Some Shifts in the Distributions of Revenue Shares and Some Implications for the Estimation of Markups*

Steve Bond† Giulio Gottardo‡

February 28, 2024

Abstract

We highlight two trends in the distributions of input costs as shares of sales revenue for publicly traded US firms. We note that these trends result in overestimation of the rise in price-cost markups using the approach of De Loecker et al. (2020). We suggest a different strategy which exploits the relation between the returns to scale in the production technology, the returns to scale in the revenue production function, and the price-cost markup for profit-maximising firms. This requires estimates only of average revenue elasticities for each of the inputs, and we show that observed revenue shares can be used to obtain informative bounds for these average revenue elasticities under a set of standard assumptions. Assuming constant returns to scale production technologies, our estimated bounds are consistent with a more modest increase in the weighted average markup. They are consistent with a substantial increase in the cross-section dispersion of markups.

JEL Codes: D2, D4, L1, L2

Keywords: Markups, Revenue Shares, Returns to Scale, Revenue Production Functions.

*We thank Giovanni Cattani, Andrea Chiavari, Ilario Ciccolella, Michael P. Devereux, Manos Perdikakis, Howard Smith, Måns Söderbom, and John Van Reenen for helpful discussions.

†University of Oxford; steve.bond@nuffield.ox.ac.uk

‡University of Oxford; giulio.gottardo@pmb.ox.ac.uk

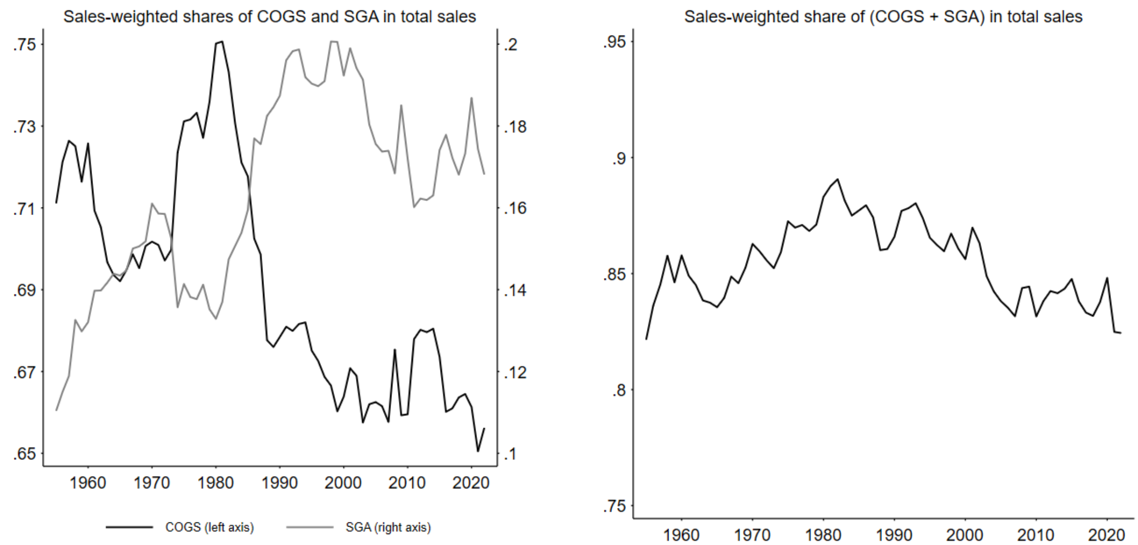
1 Introduction

Other than costs which are associated with the ownership and use of capital, company accounts disaggregate input costs into two components. The 'Cost of Goods Sold' (COGS) covers costs of labour and materials that can be assigned to the production of specific goods or services sold by the firm. 'Selling, General and Administrative expenses' (SGA) cover other costs that cannot be assigned in that way, including executive compensation, R&D and HR expenses, and advertising and marketing expenses. The difference between sales revenue and the sum of COGS and SGA gives the familiar measure of 'Earnings Before Interest, Taxes, Depreciation and Amortisation' (EBITDA).

Two trends in the distributions of these input costs are evident in the Compustat company accounts data for publicly traded US firms studied in De Loecker et al. (2020). First, over the period since 1980, there was a shift away from costs which are classified as COGS and towards costs which are classified as SGA. Figure 1 plots weighted average measures of both components of input costs expressed as shares of sales revenue, and their sum.[1] Since the early 1980s, COGS has declined from around 75% of sales to around 65% of sales, while SGA has increased from around 13% of sales to around 17% of sales on this weighted average basis. Over the same period, the sum of these two cost items accounted for a relatively stable share of sales, declining from around 88% to around 82% on this weighted average basis.

[1] Here we report weighted averages of revenue shares using sales weights, and we follow the data cleaning procedures used in De Loecker et al. (2020). In Section 2.2 we show that these patterns are highly robust to using different weights.

Figure 1: Weighted average revenue shares



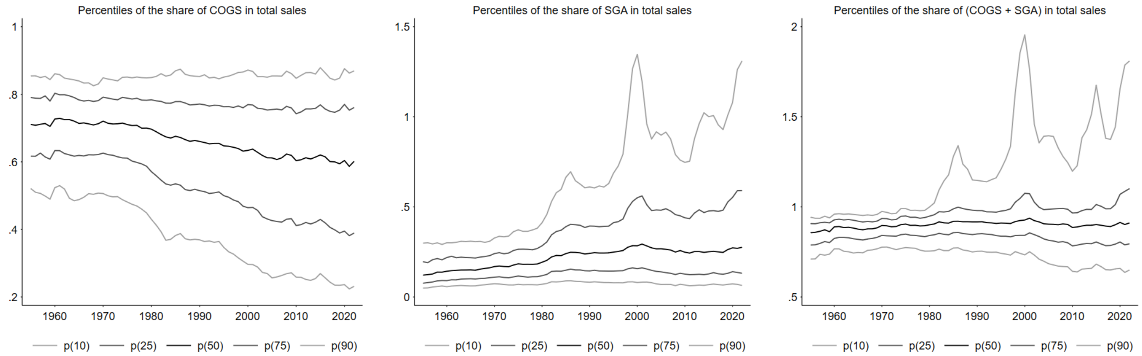
Second, there was a trend increase in the dispersion of each of these revenue shares between firms in the same year. Figure 2 plots the evolution of various quantiles of the cross-section distribution for each of these revenue shares. The median share declined for COGS, increased for SGA, and was relatively stable for the sum of these two cost items, in line with the trends in the corresponding weighted averages.[2] In each case, the interquartile range and the gap between the 90th and 10th percentiles increased substantially, particularly after 1980, as did any other reasonable measure of cross-section dispersion. These patterns remain if we account for sector, firm size and firm age fixed effects in these revenue shares, if we consider dispersion within sectors, and if we consider the residual variance in simple regressions of the revenue shares on the measured expenditures and the book value of capital (see Section 2.3 for further details).

In this paper we note that these shifts in the distributions of these revenue shares help to account for an increase in the gap between the revenue share for COGS and estimates of the revenue elasticity for COGS obtained from revenue production functions.[3] Figure 3 plots a

[2]For the sum of COGS and SGA, the median revenue share shown in Figure 2 is more stable than the weighted average revenue share shown in Figure 1. The decline after 1980 in this revenue share is mainly a feature of the data for larger firms. Note also that firms which report a sum of COGS and SGA which exceeds sales revenue are reporting negative values for EBITDA.

[3]Revenue production functions differ from production functions by being models for revenue, rather than models for output. The specifications considered in De Loecker et al. (2020), and throughout this paper, have the (natural) logarithm of sales revenue as the dependent variable. Deflating sales revenue using a common price deflator does not produce a measure of output under conditions of imperfect competition when firms differ in their market power.

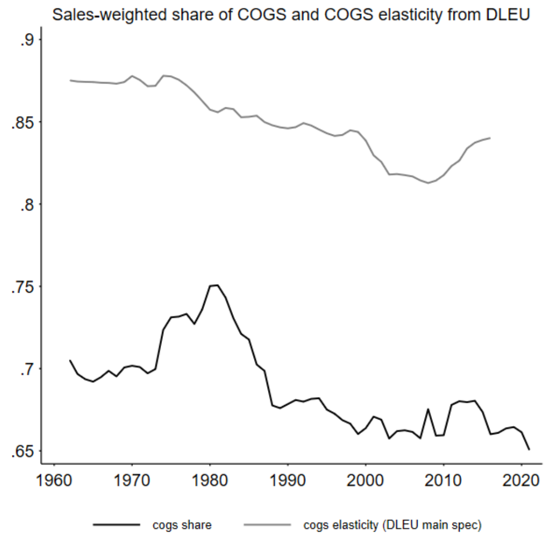
Figure 2: Distributions of revenue shares



weighted average across sectors of the elasticities for COGS reported by De Loecker et al. (2020), and the corresponding weighted average revenue share for COGS.[4] The estimated revenue elasticity exceeds the revenue share by a considerable margin, and the gap between the two series increased after the early 1980s. The same pattern is found using a range of other methods to estimate sector-specific revenue elasticities from a revenue production function specification (see Section 3 for further details). If we ignore the distinction between revenue elasticities and output elasticities and assume that the input bundle represented by expenditure on COGS is perfectly flexible and optimally chosen, then the ratio of the elasticity to the revenue share estimates the markup of price over marginal cost (De Loecker and Warzynski, 2012). This growing gap underpins the dramatic rise in estimated price-cost markups illustrated in Figure 1 of De Loecker et al. (2020), particularly in the period since 1980.

[4]For comparability, we also use sales weights here, although other weights may be more appropriate in this context (Yeh et al. (2022); Edmond et al. (2023)). We present versions of this figure using different weights in Appendix B. The elasticities are estimated using rolling 8 year sample periods which end in the year for which the elasticities are plotted.

Figure 3: COGS revenue share and revenue elasticity



For inputs which are perfectly flexible and optimally chosen, the true revenue elasticity should coincide with the true revenue share under quite general conditions (Bond et al., 2021). If we assume that COGS is such an input, this growing gap suggests that we are overestimating true revenue elasticities, and that this upward bias has increased over time. One contributory factor is the omission of the input bundle represented by expenditure on SGA from the main specifications of the revenue production function used in De Loecker et al. (2020). We show that the omitted SGA input is both relevant and positively correlated (in log levels) with the included COGS input. Included COGS then proxies for the omitted variable, and estimates of the revenue elasticity for COGS that we obtain from specifications which include SGA are substantially lower.

The increase in the cross-section dispersion of revenue shares also suggests a similar increase in the cross-section dispersion of revenue elasticities. Importantly, this pattern is found within sectors, and not only between sectors. For Cobb-Douglas specifications, these heterogeneous revenue elasticities are the slope parameters that we seek to estimate from log-linear revenue production functions.[5] Standard estimation methods require that these slope parameters are common to all firms in the estimation sample if we are to obtain

[5]For more general functional forms, such as the translog, heterogeneity in revenue elasticities may also reflect non-zero coefficients on higher-order terms. To the extent that the increase in the cross-section dispersion of revenue elasticities reflects an increase in the cross-section dispersion of markups, the slope parameters in these more general specifications will also be heterogeneous.

consistent estimates. With heterogeneous slope parameters, we obtain biased estimates of average revenue elasticities. This bias has also increased over time, in line with the increase in the dispersion of revenue shares illustrated in Figure 2.

Can we learn anything about the evolution of price-cost markups from the estimation of revenue elasticities using company accounts data of this kind? We suggest a different strategy which exploits the relation between the returns to scale in the production technology, the returns to scale in the revenue production function, and the price-cost markup for profit-maximising firms. To illustrate, suppose that the production function is Cobb-Douglas and that demand is constant elasticity of substitution (CES). The revenue production function remains log-linear, with revenue elasticity parameters which can be expressed as the ratio of the output elasticity for each input to the price-cost markup.[6] The returns to scale in the revenue production function is the sum of the revenue elasticities for each of the inputs. This sum is simply the returns to scale in the output production function (i.e. the sum of the output elasticities) divided by the price-cost markup. Then if we are willing to assume constant returns to scale (CRS) in the output production function, the returns to scale of the revenue production function is the reciprocal of the markup. We show that these results extend to any admissable forms for the production function and the demand schedule.

The challenge in implementing this approach is to obtain reliable estimates of revenue elasticities for each of the inputs, allowing for unobserved heterogeneity in the true revenue elasticities, for measurement error in revenue,[7] and, at least in the case of capital, for inputs which are chosen in advance of production. De Loecker et al. (2020) treat the bundle of inputs represented by expenditure on COGS as an input which is optimally chosen after the firm has learned its productivity and demand for that period and in the absence of any adjustment costs. In that case, Bond et al. (2021) show that the true revenue elasticity coincides with the true revenue share for profit-maximising firms. If the true revenue elasticity is a common parameter within a sample, and if there is no measurement error in the data, then the observed revenue shares would be the same for all firms in the sample. In that case, both the arithmetic mean of the observed revenue shares and the geometric mean of the observed revenue shares would coincide with the common parameter of interest, and with each other.

In practice, we have considerable dispersion in the observed revenue shares, even within sector-year samples. If that dispersion accurately reflects dispersion in the true revenue elasticities, with no measurement error in the data, then the arithmetic mean of the observed revenue shares still coincides with the arithmetic mean of the true revenue elasticities; while

[6] See, for example, Klette and Griliches (1996).

[7] More specifically, we should allow for the discrepancy between sales revenue and the value of production which arises if there is any net change to inventories of final products.

the geometric mean of the observed revenue shares is necessarily lower than their arithmetic mean, and so underestimates the arithmetic mean of the true revenue elasticities. Allowing for symmetric, mean zero, additive measurement error in the log of revenue - which is the form of measurement error that is commonly considered in the literature on the estimation of production functions - we show that the arithmetic mean of the observed revenue shares then overestimates the arithmetic mean of the true revenue elasticities. The geometric mean of the observed revenue shares is not affected by this form of measurement error, and so continues to underestimate the arithmetic mean of the true elasticities. These results indicate that for inputs which are flexible in the sense required by De Loecker et al. (2020), we can obtain upper and lower bounds for the average revenue elasticity in a sample of firms using the arithmetic mean and the geometric mean of the observed revenue shares, respectively. Similar bounds are available for size-weighted average revenue elasticities in a sample of firms, using the corresponding size-weighted arithmetic mean and size-weighted geometric mean of the observed revenue shares.

In our main results, we apply this approach to obtain upper and lower bounds for the size-weighted average revenue elasticities for COGS and SGA in each year of the sample. For capital, we follow De Loecker et al. (2020) in assuming that the stock of capital used in production in the current year was determined previously, based on information that was available in the previous year. Given that, we cannot assume that capital is optimally chosen after the firm has learned its productivity and demand for the current year, unless we are willing to assume that the firm has perfect foresight. For capital, or more generally for inputs which are chosen before the firm knows its productivity and demand for the current year, we show that the harmonic mean of the observed revenue shares provides a lower bound for the average revenue elasticity.[8] A further complication is that capital costs are not reported in company accounts. We impute capital costs in two ways: one following the method used by De Loecker et al. (2020), and one incorporating a firm-year-specific risk premium component in the cost of capital, using firm-year-specific betas obtained from WRDS Beta Suite. In practice, the implied revenue shares for capital are small enough, relative to the sum of the observed revenue shares for COGS and SGA (see Figure 1), for these complications to be of limited significance for the trends in the upper and lower bounds for the weighted average returns to scale in the revenue production functions, which is our main object of interest.

Henceforth we abbreviate the weighted average returns to scale in the revenue production functions to the weighted average revenue returns to scale, where the meaning is clear. To obtain our upper bound, we sum the upper bounds for the weighted average revenue elasticity for each of the inputs, and we obtain our lower bound similarly. Figure 4 plots

[8] The arithmetic mean of the observed revenue shares continues to provide an upper bound.

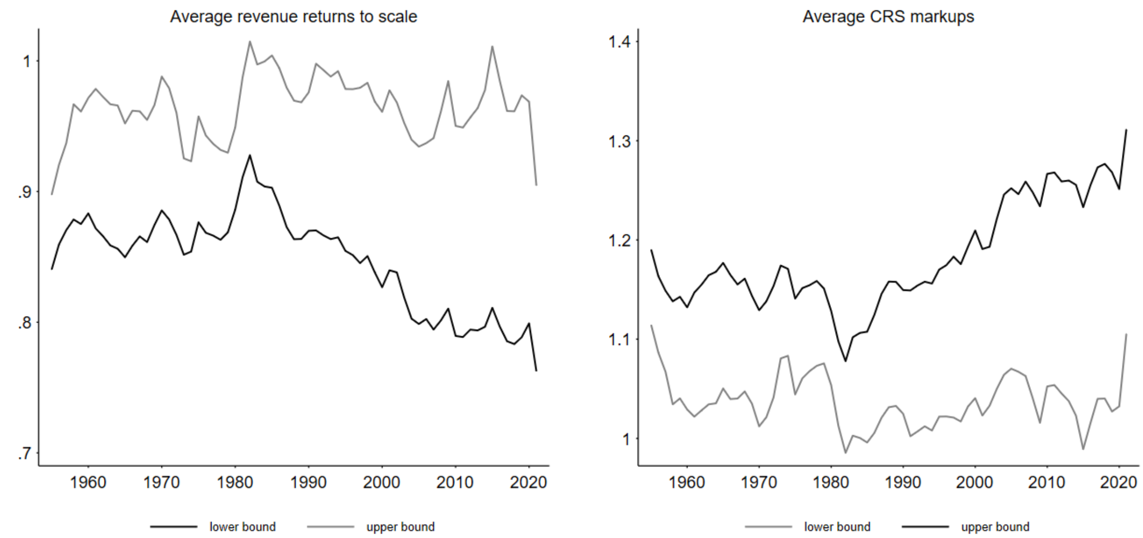
the resulting upper and lower bounds for the weighted average revenue returns to scale for publicly traded US firms.[9] These bounds are consistent with a modest decrease in the weighted average revenue returns to scale after 1980, from an interval between 0.92-1 in the early 1980s to an interval between 0.77-0.92 at the end of our sample period in 2022. Assuming that the underlying production technologies satisfy constant returns to scale, the reciprocal of our upper bound in Figure 4 gives a lower bound for the weighted average markup, and the reciprocal of our lower bound in Figure 4 gives an upper bound for the weighted average markup. Under the CRS assumption, the weighted average markup lay within the interval 1-1.09 in the early 1980s, and within the interval 1.09-1.3 in 2022. The increase over time in the width of these intervals is driven by the increase in the cross-section dispersion of revenue shares illustrated in Figure 2, and consistent with a substantial increase in the cross-section dispersion of markups.

Assuming constant returns to scale production technologies, our upper bound for the weighted average markup is considerably lower than the values reported in Figure 1 of De Loecker et al. (2020). The trends in our bounds also suggest a more modest increase in the weighted average markup from 1980 onwards. A striking feature of our results is the likely increase in the cross-section dispersion of markups over the last 40 years. That is consistent with the emergence of 'superstar' firms with considerable market power over this period, as suggested by Autor et al. (2020).

The remainder of the paper is organised as follows. Section 2 outlines the data that we use, and presents further evidence on both the shift in the composition of input costs away from COGS and towards SGA, and on the increase in the cross-section dispersion of these revenue shares, both within-sectors and overall. Section 3 presents further evidence on the gap between the revenue share for COGS and econometric estimates of the revenue elasticity for COGS obtained from revenue production functions, considering a range of econometric methods, and specifications with and without the inclusion of SGA as an additional input. Section 4 outlines the key relation between the returns to scale in the output production function, the returns to scale in the revenue production function, and the price-cost markup, which is at the core of the approach to studying the evolution of markups that we develop in this paper. Section 5 outlines a set of assumptions under which we can obtain informative upper and lower bounds for the (weighted) average revenue elasticity for each input in a sample of firms, using arithmetic, geometric, and harmonic means of the observed revenue

[9]The size weights used to obtain the bounds in Figure 4 are total costs, measured as the sum of COGS, SGA, and capital costs imputed using firm-year-specific risk premia. The sample used here additionally drops observations which have extreme values for SGA or imputed capital costs as shares of sales. Appendix A provides details of this additional data cleaning. We present versions of this figure using different samples, different size weights, different imputed capital costs, and more conservative bounds in Section 6 and Appendix G.

Figure 4: Revenue returns to scale and CRS markup bounds



shares. Section 6 presents and discusses our main results from applying this approach to accounting data for publicly traded US firms over the period 1955-2022. Section 7 concludes.

2 Data and descriptive evidence

2.1 Sample

We use company accounts data compiled by Compustat North America for publicly traded US firms, in the sense of firms which are required to file accounts with the Securities and Exchange Commission (SEC). We use data for all available years from 1955 to 2022. This period saw important changes in the composition of the sample, with a decline in the proportion of firms that are listed on the New York Stock Exchange (NYSE), and a rise in the proportion of firms that are listed on the NASDAQ or whose shares are traded on an over-the-counter exchange. Appendix A provides further detail on the composition of the sample. We follow De Loecker et al. (2020) in presenting results for all publicly traded firms in the main sections of this paper.

The main variables used in our analysis are net sales/turnover (sales), cost of goods sold (COGS), selling, general and administrative expenses (SGA), and gross property, plant, and equipment (PPE). We follow De Loecker et al. (2020) in using the gross book value of PPE at the end of the year as the measure of the capital input in our revenue production function specifications.

For the results presented in Sections 2 and 3, we follow the same data cleaning steps as De Loecker et al. (2020), dropping observations which have missing or zero values for sales, COGS, SGA, or PPE, or extreme values for the revenue share of COGS. For the results presented in Section 6, we additionally drop observations with extreme values for the revenue shares of SGA or imputed capital costs, using the same criteria that De Loecker et al. (2020) use for the revenue share of COGS. The methods used to impute capital costs are discussed in Section 5 and detailed in Appendix A. Appendix A also provides further detail on these data cleaning procedures.

2.2 The shift from COGS to SGA

Figures 5 and 6 present weighted average revenue shares for COGS, SGA, and their sum, using as size weights either COGS or total costs, measured as the sum of COGS, SGA, and capital costs imputed using firm-year-specific risk premia. The trends in these weighted average series are very similar to those constructed using sales weights, as shown in Figure 1 and discussed in the Introduction.

Figure 5: Total cost-weighted average revenue shares

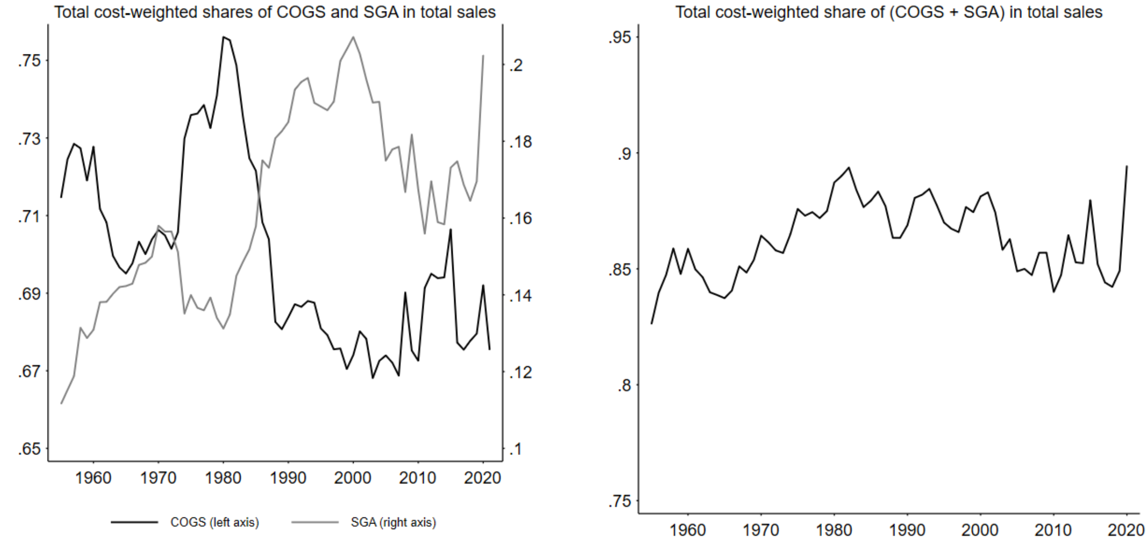
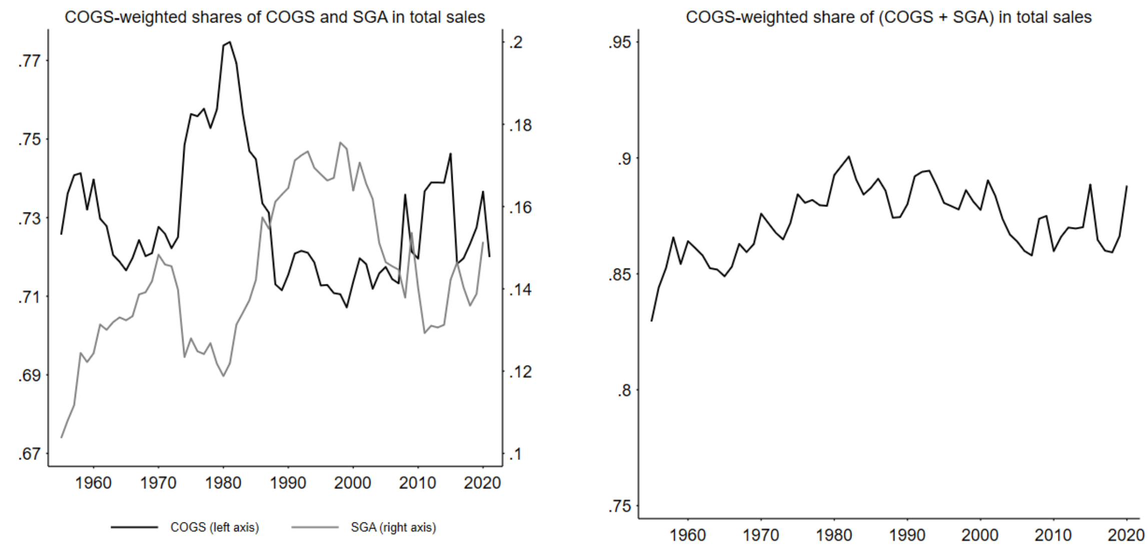


Figure 6: COGS-weighted average revenue shares

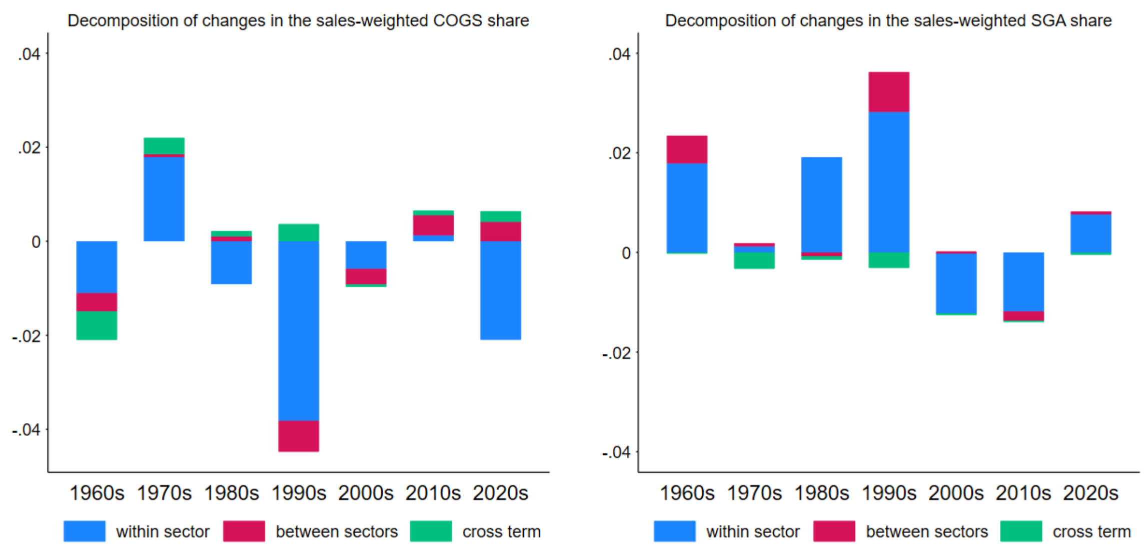


To investigate whether the aggregate shift from COGS to SGA was driven by reallocation of economic activity towards more SGA-intensive sectors, or by within-sector dynamics, we employ a standard decomposition at the sectoral level:

$$\Delta\bar{\alpha}_t^X = \sum_s w_{s,t-1}\Delta\bar{\alpha}_{st}^X + \sum_s \Delta w_{st}\bar{\alpha}_{s,t-1}^X + \sum_s \Delta w_{st}\Delta\bar{\alpha}_{st}^X$$

where $\Delta\bar{\alpha}_t^X$ is the overall change in the sales-weighted revenue share of input X between period $t-1$ and t, $\bar{\alpha}_{st}^X$ is the sales-weighted revenue share of input X in sector s in period t, and w_{st} are sector-level sales weights in period t. The first term captures within-sector changes, since it holds the sector weights constant at their values in the previous period. The second term captures between-sector reallocation, as it accounts for changes in the sector weights while holding the sector shares constant. Finally, the last term measures the residual variation, which comes from joint changes to the sector shares and weights. Since both shares and weights are quite persistent, the contribution of this cross-term will be small. Using this decomposition for changes between each of the decades covered by our sample, Figure 7 documents that most of the shift from COGS to SGA has happpened within NAICS 2 digit sectors. Shifts to more SGA-intensive sectors did happen, particularly between the 1980s and 1990s, but they account for less than one quarter of the overall change even over that period.

Figure 7: Decomposition of the changes in aggregate shares



2.3 Increasing dispersion

We first establish that the trend increase in the cross-section dispersion of these revenue shares remains after accounting for sector, firm size, and firm age fixed effects. Separately for each year, we regress the (natural) logarithm of the revenue share for COGS on sets of dummy variables for:

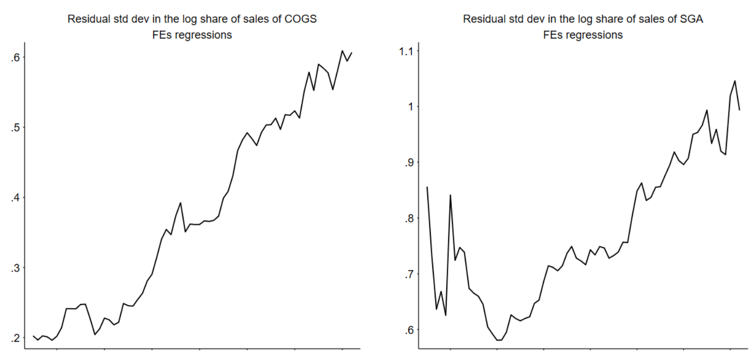
- each 2 digit NAICS sector covered by the sample

- each decile of the firm size distribution, where size is measured by sales revenue in constant prices
- each decile of the firm age distribution, where age is measured by years since the IPO
- all pairwise interactions between these sector, size and age dummies
- triple interactions between these sector, size and age dummies

We also estimate the same specification for the (natural) logarithm of the revenue share for SGA. Figure 8 plots the standard deviation of the residuals from these regressions. Accounting for sector, firm size and firm age, we still see a strong upward trend in the dispersion of these revenue shares, particularly after the mid-1970s.

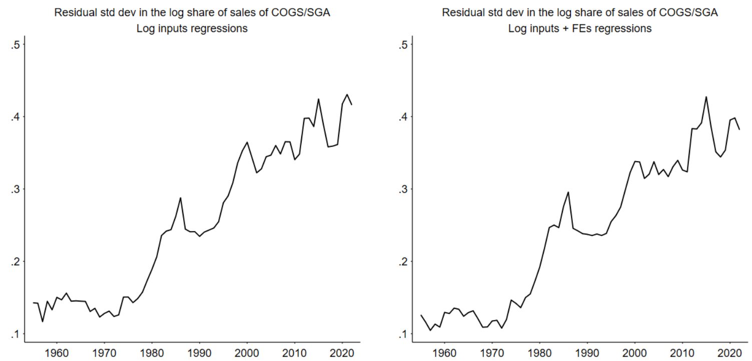
Dispersion in these revenue shares could also be explained by changes in the mix of inputs used by the firms in our sample, for example if the underlying production functions are not Cobb-Douglas. To explore this, separately for each year, we regress the (natural) logarithm of the revenue share for COGS on the (natural) logarithms of COGS, SGA, and PPE, all in constant prices.[10] Figure 9 plots the standard deviation of the residuals from these regressions, and that from a more general specification which additionally includes the full set of sector, size and age dummies outlined above, and their interactions. Accounting for the observed mix of inputs, as well as for sector, firm size and firm age, we still see a strong upward trend in the dispersion of these revenue shares, particularly after the mid-1970s.

Figure 8: Dispersion, accounting for sector, size and age



[10] Given the log-log specification that we use here, the residual variance is identical for specifications which have the (natural) logarithm of the revenue share for SGA as the dependent variable.

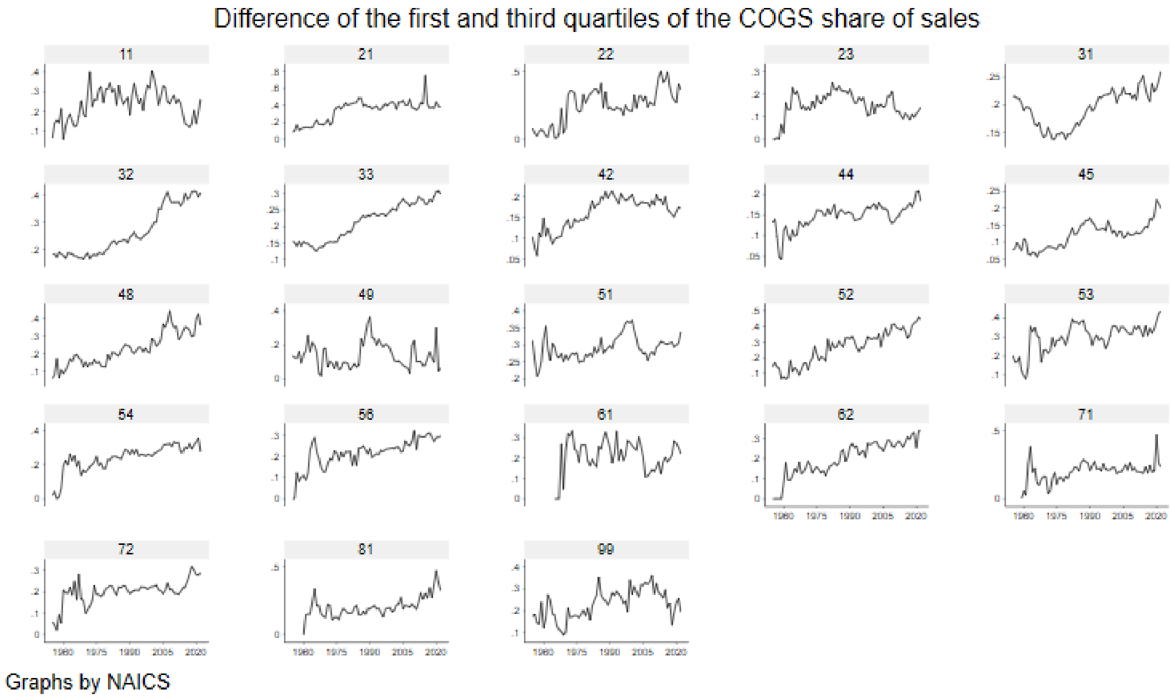
Figure 9: Dispersion, accounting for observed inputs



Finally we illustrate that there is an increase in the cross-section dispersion of the revenue share for COGS within most of the 2 digit NAICS sectors covered by the sample. Figure 10 plots the interquartile range separately for each of the 23 sectors for which we have data for more than one firm. The interquartile range is higher at the end of the period than at the start of the period in all but one sector.[11] There is a clear upward trend in this measure of within-sector cross-section dispersion, at least since the mid-1970s, in most of the sectors for which we have many observations on publicly traded firms. This includes the three manufacturing sectors (NAICS 31-33), mining, quarrying, and oil and gas extraction (21), the two retail trade sectors (44 and 45), and professional, scientific, and technical services (54).

[11] The exception is Transportation and Warehousing (postal services, couriers, warehousing and storage), NAICS 49, which accounts for 0.02% of the observations in our sample (see Table 1 in Appendix A).

Figure 10: Dispersion of COGS shares: by sectors



3 Estimated elasticities from revenue production functions

3.1 Specifications which omit SGA

We initially follow De Loecker et al. (2020) in considering a log-linear specification of the revenue production function which relates sales revenue to COGS and PPE only. We estimate models of the form

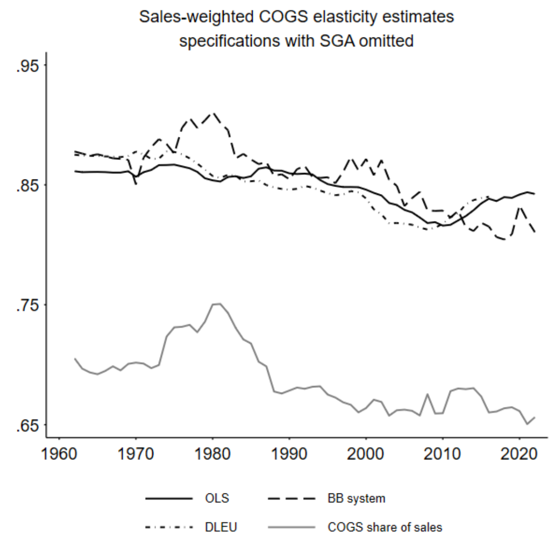
$$r_{it} = \gamma_0 + \gamma_X x_{it} + \gamma_K k_{it} + u_{it}$$

where r_{it} is the log of sales revenue for firm i in year t, x_{it} is the log of COGS, and k_{it} is the log of PPE, all in constant prices. We estimate separate models for each of the 23 NAICS 2 digit sectors covered by our sample, and for each of the rolling 8 year windows 1955-62, 1956-63, ... , 2015-22. We consider both OLS and system GMM estimates of the revenue elasticity parameters (γ_X, γ_K). The system GMM estimator (Blundell and Bond (1998, 2000)) allows for time-specific, firm-specific, and AR(1) components of the error term (u_{it}), and for correlation between the flexible input (x_{it}) and the innovations to the AR(1) component of the error term in the same period. Appendix C provides more detail of our implementation of this estimator.

Figure 11 plots a weighted average across sectors of our OLS and system GMM estimates of sector-specific revenue elasticities for COGS (γ_X) obtained from this specification, together with the same weighted average of the sector-specific COGS elasticities reported by De Loecker et al. (2020).[12] We also include the sales-weighted average of the revenue share for COGS for each year from 1962-2022. For COGS, the econometric estimates of the revenue elasticities are strikingly similar to each other, and considerably higher than the revenue share, particularly in the period after 1980. If we maintain that COGS is a flexible input, this gap suggests a source of upward bias which is common to these three estimators.

[12] Here we use sales weights, and allocate estimated elasticities to the final year of the sample period for which they are estimated. For example, the elasticities plotted for 1962 are estimated using data for the period 1955-62.

Figure 11: Econometric estimates of the COGS elasticity when SGA is omitted



3.2 Specifications which include SGA

We now augment the log-linear revenue production function specification to consider SGA as an additional input.[13] We estimate models of the form

$$r_{it} = \gamma_0 + \gamma_X x_{it} + \gamma_S s_{it} + \gamma_K k_{it} + u_{it}$$

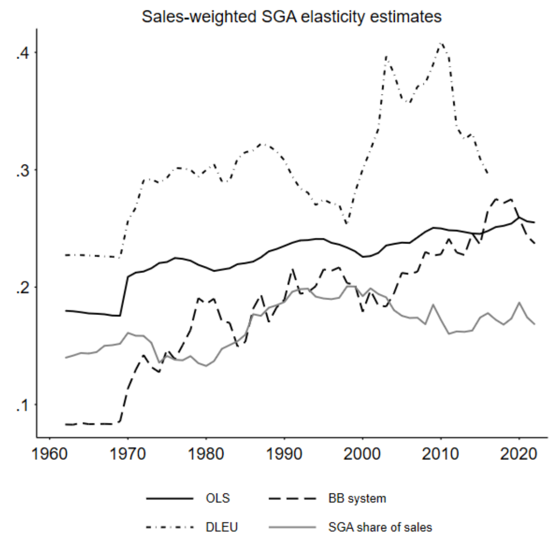
where s_{it} is the log of SGA in constant prices. As before, we estimate separate models for each 2 digit sector and for each 8 year rolling window. Our system GMM estimates also allow for correlation between the SGA input (s_{it}) and the innovations to the AR(1) component of the error term in the same period.

Figure 12 plots a weighted average across sectors of our OLS and system GMM estimates of sector-specific revenue elasticities for SGA (γ_S) obtained from this specification, together with the same weighted average of the sector-specific SGA elasticities reported by De Loecker et al. (2020),[14], and the sales-weighted average of the revenue share for COGS. The econometric estimates of the revenue elasticity for SGA are increasing over most of this period. For our OLS and system GMM estimators, the revenue elasticity for SGA is estimated to be positive and significantly different from zero in almost all of the sub-samples for sectors and time periods considered.

[13]We follow Traina (2018) in considering revenue production functions with SGA included as an input.

[14]De Loecker et al. (2020) also report estimated elasticities from this more general specification, but these are not the elasticities used to obtain their preferred estimates of markups.

Figure 12: Econometric estimates of the SGA elasticity



For each of the three estimators, Figure 13 plots a weighted average across sectors of the sector-specific revenue elasticities for COGS (γ_X) obtained from this specification, together with the corresponding estimate from the specification which omits SGA, and the revenue share for COGS. In each case, the estimated elasticity for COGS is substantially lower when SGA is included in the specification. This indicates that SGA is a relevant explanatory variable in these revenue production functions, and positively correlated with COGS (in log levels, after conditioning on k_{it}).[15] The estimated elasticities for COGS reported by De Loecker et al. (2020) using this more general specification are quite similar to the revenue share of COGS;[16] for our OLS and system GMM estimates, the gap between the estimated elasticities and the revenue share is reduced but not eliminated.

The returns to scale in these log-linear revenue production functions is given by the sum of the three revenue elasticities ($\nu^{REV} = \gamma_X + \gamma_S + \gamma_K$). Figure 14 plots a weighted average across sectors of the sector-specific returns to scale parameters obtained from this specification, using each of the three estimators that we have considered in this section. The OLS and system GMM estimates are similar to each other, while the returns to scale implied

[15] For the OLS estimator, we have a textbook example of omitted variable bias, given that $\gamma_S > 0$ and the partial correlation between s_{it} and x_{it} is positive, after conditioning on k_{it}.

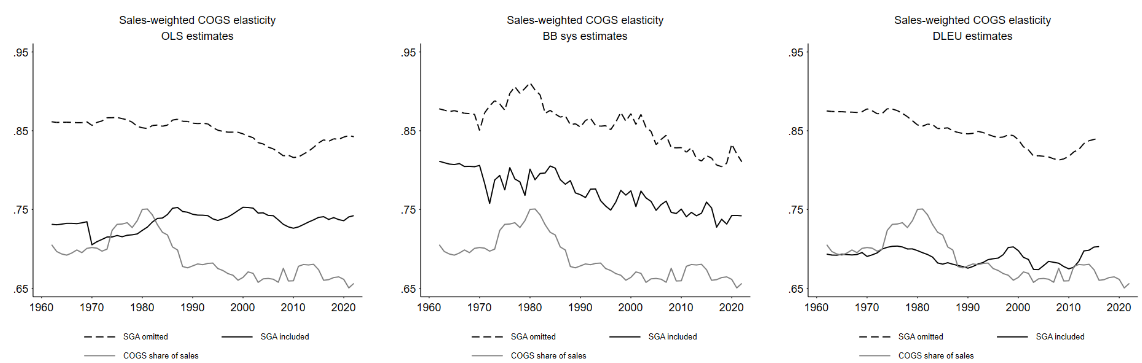
[16] As a result, combining their estimated elasticities for COGS from this more general specification with the revenue share for COGS and using their approach to estimate markups would result in lower markups, and a less dramatic rise in the weighted average markup, than that suggested by their preferred estimates.

by the elasticities reported by De Loecker et al. (2020) are somewhat higher. From 1980 onwards, these econometric estimates of the returns to scale in these revenue production functions are either close to or above one, and increasing over this period.

In the next section we show that for profit-maximising firms with constant returns to scale production technologies, the returns to scale in the revenue production function is the reciprocal of the price-cost markup. Taken at face value, these econometric estimates would suggest that if production is characterised by constant returns to scale, the weighted average markup is either close to or below one, and falling over the period since 1980. Alternatively, this could suggest a remaining source of upward bias in the estimated revenue elasticities, which is broadly common to these three estimators.

Both the system GMM estimator of Blundell and Bond (1998, 2000) and the control function estimator of De Loecker et al. (2020) rely on moment conditions which are valid if and only if the slope parameters of the model are common to all the observations in the sample used for estimation. The increase in the within-sector dispersion of revenue shares documented in Figure 10 suggests that these revenue elasticities may not be common parameters, and hence that the moment conditions exploited by these estimators may be invalid, particularly in the later sample periods. Consistent with this, we find an increase over time in the proportion of the sectors for which the validity of the moment conditions exploited by the system GMM estimator is rejected by the Hansen test of overidentifying restrictions.[17] This motivates us to develop bounds for average revenue elasticities which allow for unobserved heterogeneity in the true revenue elasticities, rather than relying on econometric estimates obtained from the specifications considered in this section.[18]

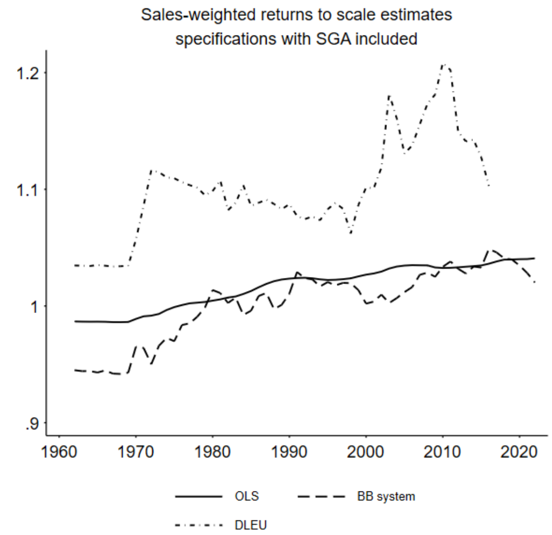
Figure 13: Econometric estimates of the COGS elasticity when SGA is included



[17]The proportion of sectors for which the null hypothesis is rejected at the 5% and 10% significance levels is reported for each sample period in Figure 20 in Appendix D. We can also note that the rejection frequencies are higher for the specification of the revenue production function which omits SGA.

[18]Consideration of translog specifications, or more general approximations to the form of the revenue production function, does not change the main results that we have highlighted in this section.

Figure 14: Econometric estimates of the returns to scale



4 Markups and the returns to scale in revenue production functions

The key to our empirical strategy is the relation between the returns to scale in the output production function, the returns to scale in the revenue production function, and the price-cost markup for profit-maximising firms. The returns to scale in the revenue production function is the sum of the revenue elasticities for each of the inputs. If we can obtain informative bounds for each of these revenue elasticities, we can then derive informative bounds for the markup if we are willing to make an assumption about the returns to scale in the underlying production technology.

To illustrate this relation, suppose that the firm produces output using the Cobb-Douglas production function

$$q_{it} = \beta_0 + \beta_X x_{it} + \beta_S s_{it} + \beta_K k_{it} + \omega_{it}$$

in which q_{it} is the log of output for firm i in year t, ω_{it} is the log of the idiosyncratic component of total factor productivity, and the output elasticities for each of the inputs are $(\beta_X, \beta_S, \beta_K)$. Suppose also that the firm faces a Constant Elasticity of Substitution (CES) inverse demand function

$$p_{it} = \delta - \left(\frac{1}{\eta}\right) q_{it} + \zeta_{it}$$

in which p_{it} is the log of the output price, ζ_{it} is the log of the idiosyncratic component of a demand shifter, and $\eta > 1$ is the absolute value of the price elasticity of demand. Simply adding p_{it} to both sides of the production function gives the expression

$$r_{it} = p_{it} + q_{it} = \beta_0 + \beta_X x_{it} + \beta_S s_{it} + \beta_K k_{it} + p_{it} + \omega_{it}$$

in which r_{it} is the log of revenue. Now substituting for p_{it} from the inverse demand function and rearranging gives the revenue production function

$$\begin{aligned} r_{it} &= \left(\left(\frac{\beta_0}{\mu}\right) + \delta\right) + \left(\frac{\beta_X}{\mu}\right) x_{it} + \left(\frac{\beta_S}{\mu}\right) s_{it} + \left(\frac{\beta_K}{\mu}\right) k_{it} + \left(\left(\frac{\omega_{it}}{\mu}\right) + \zeta_{it}\right) \\ &= \gamma_0 + \gamma_X x_{it} + \gamma_S s_{it} + \gamma_K k_{it} + u_{it} \end{aligned}$$

where $\mu = \left(1 - \left(\frac{1}{\eta}\right)\right)^{-1} \geq 1$, and the revenue elasticities for each of the inputs are $(\gamma_X, \gamma_S, \gamma_K)$.

This is the log-linear form of the revenue production function considered in Section 3. For a profit-maximising firm which chooses its level of output in each period to equate marginal revenue and marginal cost, the parameter $\mu = \left(1 - \left(\frac{1}{\eta}\right)\right)^{-1}$ is the markup of the level of the output price ($P_{it} = \exp(p_{it})$) over marginal cost (MC_{it}; i.e. $\mu = \frac{P_{it}}{MC_{it}}$). The revenue elasticity for each input is thus the ratio of the corresponding output elasticity to the price-cost markup.

The returns to scale in the output production function is the percentage change in output for a given percentage change in each of the inputs. For any form of the production function, the returns to scale is given by the sum of the output elasticities (Frisch, 1964). For the Cobb-Douglas case in our example, we have

$$\nu = \beta_X + \beta_S + \beta_K.$$

Similarly, the returns to scale in the revenue production function, or the percentage change in revenue for a given percentage change in each of the inputs, is the sum of the revenue elasticities. Simple arithmetic then establishes that in our example we have

$$\nu^{REV} = \gamma_X + \gamma_S + \gamma_K = \frac{\beta_X + \beta_S + \beta_K}{\mu} = \frac{\nu}{\mu}.$$

The returns to scale in the revenue production function is the ratio of the returns to scale in the production function to the price-cost markup for a profit-maximising firm.

For firms with market power and markups strictly greater than one, revenue elasticities are strictly lower than output elasticities. These firms face downward-sloping demand curves, and have to lower their output price to sell the additional output produced by using additional inputs. This intuition carries over to the sums of these elasticities, so that the returns to scale in the revenue production function is strictly lower than the returns to scale in the output production function.

Both the amount by which the firm has to lower its output price to sell additional output, and the profit-maximising markup of price over marginal cost, depend on the price elasticity of demand. Hence it is not surprising that the returns to scale in the revenue production function depends on the markup, as well as the returns to scale in the underlying production technology. The exact form of this relationship is simple to derive for the example considered in this section, but also holds for any admissable forms of the production and demand functions. The key is to show that the revenue elasticity for each input is, in general, the ratio of the corresponding output elasticity to the price-cost markup for a profit-maximising firm. In Appendix E we derive this key relation without restricting the functional forms of the production function or the inverse demand function.

We can apply this result to learn about the evolution of markups if we can learn about the returns to scale in the revenue production function from the available data, and if we are willing to make an assumption about the returns to scale in the underlying production technology. Without data on either firm-specific output prices or quantities, we cannot consistently estimate the output elasticity for a flexible input (see Bond et al., 2021), and so we cannot consistently estimate the returns to scale in the output production function. For the empirical analysis in this paper, we assume constant returns to scale (CRS) production technologies for publicly traded US firms. One subtle point is that we assume CRS in the sense of output elasticities which sum to one ($\nu = 1$). In that case, the markup is simply the reciprocal of the returns to scale in the revenue production function ($\mu = 1/\nu^{REV}$). Note that we do not assume CRS in the stronger sense of constant marginal cost, which additionally requires that all inputs are perfectly flexible.

In the next section we derive bounds for the (weighted) average revenue elasticity for each of the inputs in a sample of firms. We follow De Loecker et al. (2020) in treating capital as a quasi-fixed input which cannot be adjusted in year t in response to new information about either productivity or demand. Consequently we allow for increasing short-run marginal cost, which follows from diminishing returns to the flexible inputs, even with a CRS production function.

5 Bounds for average revenue elasticities

5.1 Flexible inputs

For inputs which are chosen after the firm has learned its productivity and demand for the current year, in the absence of any adjustment costs, the starting point for our analysis is the key result of Bond et al. (2021), Section 2.1: for firms which maximise profits by choosing output to equate marginal revenue and marginal cost in each period, the revenue elasticity and the revenue share are equal for flexible inputs. If, following De Loecker et al. (2020), we assume that COGS is such an input, we have

$$\gamma_{Xit} = \frac{X_{it}}{R_{it}}$$

where $X_{it} = \exp(x_{it})$ is the level of expenditure on the bundle of inputs represented by COGS, $R_{it} = \exp(r_{it})$ is the level of revenue, and γ_{Xit} is the revenue elasticity with respect to COGS for firm i in year t, which may take different values for different firms and different years. Dispersion in the revenue elasticities may reflect dispersion in the output elasticities - for example, if the production function is Cobb-Douglas with heterogeneous parameters, or is translog even with common parameters - or dispersion in the price-cost markups - for example, if the demand function is CES with heterogeneous parameters, or is not CES.

Suppose initially that the revenue elasticity is a common parameter in the sample of firms we are considering, and that there is no measurement error in the observed revenue shares. In that case, the observed revenue shares should be the same for all observations in the sample. Trivially, the arithmetic mean and the geometric mean of the observed revenue shares coincide with each other, and with the common parameter of interest.

Now suppose that the revenue elasticity takes different values for different firms and/or for different years in the sample we are considering, while there is still no measurement error in the observed revenue shares. We now observe dispersion in the observed revenue shares, but for each observation, the observed revenue share still coincides with the true revenue elasticity. If, for example, we are interested in the arithmetic mean of the true revenue elasticities in our sample, the arithmetic mean of the observed revenue shares still coincides with our object of interest. The geometric mean of the observed revenue shares is necessarily lower than the arithmetic mean of the observed revenue shares.[19] In this case, the geometric mean necessarily underestimates our object of interest.

[19] Arithmetic, geometric, and harmonic means have the property that arithmetic mean > geometric mean > harmonic mean, if there is any variation in the object that we are averaging.

We now extend the analysis to allow for symmetric, mean zero, additive measurement error in the log of revenue. We consider measurement error in revenue not because we expect that the data on sales in audited accounts is likely to be inaccurate, but to recognise the difference between the value of production and the value of sales in a given period. For firms which hold inventories of final products, sales this year may partly reflect a reduction in their holding of such inventories; conversely, part of this year's production may be sold in later years. While there are no inventories in the model considered by Bond et al. (2021), it is clear that their concept of revenue refers to the value of production. The assumption that the log of output or revenue is measured with error of this form is also common in the literature on the estimation of production functions; see, for example, Blundell and Bond (2000) and Ackerberg et al. (2015).

To consider the effects of this form of measurement error, we define r_{it} to be the log of the value of production and $\widetilde{r}_{it}$ to be the log of observed sales, and assume that $\widetilde{r}_{it} = r_{it} + \varepsilon_{it}$, where ε_{it} is a random variable drawn from a symmetric distribution with an expected value of zero, and independent of the true revenue elasticity γ_{it}^{X}. The level of the value of production is $R_{it} = \exp(r_{it})$, and the level of observed sales is $\widetilde{R}_{it} = \exp(\widetilde{r}_{it})$. The observed revenue shares are now $X_{it}/\widetilde{R}_{it} \neq \gamma_{Xit}$.

All else equal, measurement error introduces additional dispersion in the observed revenue shares. More seriously, the arithmetic mean of the observed revenue shares no longer coincides with the arithmetic mean of the true revenue elasticities. To illustrate that, first consider the effect of underestimating the value of production by 10%. We have $\widetilde{R}_{it} = 0.9R_{it}$ and $X_{it}/\widetilde{R}_{it} = 1.111\gamma_{Xit}$; the observed revenue share overstates the true revenue elasticity by $0.111\gamma_{Xit}$. Next, consider the effect of overestimating the value of production by 10%. We have $\widetilde{R}_{it} = 1.1R_{it}$ and $X_{it}/\widetilde{R}_{it} = 0.909\gamma_{Xit}$; the observed revenue share understates the true revenue elasticity by $0.0909\gamma_{Xit}$. Note that the absolute value of the difference between the observed revenue share and the true revenue elasticity is larger in the first example than in the second example. Then if these realisations of the measurement error are equally likely and we average many observations on the observed revenue shares, the arithmetic mean of the observed revenue shares overestimates the arithmetic mean of the true revenue elasticities. This intuition extends to any symmetric, mean zero, additive form of measurement error in the log of revenue which is independent of the true revenue elasticities.

In large samples, this form of measurement error has no effect on the geometric mean of the observed revenue shares. The log of the observed revenue shares is

$$x_{it} - \widetilde{r}_{it} = x_{it} - r_{it} - \varepsilon_{it}.$$

The geometric mean can be obtained by taking the arithmetic mean of these log shares, and then taking the exp of the result. In large samples, the arithmetic mean of the additive measurement error converges to its expected value of zero, so that the arithmetic mean of the log of the observed revenue shares converges to the arithmetic mean of the log of the true revenue shares. If the object of interest is the arithmetic mean of the true revenue elasticities, the geometric mean of the observed revenue shares will still underestimate our object of interest in large samples, with this form of measurement error.

Our annual samples for publicly traded firms contain between 353 observations (in 1955) and 6,776 observations (in 1996), with 3,628 observations on average over the period 1955-2022. We assume that these sample sizes are high enough for these large sample results to apply.

These results indicate that for flexible inputs, we can treat the arithmetic mean of the observed revenue shares as an upper bound for the arithmetic mean of the true revenue elasticities, and we can treat the geometric mean of the observed revenue shares as a lower bound for the arithmetic mean of the true revenue elasticities. These bounds are robust to any form of unobserved heterogeneity in the true revenue elasticities, and to additive measurement error in the data on log revenue, at least in large samples. Similar results are available for any size-weighted arithmetic mean of the true revenue elasticities, using the corresponding size-weighted arithmetic mean of the observed revenue shares as the upper bound, and the corresponding size-weighted geometric mean of the observed revenue shares as the lower bound. We focus on weighted averages in our empirical results, partly for comparability with De Loecker et al. (2020), and partly to reduce the influence of extreme values for some of the observed revenue shares, which tend to be a feature of the data mainly for younger and smaller firms.

5.2 Quasi-fixed inputs

We follow De Loecker et al. (2020) in assuming that the capital input to production is chosen before the firm has learned its productivity and demand for the current year, based on information that was available in the previous period. Assuming that the revenue elasticity and the cost of capital were known when the level of capital was chosen, the generalisation of the key result from Bond et al. (2021), Section 2.1, for profit-maximising firms is

$$\gamma_{Kit} = \frac{\rho_{it} K_{it}}{E_{t-1}[R_{it}]}$$

where ρ_{it} is the user cost of capital for firm i in year t, $K_{it} = \exp(K_{it})$ is the level of the capital input, $E_{t-1}[R_{it}]$ is the expected value of revenue in year t given information available in year $t-1$, and γ_{Kit} is the revenue elasticity with respect to capital. In Appendix F we illustrate this result for an example with a Cobb-Douglas production function and a CES demand function, as considered in Section 4.

For such quasi-fixed inputs, the revenue elasticity is equal to the expected value of the revenue share given information available when the level of the input was chosen. The realised level of revenue in year t will differ from the firm's prior expectation as a result of innovations to productivity and demand in year t. Assuming rational expectations, we have $R_{it} = E_{t-1}[R_{it}] + e_{it}$, where e_{it} is a forecast error with $E_{t-1}[e_{it}] = 0$. For quasi-fixed inputs, these forecast errors generate variation in the realised revenue shares, even if the revenue elasticity is a common parameter in the sample of firms we are considering, and there is no measurement error in the data.

As a result, the arithmetic mean of the realised revenue shares will overestimate the common revenue elasticity in large samples, even in this special case. To illustrate, suppose that the forecast error takes the value e or $-e$, each with probability 0.5. The effect on the ratio $\rho_{it}K_{it}/E_{t-1}[R_{it}]$ of understating the denominator by a fixed amount e, when we have $R_{it} < E_{t-1}[R_{it}]$, is larger than the effect on the ratio of overstating the denominator by a fixed amount e, when we have $R_{it} > E_{t-1}[R_{it}]$. Averaging many observations on the realised revenue shares, the arithmetic mean then overestimates the common parameter of interest. This intuition extends to any symmetric, mean zero distribution for the forecast errors (e_{it}).

To obtain a consistent estimator of the common revenue elasticity in this special case we can use the reciprocal of the realised revenue shares. Taking reciprocals, we have

$$\frac{1}{\gamma_{Kit}} = \frac{E_{t-1}[R_{it}]}{\rho_{it}K_{it}} = \frac{R_{it}}{\rho_{it}K_{it}} - \frac{e_{it}}{\rho_{it}K_{it}}.$$

The arithmetic mean of the second term converges to its expected value of zero, under our assumption that the user cost of capital (ρ_{it}) was known in period $t-1$. The arithmetic mean of the first term then converges to $(1/\gamma_{Kit})$, and the reciprocal of that arithmetic mean converges to the revenue elasticity (γ_{Kit}). This consistent estimator is the harmonic mean of the realised revenue shares $(\rho_{it}K_{it}/R_{it})$.

To consider the effect of dispersion in the true revenue elasticities on the properties of the harmonic mean, suppose that we happen to observe the true revenue elasticities for each observation in a sample of firms. The harmonic mean is necessarily lower than the arithmetic mean, so that the harmonic mean of the true revenue elasticities will underestimate their arithmetic mean. In large samples, the harmonic mean of the realised revenue shares converges to the harmonic mean of the true revenue elasticities, and so underestimates the arithmetic mean of the true revenue elasticities.

Finally, allowing for symmetric, mean zero, additive measurement error in the log of revenue also results in underestimation of the arithmetic mean of the true revenue elasticities using the harmonic mean of the observed revenue shares. This results from the concavity of the log function. If the log of revenue is measured with this form of error, the arithmetic mean of the level of observed revenue ($\widetilde{R}_{it}$) will overestimate the arithmetic mean of the value of production (R_{it}) in large samples. Replacing R_{it} by $\widetilde{R}_{it}$, the arithmetic mean of the reciprocal of the observed revenue shares will then overestimate the reciprocal of the revenue elasticity, in the special case in which the true revenue elasticity is a common parameter. The reciprocal of that arithmetic mean, which is the harmonic mean of the observed revenue shares, will underestimate the common revenue elasticity (γ_{Kit}). The effect of this form of measurement error is in the same direction as the effect of dispersion in the true revenue elasticities, and reinforces the tendency of the harmonic mean of the observed revenue shares to underestimate the arithmetic mean of the true revenue elasticities when we have both dispersion in the true revenue elasticities and measurement error of this form.

These results indicate that for quasi-fixed inputs, we can treat the arithmetic mean of the observed revenue shares as an upper bound for the arithmetic mean of the true revenue elasticities, and we can treat the harmonic mean of the observed revenue shares as a lower bound for the arithmetic mean of the true revenue elasticities. Similar results are available for any size-weighted arithmetic mean of the true revenue elasticities, using the corresponding size-weighted arithmetic mean of the observed revenue shares as the upper bound, and the corresponding size-weighted harmonic mean of the observed revenue shares as the lower bound.

5.3 Imputing capital costs

A further complication in applying these results to obtain bounds for the (weighted) average revenue elasticity for capital is that we do not observe the cost of owning and using capital in company accounts, in a way which is equivalent to observing expenditures on flexible inputs. The book value of PPE is a valuation of assets which have been purchased by the firm and which are still owned at the end of the accounting period. To convert this measure of the capital stock to an equivalent flow cost in year t, we require a measure of the user cost of capital. For example, if assets purchased by the firm in year $t-1$ contribute to production from year t onwards, and depreciate at the geometric rate d, and if the real price of capital assets is constant and normalised to one, the user cost of capital for investment which is financed by equity has the form $\rho_{it} = r^E_{it} + d$, where r^E_{it} is the required real rate of return on the firm's equity between years $t-1$ and t.

De Loecker et al. (2020) impute capital costs as $\rho_{it} K_{it}$ using an expression of this form. More specifically, they construct the user cost as a real interest rate plus a premium of 12%, which could be decomposed as an equity risk premium of 7% plus a depreciation rate of 5%, for example. We present results which impute capital costs in this way as a robustness check. Our main results incorporate a firm-year-specific equity risk premium. More specifically, we construct the user cost for firm i in year t as $\rho_{it} = r_t + 0.07\beta_{it} + 0.05$, where r_t is a real interest rate and β_{it} is a CAPM beta coefficient. The market risk premium is assumed to be 7% and the depreciation rate is assumed to be 5%. We use firm-year-specific beta coefficients obtained from WRDS Beta Suite where these are available, and we impute firm-year-specific beta coefficients on the basis of sector, firm size, firm age, and year, otherwise. Further details are provided in Appendix A.[20]

[20] There is a vast literature on the measurement of the cost of capital, and our approach could be further refined in several ways. For example, we could incorporate information available in company accounts about firm-specific variation in depreciation rates or the use of debt finance. In principle we could also incorporate tax adjustments, although in practice that would be challenging for multinational corporations with complex tax planning arrangements. We leave these extensions for future research.

6 Results and implications for the evolution of markups

We first present the results from using the approach outlined in Section 5 to obtain bounds for the weighted average revenue elasticities for COGS, SGA, and capital, for each year from 1955 to 2022. We then combine these results to obtain bounds for the weighted average returns to scale in the revenue production functions, and the corresponding bounds for the weighted average markup under the assumption of constant returns to scale production technologies, as outlined in Section 4.

In this section we present results for the Compustat sample of publicly traded firms, after additionally dropping observations with extreme values for the shares of either SGA or imputed capital costs in sales revenue (see Appendix A). In Appendix G.1 we present corresponding results for the sample considered in Sections 2 and 3. Based on arguments in Edmond et al. (2023), we focus in this section on results which use total costs as the size weights. Total costs are measured as the sum of COGS, SGA, and capital costs imputed using our firm-year-specific risk premia. In Appendix G.2, we also present corresponding results using different size weights, and imputing capital costs using a common risk premium for all observations, as in De Loecker et al. (2020). In Appendix G.3, we report results when companies listed on exchanges other than the NYSE and NASDAQ are dropped from the sample.

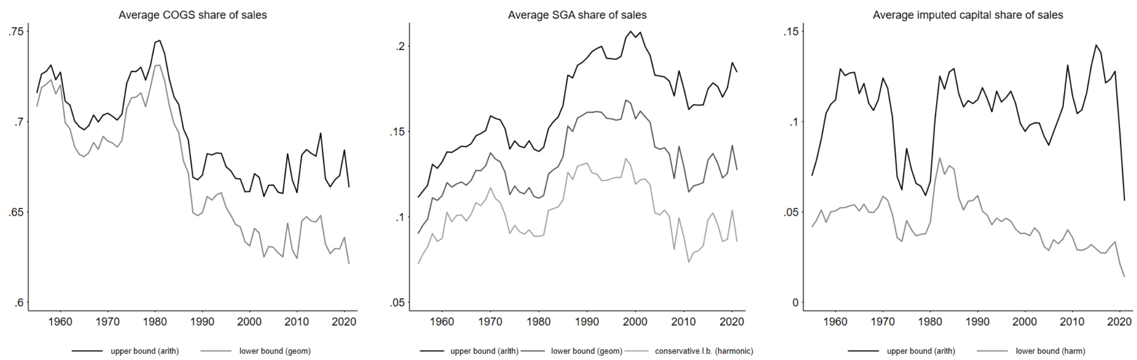
6.1 Revenue elasticities

We assume that the bundle of inputs represented by expenditure on COGS is a flexible input, in the sense required for the ratio estimator of markups used by De Loecker et al. (2020). In that case, the weighted arithmetic average of the observed revenue shares provides an upper bound for the weighted arithmetic average of revenue elasticities, and the weighted geometric average of the observed revenue shares provides a lower bound.[21] Figure 15 plots these upper and lower bounds, weighted by total costs.[22]

[21] We compute the weighted geometric mean as the exp of the estimated intercept parameter from a weighted least squares regression of the log revenue shares on a constant.

[22] Figure 21 in Appendix G.1 presents the corresponding figure using sales weights rather than total cost weights to construct each of the weighted averages.

Figure 15: Bounds for revenue elasticities



Our upper bound for the weighted average revenue elasticity for COGS falls from around 0.74 in the early 1980s to around 0.66 in 2022. Comparing this series to the corresponding weighted average of COGS revenue shares shown in Figure 5, the effect of dropping observations with extreme values for the remaining revenue shares is a modest reduction in the level of the implied upper bound, with very little impact on the trend. Our lower bound for the weighted average revenue elasticity for COGS falls from around 0.72 in the early 1980s to around 0.62 in 2022. The gap between these weighted arithmetic and geometric means increases over this period, in line with the increase in the dispersion of the COGS revenue shares, illustrated in Figure 2.

Referring back to Figure 11, we can now interpret the series for the weighted arithmetic average of the observed revenue shares as an upper bound for the weighted average of revenue elasticities. The considerable gap between that upper bound, and the weighted average of revenue elasticities obtained from the estimation of revenue production functions, is consistent with our view that those econometric estimates of revenue elasticities are likely to be affected by important sources of upward bias.

For the bundle of inputs represented by SGA, Figure 15 plots the weighted arithmetic, geometric, and harmonic means of the observed revenue shares.[23] Our upper bound for the weighted average revenue elasticity for SGA increases from around 0.13 in the early 1980s to around 0.18 in 2022. Comparing this series to the corresponding weighted average of SGA revenue shares shown in Figure 5, the effect of dropping observations with extreme values is a modest reduction in the implied upper bound, particularly in more recent years. If we treat SGA as a flexible input, the weighted geometric mean again provides a lower bound for the weighted average revenue elasticity. This lower bound increases from around 0.11 in the early 1980s to around 0.13 in 2022, having been somewhat higher in the 1990s.

[23] We compute the weighted harmonic mean as the reciprocal of the estimated intercept parameter from a weighted least squares regression of the reciprocal of the revenue shares on a constant.

If we prefer to treat SGA as a quasi-fixed input, the weighted harmonic mean provides a more conservative lower bound for the weighted average revenue elasticity. In that case the lower bound at the end of the period is marginally below its level in the early 1980s. Again the gap between these different weighted means increases over the period, in line with the increase in the dispersion of the SGA revenue shares, illustrated in Figure 2.

For capital, Figure 15 plots the weighted arithmetic and harmonic means of the imputed revenue shares. We follow De Loecker et al. (2020) in treating capital as a quasi-fixed input, in which case these series provide upper and lower bounds for the weighted average revenue elasticity. Fluctuations in these bounds are strongly influenced by fluctuations in the real interest rate over this period, with no clear trend implied for the weighted average revenue elasticity.[24] Our upper bound for the weighted average revenue elasticity for capital fluctuates between 0.06 and 0.14. The imputed revenue shares for capital are small enough, relative to the sum of the observed revenue shares for COGS and SGA, for details of the imputation to be of limited significance for the trends in our bounds for the weighted average returns to scale in the revenue production functions.

6.2 Revenue returns to scale, and implied markups

To obtain our upper and lower bounds for the weighted average revenue returns to scale, we sum the corresponding bounds for the weighted average revenue elasticity across these three inputs. Treating both COGS and SGA as flexible inputs, the implied upper and lower bounds for the weighted average revenue returns to scale in our sample of publicly traded firms are presented in Figure 4. The upper bound was around one in the early 1980s, and has declined to around 0.9 over the last 40 years. The lower bound was around 0.9 in the early 1980s, and has declined to around 0.75 by the end of our sample period. The increase in the interval between these upper and lower bounds reflects the trend increase in the cross-section dispersion of the observed revenue shares in the data for publicly traded firms.

Referring back to Figure 14, we can compare our upper bound for the weighted average revenue returns to scale to the estimates obtained from the revenue production function specifications considered in Section 3.2. From around 1980 onwards, the econometric estimates obtained from these revenue production functions exceed our upper bound. This difference is not explained by the use of different size weights to compute these weighted

[24]The sharp fall in imputed capital costs as a share of sales at the end of our sample period is due to the sharp increase in inflation in 2021-2022. The Federal funds effective rate and the yield on 10-year Treasuries (used as nominal interest rates in our user cost of capital measures) increased much less than the inflation rate over the same period, resulting in lower *ex post* real interest rates.

averages.[25] This pattern is again consistent with our view that the econometric estimates are likely to be biased upwards, and that the sources of this upward bias have become more important in more recent years.

For a profit-maximising firm with a constant returns to scale production function, in the sense of output elasticities which sum to one, the price-cost markup is the reciprocal of the returns to scale in the revenue production function. Assuming CRS technologies for publicly traded firms, we can obtain a lower bound for the weighted average markup as the reciprocal of our upper bound for the weighted average revenue returns to scale, and we can obtain an upper bound for the weighted average markup as the reciprocal of our lower bound for the weighted average revenue returns to scale. These bounds for the weighted average markup in our sample are also presented in Figure 4. The lower bound was around one in the early 1980s, and increased to around 1.1 by 2022. The upper bound was around 1.1 in the early 1980s, and increased to around 1.3 over the same period. The trends in these bounds are consistent with a modest rise in the weighted average markup since the early 1980s. The increase over time in the width of these intervals is consistent with a substantial rise in the cross-section dispersion of markups.

For comparison, De Loecker et al. (2020) estimate a rise in the weighted average markup for publicly traded firms from around 1.2 in the early 1980s to around 1.6 in 2016.[26] Their estimate is considerably higher than our upper bound in each year of their sample period, and particularly so in the later years of their sample period.

If we treat SGA as a quasi-fixed input, we obtain a more conservative lower bound for the weighted average revenue elasticity for SGA, and hence a more conservative lower bound for the weighted average revenue returns to scale. In turn, this implies a more conservative upper bound for the weighted average markup. These more conservative bounds are presented in Figure 16. The upper bound for the weighted average markup in this case is naturally higher throughout the sample period, but this has little effect on the trend since the early 1980s.[27]

If we are willing to assume that measured sales provides accurate data on the value of production, the weighted average of the observed revenue shares then coincides with the weighted average revenue elasticity for flexible inputs. Treating both COGS and SGA as flexible inputs, we then require bounds only for the weighted average revenue elasticity for

[25] Figure 31 in Appendix G.4 directly compares our upper bound with the econometric estimates, using either sales or total costs to construct the respective weighted averages.

[26] See De Loecker et al. (2020), Figure 1. Again this difference is not explained by their use of different size weights. Figure 22 in Appendix G.1 presents our bounds for the sales-weighted average markup.

[27] Even in this case, our upper bound suggests a weighted average markup of around 1.3 in 2016, or output prices which are around 30% higher than marginal costs. This is roughly half of the 60% markup suggested in De Loecker et al. (2020).

capital. These assumptions would imply less conservative bounds for the weighted average revenue returns to scale, and hence for the weighted average markup. We present these less conservative bounds for comparison in Figure 17.

If we abstract from measurement error in revenue and further assume that all three inputs are flexible, the revenue share for each input then coincides with the revenue elasticity, and the sum of the revenue shares then coincides with the revenue returns to scale. In that case, the sum of the weighted average revenue shares for each of the inputs would coincide with the weighted average revenue returns to scale; this is the upper bound for the weighted average revenue returns to scale shown in Figures 4, 16, and 17. Assuming CRS production technologies, the reciprocal of the sum of the weighted average revenue shares for each of the inputs would then coincide with the weighted average markup; this is the lower bound for the weighted average markup shown in Figures 4, 16, and 17. Note that if all inputs are flexible, CRS in the sense of output elasticities which sum to one implies CRS in the stronger sense of constant marginal costs.

These results are helpful in relating our approach to using data from company accounts to study the evolution of markups to the 'accounting approach' discussed in Autor et al. (2020), for example. Assuming CRS in the stronger sense of constant marginal costs, we can infer the markup for each firm as the ratio of revenue to total costs, and hence we can infer the weighted average markup in any sample.[28] This is closely related to the use of profit margins, where profits are measured as the difference between sales and total costs, and the profit margin is the ratio of profits to sales. We generalise this approach by allowing for quasi-fixed inputs and accounting for the difference between sales revenue and the value of production. A consequence is that we can only provide bounds for the evolution of the weighted average markup.

[28]CRS in the sense of constant marginal costs is also needed to infer the output elasticity for COGS from the share of COGS in total costs. Dividing the cost share of COGS by the revenue share of COGS returns the ratio of revenue to total costs. A variant of the ratio estimator used by De Loecker et al. (2020) is then equivalent to inferring the markup as the ratio of revenue to total costs.

Figure 16: Revenue returns to scale and CRS markups with conservative bounds

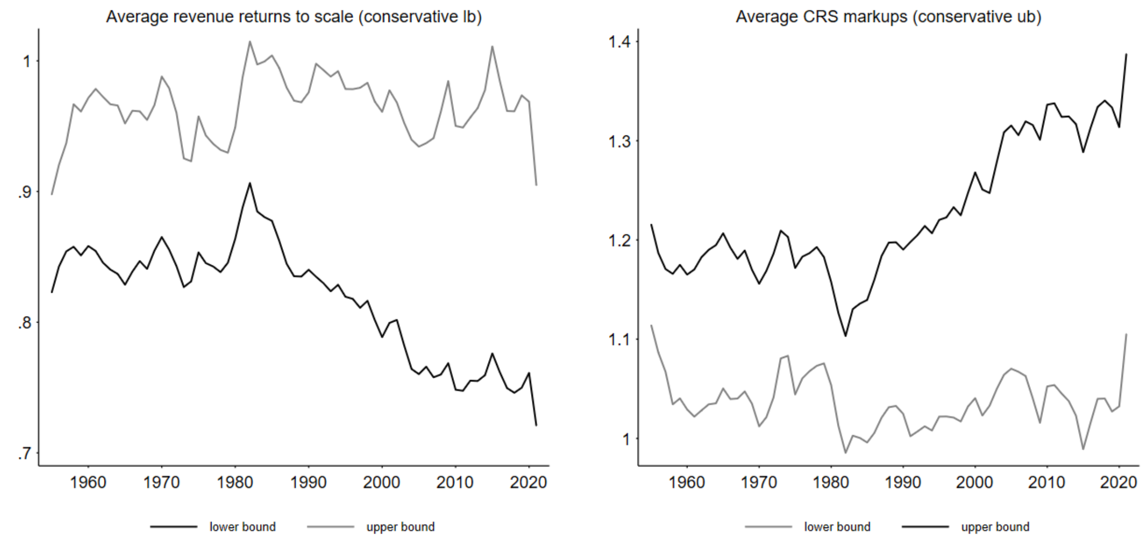
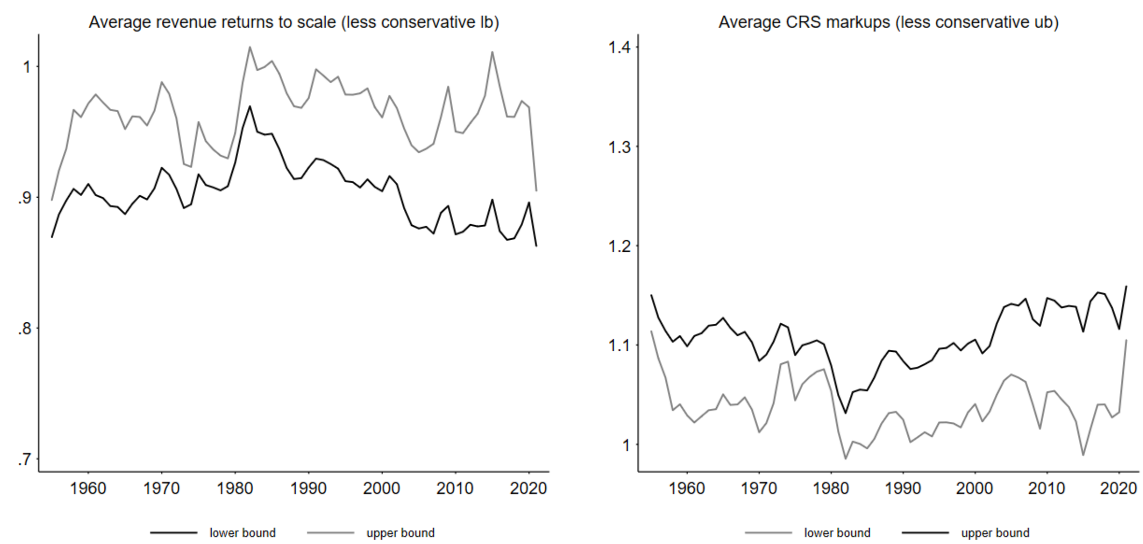


Figure 17: Revenue returns to scale and CRS markups with less conservative bounds



7 Conclusion

In this paper we have outlined how both within-sector dispersion in revenue elasticities and the omission of the relevant SGA input contribute to overestimation of the revenue elasticity for COGS from revenue production functions of the form used in De Loecker et al. (2020). The rise in the revenue share of SGA and the rise in the within-sector dispersion of revenue shares since the early 1980s then contribute to overestimation of the rise in the weighted average price-cost markup for publicly traded firms, using the ratio estimator of De Loecker et al. (2020).

We develop a different approach which requires estimates only of revenue elasticities. Assuming constant returns to scale in the underlying production technology, the markup for a profit-maximising firm can be inferred from the sum of these revenue elasticities, or equivalently from the returns to scale in the revenue production function. We show that informative bounds for average revenue elasticities can be obtained using available data on revenue shares, allowing for any form of heterogeneity in the true revenue elasticities, and for the presence of quasi-fixed inputs, as well as a standard form of measurement error in the revenue data. Assuming constant returns to scale only in the sense of output elasticities which sum to one, this allows us to infer bounds for the weighted average markup in our sample. We discuss how this approach generalises the 'accounting approach' to markup estimation discussed in Autor et al. (2020), for example.

Applying this approach to company accounts data for publicly traded firms, our upper bound for the weighted average markup is considerably lower than the estimates presented in De Loecker et al. (2020). Although revenue elasticities are necessarily lower than output elasticities for firms with market power, our results suggest that effect is dominated in practice by overestimation of the revenue elasticity for COGS, using the approach of De Loecker et al. (2020). Over the last 40 years, our empirical bounds are consistent with a modest rise in the weighted average markup, and with a more substantial rise in the cross-section dispersion of markups.

References

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer, "Identification properties of recent production function estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms*," *The Quarterly Journal of Economics*, 02 2020, *135* (2), 645–709.

Blundell, Richard and Stephen Bond, "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 1998, *87* (1), 115–143.

_ , _ , Michael Devereux, and Fabio Schiantarelli, "Investment and Tobin's Q: Evidence from company panel data," *Journal of Econometrics*, 1992, *51* (1), 233–257.

Blundell, Richard W and Stephen R Bond, "GMM estimation with persistent panel data: an application to production functions," *Econometric Reviews*, 2000, *19* (3), 321–340.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch, "Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data," *Journal of Monetary Economics*, 2021, *121*, 1–14.

Damodaran, Aswath, "Variables used in Datasets," https://pages.stern.nyu.edu/ adamodar/New_Home_Page/datafile/variable.htm 2023.

De Loecker, Jan and Frederic Warzynski, "Markups and firm-level export status," *American Economic Review*, 2012, *102* (6), 2437–71.

_ , Jan Eeckhout, and Gabriel Unger, "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "How Costly Are Markups?," *Journal of Political Economy*, 2023, *131* (7), 1619–1675.

Frisch, Ragnar, *Theory of Production*, 1 ed., Springer Dordrecht, 1964.

Klette, Tor Jakob and Zvi Griliches, "The inconsistency of common scale estimators when output prices are unobserved and endogenous," *Journal of Applied Econometrics*, 1996, *11* (4), 343–361.

Traina, James, "Is Aggregate Market Power Increasing? Production Trends Using Financial Statements," Technical Report, SSNR Working Paper 2018.

Yeh, Chen, Claudia Macaluso, and Brad Hershbein, "Monopsony in the US Labor Market," *American Economic Review*, July 2022, *112* (7), 2099–2138.

Appendices

A Data sources, cleaning, and summaries

The main data source is the annual version of Compustat North America, the database of US and Canadian publicly traded firms. In the same spirit as the recent literature about the long run evolution of market power in the US, we focus on the widest available time window, i.e. years from 1955 until 2022. We define sectors as 2-digit NAICS codes, as in De Loecker et al. (2020). The financials used in our analysis are net sales/turnover (sales); cost of goods sold (COGS); selling, general, and administrative expenses (SGA); and gross property, plant, and equipment (PPE). We follow De Loecker et al. (2020) and use the gross book value of PPE as the capital stock measure in our revenue production function specifications. All these measures are reported in current US Dollars; we deflate them using the annual US GDP implicit price deflator from FRED to make them comparable across years. While COGS and SGA are flow measures of expenditure, PPE is a valuation of the stock of tangible capital, the flow cost of which is not observed in the data. Since our approach to markup estimation requires revenue shares for all inputs, we need measures of the user cost of capital.

We follow two approaches for constructing the user cost of capital. (i) We follow De Loecker et al. (2020) closely and build an annual measure, common across sectors, as $\rho_t = i_t^{FED} - \pi_t + 0.12$, where i_t^{FED} is the average annual Federal funds effective rate, π_t is the average annual inflation rate using the US GDP implicit price deflator (both from FRED), and 0.12 accounts jointly for depreciation and the equity premium. Two values that are consistent with this joint term would be, respectively, 0.05 and 0.07. (ii) We construct a firm-year-specific user cost of capital using betas from WRDS Beta Suite. This source reports annual betas for a large sample of North American firms with ticker symbols among the available identifiers. We match WRDS Beta Suite and Compustat by ticker symbol and year. Of the 254,903 firm-year observations available from Compustat after implementing our basic cleaning routine (see below), 78,315 are matched to a WRDS Beta Suite ticker-year obervation. For the remaining 176,588 Compustat firm-year observations, we use the average beta from WRDS among observations belonging to the same year, sector, sales decile, and age decile. We measure the age of a company as the number of years since its data first appeared in Compustat. For most firms, this is the same as the number of years they have been filing accounts with a regulatory body. We then construct a firm-year-specific cost of equity using the operational definition from Damodaran (2023), i.e. $r_{it}^E = i_t^{10YR} - \pi_t + 0.07\beta_{it}$, where i_t^{10YR} is the average annual yield of 10-year US Treasuries, β_{it} is the firm-year-specific beta, and 0.07 is the assumed market risk premium; and we construct our firm-year-specific user cost of capital as $\rho_{it} = r_{it}^E + 0.05$, where 0.05 is the assumed depreciation rate. Using (i) and (ii) we then calculate two alternative flow measures of capital costs as $\rho_t K_{it}$ and $\rho_{it} K_{it}$. Our main results use the firm-year-specific measure of the user cost of capital, but using the common annual measure does not change those results qualitatively (see Appendix G.2).

We employ two alternative cleaning routines for robustness. The first is the same as De Loecker et al. (2020) (DLEU): we drop all observations with missing, negative, or null sales, COGS, SGA, or PPE, and then those observations which are below the first or above the last percentile of the COGS-sales ratio, separately for each year. The second routine is an enhanced version of the first, as we additionally drop those observations which are below the first or above the last percentile of the SGA-sales, PPE-sales and imputed capital costs-sales ratios, separately for year. This enhanced cleaning is preferred for our main results in Section 6 because our bounds for the weighted average markup also use these revenue shares. Since we focus on weighted averages, and extreme values are more common for younger and smaller firms, our results are not qualitatively affected by the choice of cleaning routines (see Appendix G.1).

Our sample has 254,903 firm-year observations from 1955 until 2022 with data on sales, COGS, SGA, PPE, and imputed capital costs after the applying the basic cleaning routine. The number of observations drops to 243,071 with enhanced cleaning (Table 1). All observations belong to 24 2-digit NAICS sectors. NAICS 92 (Public Administration) only has one observation which is thus automatically excluded from the analysis despite appearing in the basic cleaning sample.

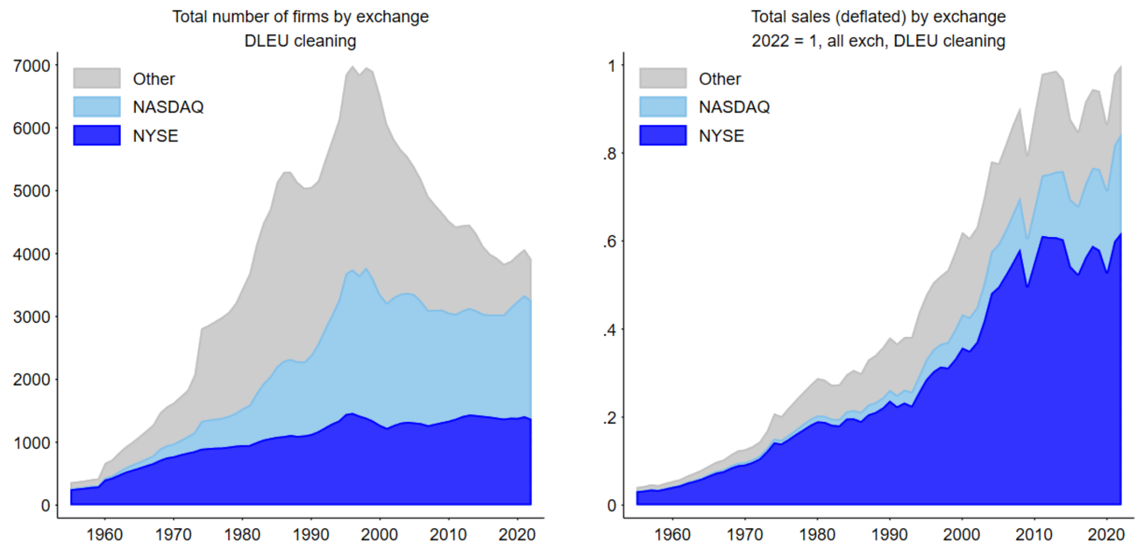
Table 1: Sectors and observations

NAICS	obs [1]	obs [2]	sector description
11	1,080	1,032	Agriculture, Forestry, Fishing and Hunting
21	14,961	13,442	Mining, Quarrying, and Oil and Gas Extraction
22	1,185	1,043	Utilities
23	4,235	3,718	Construction
31	14,224	13,759	Manuf (food, beverages, tobacco, clothing)
32	32,665	31,215	Manuf (wood, paper, oil and coal, chemicals, non-metal minerals)
33	78,651	76,677	Manuf (metal, machinery, electronics, electronic equipment, transport, furniture, miscellanea)
42	11,719	10,582	Wholesale Trade
44	8,319	8,153	Retail Trade (motor v, building mat, garden eq, supplies, grocery, furniture and hpousehold appl)
45	7,165	6,899	Retail Trade (gen merch and warehousing, health and care, gas stations, clothing, sports and hobbies)
48	5,978	5,495	Transportation and Warehousing (air, rail, water, truck, transit, pipeline, tourism, support activities)
49	415	394	Transportation and Warehousing (postal services, couriers, warehousing and storage)
51	27,434	26,225	Information
52	5,137	4,712	Finance and Insurance
53	5,460	5,094	Real Estate and Rental and Leasing
54	11,776	11,298	Professional, Scientific, and Technical Services
56	6,079	5,751	Adm. and Supp. and Waste Mgmt and Remediation Svcs
61	1,336	1,260	Educational Services
62	4,435	4,197	Health Care and Social Assistance
71	1,846	1,730	Arts, Entertainment, and Recreation
72	6,273	6,071	Accommodation and Food Services
81	1,358	1,307	Other Services (except Public Administration)
92	1	0	Public Administration
99	3,171	3,017	Other

Note: [1] DLEU cleaning routine; [2] enhanced cleaning.

The number of observations changes substantially over time, with numerous firms entering and exiting the panel. Entry happens when a company begins filing with a regulatory body (e.g. the SEC), while exit from the panel can signal the cessation of activities (e.g. bankruptcy), or a merger/acquisition by another company. Delisting or changing stock exchange does not warrant the removal of a company from the panel, since data is still collected for companies that do not have to file with the regulatory body anymore but have done so in the past. The total number of companies in our basic cleaning sample starts at 364 in 1955, peaks at 6,992 in 1996, and falls to 3,909 in 2022. Despite a substantial fall in the total number of companies after the Dotcom bubble, aggregate sales in constant prices continued to grow until the early 2010s, and then remained broadly stable, with some fluctuations. The same patterns hold for the enhanced cleaning sample, and also for the entire dataset before any cleaning is performed. The evolution in the number of firms and their total sales is driven both by within-stock exchange dynamics (e.g. the NYSE growing in size), and by the change in the relative weights of different stock exchanges. More specifically, whereas the NYSE made up almost 69% of total observations in 1955, its share was only around 35% in 2022. This was driven by the rise of the NASDAQ, and, to a larger extent up to the late 1990s, by the growth of other security exchanges, mostly the now-defunct OTC Bulletin Board and other over-the-counter

Figure 18: Observations and total sales by year and exchange

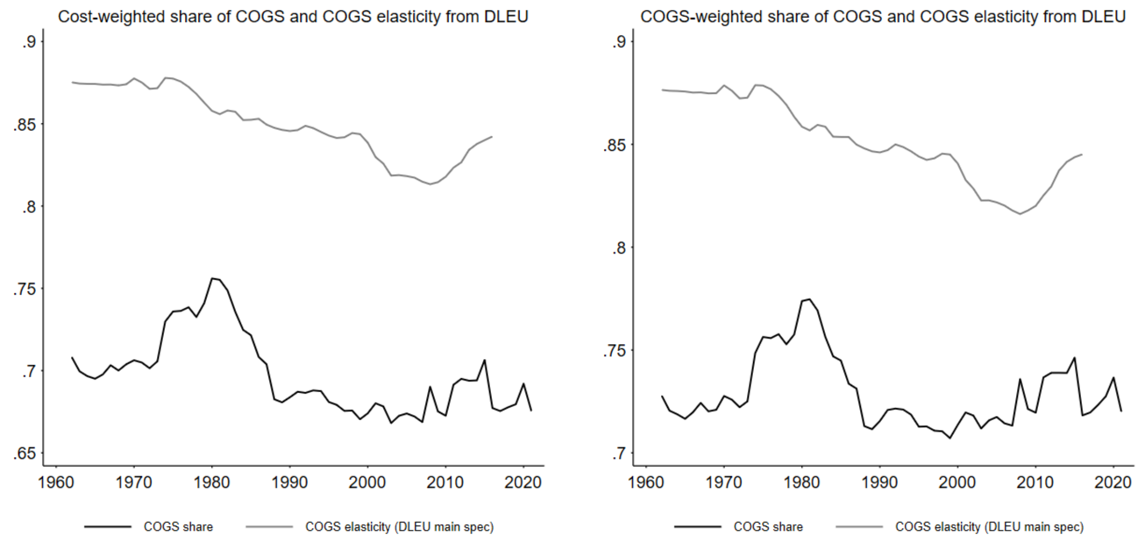


exchanges. The share of total sales of exchanges other than the NYSE also grew over time, but not by as much as the share of firms, with the NYSE still accounting for roughly 62% of total sales in 2022 (Figure 18). The same holds when looking at total costs instead of total sales. For robustness, we also perform our analysis on the sub-sample of NYSE listed firms, and on the sub-sample of NASDAQ and NYSE listed firms (i.e. excluding the "Other" exchange category).

B COGS elasticity from DLEU and COGS revenue share

Figure 19 replicates Figure 3 using total costs and COGS weights, respectively. The underlying sample for both figures is the one resulting from the DLEU cleaning routine.

Figure 19: COGS revenue share and revenue elasticity



C Econometric estimation of the revenue production function

We estimate the revenue production functions using two estimators: OLS (as a benchmark), and the dynamic panel estimator of Blundell and Bond (1998, 2000). This is done both for specifications that omit and include SGA as a revenue input. We illustrate only the latter, as the same input choice and timing assumptions are made when SGA is omitted.

For the OLS benchmark, we estimate the following model with ordinary least squares, separately for each sector, using 8-year overlapping windows:

$$r_{it} = \gamma_0 + \gamma_X x_{it} + \gamma_S s_{it} + \gamma_K k_{it} + u_{it}.$$

Revenue and the three inputs are expressed in log USD in constant prices. The OLS estimate of the revenue elasticity of input V is simply the coefficient γ_V. For sectors that lack a sufficient number of observations at the start of the sample period, the initial window is widened to ensure that estimates can be produced.

The dynamic panel estimator follows Blundell and Bond (2000) closely by allowing for time-specific and firm-specific components of the error term (u_{it}), specifying an AR(1) process for the remaining component of the error term, and further allowing for additive measurement error in log revenue, which is assumed to be serially uncorrelated. The resulting dynamic specification is:

$$\begin{aligned}(r_{it} - \gamma_X x_{it} - \gamma_S s_{it} - \gamma_K k_{it}) - \rho(r_{i,t-1} - \gamma_X x_{i,t-1} - \gamma_S s_{i,t-1} - \gamma_K k_{i,t-1}) \\ = \gamma_{0t} - \rho\gamma_{0,t-1} + (1-\rho)\omega_i + \xi_{it} + \varepsilon_{it} - \rho\varepsilon_{i,t-1}\end{aligned} \tag{1}$$

where ρ is the AR(1) coefficient in the revenue TFP process, ξ_{it} is the innovation in the same process, ω_i is the firm fixed effect, γ_{0t} is the time fixed effect, and ε_{it} is measurement error in log revenue.

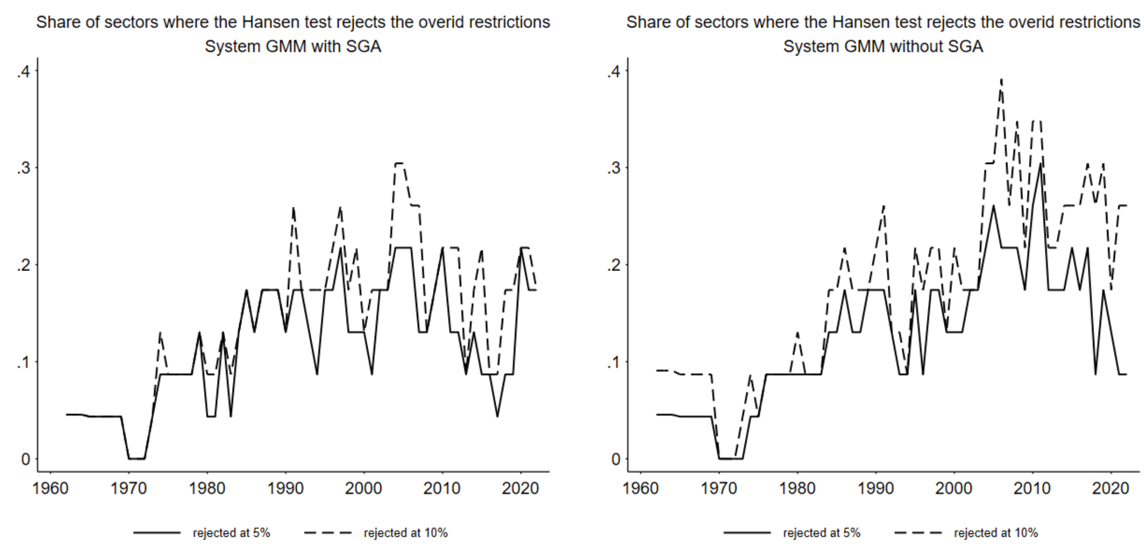
We assume that COGS and SGA are chosen by the firm in year t after observing revenue TFP (i.e. x_{it} and s_{it} are correlated with ξ_{it}), whereas PPE used in production in year t is chosen in year $t-1$

(i.e. k_{it} is uncorrelated with ξ_{it}, but correlated with $\xi_{i,t-1}$). We include year dummies to account for the year fixed effects. Then let $\epsilon_{it} = (1-\rho)\omega_i + \xi_{it} + \varepsilon_{it} - \rho\varepsilon_{it-1}$. The moment conditions used for the equations in first-differences are: $\mathrm{E}[x_{i,t-s}\Delta\epsilon_{it}] = 0$, $\mathrm{E}[s_{i,t-s}\Delta\epsilon_{it}] = 0$, and $\mathrm{E}[r_{i,t-s}\Delta\epsilon_{it}] = 0$ for $s = 2, 3$, and $\mathrm{E}[k_{i,t-s}\Delta\epsilon_{it}]$ for $s = 1, 2, 3$. The moment conditions used for the equations in levels are: $\mathrm{E}[\Delta x_{i,t-s}\epsilon_{it}] = 0$, $\mathrm{E}[\Delta s_{i,t-s}\epsilon_{it}] = 0$, and $\mathrm{E}[\Delta r_{i,t-s}\epsilon_{it}] = 0$ for $s = 1, 2, 3$, and $\mathrm{E}[\Delta k_{i,t-s}\epsilon_{it}]$ for $s = 0, 1, 2, 3$. The system GMM estimator combines these sets of moment conditions. The model is first estimated separately for each sector using 8-year overlapping windows, without imposing the non-linear restrictions on the coefficients in (1) that are implied by the assumed AR(1) revenue TFP process. We then impose those non-linear restrictions as in Blundell et al. (1992), so that the revenue elasticity of input V is the coefficient γ_V, like in the OLS case. For sectors that lack a sufficient number of observations at the start of the sample period, the initial window is widened to ensure that estimates can be produced.

D Hansen tests

For each sample period, and for specifications with and without the inclusion of SGA, Figure 20 reports the proportion of sectors in which the moment conditions used to obtain the system GMM estimates of our unrestricted dynamic specifications are rejected by the Hansen test of overidentifying restrictions.

Figure 20: Hansen tests of overidentifying restrictions



E Theory: revenue elasticities and markups

The production function takes the general form:

$$q_{it} = f_{it}(x_{it}, s_{it}, k_{it}) \tag{2}$$

where q_{it} is log output, $f_{it}(.)$ is the production function, and x_{it}, s_{it}, and k_{it}, are the logs of COGS, SGA, and PPE. We allow the production function to differ across firms and years (e.g. to accomodate non-Hicks neutral productivity), and do not restrict the production function to be homogeneous.

The log of the inverse demand schedule takes the general form:

$$p_{it} = p_{it}(q_{it}) \tag{3}$$

where p_{it} is log output price. We allow the inverse demand function to differ across firms and years, which introduces markup heterogeneity, potentially even for firms with the same level of output. This formulation allows for both monopolistic and oligopolistic competition, as $p_{it}(.)$ can capture the strategic choices of competitors. Also, we do not need to assume that inverse demand is homogenous.

We define the log revenue function by combining (2) and (3):

$$r_{it} = p_{it}(f_{it}(x_{it}, s_{it}, k_{it})) + f_{it}(x_{it}, s_{it}, k_{it})$$

By definition, the revenue elasticity of any input V_{it} is:

$$\frac{\partial r_{it}}{\partial v_{it}} = \frac{\partial f_{it}}{\partial v_{it}} \left(p'_{it}(q_{it}) + 1\right) \tag{4}$$

The first term of the right hand side is the output elasticity of the input, while the second term depends only on the elasticity of demand at output level q_{it} (since demand is not necessarily CES, $p'_{it}(q_{it})$ is not necessarily a constant). Notice that this expression is obtained by differentiating the log revenue function, so it does not require assumptions on the timing of input choices.

We can then use static profit maximization to show that the term $(p'_{it}(q_{it}) + 1)$ is the inverse of the markup. To do so, we only require the further assumption that the firm can choose output q_{it} in period t. This is always the case when at least one input is flexible. The firm's objective is:

$$\max_{Q_{it}} P_{it}(Q_{it})Q_{it} - C_{it}(Q_{it})$$

where uppercases denote levels, and $C_{it}(Q_{it})$ is the cost function, which can be obtained analytically through cost minimization.[29]

The first order condition with respect to the only choice variable, Q_{it}, after minimal manipulation, is:

$$\frac{P_{it}}{C'_{it}(Q_{it})} = \left(p'_{it}(q_{it}) + 1\right)^{-1}$$

The left hand side is just the ratio of output price to marginal cost, i.e. the markup, μ_{it}. Hence, at the profit maximizing quantity, we can substitute this into (4) to obtain:

$$\frac{\partial r_{it}}{\partial v_{it}} = \frac{\partial f_{it}}{\partial v_{it}} \frac{1}{\mu_{it}}. \tag{5}$$

In words, the revenue elasticity of any input is the output elasticity divided by the markup. This relation holds also for dynamic inputs, provided the firm maximizes profits each period and is able to adjust output (by adjusting other, e.g. flexible, inputs). Notice that this result does not rely on any assumption about the sum of output elasticities.

If we then introduce the assumption from Section 4 that the sum of output elasticities is common across firms and known, i.e. $\sum_v \frac{\partial f_{it}}{\partial v_{it}} = \nu$, we can sum both sides of (5) over all inputs and solve for the markup:

$$\mu_{it} = \frac{\nu}{\sum_v \frac{\partial r_{it}}{\partial v_{it}}}$$

[29] As long as (2) is well behaved, and there is at least one input that can be chosen at time t, a solution to the cost minimization problem exists. It follows that we can write total cost as a function of Q_{it} only.

where the denominator is the sum of revenue elasticities. This result generalizes the example discussed in Section 4 for Cobb-Douglas production and CES demand.

F Theory: revenue elasticities for quasi-fixed inputs

Consider a firm with a single input (X_{it}) and a Cobb-Douglas revenue production function $R_{it} = A_{it}X_{it}^{\gamma}$. This is of the form derived in Section 4, assuming an underlying Cobb-Douglas production function and CES demand, with $A_{it} = \exp(\gamma_0 + u_{it})$ and $\gamma = \beta/\mu$, where β is the output elasticity and μ is the price-cost markup for a profit-maximising firm. The firm hires or rents the input X_{it} at a known unit price W, so that net revenue is $\Pi_{it} = A_{it}X_{it}^{\gamma} - WX_{it}$.

Recall that $u_{it} = (\omega_{it}/\mu) + \zeta_{it}$ depends on the idiosyncratic components of total factor productivity and demand. If the level of the input is chosen to maximise net revenue after the firm has observed productivity and demand in period t, we have the usual first-order condition

$$\frac{\partial \Pi_{it}}{\partial X_{it}} = \gamma A_{it}X_{it}^{\gamma-1} - W = \gamma\left(\frac{R_{it}}{X_{it}}\right) - W = 0$$

and hence

$$\gamma = \frac{WX_{it}}{R_{it}}.$$

This is the result of Bond et al. (2021), Section 2.1, for flexible inputs, in this special case.

Now suppose that the level of the input is chosen in the previous period, to maximise the expected value of net revenue in period t given information available in period $t-1$, i.e. to maximise $E_{t-1}[\Pi_{it}] = E_{t-1}[A_{it}]X_{it}^{\gamma} - WX_{it}$. The first-order condition is then

$$\frac{\partial E_{t-1}[\Pi_{it}]}{\partial X_{it}} = \gamma E_{t-1}[A_{it}]X_{it}^{\gamma-1} - W = \gamma\left(\frac{E_{t-1}[A_{it}]X_{it}^{\gamma}}{X_{it}}\right) - W = \gamma\left(\frac{E_{t-1}[R_{it}]}{X_{it}}\right) - W = 0$$

and hence

$$\gamma = \frac{WX_{it}}{E_{t-1}[R_{it}]}.$$

This is the corresponding relation between the revenue elasticity and the expected revenue share for quasi-fixed inputs that we exploit in Section 5.2.

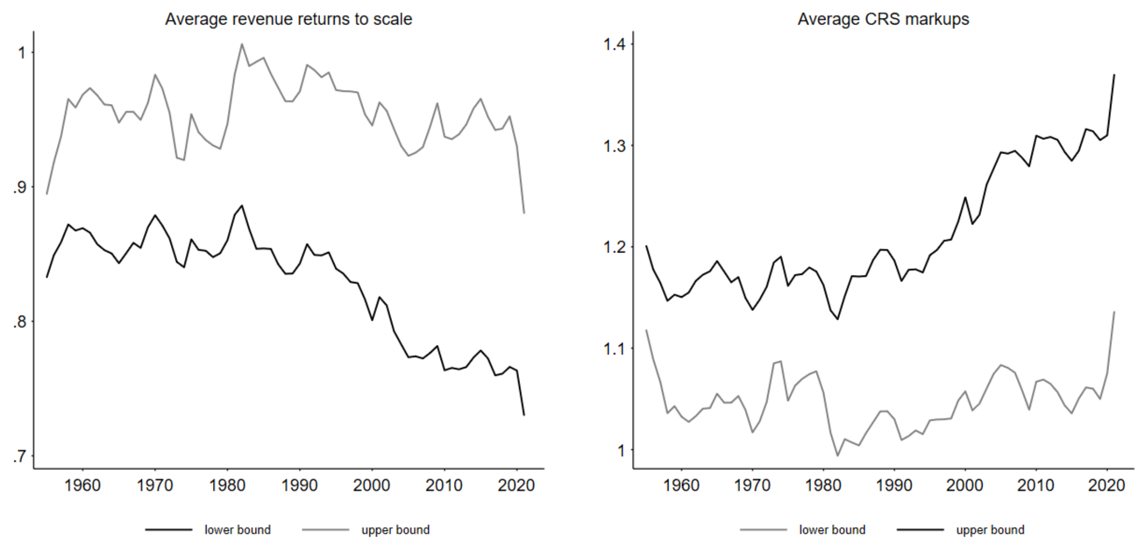
G Robustness: average revenue elasticities, returns to scale, and markups

G.1 Sales weights and basic cleaning sample

Figure 21 reports bounds for the weighted average revenue elasticities of COGS, SGA, and PPE as outlined in Section 5, when the sample is cleaned following the basic cleaning routine (see Appendix A) instead of the enhanced one, and when sales weights are used in place of total cost weights. Figure 22 reports the corresponding weighted average revenue returns to scale, and the weighted average markups implied under the assumption of constant returns to scale production technologies.

Figure 21: Bounds for revenue elasticities

Figure 22: Revenue returns to scale and CRS markups



G.2 Common user cost of capital

Figures 23 and 25 report bounds for the weighted average revenue elasticities of COGS, SGA, and PPE as outlined in Section 5, when the sample is cleaned following the enhanced cleaning routine, and a common risk permium is used to impute the user cost of capital, instead of a firm-year-specific risk premium (see Appendix A). Figures 24 and 26 report the corresponding weighted average revenue returns to scale, and the weighted average markups implied under the assumption of constant returns to scale production technologies.

Figure 23: Bounds for revenue elasticities (total cost weights)

Figure 24: Revenue returns to scale and CRS markups (total cost weights)

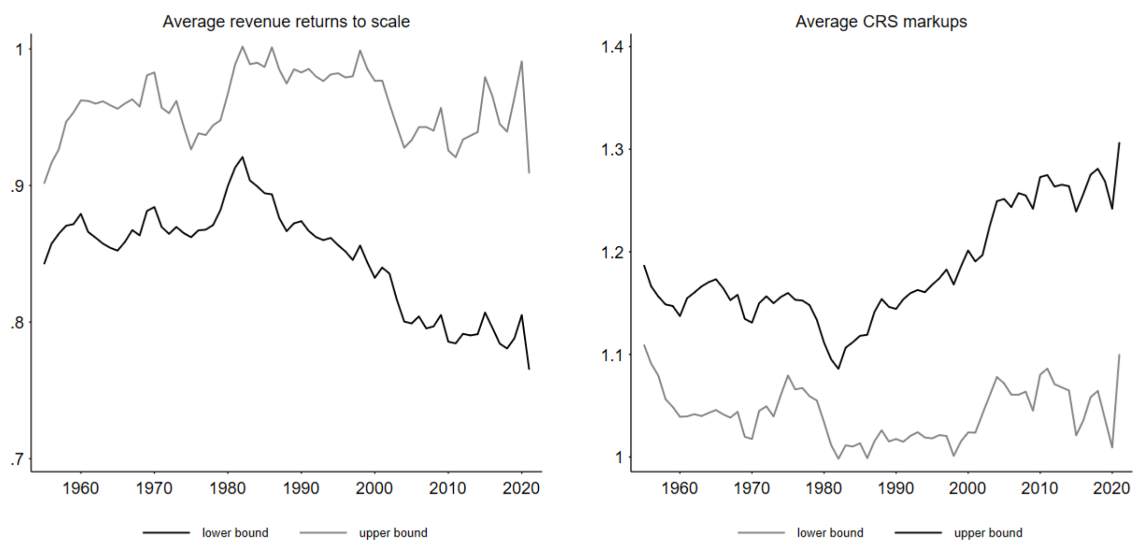


Figure 25: Bounds for revenue elasticities (sales weights)

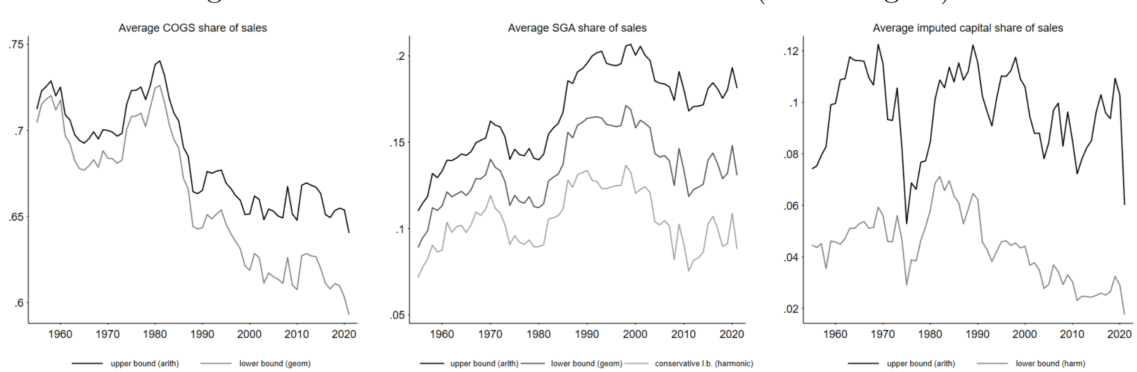
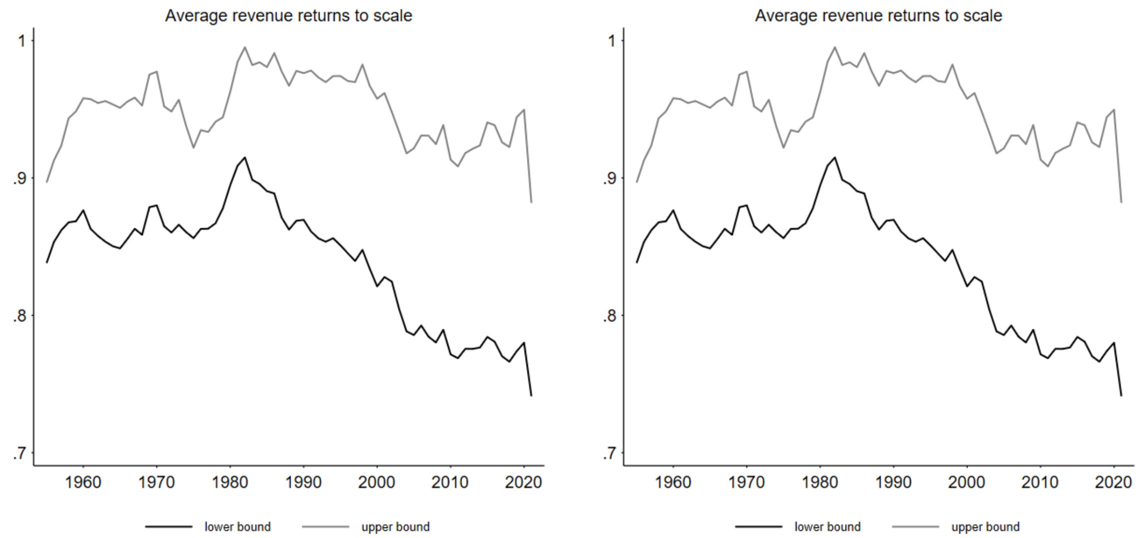


Figure 26: Revenue returns to scale and CRS markups (sales weights)



G.3 NYSE and NASDAQ listed companies only

Figures 27 and 29 report bounds for the weighted average revenue elasticities of COGS, SGA, and PPE as outlined in Section 5, when the sample is cleaned following the enhanced cleaning routine, and companies that are not listed on either the NYSE or NASDAQ are dropped. Figures 28 and 30 report the correspoding weighted average revenue returns to scale, and the weighted average markups implied under the assumption of constant returns to scale production technologies.

Figure 27: Bounds for revenue elasticities (sales weights)

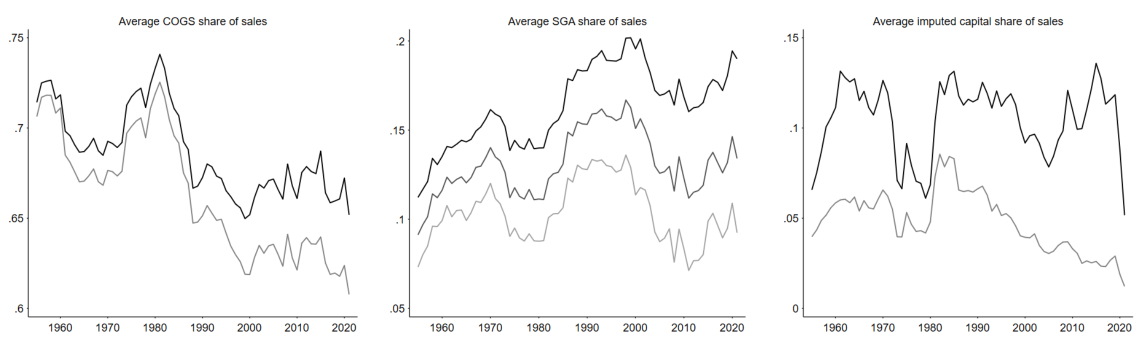


Figure 28: Revenue returns to scale and CRS markups (total cost weights)

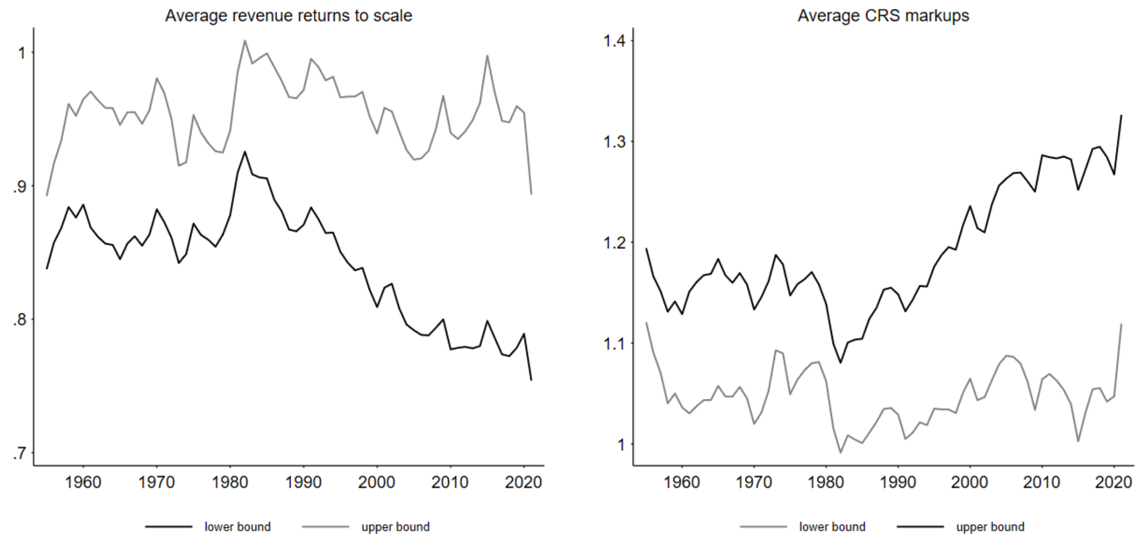


Figure 29: Bounds for revenue elasticities (sales weights)

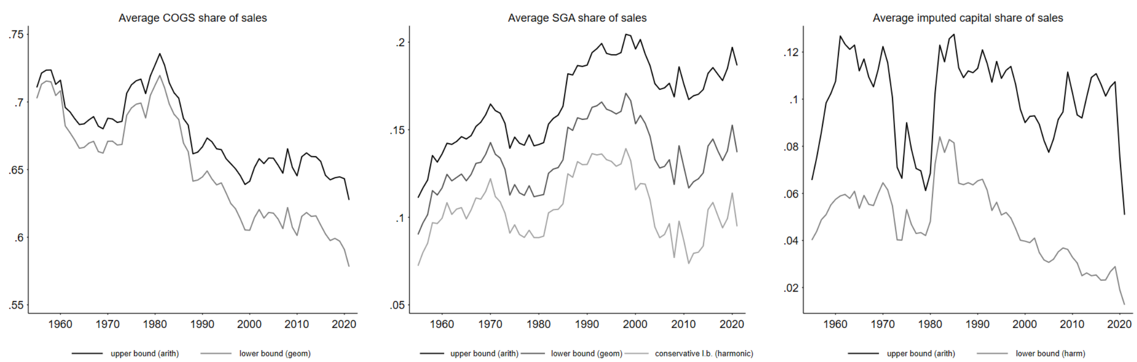
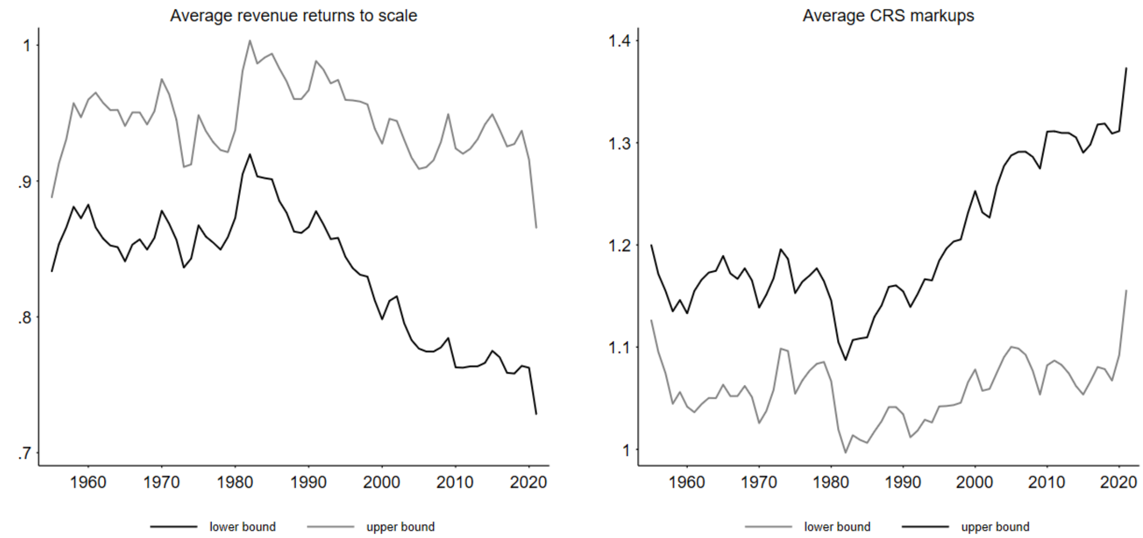


Figure 30: Revenue returns to scale and CRS markups (sales weights)



G.4 Comparison of upper bounds and econometric estimates of revenue returns to scale

Figure 31 plots our upper bound for the weighted average revenue returns to scale of the revenue function, together with the corresponding weighted average across sectors of the revenue returns to scale implied by econometric estimates of the revenue production function specification with SGA included. Here we consider both total cost weights and sales weights.[30]

[30]Figure 4 shows the upper bound using total cost weights, while Figure 14 shows the econometric estimates using sales weights.

Figure 31: Upper bound and econometric estimates of the revenue returns to scale