{\varphi_{it}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} & \left[\left(\frac{\beta_\ell}{1-\beta_m} \right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} - \left(\frac{\beta_\ell}{1-\beta_m} \right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}} \right] \\
& \times \sum_{\tau=0}^{\infty} [\beta(1-\delta)]^{\tau} \exp \Bigg\{ \frac{\rho^{\tau+1}\omega_{it}}{1-\beta_m-\beta_\ell} - \frac{\beta_\ell \cdot \varphi_W^{\tau+1}}{1-\beta_m-\beta_\ell} \ln\left(W_{it}\right) - \frac{\beta_m \cdot \varphi_M^{\tau+1}}{1-\beta_m-\beta_\ell} \ln\left(P_{it}^{M}\right) \\
& + \frac{1}{2} \left[\frac{\beta_\ell}{1-\beta_m-\beta_\ell} \right]^2 \sigma_W^2 \sum_{s=0}^{\tau} \varphi_W^{2(\tau-s)} + \frac{1}{2} \left[\frac{\beta_m}{1-\beta_m-\beta_\ell} \right]^2 \sigma_M^2 \sum_{s=0}^{\tau} \varphi_M^{2(\tau-s)} \\
& + \frac{1}{2} \left[\frac{1}{1-\beta_m-\beta_\ell} \right]^2 \rho^2 \sum_{s=1}^{\tau} \rho^{2(\tau-s)} \sigma_\omega^2 + \frac{\sigma_{\xi^A}^2}{2} \cdot \rho^{2b} \cdot \left(\frac{1}{1-\beta_m-\beta_\ell} \right)^2 \\
& + \left(\frac{1}{1-\beta_m} \right) \left(\frac{1}{1-\beta_m-\beta_\ell} \right) \frac{1}{2} \sigma_{\xi^B}^2 \Bigg\}
\end{aligned} \tag{61}
$$

which is exactly what we wanted to show. Note that the case in Ackerberg, Caves and Frazer (2015) can be derived as a limit of $\beta_m \to 0$. Then, we get:

$$
\begin{aligned}
\lim_{\beta_m \to 0} \hat{I}^*_{it}(\beta_m) = \frac{\beta}{\varphi_{it}} \sum_{\tau=0}^{\infty} \Bigg\{ & [\beta(1-\delta)]^{\tau} \left[\beta_\ell^{\frac{\beta_\ell}{1-\beta_\ell}} - \beta_\ell^{\frac{1}{1-\beta_\ell}} \right] \\
& \times \exp\Bigg[\frac{\rho^{\tau+1}\omega_{it}}{1-\beta_\ell} - \frac{\beta_\ell \cdot \varphi_W^{\tau+1}}{1-\beta_\ell} \ln\left(W_{it}\right) \\
& + \sum_{s=0}^{\tau} \frac{\sigma_W^2}{2} \left[\frac{\beta_\ell}{1-\beta_\ell} \right]^2 \cdot \varphi_W^{2(\tau-s)} + \frac{1}{2} \left(\frac{1}{1-\beta_\ell} \right)^2 \rho^2 \cdot \sum_{s=1}^{\tau} \sigma_\omega^2 \cdot \rho^{2(\tau-s)} \\
& + \frac{1}{2} \left(\frac{1}{1-\beta_\ell} \right)^2 \rho^{2b} \cdot \sigma_{\xi^A}^2 + \left(\frac{1}{1-\beta_\ell} \right) \frac{\sigma_{\xi^B}^2}{2} \Bigg] \Bigg\}
\end{aligned} \tag{62}
$$

which is the equivalent of the expression for investment in Ackerberg, Caves and Frazer (2015) on their page 2446.[13] Our expression in (62) becomes identical to the one in Ackerberg, Caves and Frazer (2015) whenever $\sigma_{\xi^\ell}^2 \to 0$ and we have $\beta_0 = 1$. Note that it is straightforward to allow for measurement error in labor. Whenever we have $L_{it}^{\text{err}} = L_{it}^* \exp(\xi_{it}^\ell)$ such that $\mathbb{E}_{it}\left[\exp(\xi_{it}^\ell)\right] = \sigma_{\xi^\ell}^2/2$, then the optimal investment function is characterized as:

$$
\begin{aligned}
I_{it}^* = \frac{\beta}{\varphi_{it}} \beta_m^{\frac{\beta_m}{1-\beta_m-\beta_\ell}} & \left[\left(\frac{\beta_\ell}{1-\beta_m} \right)^{\frac{\beta_\ell}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\beta_\ell^2 \sigma_{\xi^\ell}^2}{2} \right) - \left(\frac{\beta_\ell}{1-\beta_m} \right)^{\frac{1-\beta_m}{1-\beta_m-\beta_\ell}} \exp\left(\frac{\sigma_{\xi^\ell}^2}{2} \right) \right] \\
\times \sum_{\tau=0}^{\infty} [\beta(1-\delta)]^{\tau} \exp\Bigg\{ & \frac{\rho^{\tau+1}\omega_{it}}{1-\beta_m-\beta_\ell} - \frac{\beta_\ell \cdot \varphi_W^{\tau+1}}{1-\beta_m-\beta_\ell} \ln\left(W_{it}\right) - \frac{\beta_m \cdot \varphi_M^{\tau+1}}{1-\beta_m-\beta_\ell} \ln\left(P_{it}^M\right) \\
+ \frac{1}{2} \left[\frac{\beta_\ell}{1-\beta_m-\beta_\ell} \right]^2 & \sigma_W^2 \sum_{s=0}^{\tau} \varphi_W^{2(\tau-s)} + \frac{1}{2} \left[\frac{\beta_m}{1-\beta_m-\beta_\ell} \right]^2 \sigma_M^2 \sum_{s=0}^{\tau} \varphi_M^{2(\tau-s)} \\
+ \frac{1}{2} \left[\frac{1}{1-\beta_m-\beta_\ell} \right]^2 & \rho^2 \sum_{s=1}^{\tau} \rho^{2(\tau-s)} \sigma_\omega^2 + \frac{\sigma_{\xi^A}^2}{2} \cdot \rho^{2b} \cdot \left(\frac{1}{1-\beta_m-\beta_\ell} \right)^2 \\
+ \left(\frac{1}{1-\beta_m} \right) & \left(\frac{1}{1-\beta_m-\beta_\ell} \right) \frac{1}{2} \sigma_{\xi^B}^2 \Bigg\}
\end{aligned} \tag{63}
$$

[13] The only components that are different are the terms containing $\sigma_{\xi^A}^2$ and σ_ω^2. Ackerberg, Caves and Frazer (2015) find $\frac{1}{2}\left(\frac{1}{1-\beta_\ell}\right)^2 \rho_b^2 \cdot \sum_{s=1}^{\tau} \sigma_\omega^2 \cdot \rho_b^{2(\tau-s)}$ and $\frac{1}{2}\left(\frac{1}{1-\beta_\ell}\right)^2 \rho_b^2 \rho^{2\tau} \cdot \sigma_{\xi^A}^2$ instead $\frac{1}{2}\left(\frac{1}{1-\beta_\ell}\right)^2 \rho^2 \cdot \sum_{s=1}^{\tau} \sigma_\omega^2 \cdot \rho^{2(\tau-s)}$ and $\frac{1}{2}\left(\frac{1}{1-\beta_\ell}\right)^2 \rho^{2b} \cdot \sigma_{\xi^A}^2$. Given that ρ_b is not a well-defined object and ρ^τ cannot appear in those terms associated with $\sigma_{\xi^A}^2$, it is relatively safe to assume that the expressions in Ackerberg, Caves and Frazer (2015) are typos.

Given these observations, we remain. $\square$

Note that DGP1 in Ackerberg, Caves and Frazer (2015) with optimization error in labor is equivalent to their DGP3. To complete the description of the data-generating process, we need to specify how we initalize capital, productivity and wages.

$$K_{i0} = \text{exp}(-10) \simeq 0.0000454 \tag{64}$$

$$\omega_{i0} = \sigma_\omega \cdot \varepsilon_{i0} \tag{65}$$

$$W_{i0} = \sigma_W \cdot \varepsilon_{i0} \tag{66}$$

where $\varepsilon \sim N(0,1)$. Note that all firms start with almost zero stock of capital. Finally, we follow Ackerberg, Caves and Frazer (2015) and Collard-Wexler and De Loecker (2020), and inject measurement error in output and material inputs. More precisely, we have:

$$\text{ln}(Y_{it}) = \text{ln}(Y_{it}^*) + \varepsilon_{it}^Y \tag{67}$$

$$\text{ln}(M_{it}) = \text{ln}(M_{it}^*) + m_E \cdot \varepsilon \tag{68}$$

where $\varepsilon_{it}^Y \sim N(0, \sigma_Y^2)$ and m_E^2 is the cross-sectional variance of demeaned levels of M_{it}^*.

O.7 Counterfactual exercises

Our baseline estimates in section II imply median markdowns of 1.53. This is well in line with the meta-study by Sokolova and Sorensen (2020): our results fall around the median in their distribution of estimates for the elasticity of labor supply. Nevertheless, we further investigate the magnitude of our estimates with two sets of counterfactual exercises in the spirit of Brooks et al. (2021).[14]

PROFIT SHARE. In our first exercise, we verify that the majority of variable profits are not accounted for by markdowns, ensuring that our markdowns are not implausibly large. To do this, note that variable profits as a fraction of revenues (also referred to as the profit share) s_π are defined as:

$$\begin{aligned} s_\pi &\equiv 1 - \alpha_K - \alpha_\ell - \alpha_M - \alpha_E \\ &= 1 - \alpha_K - \theta_\ell \cdot \nu^{-1} \cdot \mu^{-1} - \theta_M \cdot \mu^{-1} - \alpha_E \end{aligned} \tag{69}$$

where we applied our results from proposition 1 in the second equality. Then, conditional on profits only stemming from labor market power, the counterfactual profit share satisfies:

$$s_{\pi|\mu=1} = 1 - \alpha_K - \theta_\ell \cdot \nu^{-1} - \theta_M - \alpha_E \tag{70}$$

Summary statistics on profit shares and their counterfactual counterparts can be found in table 14.

Table 14: Actual and counterfactual profit shares.†

PROFIT SHARE	Median	Mean	25%	75%	SD
Actual	0.203	0.190	0.101	0.303	0.227
Counterfactual	0.081	0.072	0.004	0.159	0.203
Sample size	$1.393 \cdot 10^6$				

†Markdowns are estimated under the assumption of a translog specification for gross output. The flexible input is materials. Each industry group in manufacturing corresponds to the manufacturing categorization of the BEA which approximately follows a 3-digit NAICS specification. Actual profit shares are defined as variable profits relative to revenues whereas counterfactual profit shares are constructed by setting markups to unity. By doing so, we follow the counterfactual experiments of Brooks et al. (2021). Source: Authors' calculations from ASM/CM data in 1976–2014.

[14] We thank an anonymous referee for these helpful suggestions.

We find that, for the median plant, the majority of variable profits are actually accounted for by markups. Approximately $0.081/0.203 = 40$ percent of the median plant's profits are due to labor market power which we deem as reasonable.

AGGREGATE LABOR SHARE. In the following, we evaluate the time evolution of the aggregate labor share in the absence of labor market power. Following Brooks et al. (2021) and Kehrig and Vincent (2021), we define the labor share as payments to labor relative to value added:

$$\eta_t^\ell \equiv \frac{\sum_{i \in F_t} w_{it} \ell_{it}}{\sum_{i \in F_t} p_{it} y_{it} - p_{it}^M m_{it} - p_{it}^E e_{it}} \tag{71}$$

For this empirical exercise, we implement the definition for value added from Kehrig and Vincent (2021). The key difference, when compared to the standard definition from the Census Bureau, lies in the use of inventories for material inputs and purchased services used as intermediate inputs. These components are included by Kehrig and Vincent (2021) but not by the Census Bureau. By construction, therefore, the Census Bureau's definition for value added is smaller than that of Kehrig and Vincent (2021) which immediately implies that labor shares under the latter must be larger. However, intermediate services are not available at the plant level. Instead, Kehrig and Vincent (2021) impute the ratio of purchased services to sales at the industry level. They show that including intermediate services only has an impact on the level of the labor share and does not affect its time evolution. As a result, we will simply ignore purchased services for intermediate use.

Let a firm i's wage bill share (relative to the national economy) be equal to:

$$\omega_{it}^\ell \equiv \frac{w_{it} \ell_{it}}{\sum_{k \in F_t} w_{kt} \ell_{kt}} \tag{72}$$

Then, given our definitions of markups and markdowns, we can use equations (71) and (72) to derive the economy's aggregate labor share:

$$\begin{aligned}
\left(\eta_t^\ell\right)^{-1} &= \sum_{i \in F_t} \frac{p_{it} y_{it} - p_{it}^M m_{it} - p_{it}^E e_{it}}{\sum_{i \in F_t} w_{it} \ell_{it}} \\
&= \sum_{i \in F_t} \frac{p_{it} y_{it} - p_{it}^M m_{it} - p_{it}^E e_{it}}{w_{it} \ell_{it}} \cdot \omega_{it}^\ell
\end{aligned}$$

$$= \sum_{i \in F_t} \left(\frac{1 - \alpha_{it}^M - \alpha_{it}^E}{\alpha_{it}^L} \right) \cdot \omega_{it}^\ell$$

$$= \sum_{i \in F_t} \left(\frac{\nu_{it}\mu_{it}}{\theta_{it}^L} - \frac{\nu_{it}\theta_{it}^M}{\theta_{it}^L} - \frac{\alpha_{it}^E}{\alpha_{it}^L} \right) \cdot \omega_{it}^\ell$$

Hence, we can write the labor share η_t^ℓ as:

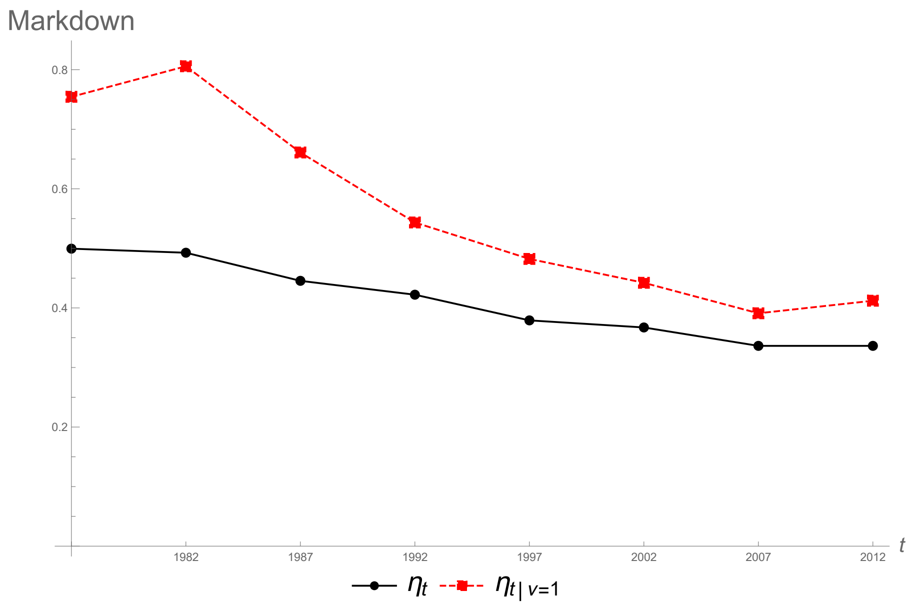
$$\eta_t^\ell = \left(\sum_{i \in F_t} \left[\nu_{it} \cdot \left(\frac{\mu_{it} - \theta_{it}^M}{\theta_{it}^L} \right) - \frac{\alpha_{it}^E}{\alpha_{it}^L} \right] \cdot \omega_{it}^\ell \right)^{-1} \tag{73}$$

According to Brooks et al. (2021), the counterfactual labor share without *monopsony* power would then be equal to:

$$\eta_{t|\nu=1}^\ell = \left(\sum_{i \in F_t} \left[\frac{\mu_{it} - \theta_{j(i)t}^M}{\theta_{j(i)t}^L} - \frac{\alpha_{it}^E}{\alpha_{it}^L} \right] \cdot \omega_{it}^\ell \right)^{-1} \tag{74}$$

Our results are displayed in figure 8. We find that our constructed labor share declines from about 50 percent to 33 percent from 1977 to 2012. The counterfactual series implies that the labor share declined from about 75 percent in 1977 to 41 percent in 2012. Hence, the fall in the labor share would be even more pronounced in the absence of monopsony power through the lens of the counterfactual exercise in Brooks et al. (2021). If markdowns were implausibly large, then we would expect the counterfactual labor share to be unreasonably high as well. Our counterfactual exercise does not seem to indicate that this is the case.

Figure 8: Actual and counterfactual aggregate labor shares.



Actual labor shares are defined as the aggregate wage bill divided by total value added. Counterfactual labor shares are calculated according to equation (74) in which markdowns are set to unity following Brooks et al. (2021). Source: Authors' own calculations from quinquennial CM data from 1977–2012.

O.8 Labor market models with $\varepsilon_S \geq 0$

O.8.1 Wage posting à la Burdett-Mortensen

For ease of notation, we drop a particular firm f's index. In the wage posting model of Burdett and Mortensen (1998), a firm's law of motion for its stock of labor is given by:

$$L_t = (1 - s(w_t))L_{t-1} + R(w_t) \tag{75}$$

where $s(\cdot)$ and $R(\cdot)$ denote the separation and recruiting functions, respectively. Note that these are allowed to explicitly depend on the posted wage. In a stationary setting, we must have $L_t = \frac{R(w_t)}{s(w_t)}$. Assuming that these functions are differentiable, it is straightforward to show that labor supply elasticities satisfy:

$$\varepsilon_S = \varepsilon_{Rw} - \varepsilon_{sw} > 0$$

where $\varepsilon_{Rw,t}$ and $\varepsilon_{sw,t}$ denote separation and recruiting elasticities, respectively. The above object is strictly positive since higher wages encourage hiring and lead to fewer separations, i.e. $\varepsilon_{Rw} > 0$ and $\varepsilon_{sw} < 0$.

Formally, the separation rate is induced by some exogenous job destruction process and poaching. In particular, we have $s(w) = \delta + \lambda(1 - F(w))$. Then, $-\varepsilon_{sw} = \lambda f(w) > 0$ follows directly from the fact that probability distribution functions are non-negative. Recall that the equilibrium wage distribution function has full support on $[0, \overline{w}]$ in the baseline framework of Burdett and Mortensen (1998). Furthermore, recruitment satisfies $R(w) = R^u + \lambda \cdot \int_0^w L(x)dF(x)$ where R^u is the stock of recruits from the pool of unemployment. Note that this does not vary across wage levels w since workers' values of unemployment are normalized to zero in Burdett and Mortensen (1998). Hence, unemployed workers accept any given offer. Given this structure, it is straightforward to derive that $\varepsilon_R = \lambda \cdot \frac{f(w)L(w)w}{R(w)} > 0$. While we focus here on the canonical model of Burdett and Mortensen (1998), upward-sloping labor supply curves are also present in more generalized settings such as Bontemps, Robin and Van den Berg (2001) and Mortensen (2003).

O.8.2 Additive Random Utility Models (ARUM)

In this section, we consider a class of additive random utility models as described in Chan, Kroft and Mourifie (2019). We do so because their setup nests a variety of labor market models which we will discuss below. There are K types indexed by k which each have a mass of m_k such that $\sum_{k=1}^{K} m_k = 1$. An individual worker i with type k (which is allowed to be multidimensional) is faced with the problem of choosing among a set of employers $\mathcal{J} = \{1, 2, \ldots, J\}$. Worker choice is informed by non-pecuniary benefits, wage compensation, and some idiosyncratic term. A worker's outside option is denoted by "employer" 0. Its maximization problem is characterized by:

$$\max_{j \in \mathcal{J} \cup \{0\}} u_{kj} + w_{kj} + \varepsilon_{ij} = \max_{j \in \mathcal{J} \cup \{0\}} v_{kj} + \varepsilon_{ij}$$

The surplus function is defined as:

$$\mathcal{S}(\mathbf{v}_k) = \mathbb{E}\left[\max_{j \in \mathcal{J} \cup \{0\}} v_{kj} + \varepsilon_{ij}\right]$$

Then, Chan, Kroft and Mourifie (2019) characterize the labor supply function as:

$$\begin{aligned} L_{kj} &= m_k \cdot \text{Pr}\left(v_{kj} + \varepsilon_{kj} \geq v_{kj'} + \varepsilon_{ij'}, \text{ for all } j' \in \mathcal{J} \cup \{0\}\right) \\ &= m_k \cdot \frac{\partial \mathcal{S}(\mathbf{v}_k)}{\partial v_{kj}} \end{aligned} \tag{76}$$

Chan, Kroft and Mourifie (2019) show that this object exists whenever ε_{ij} is independent of v_{kj} and is absolutely continuous with respect to the Lebesgue measure. Furthermore, the surplus function is convex in $\mathbf{v}_k$ under those assumptions. Hence, labor supply schedules are non-decreasing. Therefore, we have:

$$\varepsilon_S^{kj} = \frac{m_k}{L_{kj}} \frac{\partial^2 \mathcal{S}(\mathbf{v}_k)}{\partial^2 v_{kj}} w_{kj} \geq 0$$

The generalized setting of Chan, Kroft and Mourifie (2019) is quite convenient as it nests the setups of Card et al. (2018) and Lamadon, Mogstad and Setzler (2022). This can be done by appropriately defining worker types and assuming that idiosyncratic shocks are drawn from an Extreme Value Type I distribution.

O.8.3 Monopsonistic competition

In the simplest setting, upward-sloping labor supply curves are generated purely through preferences, even in the absence of strategic complementarities across firms. For instance, this would be true in a setting in which a representative household supplies a bundle of differentiated labor $\mathbf{L}_t = \{L_{it}\}_{i=1}^{K}$ and has preferences over some composite consumption bundle C_t.

Suppose the household's preferences are summarized by some function $u(C_t, \mathbf{L}_t)$ that is continuously differentiable in its arguments. Then, the schedule of labor supply functions is determined by a system of non-linear equations consisting of $\frac{(K+1)K}{2} + 1$ equations. Intuitively, labor supply schedules are upward sloping whenever substitution effects dominate their income counterparts.

Horizontal job differentiation. Under this class of models, workers are heterogeneous in their preferences over non-wage characteristics of a job. A simple way to capture this idea is to assume that a worker's utility is increasing in wages and decreasing in distance to work. Then, wages act as a compensating differential. Examples are Bhaskar and To (1999) and Staiger, Spetz and Phibbs (2010) who adopt frameworks in the spirit of Salop (1979).[15]

Double-nested CES preferences (Atkeson-Burstein). Berger, Herkenhoff and Mongey (Forthcoming) consider a monopsonistic environment in the tradition of Atkeson and Burstein (2008). With some abuse of notation, preferences are characterized by:

$$u\left(C_t - \frac{1}{\varphi^{\frac{1}{\varphi}}}\frac{\mathbf{L}_t^{1+\frac{1}{\varphi}}}{1+\frac{1}{\varphi}}\right) \text{ with } \mathbf{L}_t = \left(\int_0^1 \mathbf{L}_{jt}^{\frac{\theta+1}{\theta}}\, dj\right)^{\frac{\theta}{\theta+1}} \text{ and } \mathbf{L}_{jt} = \left(\sum_{f=1}^{F_j} n_{fjt}^{\frac{\eta+1}{\eta}}\right)^{\frac{\eta}{\eta+1}}$$

Thus, preferences follow the GHH specification in consumption and labor whereas labor is a double-nested CES composite. This gives rise to labor supply elasticities of the

[15] In particular, Staiger, Spetz and Phibbs (2010) assume that firms are uniformly distributed around a circle of measure one. Whenever the measure of firms N is fixed and workers' utility is increasing (decreasing) in their wage (distance to work), a firm i's labor supply function can be characterized as $L_i = \alpha + \tau^{-1}\left[w_i - \left(\frac{w_{i-1}+w_{i+1}}{2}\right)\right]$ where $\tau > 0$ denote travel costs (denoted in units of utility) per unit distance. Given this structure, we must have $\varepsilon_S > 0$.

form:

$$\varepsilon_S = \frac{1}{\eta} + \left(\frac{1}{\theta} - \frac{1}{\eta}\right) \cdot s > 0$$

where $s \in [0, 1]$ is a firm's share of the industry's total payroll. The latter is guaranteed to be positive whenever $\eta > \theta$ which is the more natural assumption.

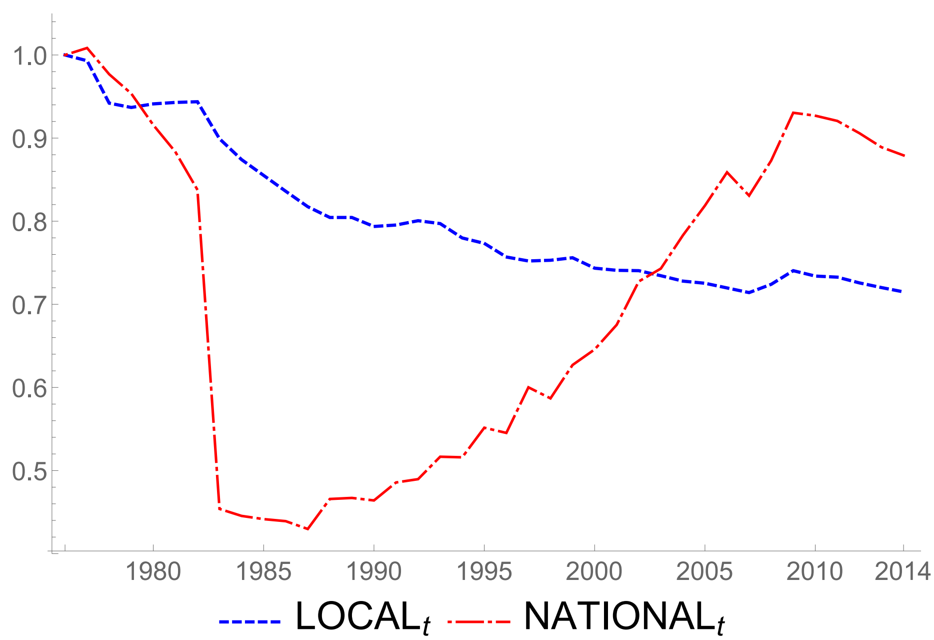
O.9 Concentration indices

NATIONAL CONCENTRATION. We construct national employment concentration, following Autor et al. (2020), as follows:

$$\begin{aligned} \text{NATIONAL}_t &= \sum_{j \in J} \omega_{jt} \text{HHI}_{jt} \\ &= \sum_{j \in J} \omega_{jt} \left[\sum_{f \in F_t(j)} \left(\frac{x_{ft}}{X_{F(j)t}} \right)^2 \right] \quad \text{s.t. } X_{F(j)t} = \sum_{f' \in F_t(j)} x_{f't} \end{aligned} \tag{77}$$

Hence, national concentration is a weighted average of industry-level HHIs. We implement this measure by using employment weights and by calculating HHI_{jt} at the 3-digit NAICS-year level. The results are displayed in the figure below.

Figure 9: National employment concentration has been increasing since the early 1980s.



HHI levels are normalized relative to their initial value in 1976. Source: Authors' own calculations from LBD data from 1976–2014.

Consistent with Autor et al. (2020), we find that national employment concentration has been rising since the early 1980s. If we look at the whole available period of 1976 – 2014, then it is clear that national concentration has not been rising monotonically. In fact, it was declining from 1976 till 1981 with a particularly sharp drop in 1982 which is consistent with Rinz (2020). While it is tempting to explain this almost continuous drop as measurement error, it is unlikely to be the case with administrative data. Furthermore, Rinz (2020) has argued that it is mainly driven by telecommunications industries and refers to a Department of Justice case in 1982 in which AT&T was required to divest itself of local telephone companies.

Regardless of the rationale behind this drop, it is clear that the time series for national employment concentration does not follow the patterns of our constructed markdown $\mathcal{V}_t$ in the least. Hence, we conclude that caution should be exercised when proxying market power with measures of concentration.

CONCENTRATION IN VACANCIES. We use two sources of data to investigate labor market concentration: employment data from the Longitudinal Business Database (LBD)—as seen in the main body—and vacancy data from Burning Glass Technologies (BGT).

The BGT data is a unique source of micro-data that contains approximately 160 million electronic job postings in the U.S. economy spanning the years 2007 and 2010–2017. These job postings were collected and assembled by BGT, an employment analytics and labor market information company, that examines over 40,000 online job boards and company websites to aggregate the job postings, parse, and deduplicate them into a systematic, machine-readable form, and create labor market analytics products. With the breadth of this coverage, the resulting database purportedly captures the near-universe of jobs posted online, estimated to be near 80 percent of total job ads. Using BGT vacancy data allows us to compute the concentration of job openings, thus zeroing in on concentration in local labor demand and computing an index of concentration that reflects how many employers are active in the hiring process in a local market.

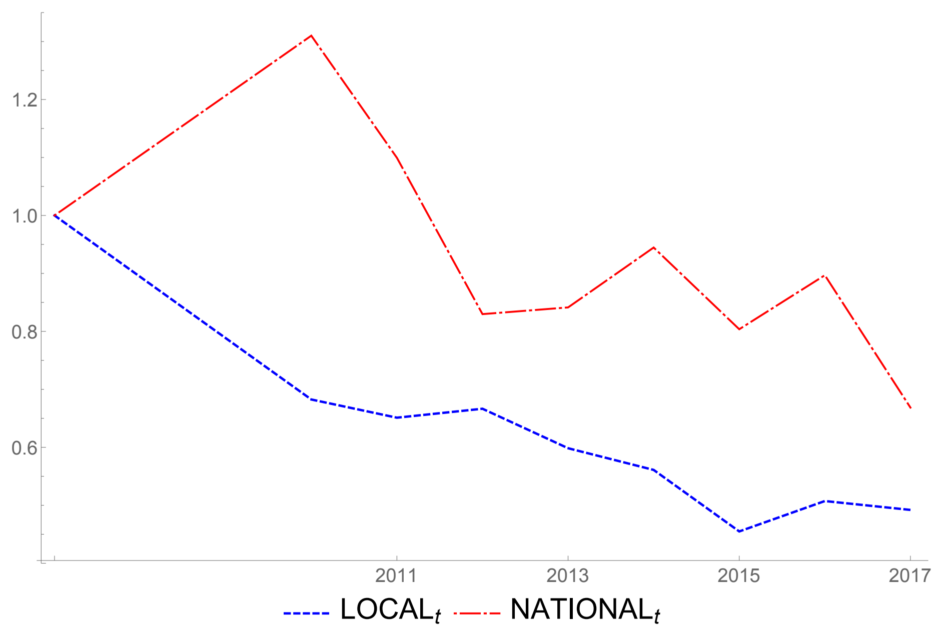
The BGT data has both extensive breadth and detail. Unlike sources of vacancy data that are based on a single job board such as `careerbuilder.com` or `monster.com`, BGT data span multiple job boards and company sites. The data are also considerably richer than sources from the Bureau of Labor Statistics (BLS), such as JOLTS (Job Openings and Labor Turnover Survey).[16] In addition to detailed information on occupation, geography,

[16]Although JOLTS asks a nationally representative sample of employers about vacancies they wish to fill in the near term, the data are typically available only at aggregated levels, and do not allow for a detailed

and employer for each vacancy, BGT data contain thousands of specific skills standardized from open text in each job posting. BGT data thus allow for a detailed analysis of vacancy flows within and across occupations, firms, and labor market areas, enabling us to document trends in employers' concentration at a very granular level.

The data, however, is not perfect. Although roughly two-thirds of hiring is replacement hiring, we expect vacancies to be somewhat skewed towards growing areas of the economy (Davis, Faberman and Haltiwanger, 2012; Lazear and Spletzer, 2012). Additionally, the BGT data only covers online vacancies. Even though vacancies for available jobs have increasingly appeared online rather than in traditional sources, it is a valid concern that the types of jobs posted online are not representative of all openings. Hershbein and Kahn (2018) provide a detailed description of the industry-occupation mix of vacancies in BGT relative to JOLTS: although BGT postings are disproportionately concentrated in occupations and industries that require greater skill, the distributions are stable across time, and the aggregate and industry trends in BGT track BLS sources closely.

Figure 10: National and local trends in the concentration of job postings.



HHI levels are normalized relative to their initial value in 2007. Observations from the Great Recession (2008–2009) are not available and are interpolated from 2007 to 2010. Source: BGT (2007, 2010–2017).

In the BGT data, we define a local labor market as an occupation-metro area pair. We define

taxonomy of local labor markets.

occupations at the 4-digit SOC level, for a total of 108 groups derived from the BLS 2010 SOC system, which aggregates "occupations with similar skills or work activities" (BLS, 2010). While our definition of occupations is considerably less detailed than the job titles available in the BGT data, we believe it offers an appropriate balance between accurately capturing the competitiveness of a market and identifying the demand for different bundles of skills.[17] Nevertheless, our results hold true for other classifications.[18] Metropolitan areas correspond to the 2013 Core-Based Statistical Areas (CBSA) with a population over 50,000. As a result, there are 382 metro areas in our final BGT dataset. In the end, we identify 41,256 local labor markets in the BGT data.

We regard vacancies concentration as the closest measure to the concentration faced by job seekers in a specific (local or national) labor market. We construct local and national concentration measures of vacancies using BGT data. Market-level HHIs are aggregated through their respective vacancy shares.[19] Figure 10 plots the time series of the aggregate local and national concentration of vacancies and shows that local concentration is markedly decreasing over time. Specifically, the local HHI of vacancies drops in the post recession period 2010–2017 by approximately 20 percent. The decrease is even more dramatic if we consider the change between 2007 and 2017—though it is to be noted that the BGT data is not available during 2008–09. Note that the pattern for the national concentration of vacancies is comparable to its employment counterpart.

[17]Indeed, too fine an occupational classification would mechanically lead to a small number of firms posting jobs in each market. This would bias our estimates of labor market concentration upward. On the other hand, too broad an occupational classification would erase important distinctions between heterogeneous skills used in different occupations. Even though many studies find that broad occupational changes are not uncommon in U.S. labor markets (Huckfeldt, 2017; Macaluso, 2019), especially for laid-off workers, we choose the 4-digit SOC level as a useful compromise.

[18]Examples of 4-digit SOC occupations among Production ones are Food Processing Workers (5130), Assemblers and Fabricators (5120), Textile, Apparel, and Furnishings Workers (5160), and Plant and System Operators (5180).

[19]Our results are quantitatively unaffected whenever we use employment shares instead.

References

Ackerberg, D., K. Caves, and G. Frazer, "Identification Properties of Recent Production Function Estimators," *Econometrica*, 2015, *83* (6), 2411 – 2451.

Atalay, E., "Materials Prices and Productivity," *Journal of the European Economic Association*, 2014, *12* (3), 575 – 611.

Atkeson, A. and A. Burstein, "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, 2008, *98* (5), 1998 – 2031.

Autor, D., D. Dorn, L. Katz, C. Patterson, and J. Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics*, 2020, *135* (2), 645–709.

Basu, S. and J.G. Fernald, "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, 1997, *105* (2), 249 – 283.

Berger, D., K. Herkenhoff, and S. Mongey, "Labor Market Power," *American Economic Review*, Forthcoming.

Bhaskar, V. and T. To, "Minimum Wages for Ronald McDonald Monopsonies: A Theory of Monopsonistic Competition," *Economic Journal*, 1999, *109*, 190–203.

Blundell, B. and S. Bond, "GMM Estimation with Persistent Panel Data: An Application to Production Functions," *Econometric Reviews*, 2000, *19* (3), 321–340.

Bond, S., A. Hashemi, G. Kaplan, and P. Zoch, "Some Unpleasant Markup Arithmetic: Production Function Elasticities and Their Estimation from Production Data," *Journal of Monetary Economics*, 2021, *121*, 1–14.

Bontemps, C., J.-M. Robin, and G.J. Van den Berg, "An Empirical Equilibrium Job Search Model with Search on the Job and Heterogeneous Workers and Firms," *International Economic Review*, 2001, *40* (4), 1039–1074.

Brooks, W.J., J.P. Kaboski, Y.A. Li, and W. Qian, "Exploitation of Labor? Classical Monopsony Power and Labor's Share," *Journal of Development Economics*, 2021, *150* (102627).

Burdett, K. and D.T. Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 1998, *39* (2), 257–273.

Card, D., A.R. Cardoso, J. Heining, and P. Kline, "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 2018, *36* (1), 13–70.

Chan, M., K. Kroft, and I. Mourifie, "An Empirical Framework for Matching with Imperfect Competition," Working Paper, 2019.

Collard-Wexler, A. and J. De Loecker, "Production Function Estimation and Capital Measurement Error," Working Paper, 2020.

Cooper, R., J. Haltiwanger, and J.L. Willis, "Search Frictions: Matching Aggregate and Establishment Observations," *Journal of Monetary Economics*, 2007, *54* (S1), 56–78.

Davis, S. J., C. Grim, J. Haltiwanger, and M. Streitwieser, "Electricity Unit Value Prices and Purchase Quantities: U.S. Manufacturing Plants, 1963 - 2000," *Review of Economics and Statistics*, 2013, *95* (4), 1150–1165.

Davis, S.J., R.J. Faberman, and J. Haltiwanger, "Labor market flows in the cross-section and over time," *Journal of Monetary Economics*, 2012, *59* (1), 1–18.

Dey, M., S.N. Houseman, and A.E. Polivka, "Manufacturers' Outsourcing to Staffing Services," *ILR Review*, 2012, *65* (3), 533 – 559.

Dobbelaere, S. and J. Mairesse, "Panel Data Estimates of the Production Function and Product and Labor Market Imperfections," *Journal of Applied Econometrics*, January 2013, *28* (1), 1–46.

Flynn, Z., A. Gandhi, and J. Traina, "Identifying Market Power in Production Data," Working Paper, 2019.

Foster, L., J. Haltiwanger, and C.J. Krizan, "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," in Charles R. Hulten, Edwin R. Dean, and Michael J. Harper, eds., *New Developments in Productivity Analysis*, Chicago: University of Chicago Press, 2001, chapter 8, pp. 303–372.

Gandhi, A., S. Navarro, and D.A. Rivers, "On the Identification of Gross Output Production Functions," *Journal of Political Economy*, 2020, *128* (8), 2973–3016.

Hall, R.E., "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 1988, *96* (5), 921–947.

_ , "Measuring Factor Adjustment Costs," *Quarterly Journal of Economics*, 2004, *119*, 899–927.

Hu, Y., G. Huang, and Y. Sasaki, "Estimating production functions with robustness against errors in the proxy variables," *Journal of Econometrics*, 2020, *215* (2), 375–398.

Huckfeldt, C., "Understanding the Scarring Effects of Recessions," Working Paper, 2017.

Kehrig, M., "The Cyclical Nature of the Productivity Distribution," Working Paper, 2015.

_ **and N. Vincent**, "The micro-level anatomy of the labor share decline," *Quarterly Journal of Economics*, 2021, *136* (2), 1031–1087.

Klette, T.J. and Z. Griliches, "The Inconsistency of Common Scale Estimators When Output Prices are Unobserved and Endogenous," *Journal of Applied Econometrics*, 1996, *11* (4), 343–361.

Lamadon, T., M. Mogstad, and B. Setzler, "Imperfect Competition, Compensating Differentials and Rent Sharing in the U.S. Labor Market," *American Economic Review*, 2022, *112* (1), 169–212.

Lazear, E.P. and J.R. Spletzer, "Hiring, Churn, and the Business Cycle," *American Economic Review*, 2012, *102* (3), 575–579.

Levinsohn, J.A. and A. Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 2003, *70* (2), 317–340.

Loecker, J. De and F. Warzynski, "Markups and Firm-Level Export Status," *American Economic Review*, 2012, *102* (6), 2437–2471.

Macaluso, C., "Skill Remoteness and Post-Layoff Labor Market Outcomes," Working Paper, 2019.

McDonald, I.M. and R.M. Solow, "Wage Bargaining and Employment," *American Economic Review*, 1981, *71* (5), 896–908.

Mortensen, D., "How Monopsonistic is the (Danish) Labor Market?," in P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, eds., *Knowledge, Information and Expectations in Modern Macroeconomics*, Princeton, NJ: Princeton University Press, 2003.

Olley, S. and A. Pakes, "The Dynamics of Productivity in the Telecommunications Industry," *Econometrica*, 1996, *64*, 1263–1298.

Rinz, K., "Labor Market Concentration, Earnings, and Inequality," *Journal of Human Resources*, 2020, pp. 0219–10025R1.

Rossi-Hansberg, E., P.-D. Sarte, and N. Trachter, "Diverging Trends in National and Local Concentration," *NBER Macroeconomics Annual*, 2020, *35*.

Salop, S., "Monopolistic Competition with Outside Goods," *Bell Journal of Economics*, 1979, *10*, 141–156.

Sokolova, A. and T. Sorensen, "Monopsony in Labor Markets: A Meta-Analysis," *ILR Review*, 2020, *74* (1), 27–55.

Staiger, D.O., J. Spetz, and C.S. Phibbs, "Is There Monopsony in the Labor Market? Evidence from a Natural Experiment," *Journal of Labor Economics*, 2010, *28* (2), 211–236.

Syverson, C., "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 2004, *112* (6), 1181–1222.

_ , "Product Substitutability and Productivity Dispersion," *Review of Economics and Statistics*, 2004, *86* (2), 534–550.